Rudolf Vetschera

Experimentation and Learning in Repeated Cooperation

Abstract

We study an agency model, in which the principal has only incomplete information about the agent's preferences, in a dynamic setting. Through repeated interaction with the agent, the principal learns about the agent's preferences and can thus adjust the inventive system. In a dynamic computational model, we compare different learning strategies of the principal when facing different types of agents. The results indicate the importance of a correct specification of the agent's preferences.

1 Introduction

Innovative organizational forms like virtual organizations pose new questions to organization theory, which challenge traditional forms of analyzing organizations. Many of these new forms of organization rely on cooperation between partners, who are located at considerable distance and communicate with each other only via technical means. Yet it is important for the survival of such organizations that each partner is able to predict, and to a certain extent control, the behavior of the other partners. Prediction and control of the actions of another economic agent requires knowledge about the preferences of that agent. This necessity can clearly be demonstrated in classical models of agency theory (Mirrlees, 1976; Harris/Raviv, 1979; Spremann, 1987). In these models, the principal is assumed to have perfect knowledge of the agent's utility function. This makes it possible for the principal to design an incentive system controlling the agent's behavior in a way that is optimal for the principal.

Even in a traditional hierarchical organization, the assumption that a superior has perfect knowledge of her subordinates' preferences can be considered as problematic and has been criticized in the literature. For example, (Rose/Willemain, 1996b, p.142) stated that „The principal does not know the agent’s utility function and may have only a vague articulation of her own“. In a setting where both partners are geographically dispersed and communicate only via electronic media, this assumption is even more unrealistic. In such a situation, one can only assume (highly) incomplete knowledge about a network partner's preferences. However, once we consider information to be incomplete, we also have to take into account another phenomenon: learning. In each interaction, a network partner will also reveal some information about his preferences, allowing the other partner to build a continuously improving representation of his preferences.

The relevance of learning for a principal – agent relationship has been pointed out for example by (Eisenhardt, 1989), but so far there is little literature which explicitly takes into account that a principal improves her model of an agent's behavior by observing the agent's reactions to various incentive systems. In the present paper, we study the continuous learning that takes place between two partners engaged in repeated cooperation. The paper is structured as follows: in section two, we give a brief exposition of the decision problems faced by the two partners and introduce a formal model describing this situation based on the work in (Vetschera, 2000). Considering the specific situation of partners in geographically dispersed networks, this model deviates from traditional agency models also by using ex-ante instead of ex-post incentives.

Section three introduces the dynamic extension of this model. Since we are dealing with a situation of incomplete information about a remote network partner, we must also take into account the possibility that even the structure of the preference model about the partner might contain errors. Section four thus introduces a simulation model in which various levels of "misunderstandings" between the remote partners can be analyzed. Section five presents some results from experiments with this model. Section six concludes the paper by providing an outlook on ongoing and future research activities.

2 The Ex ante Incentive Model

Agency models usually consider an incentive system that can be describes as ex post incentive system. In an ex post incentive system, payments are made by the principal to the agent only after the agent has undertaken his effort and outcomes (or whatever information is used to determine the payment) are observed.

In transactions with a remote partner, the situation might be different. It could be necessary for one partner to invest into the relationship *before* the other partner performs any activity or outcomes are obtained. In transaction cost theory, such investments might be considered under the category of costs of agreement.

One might ask why such an ex ante investment should influence the transaction partner's behavior at all. If he already has received the reward, why should he then exert extra effort or in some other way adapt his behavior to the principal's wishes? Apart from ethical considerations, this argument is also not valid when the time of the investment does not coincide with the effect the investment has on the agent. Ex ante investment by the principal can lead to consequences for the agent which occur only after he has performed his activities and which are contingent on how these actions have been performed. Consider for example expenditures for publicizing a new strategic alliance. The more such an alliance is publicized, the larger would be the damage to reputation if one partner later defects. Thus ex ante expenditures might influence the subsequent behavior of the partner by altering consequences at a still later stage.

Given that there is only incomplete information on the transaction partner's preferences, we model his decision as a random variable. In order to keep the model simple, we will only consider two possible actions of the transaction partner: cooperation and defection. Figure 1 illustrates the decision problem from the point of view of the partner setting the ex ante incentive.

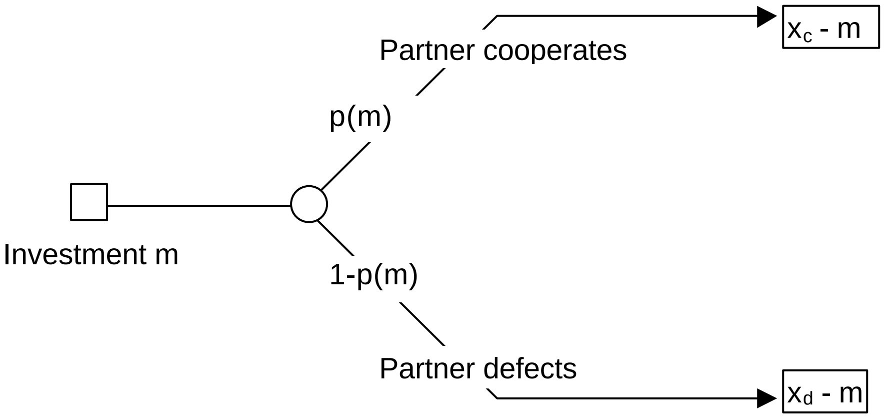

Figure 1: Decision problem of the investor

We thus consider a situation in which a transaction with the partner definitely takes place, leaving the relationship is not an option. The only decision to be made thus concerns the level of ex ante investment *m*. The outcome of the transaction depends on the partner's decision to cooperate or defect. If the partner cooperates, the benefit from the transaction is x_c, if he defects, the benefit is x_d. The probability of cooperation depends (monotonically) on the ex ante investment *m*.

From the point of view of the transaction partner, we have to consider two types of outcome, those which are affected by the investment and those which are not. The decision problem of the transaction partner is shown in figure 2.

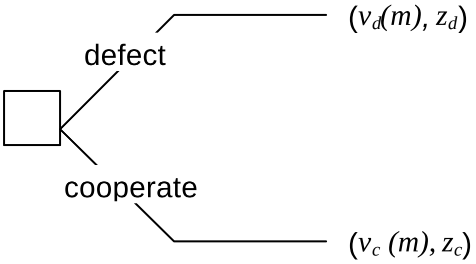

Figure 2: Decision problem of the transaction partner

We denote the consequences which are influenced by the investment by *v(m)*, those which are not influenced by *z*. Both types of consequences also depend on the action taken. Furthermore, we assume that the agent evaluates these two types of consequences according to a linear utility function of the form

$$u(v(m), z) = w \cdot v(m) + (1 - w) \cdot z \tag{1}$$

In utility function (1), parameter *w (*0 = *w* = 1) represents the weight which the agent assigns to the benefits from cooperation. This weight, however, is unknown to the principal.

The transaction partner will cooperate whenever his utility from cooperation exceeds the utility obtained when defecting, i.e. when

$$w \cdot v_c(m) + (1 - w) \cdot z_c \geq w \cdot v_d(m) + (1 - w) \cdot z_d \tag{2}$$

To simplify the notation, we rewrite (2) as

$$w(z + v(m)) > z \tag{3}$$

where $z = z_d - z_c$ and $v(m) = v_c(m) - v_d(m)$. We further assume that v is a linear function of m, i.e. that

$$v(m) = a{\cdot}m \tag{4}$$

Linearity of (4) is not a crucial assumption for the following analysis, however, $v(m)$ must be a concave function of m for a unique solution to the principal's optimization problem to exist.

Following the literature on decision making under incomplete information (Weber, 1987), we assume that the only information the principal has on the transaction partner's preferences is an interval $(\underline{w}, \overline{w})$ in which the true weight falls, and that weights are uniformly distributed within that interval.

Under these assumptions, the probability that the transaction partner cooperates is given by

$$p_c = \begin{cases} 1 & \text{for } z/(a \cdot m + z) \leq \underline{w} \\ \dfrac{\overline{w} - z/(a \cdot m + z)}{\overline{w} - \underline{w}} & \text{for } \underline{w} \leq z/(a \cdot m + z) \leq \overline{w} \\ 0 & \text{for } \overline{w} \leq z/(a \cdot m + z) \end{cases} \tag{5}$$

i.e. by the fraction of the total interval of the parameter space $(\underline{w}, \overline{w})$ in which cooperation takes place.

Focussing for the moment on the middle term of (5), we obtain the expected profit of the principal as

$$\begin{aligned} g &= p_c \cdot x_c + (1 - p_c) \cdot x_d - m \\ &= \frac{\overline{w} - z/(a \cdot m + z)}{\overline{w} - \underline{w}} x_c + \left(1 - \frac{\overline{w} - z/(a \cdot m + z)}{\overline{w} - \underline{w}}\right) x_d - m \end{aligned} \tag{6}$$

and the optimum investment m^* of the principal as

$$m^* = \frac{-z(\overline{w} - \underline{w}) + \sqrt{az(\overline{w} - \underline{w})(x_c - x_d)}}{a(\overline{w} - \underline{w})} \tag{7}$$

However, m^* will only be non-negative iff

$$a(x_c - x_d) > z(\overline{w} - \underline{w}) \tag{8}$$

If this condition is not fulfilled, the optimum for the principal consists in not providing any incentives at all and accept the likely defection of the transaction partner. The same is true in the last row of equation (5), thus whenever m^* falls below a lower threshold of

$$\underline{m} = \frac{z(1 - \overline{w})}{a\overline{w}} \tag{9}$$

no investment should take place at all. Conversely, since it is not possible to increase the probability of cooperation beyond 1, the first row of equation (5) provides an upper bound for the meaningful range of *m** as

$$\overline{m} = \frac{z(1-\underline{w})}{a \cdot \underline{w}} \tag{10}$$

Whenever *m** according to (7) exceeds $\overline{m}$, only the amount $\overline{m}$ should be invested.

3 A Dynamic Model

Each completed interaction with the agent can reveal some information about the agent's preferences to the principal. When the agent has cooperated, it follows from (3) that

$$w \geq \frac{z}{z + a \cdot m} \tag{11}$$

and thus the principal can update her lower bound estimate of *w* accordingly. Similarly, whenever the agent defects, the upper bound estimate $\overline{w}$ can be updated.

To simplify the exposition, we introduce the following notation. The interval $(\underline{w}, \overline{w})$ represents the principal's *current state of knowledge* about the agent's preferences. For a given decision problem, i.e. given payoffs x_c, x_d and *z*, as well as a given parameter α, both the optimal level of investment and the corresponding net profit depend only on the state of knowledge. We thus write them as $m^*(\underline{w}, \overline{w})$ and $g^*(\underline{w}, \overline{w})$. When we consider a multi-period problem, the values realized in period *t* are designated by $m_t^*(\underline{w}, \overline{w})$ and $g_t^*(\underline{w}, \overline{w})$.

We first consider a two-period problem. In making her decision on the incentive for the first period, the principal has to take into account the effects on the second period. Thus the profit function for the first period must be extended to consider profit in the second period:

$$g = p_c(m_1) \cdot (x_c + g_2^*(\underline{w}_c, \overline{w}_c)) + (1 - p_c)(x_d + g_2^*(\underline{w}_d, \overline{w}_d)) - m_1 \tag{12}$$

where $(\underline{w}_c, \overline{w}_c)$ and $(\underline{w}_d, \overline{w}_d)$ represent the updated state of knowledge after cooperation and defection, respectively.

Equation (12) shows the typical recursive structure of a dynamical programming problem. Thus it is conceptually easy to extend it to more than two periods. However, since function (12) is already not convex in m_1 and the complexity of the problem rapidly increases when additional periods are taken into account, we do not present analytical solutions for this problem. The simulations presented in the following section are based on a numerical solution procedure.

4 Simulation Experiments

4.1 Simulation in Learning Models

Simulation models have been used by many researchers to study phenomena of learning and strategic interaction in settings which are similar to our problem. Most widely known is probably the work of Axelrod (Axelrod, 1984) about the prisoner's dilemma game. While in Axelrod's original experiments, agents used the same strategies during the entire simulation, later studies, e.g. (Axelrod, 1987), also considered the possibility that agents learn strategies over time.

These simulation experiments, as well as other similar studies (Watanabe/Yamagishi, 1999; Hoffmann, 2001), differ in several important aspects from the present study. Firstly, in these models, agents learn strategies and do not explicitly model their opponent's preferences or behavior. Strategies are often represented as finite automata, following the approach developed by (Rubinstein, 1986), and are modified using genetic algorithms. The "genes" used in these algorithms can also be considered as representations of explicit knowledge or organizational routines, which are learned over time. While this interpretation was used in some studies of organizational learning (Bruderer/Singh, 1996), those experiments lack the specific focus on interaction with other agents.

Furthermore, agents in these models usually do not play against a fixed partner, but against a whole population (or a random sample of the population). Thus, strategies are not adapted to a specific partner but are evaluated in a more general framework. One exception is the paper by (Meng/Pakath, 2001), where agents learn strategies in an iterated prisoner's dilemma game against a specific partner. However, the focus of that paper is on design issues of the classifier system used to represent the agents' knowledge.

The work which is probably most closely related to our problem is (Rose/Willemain, 1996b; Rose/Willemain, 1996a). In this model, a principal learns to use various types of incentive systems vis-a-vis a population of different agents. Our model differs from this model by allowing for the incentive level *m* to be varied continuously rather than to be set at prespecified levels. More importantly, in our model, one partner formulates and refines an explicit model of the preferences of the other partner, rather than learning which incentive strategy works best in a general environment.

4.2 Agent Types

An optimal learning strategy based on equation (12) rapidly becomes very hard to compute. This effort might not be worthwhile when explicit consideration of learning will improve the principal's profit only slightly compared to the one-period model (7). Furthermore, it is based on two rather rigid assumptions:

1. That the linear model (1) is a correct representation of the agent's preferences
2. That the agent does not anticipate the learning effect and modify his behavior accordingly.

To check for the robustness of a dynamic model against mis-specification of the agent's preferences or deliberate manipulation by the agent, we confront our agents with various types of opponents similar to the model of (Meng/Pakath, 2001).

Four variants for the principal were used in the simulations. The first two variants were based on the dynamic optimization model of equation (12) ("Dynamic") and on the single-period optimization of equation (7) ("One-Shot"). To study whether such optimization models have a significant value at all, they were compared to two "naive" strategies, in which incentives were increased by a random amount after defection and decreased after cooperation. The two naive strategies differed in the reference value that was updated. In the first strategy ("Naive/f" for "fixed"), the absolute amount of compensation payments was adapted. In the second strategy ("Naive/r" for "relative"), incentives were set relatively to *z*, the defection payoff and the ratio of *m* to *z* was adapted to better relate compensation to the actual decision problem of the partner.

To analyze the second research question, these four types of principals were tested against four types of transaction partner: The first type ("Honest") based its decision strictly on the utility criterion (2). The second type ("Noisy") also reacted according to equation (2), but its perception of the cooperation benefits *v(m)* was disturbed by a random term, thus it sometimes made a decision inconsistent with its true utility function.

For the transaction partner, it is an advantage if the principal underestimates the true weight *w*, since the principal will then provide higher incentives for cooperation. Underestimation of *w* will be achieved when the transaction partner defects in situations in which, according to his true utility function, he should not defect. This behavior is modeled in the third type of partner ("Bias"). Similarly to "Noisy", this partner deviates from the original utility function by introducing a random disturbance to the benefits of cooperation, but unlike "Noisy", the benefit is only modified downwards, thus inducing additional defections.

The fourth type of partner ("Random") randomly chooses between cooperation and defection without considering the problem parameters at all.

4.3 Experimental Setup and Parameter Values

From the four types of models for the principal and four types of models for the transaction partner, 16 possible pairs can be formed. Experiments were performed using all these pairs.

In a stable environment of repeated identical problems, many of those pairs, especially those with the optimizing principal and/or "Honest" transaction partners, would quickly reach a steady state in which nothing is learned, although the principal's state of knowledge about the transaction partner might be far less than perfect. To introduce a certain level of environmental uncertainty necessary for learning, experiments used sequences of randomly generated decision situations. A decision situation in this context is characterized by values of the payoff levels x_c and *z*. Each experiment consisted of 30 simulated interactions (time periods), and for each period values of x_c and *z* were randomly generated from uniform distributions. The upper and lower bounds of the uniform distribution can be varied to analyze the impact of uncertainty of model outcomes. To simplify comparison of results between different types of principals and transaction partners, all pairs of principal and transaction partner went through the same sequence of decision situations in each experiment. In each sequence, the principal started with full uncertainty about *w*, i.e. the state of knowledge was initialized to the whole interval (0,1). To obtain adequate data for statistical analysis, 1000 such experiments were performed for each parameter setting. Table 1 summarizes the parameter values used.

Parameter	Range	Description/Remarks
x_c	0.1 – 1.0	Principal’s payoff in case of cooperation
x_d	0	Principal’s payoff in case of defection
z	0.1 – 0.9	Partner’s payoff from defection
a	1	Efficiency of incentive (constant for all experiments)
w_{true}	0.6	True weight (constant for all experiments)
$\underline{w}_0$	0	Starting value for lower bound of w
$\overline{w}_0$	1	Starting point for upper bound on w

Table 1: Simulation parameters

The simulation was implemented in Object Pascal using the Borland Delphi compiler.

4.4 Hypotheses

Our research questions lead to the formulation of several hypotheses, which can be statistically tested using the simulation results:

Hypothesis **H1**: The dynamic model will perform better than the static model.

This hypothesis is a direct result of the first research question, where we wanted to ask whether the considerably more complex dynamic model is worth the extra computational effort. Obviously, the dynamic model is designed to generate an optimal solution for a multi-period problem. However, since the model does not have perfect information about future decision situations, it might miscalculate the effect of learning. This hypothesis thus tests whether it indeed performs better than the static model in an uncertain environment.

Hypothesis **H2**: Optimizing strategies perform better than naive strategies

This hypothesis is based on the same argument as **H1**, but compares the two “optimal” strategies to the much simpler adaptive ones.

Hypothesis **H3**: The dynamic strategy will gain more payoff from learning then the one-shot strategy.

Since the dynamic strategy was designed to explicitly consider learning effects, it should also benefit more from learning.

Hypothesis **H4**: The dynamic strategy will gain more knowledge from learning than the one-shot strategy.

The effect of learning can be measured in several ways. While hypothesis **H3** is formulated in terms of payoff values, this hypothesis focuses on the information obtained, i.e. the size of the interval $(\underline{w}, \overline{w})$.

Hypothesis **H5**: The dynamic strategy will encounter more defections (in the early periods).

Since the dynamic strategy tries to learn about the transaction partner’s preferences, we expect it to “experiment” more, especially in the early periods of interaction. Thus it will more likely encounter defections than the static strategy.

5 Results

5.1 Statistical tests

In this section, we present the results of the statistical analysis of data obtained from the simulation experiments. All hypotheses to be tested basically have the same structure: outcomes for one type of strategy are assumed to be significantly better than outcomes from another strategy. Therefore, we can perform tests independently for each of the four different types of opponents and discuss robustness of results with respect to different environmental conditions by comparing the results for the different opponents.

Since all pairs of principal/partner were tested using the same sequence of decision situations, we can directly compare results between pairs in each experiment. To test whether one strategy outperforms another strategy, we therefore can test the hypothesis whether, on average, the difference in outcomes between the two strategies is significantly greater than zero. Since these differences were not always normally distributed, we not only performed a t-test but also a nonparametric sign test.

5.2 Results for H1

Hypothesis **H1** stated that the dynamic strategy should perform better than the static strategy. To test this hypothesis, we analyzed the total profit obtained by these strategies in their interactions with the different types of opponents over all 30 periods. The following figures show box plots representing the distribution of total profits across all experiments.

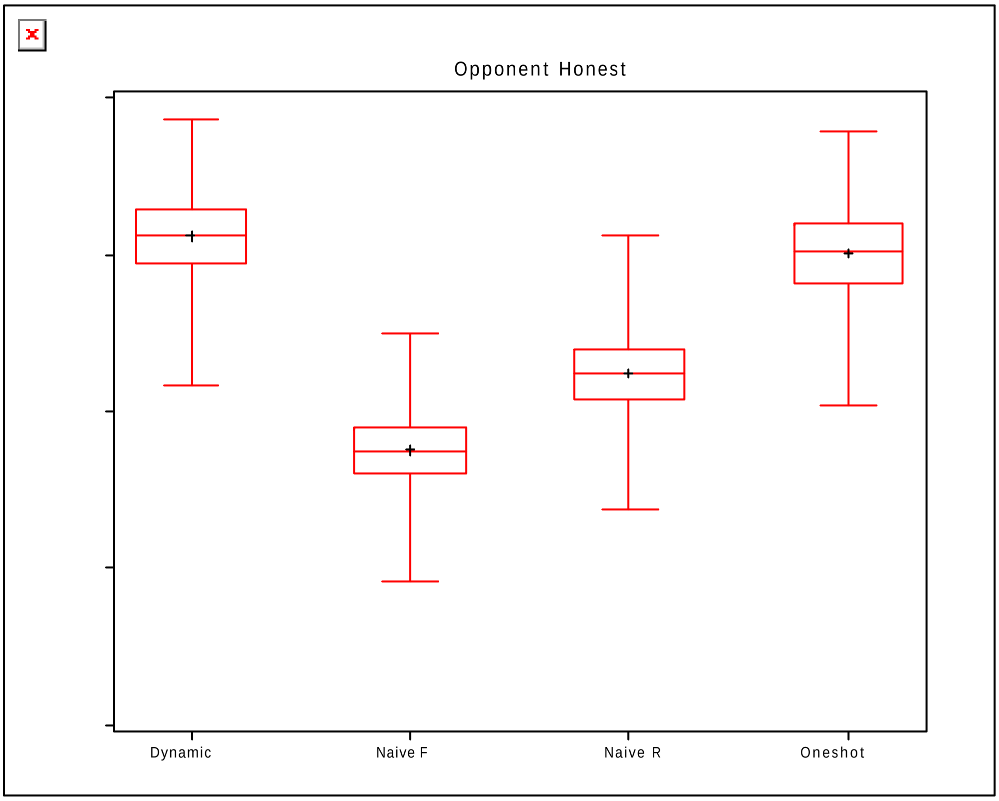

Figure 1: Profit against "Honest" opponent

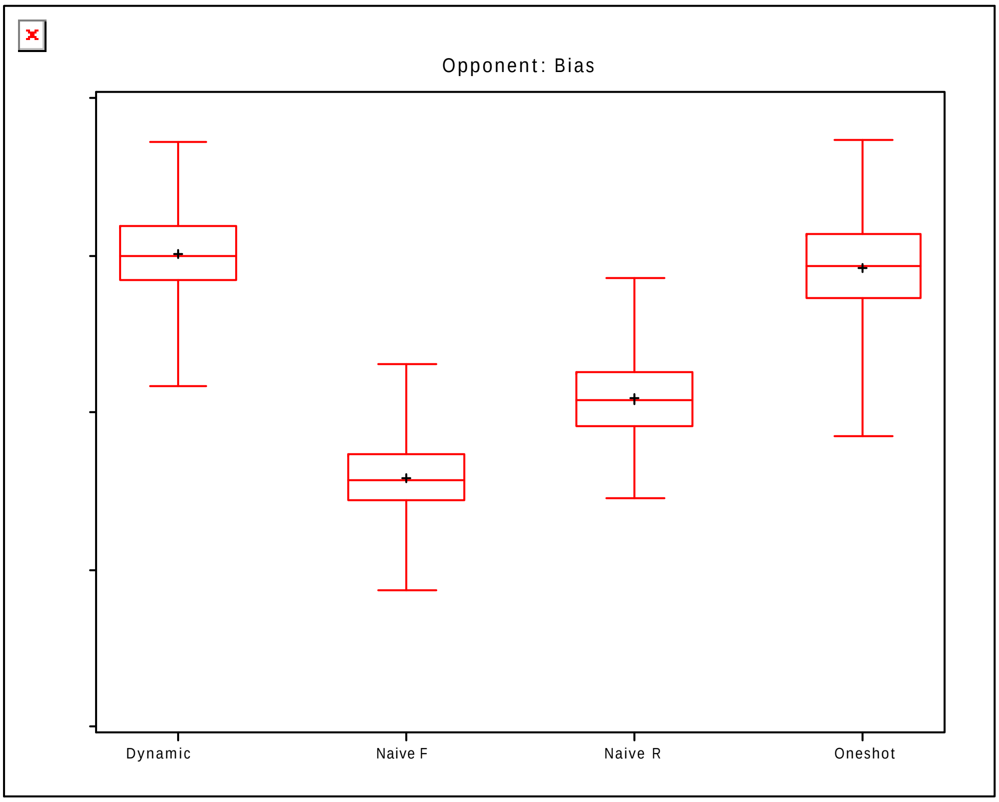

Figure 2: Profit against "Bias" opponent

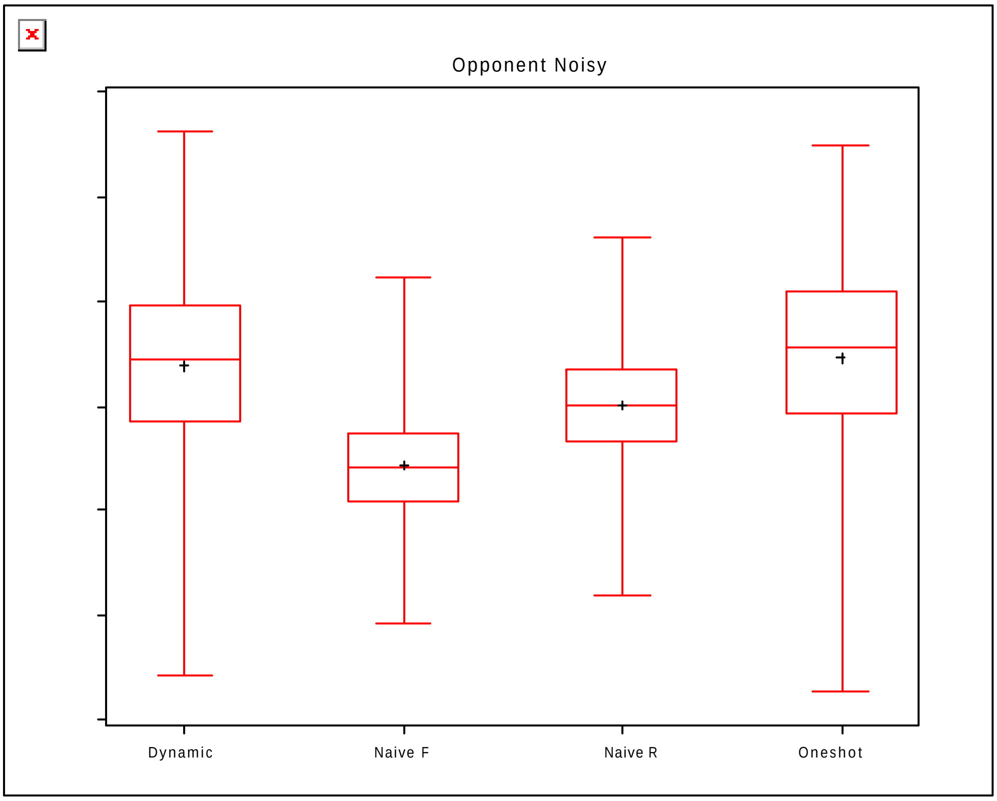

Figure 3: Profit against "Noisy" opponent

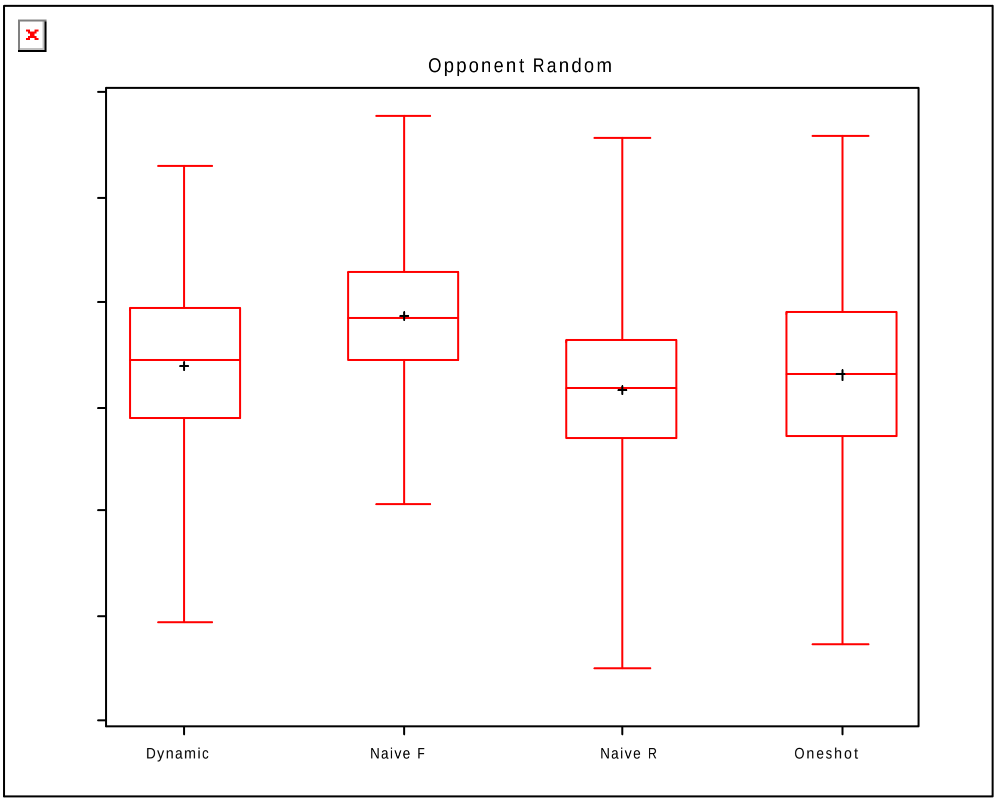

Figure 4: Profit against "Random" opponent

The results of the statistical analysis comparing the performance of the two strategies experiment by experiment are shown in table 2

Opponent	Honest	Noisy	Bias	Random
Mean	0.11147	-0.03782	0.08997	0.03969
SD	0.15899	0.47800	0.19899	0.48873
Median	0.07789	-0.05146	0.04877	0.03373
t-test	22.171	-2.5020	14.29718	2.56787
p	<0.001	0.0125	<.0001	0.0104
N > 0	747	454	659	535
Sign test	247	–46	159	35
p	<0.0001	0.0040	<.0001	0.0291
Kolmogorov-Smirnov	0.09070	0.03117	0.10562	0.03083
p	<0.01	0.019	<0.010	0.021

Table 2: Hypothesis H1, Experiment by Experiment Comparison

Table 2 shows the difference between total profit from the dynamic strategy and the static strategy for the various types of transaction partners. The first block of rows provides descriptive data on the difference. The second blocks contains results of a t-test testing against the hypothesis that this difference is zero. The third block contains results of a nonparametric sign test for the same hypothesis and the last block provides the Kolmogorov-Smirnov statistic, which in this case indicates that the difference is not normally distributed in several instances.

As could be expected, the dynamic strategy indeed performs better against the "Honest" transaction partner, who always reacts according to his true utility function. However, this superior performance is not robust against random distortions in the reactions of the partner, as against the "Noisy" transaction partner, the static strategy performs better. However, this result is not significant at the 1% confidence level according to the t-test.

Surprisingly, the dynamic strategy also performed better against the "Bias"ed transaction partner, who deliberately tries to influence learning. But the results also show that this partner is indeed able to exploit the principal to a certain extent and lower her profit. Against the "Random" partner, the difference is again not significant.

5.3 Results for H2

Hypothesis **H2**: optimizing strategies perform better than naive strategies

For this hypothesis, we have to compare both optimizing strategies (dynamic and static) to both naive strategies (fixed and relative payments). The following tables present the results of these comparisons.

5.3.1 Dynamic vs. Naive fixed strategies

Opponent	Honest	Noisy	Bias	Random
Mean	1.36476	0.47398	1.42924	-0.24254
SD	0.23623	0.38598	0.24462	0.43106
Median	1.38027	0.51694	1.42971	-0.20528
t-test	182.696	38.833	184.762	-17.793
p	<0.0001	<0.0001	<0.0001	<0.0001
N > 0	1000	879	1000	295
Sign test	500	379	500	–205
p	<0.0001	<0.0001	<0.0001	<0.0001
Kolmogorov-Smirnov	0.03631	0.04948	0.02231	0.04563
p	<0.010	<0.01	>0.150	<0.010

Table 3: Dynamic vs. Naive fixed strategies, Experiment by Experiment Comparison

Both tests clearly confirm the superiority of the optimizing strategy versus a rational transaction partner. However, when faced with an irrational partner, the simpler strategy performs significantly better.

5.3.2 Dynamic vs. Naive relative strategies

Opponent	Honest	Noisy	Bias	Random
Mean	0.87148	0.18324	0.92134	0.10884
SD	0.19133	0.37618	0.20533	0.46545
Median	0.86543	0.22122	0.91337	0.13245
t-test	144.039	15.403	141.892	7.394
p	<0.0001	<0.0001		<.0001
N > 0	1000	712	1000	604
Sign test	500	212	500	104
p	<0.0001	<0.0001		<0.0001
Kolmogorov-Smirnov	0.02135	0.05202	0.02349	0.02776
p	>0.150	<0.010	>0.150	0.061

Table 4: Dynamic vs. Naive relative strategies, Experiment by experiment comparison

Surprisingly, the naive strategy based on relative payoffs was outperformed by the dynamic strategy even in the case of the pure random transaction partner. It seems that its apparently higher level of rationality led it to fall into the same ‘trap’ as the optimizing strategies in trying to adapt to a pattern that in reality did not exist in the opponent’s behavior.

5.3.3 Static vs. Naive fixed strategies

Opponent	Honest	Noisy	Bias	Random
Mean	1.25328	0.51180	1.33928	-0.28222
SD	0.27347	0.41311	0.30074	0.45118
Median	1.27026	0.56137	1.36176	-0.27591
t-test	144.925	39.178	140.826	-19.781
p	<.0001	<.0001	<.0001	<.0001
N > 0	1000	882	1000	289
Sign test	500	382	500	–211
p	<0.0001	<0.0001	<0.0001	<0.0001
Kolmogorov-Smirnov	0.03056	0.05752	0.03759	0.03164
p	0.023	<0.010	<0.010	0.016

Table 5: Static vs. Naive fixed strategies, Experiment by experiment comparison

The results for the static strategy are very similar to the results for the dynamic strategy. Again, the optimizing strategy is clearly superior to the naive heuristic against a rational opponent, even if that opponent purposefully tries to misrepresent his preferences. But when faced with a purely random opponent, the simple strategy performs better. Results against a rational agent with random noise are also almost identical to those obtained by the dynamic optimization strategy.

5.3.4 Static vs. Naive relative strategies

Opponent	Honest	Noisy	Bias	Random
Mean	0.76001	0.22106	0.83137	0.06915
SD	0.22800	0.40205	0.25856	0.49118
Median	0.76049	0.29930	0.85182	0.07078
t-test	105.412	17.38688	101.678	4.452
p	<.0001	<0.0001	<0.0001	<0.0001
N > 0	1000	748	997	562
Sign test	500	248	497	62
p		<0.0001	<0.0001	<0.0001
Kolmogorov-Smirnov	0.03189	0.09081	0.041551	0.02140
p	0.015	<0.010	<0.010	>0.150

Table 6: Static vs. Naive relative strategy, Experiment by experiment comparison

These results again confirm those for the dynamic strategy. Here the optimizing strategy again performs better than the naive heuristic even in case of a purely random opponent, although the difference is smaller than for the dynamic strategy.

Summarizing the results for hypothesis **H2**, we note that this hypothesis was confirmed for rational opponents. This is not very surprising, since the optimizing strategies were specifically designed to lead to optimal performance against such an opponent. The more interesting part of **H2** concerns performance against an agent who violates the rationality assumptions made in formulating the optimizing strategies. Here the results are still encouraging. Against a moderately irrational opponent (the "Noisy" agent), the optimizing strategies still clearly outperformed the naive heuristics and even against a completely irrational opponent, the optimizing strategies still could do as well as the worse heuristic and were only marginally outperformed by the other one. This result is somewhat surprising, since one could argue that the "Naive relative" heuristic is a more elaborate approach than the "Naive fixed" heuristic. It seems that trying to be a bit smart is even worse than to be completely naive; one has to be very smart to reap the benefits of sophistication.

5.4 Results for H3

Hypothesis **H3** compared the benefits from learning for the two optimizing strategies. We expect that the dynamic strategy will gain more from learning then the one-shot strategy.

To test this hypothesis, we compare the average profit of the first ten rounds of each experiment to the average profit of rounds 21-30. The gain from learning is the difference between the two average profits. According to the hypothesis, this gain should be larger for the dynamic strategy than for the one-shot strategy. Figures 5 to 8 show the distribution of these gains for the different strategies.

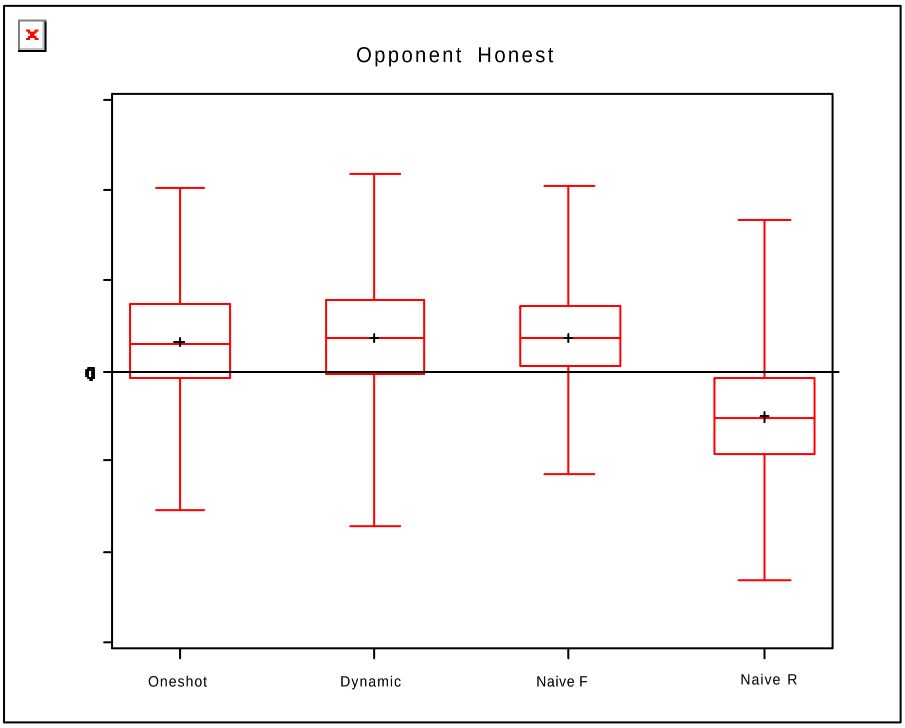

Figure 5: Gain from learning, "Honest" opponent

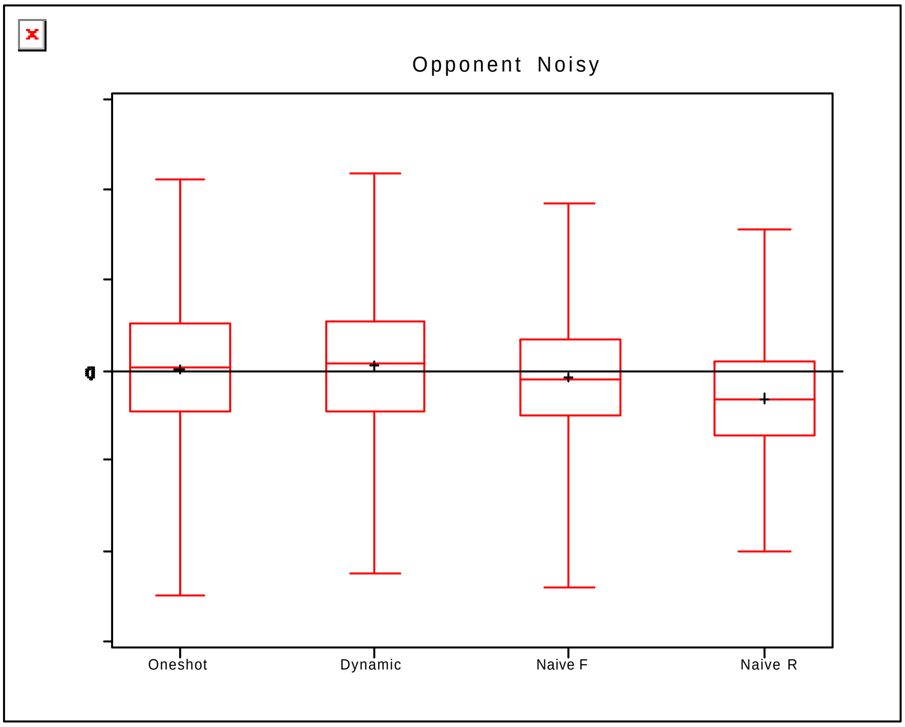

Figure 6: Gain from learning, "Noisy" opponent

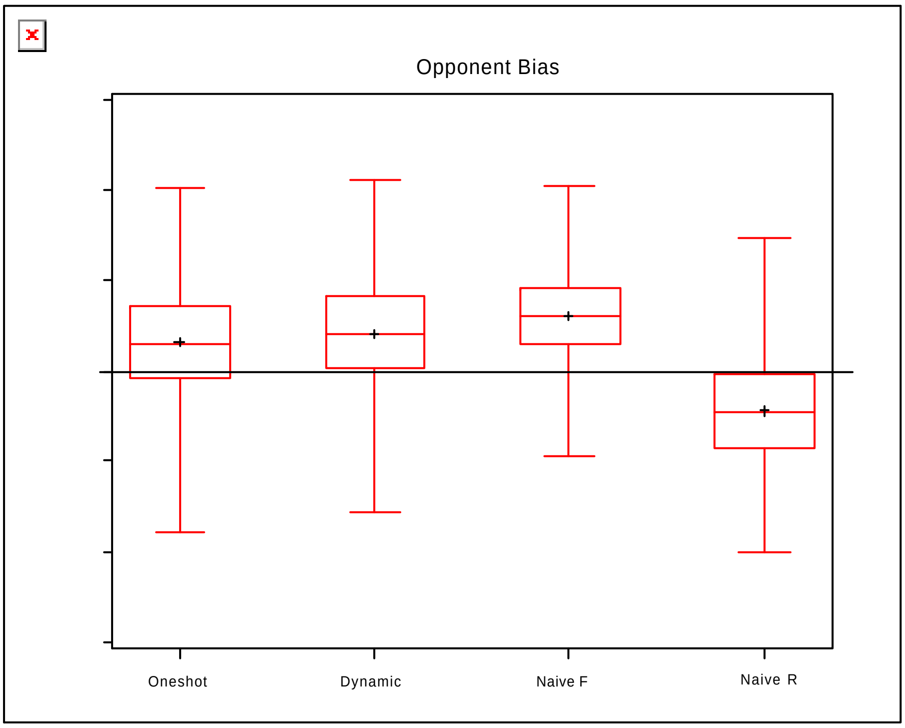

Figure 7: Gain from learning, "Biased" opponent

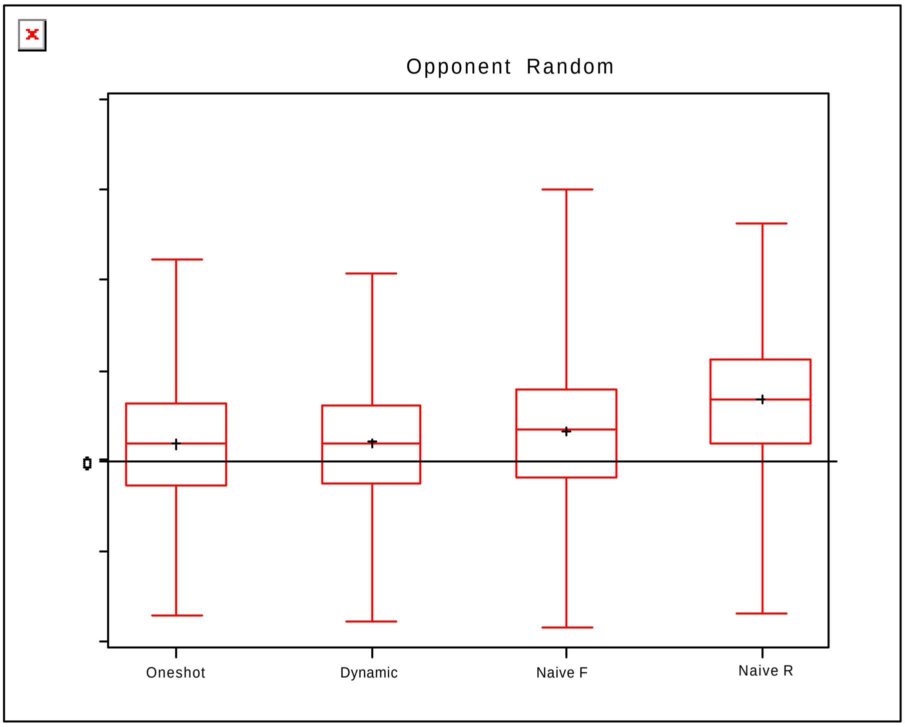

Figure 8: Gain from learning, "Random" opponent

Surprisingly, the "Random" opponent was the only one against which all strategies could gain from "learning", although this opponent did not have any preferences that could have been learned.

For the experiment-to-experiment comparison, we form the difference of the gains of the two strategies and test against the hypothesis that this difference is zero. These results are also shown in table 7:

	Opponent	Honest	Noisy	Bias	Random
Dynamic	Mean	0.37223	0.04756	0.41832	0.19743
	SD	0.58581	0.71094	0.57917	0.65334
	Median	0.35522	0.08946	0.41842	0.19606
	t-test	20.094	2.115	22.841	9.556
	p	<.0001	0.0347	<.0001	<.0001
	N > 0	731	547	766	612
	Sign test	231	47	266	112
	p	<.0001	0.0033	<.0001	<.0001
	Kolmogorov-Smirnov	0.02437	0.03538	0.01918	0.01623
	p	>0.150	<0.010	>0.150	>0.150
One shot	Mean	0.31853	0.01107	0.31299	0.18727
	SD	0.57847	0.71869	0.58187	0.65036
	Median	0.30744	0.02802	0.30213	0.19091
	t-test	17.413	0.487	17.010	9.106
	p	<.0001	0.6263	<.0001	<.0001
	N > 0	707	517	696	610
	Sign test	207	17	196	110
	p	<.0001	0.2967	<.0001	<.0001
	Kolmogorov-Smirnov	0.01505	0.02455	0.01697	0.01458
	p	>0.150	0.148	>0.150	>0.150
Difference	Mean	0.05370	0.03649	0.10534	0.01015
	SD	0.11556	0.70202	0.14548	0.70476
	Median	0.05707	0.02767	0.10264	0.00786
	t-test	14.696	1.644	22.897	0.456
	p	<.0001	0.1006	<.0001	0.6488
	N > 0	714	511	796	505
	Sign test	214	11	296	5
	p	<.0001	0.5067	<.0001	0.7760
	Kolmogorov-Smirnov	0.04224	0.05115	0.04131	0.02398
	p	<0.010	<0.010	<0.010	>0.150

Table 7: Results for hypothesis **H3**

Both strategies could indeed learn from the interaction with the opponent in a significant number of cases. The only setting in which we must reject the hypothesis that learning took place at all is the one-shot strategy facing a “Noisy” opponent.

Surprisingly, even though the dynamic strategy could significantly improve its performance against the “Noisy” opponent and the one-shot strategy could not, the difference in learning between those two strategies is not significant. This can probably be attributed to the fact that the learning effect for the dynamic strategy against this opponent is also very small, and when the positive (but insignificant) learning of the one-shot strategy is subtracted, it becomes too small to remain significant.

The difference between the strategies is also not significant for the “Random” opponent. Here both strategies could significantly improve their performance over time, although that improvement was smaller than against the more rational opponents.

Another unexpected result is that the dynamic strategy gained more from learning when faced with a “Biased” opponent than with an “Honest” opponent. However, this result might be due to the way learning is measured here. The larger difference between early and late periods for the “Biased” opponent results from the fact that during the first ten periods, the dynamic strategy obtained a much lower profit when facing a “Biased” opponent than with an “Honest” opponent. This could be an indicator that with the “Honest” opponent most learning took place already during the first few interactions. So when we measure learning by comparing average profits of the first and third set of ten interactions, a considerable amount of learning against the “Honest” opponent might be missed.

But even when we do interpret the difference between the “Honest” and “Biased” opponents only with caution, it remains obvious from these results that a deceiving, but rational opponent is probably preferable to a completely irrational one.

5.5 Results for H4

Hypothesis **H4** was similar to **H3**, but compared the knowledge gained from learning instead of the increase in profit. To measure knowledge, we look at the size of the weight interval $(\underline{w}, \overline{w})$ after learning. The more this interval can be reduced, the more has been learned about the opponent’s preferences.

To compare the speed with which the two strategies learn, we look at the status after 10 and after all 30 periods. Table 8 shows the results after 10 periods. Since it is obvious that the weight interval will be reduced and not increased during the interactions, it is not necessary to test whether learning effects are positive at all. We thus present test results only for the difference between the two strategies.

	Opponent	Honest	Noisy	Bias	Random
Dynamic	Mean	0.88068	0.93005	0.88332	0.96117
	SD	0.06636	0.09111	0.05436	0.07313
	Median	0.88557	0.99000	0.88361	0.99000
One shot	Mean	0.65411	0.74403	0.68429	0.95930
	SD	0.13029	0.20234	0.14008	0.10167
	Median	0.65705	0.71048	0.70562	0.99000
Difference	Mean	0.22657	0.18601	0.19903	0.00187
	SD	0.13266	0.20432	0.14229	0.12329
	Median	0.22099	0.20240	0.18409	0.00000
	t-test	54.007	28.789	44.232	0.478
	p	<.0001	<.0001	<.0001	0.6325
	N > 0	962	655	931	106
	Sign test	462	262.5	431	-35
	p	<.0001	<.0001	<.0001	<.0001
	Kolmogorov-Smirnov	0.04130	0.16441	0.05740	0.40504
	p	<0.010	<0.010	<0.010	<0.010

Table 8: Results for Hypothesis **H4**, 10 periods

A peculiar result is the high median value of 0.99 against a random opponent. This was the terminal value for learning, so in most cases both strategies were convinced after 10 interactions that they "knew" that opponent's true weight (which in fact does not exist at all).

Apart from the "Random" opponent, the dynamic strategy could reduce its estimate of the weight interval significantly more than the one-shot strategy during the first ten periods.

Table 9 makes the same comparison for the final situation after 30 periods.

	Opponent	Honest	Noisy	Bias	Random
Dynamic	Mean	0.92348	0.97806	0.92343	0.98985
	SD	0.03706	0.03150	0.03449	0.00381
	Median	0.91794	0.99000	0.91784	0.99000
One shot	Mean	0.75571	0.90148	0.77900	0.98966
	SD	0.10274	0.13961	0.09881	0.01114
	Median	0.74723	0.99000	0.76174	0.99000
Difference	Mean	0.16777	0.07659	0.14443	0.00020
	SD	0.10202	0.13872	0.10061	0.01177
	Median	0.17030	0.00000	0.15452	0.00000
	t-test	52.001	17.459	45.397	0.527
	p	<.0001	<.0001	<.0001	0.5981
	N > 0	943	330	911	11
	Sign test	443	123.5	411	1.5
	p	<.0001	<.0001	<.0001	0.6476
	Kolmogorov-Smirnov	0.03054	0.38556	0.05262	0.50465
	p	0.023	<0.010	<0.010	<0.010

Table 9: Results for Hypothesis **H4**, all periods

5.6 Results for H5

Hypothesis **H5** dealt with the number of defections. Especially early in the interaction, we suppose that the dynamic strategy is more likely to experiment and thus will encounter more defections than the one-shot strategy.

Table 10 shows the number of cooperations in the first ten periods.

	Opponent	Honest	Noisy	Bias	Random
Dynamic	Mean	8.40000	7.77000	8.10100	8.01600
	SD	0.78525	1.34123	0.92502	1.27411
	Median	9.00000	8.00000	8.00000	8.00000
One-shot	Mean	8.56200	8.23900	8.19100	7.99700
	SD	1.19314	1.61508	1.38001	1.24922
	Median	9.00000	9.00000	8.00000	8.00000
Difference	Mean	-0.16200	-0.46900	-0.09000	0.01900
	SD	0.86747	1.67208	1.05878	1.77589
	Median	0.00000	-1.00000	0.00000	0.00000
	t-test	-5.906	-8.870	-2.688	0.338
	p	<.0001	<.0001	0.0073	0.7352
	N > 0	160	229	184	398
	Sign test	-108.5	-141.5	-94.5	3
	p	<.0001	<.0001	<.0001	0.8588
	Kolmogorov-Smirnov	0.2659285	0.16055	0.28213	0.10895
	p	<0.010	<0.010	<0.010	<0.010

Table 10: Number of cooperations during the first 10 periods

Except for the "Random" opponent, which was programmed to defect in 20% of all interactions, the dynamic strategy indeed encountered more defections and thus less cooperation than the one-shot strategy. It seems that the dynamic strategy is particularly likely to set the incentive level too low when faced with a "Noisy" opponent, while with the more rational opponents, the results of the two strategies are more similar.

Table 11 shows the same statistics for the entire run of 30 periods.

	Opponent	Honest	Noisy	Bias	Random
Dynamic	Mean	27.02500	22.34900	26.35100	24.01500
	SD	1.33797	4.18602	1.56788	2.12869
	Median	27.00000	23.00000	27.00000	24.00000
One shot	Mean	26.60400	23.25500	25.81800	23.88800
	SD	1.95269	4.93339	2.37413	2.25000
	Median	27.00000	25.00000	26.00000	24.00000
Difference	Mean	0.42100	-0.90600	0.53300	0.12700
	SD	1.52515	5.85897	1.89697	3.10137
	Median	0.00000	-1.00000	0.00000	0.00000
	t-test	8.729	-4.890	8.885	1.295
	p	<.0001	<.0001	<.0001	0.1956
	N > 0	354	365	366	455
	Sign test	35.5	-94	40	18.5
	p	0.0055	<.0001	0.0020	0.2230
	Kolmogorov-Smirnov	0.25474	0.07355	0.24463	0.07317
	p	<0.010	<0.010	<0.010	<0.010

Table 11: Number of cooperations during all 30 periods

In the long run, the better learning of the dynamic strategy pays off and it achieves a higher rate of cooperation than the one shot strategy, at least for the more rational opponents. Only for the "Noisy" opponent, the one shot strategy achieved a higher rate of cooperation. This is consistent with table 2, which showed that for that opponent, the one shot strategy also achieved a higher profit.

6 Conclusions and Topics for Future Research

Two sets of conclusions can be drawn from our results. The first one is directly related to our initial research question concerning the efficiency of learning processes in repeated interactions. It is evident from our results concerning hypotheses **H2** and **H3** that more elaborate strategies in determining the level of incentives do lead to better results and that improving the information input to such strategies over time is indeed beneficial. This result can directly be translated into practical advice: incidents of defection, while possibly harmful at the moment, provide valuable information about the transaction partner's preferences that can be of use in later interactions. Thus testing (and thus, at a later stage, knowing) the limits of cooperation can indeed be beneficial and might even be worth taking the risk of short-run defections. As our analysis of **H5** has shown, a higher level of defections might be only a temporary phenomenon and in the long run, better information can also lead to less defections.

On the other hand, even though **H1** was also confirmed and the dynamic strategy indeed performed better than the one-shot strategy, this difference was rather small in comparison to the difference between optimizing and naive strategies. This is also in contrast to the results of **H4**, which has shown that the difference in information gained by the dynamic and one-shot strategies is substantial. Taken together, these results seem to indicate a considerably decreasing marginal benefit of information about the transaction partner's preferences. Taking into account that the dynamic strategy involves a considerably higher complexity and a computational effort which is by several orders of magnitude larger than that of the one-shot strategy, there are obvious economical limits to the level of sophistication one should use in determining the incentive levels provided to transaction partners.

These results can also be interpreted from the transaction partner's point of view. One must be aware that every decision one makes also "leaks" a certain amount of information about one's preferences to anyone who is affected by that decision or just able to observe it. This information can be exploited by the partner in future interactions.

Apart from these conclusions, which directly relate to our initial research questions, our results also have more far reaching consequences for the application of formal models to decisions involving several actors. In a hierarchical or in any other way distributed decision environment, each agent needs a model of other agents to predict their reactions to its own actions (Schneeweiss, 1999). In a realistic setting, this model is necessarily an approximation of the other agent's true decision processes. A completely accurate representation of one agent's decision process by another agent's model is only possible in asymmetric situations in which the information processing capabilities of one agent are much greater than those of the other agent. Clearly, this is not possible in a symmetric setting, because then each agent would need to be much smarter than the other agents.

Our results highlight the fact that the quality of this approximation can be of crucial importance for the success of any strategy that tries to anticipate other agents' behavior. As we have seen, adding just a moderate level of noise to the other agent's behavior can have a dramatic impact, and surprisingly, the impact of purely random disturbances might even be stronger than when the other agent deliberately tries to conceal his preferences.

This result in a way revives the discussion about game theory initiated by (Kadane/Larkey, 1982; Kadane/Larkey, 1983). They argued that the usual assumptions made in game theory about the rationality of opponents are not realistic and instead of assuming that the opponent always maximizes his utility, one should use a (subjective) probability distribution over the opponent's strategies. Our results point into a similar direction. It seems to be important to have a correct model of the opponent's decision process, whether is it a traditional, rational, utility-maximizing or an entirely different one. Alternatively, one could conclude from our results that robustness of one's own decisions with regard to the model of the opponent's decision process could be an important factor.

While our model has produced some interesting results, it still has a number of limitations. So far, it has been tested only for a limited number of parameter settings, and its validity for a wider range of parameters needs to be analyzed more thoroughly. It also deals with a rather asymmetric setting, one could also imagine a situation in which two partners provide certain incentives to each other. These questions will be addressed in future versions of the model.

References

Axelrod, R. (1984): *The Evolution of Cooperation*. Basic Books, New York.

Axelrod, R. (1987): *The Evolution of Strategies in the Iterated Prisoner's Dilemma*. In: L. Davis (Ed.): Genetic Algorithms and Simulated Annealing. Pitman, London: 32-41.

Bruderer, E.; Singh, J.V. (1996): *Organizational Evolution, Learning, and Selection: A Genetic-Algorithm-Based Model*. Academy of Management Journal 39: 1322-1349.

Eisenhardt, K.M. (1989): *Agency Theory: An Assessment and Review*. Academy of Management Review 14: 57-74.

Harris, M.; Raviv, A. (1979): *Optimal Incentive Contracts with Imperfect Information*. Journal of Economic Theory 20: 231-259.

Hoffmann, R. (2001): *The Ecology of Cooperation*. Theory and Decision 50: 101-118.

Kadane, J.B.; Larkey, P.D. (1982): *Subjective Probability and the Theory of Games*. Management Science 28: 113-120.

Kadane, J.B.; Larkey, P.D. (1983): *The Confusion of Is and Ought in Game Theoretic Contexts*. Management Science 29: 1365-1379.

Meng, C.-L.; Pakath, R. (2001): *The Iterated Prisoner's Dilemma: Early Experiences with Learning Classifier System-based Simple Agents*. Decision Support Systems 31: 379-403.

Mirrlees, J.A. (1976): *The optimal structure of incentives and authority within an organization*. The Bell Journal of Economics 7: 105-131.

Rose, D.; Willemain, T.R. (1996a): *The Principal-Agent Problem with Adaptive Players*. Computational & Mathematical Organization Theory 1: 157-182.

Rose, D.; Willemain, T.R. (1996b): *The Principal-Agent Problem with Evolutionary Learning*. Computational and Mathematical Organization Theory 2: 139-162.

Rubinstein, A. (1986): *Finite Automata Play the Repeated Prisoner's Dilemma*. Journal of Economic Theory 39: 83-96.

Schneeweiss, C. (1999): *Hierarchies in Distributed Decision Making*. Springer, Berlin.

Spremann, K. (1987): *Agent and Principal*. In: G. Bamberg and K. Spremann (Ed.): Agency Theory, Information and Incentives. Springer, Berlin: 3-37.

Vetschera, R. (2000): *Investing in Cooperative Relationships: A Simple Analytical Model*. Working Paper, OP 2000-02, UNiversity of Vienna, Department of Business Studies, Vienna.

Watanabe, Y.; Yamagishi, T. (1999): *Emergence of strategies in a selective play environment with geographic mobility: A computer simulation*. In: M. Foddy, M. Smithson, S. Schneider and M. Hogg (Ed.): Resolving social dilemmas. Psychology Press, 55-66.

Weber, M. (1987): *Decision making with incomplete information*. European Journal of Operational Research 28: 44-57.