Perceiving strategic environments

-An experimental study of strategy formation and transfer-

Andreas Nicklisch*

Max Planck Institute for Research into Economic Systems,

Strategic Interaction Group,

Kahlaische Straße 10, D-07745 Jena, Germany;
email nicklisch@mpiew-jena.mpg.de

Abstract

Within the setting of two simple two-person coordination games the formation of subjective strategies is observed experimentally. Though the structure of the game is unknown players use their actions in order to coordinate on a specific equilibrium. Strategies enable them to interpret the opponent's behavior in an appropriate way. It turns out that more informed players coordinate faster while the strategy of less informed players is more robust with respect to changes in the game structure.

Keywords: learning, strategy formation, partial reinforcement, strategy adaptation

[*APA*] 2340, 2343
[*JEL*] D83, D84

*The author would like to thank Yoella Bereby-Meyer, Gerlinde Fellner, Werner Güth and Axel Ockenfels for their support.

1 Knowledge and behavior

Human behavior is somehow mystic. Although we all behave in different ways and follow different paths, it seems that there is a specific underlying structure. One key element of the whole puzzle is learning. There is a growing number of learning models in the economic literature. Among many others, there are two very prominent models, the reinforcement learning model (RI) of Erev & Roth (1995) and the the experience-weighted attraction learning model (EWA) of Camerer & Ho (1999). Measuring learning progress as a certain action frequency, the explanatory success of those models is quite impressive (e.g., Camerer, Ho & Chong, 2003). However, there is a distinct difference between actions and strategies. Subjects do not learn actions; they choose them. But do they learn strategies? We interpret actions on the basis of strategies (e.g., the choice in a decomposed prisoner's dilemma, Pruitt, 1970). Although, very little is known about the strategy formation. The major difficulty arises from the fact that we can observe actions, but for detecting strategies we have to interpret actions. Previous studies show that players translate observations into subjective models of the game and play the game consistent to those models (Nyarko & Schotter, 2000). This study attempts to explore basic facts of strategy formation as well as the development of strategies in changing situations. In more detail, what kind of information is necessary to form the model, how long does it take to form the model and finally, what happens, if the environment changes its structure?

In a first phase of the experiment subjects face a very simple, but unknown game, in which learning a specific action does not improve the individual (monetary) performance. In the second phase of the experiment players face a different game so that players have to review strategies. Therefore, the article is organized in the following way: Section two introduces a first game, in which we observe the way of strategy formation and present the experimental results for this setting. Section three concentrates on a second game where strategies have to be reviewed. We provide some experimental evidence on knowledge transfers. Finally, section four summarizes the results and concludes the findings.

2 Strategy formation

2.1 The mutual fate control game

Experiments on detecting games test the ability of subjects to investigate the opponent's type knowing the own type of player (for example Cox, Shachat & Walker, 1997 and Oechssler & Schipper, 2003). However, subjects in our experiment play without any priors. Hence they have to develop the structure of the game themselves. The first game is known in the psychological literature as the mutual fate control game (Sidowski et al., 1956 and Kelley et al., 1962). In the mutual fate control game (hereafter MFC) each player's payoff is solely determined by the action chosen by the opponent. The MFC is symmetric in the sense that the consequences of the actions are the same for all subjects. The payoff matrix for MFC is defined as following:

Player 1, Player 2	a	b
a	$1,1$	$0,1$
b	$1,0$	$0,0$

Table 1: The payoff matrix for MFC

Under the complete information assumption (common knowledge about the game's structure) any mixed strategy vector is a weak Nash equilibrium.[1] In previous experiments under little information, i.e. players had only information on their own action and their own payoff, the mutual desirable cell (in Table 1 it is the "a,a" cell) was chosen frequently (average rate for the first ten periods 55%), but remained unstable and did not increase as one would expect by repeating the game in partner design (average rate for the last ten periods periods 65%, Mitropoulos, 2001). Mitropoulos (2001) showed that pure adaptive learning rules like RI provide little explanation for choices in the MFC. Instead, individual choices largely follow the simple strategy "win-stay/lose-change" (hereafter ws/lc). As players have very little information in the MFC, it seems reasonable to assume a simple, myopic strategy. Ws/lc describes cognitive reasoning following the learning direction theory of Selten & Stöcker (1986) in the sense that players change their choices guided by a qualitative picture of the world. Whenever they get a negative feedback (no point per round), they change the direction of their choice. Note that due to the simple game structure ws/lc is fictitious play. More precisely, fictitious play is a special parametrization of EWA, where foregone payoffs are not discounted (see, e.g., Camerer, Ho & Chong, 2003). However, we consider EWA not as a learning model, but as a strategy. The question is not, how accurate subjects follow the EWA, but how long it takes until they follow this strategy.

In our experiment we introduce players with two different levels of information on the game.

[1]Of course, the notion of equilibria, optimal strategies etc. is quite problematic in this setting since players have very limited knowledge on the game's structure. Nonetheless, we use those concepts as benchmark solutions.

- In the setting under little information players were told that they play a symmetric two-person game with two choices and a payoff of either 0 or 1. The subjects observed only their own choice and their own payoff (hereafter LI-players).
- In contrast, medium informed subjects (hereafter MI-players) observed more than LI-players. We provided the same initial information (two-person game, payoff either 0 or 1) but also the choice of the opponent in the same round.

In our setting participants first faced the MFC in a partner design for 60 periods, where both players in each group were either LI-players or MI-players. After 30 periods we changed the mutually desirable cell (see appendix for the games' payoff matrices). Players were informed that the structure of the game changed, so that they had to re-coordinate.

In principle, players face 2^8 different types of games. Eliminating the trivial games[2] and all games with a weakly dominated action for at least one player (see Mitropoulos, 2001), one ends up with four different games: the matching pennies games (where one players wins one point when he chooses the same action as the opponent, while the other wins when he chooses the opposite action of the other), a simple coordination game (where both players win a point when they both choose the same action), MFC and the fate-control/behavior-control game (hereafter FCBC; see next section). LI-players cannot disentangle those games without assuming strategies for the opponent. In contrast, MI-players can disentangle the games considering the perceived clear-cut correlation between the opponent's choice and the own payoff. Thus, they should learn much faster.

2.2 Experimental results

Experiments took place in 4 sessions in June and August 2003 in the computer laboratory of the Max Planck Institute for Research Into Economic Systems, Jena. In total, 74 undergraduate students of the University of Jena participated in groups of 24 to 26 subjects in the computerized experiment using zTree software (Fischbacher, 1999). Each subject played four phases of 30 periods, 2 times 30 periods of MFC and 2 times 30 periods of FCBC (see section three) in partner design for 120 periods. Both players in each group were either LI- or MI-players so that there were 13 groups of LI-players and 12 groups of MI-players (plus 12 groups for a control treatment, see section 3). After each phase, participants

[2]The 2 games are those, where both players get no/a payoff per period irrespectively of their choice.

were asked to describe the payoff mechanism in an open question. Subjects were informed about playing a symmetric two-person game in a pair with the same opponent for the entire duration of the experiment and a dichotomic payoff of either 0 or 1 point per period worth 0.08 euros. Maximum payoffs that could be earned was 12.10 euros including a 2.50 euros show-up fee for approximately 60 minutes. Average payments were 8.44 euros for the LI-player, 9.06 euros for MI-player.

As expected, the average rate of the mutual desirable cell choice for MI-players (see Table 2) is significantly higher than the corresponding rates for LI-players in the first phase.[3] However, comparing the performance in the second phase (periods 31 to 60) indicates that there is no significant difference between LI- and MI-players.[4] LI-players perform as good as MI-players, although analyzing the elicited knowledge on the structure of the game indicates that at least one player in each group in both phases of the MI-setting understands the payoff mechanism. In contrast, LI-players obtain more or less no structural knowledge on the game throughout the two phases.

MFC	MI	LI
avrg period $1-10$	46.6	33.8
avrg period $21-30$	65.8	46.9
avrg phase 1	63.2	41.6
avrg period $31-40$	38.3	34.6
avrg period $51-60$	59.2	60.7
avrg phase 2	50.3	47.7

Table 2: Average percentage of simultaneous choices of the desirable cell

Certainly, correlations are important elements of strategy formation. MI-players increase the speed of their strategy formation process. In the first phase, 8 of 12 groups explore only 3 or less cells of the payoff matrix. For moving only across 3 cells they need on average 4.5 periods. In the second phase, 9 of 12 groups explore 3 or less cells (which takes them on average 4.6 periods). On the other hand, LI-players cannot observe clear-cut correlations. 9 of 13 groups explore all 4 cells (second phase 8 of 13 groups). LI-players focus on the relation between their own choices and their payoff. Although the own action does not directly influence the own payoff, there is the indirect relation. In psychological literature learning on the basis of noisy signals (meaning no clear link

[3] The rate for MI-players is on average 20.9% greater ($p-value$ for the equivalence hypothesis in a two-tailed t-test < 0.001).

[4] $p-value$ for the equivalence hypothesis in a two-tailed t-test 0.086.

between individual action and its consequences) is defined as partial reinforcement. As a very robust result, partial reinforcement of actions is slower than reinforcement learning but more robust to extinction (e.g., Bereby-Meyer & Roth, 2003). But do LI-players understand this relation? And if so, does this understanding enhance their performance? Indeed, the stronger LI-players perceive a relation between own choice and own payoff, the more they earn overall. Let us define the correlation between own choice and payoffs, which subjects experience, as the perceived relation between choices and payoffs. One finds that the final payoff and this degree of relation are highly correlated. The maximum value for the correlation between final payoff and degree of relation is achieved in phase 1 (2) based on the first 7 (6) periods with $\rho = 0.935$ ($\rho = 0.80$). It seems that correlations form strategies. Clear-cut correlations are shortcuts in the sense that they lead to a fast strategy formation (MI-players). On the other hand, LI-players, who do not observe clear-cut correlations, use the entire scope of the working memory, which, according to psychological literature, is limited to a capacity of approximately 7 items (e.g., see Baddeley, 1986), to detect correlations and develop a strategy with respect to those findings.

But why do MI-players perform so badly (relative to LI-players) in the second phase? Considering that they get more structural information they should do better than LI-players. Table 3 shows the average rate of lose-stay (ls), win-change (wc) and win-stay/lose-change (ws/lc). As one can see, there is not much difference between the phases for LI-players. The level of strategy errors (ls and wc) as well as the rate of successful strategy choice (ws/lc) does not change much. But MI-players' average ws/lc-rate in phase 2 does not differ from LI-players' since the ws/lc-rate for MI-players drops due to an increase in their ls-rate. The reason lies in the shortcut function of clear-cut correlations in the process of strategy formation. Analyzing the elicited knowledge on the structure of the game, MI-players expect to play the simple coordination game in the second phase instead of the MFC. Though MI-players have in principle the ability to disentangle the possible types of games, they do not explore the games structure in the second phase (9 of 12 groups explored only 3 or less cells). Consequently, they do not revise their choice despite losing, but "wait" for the other player to choose the same action as they do. It seems that strategies of medium informed players are less robust to incremental changes in the environment due to the fastness of their formation process.

phase 1	MI	LI
avrg ls	59.3	52.5
avrg wc	8.6	17.9
avrg ws/lc	78.0	66.8
phase 2		
avrg ls	71.1	57.6
avrg wc	2.6	14.7
avrg ws/lc	69.9	68.3

Table 3: Average percentage of strategies

3 Strategy transfer

3.1 The fate control / behavior control game

What happens if the game structure changes dramatically? How persistent are strategies? It seems natural that subjects use pre-learned strategies as a starting point in a new environment. A second game is introduced, which is slightly different from the MFC. In our experiment, the same two subjects continue with two sequences of 30 periods FCBC (after playing for two times 30 periods MFC) with a change in the mutual desirable cell from one phase to the other. In a control treatment we have LI-players who play two times 30 periods of the FCBC in partner design without playing MFC before (hereafter CO-players). All subjects are informed that the game structure changed. The fate-control/behavior-control game was first introduced by Thibaut & Kelley (1959) and Rabinowitz et al. (1966). In FCBC the relationship among the players can be characterized as player 1 controlling the fate of player 2 (as in the MFC), but player 1 having a best answer to 2's pure choice. Hence, the choices of one player are no longer irrelevant for his payoff. The payoff matrix for FCBC is defined as following:

Player 1, Player 2	a	b
a	$1, 1$	$0, 1$
b	$0, 0$	$1, 0$

Table 4: The payoff matrix for FC-BC

For p indicating the probability of player 1 choosing a and q the probability of player 2 choosing a, the three vectors (p, q) denote three types of equilibria in FCBC: $(\frac{1}{2}, \frac{1}{2})$, $(1, q')$, and $(0, q'')$ for all q' and q'', respectively, such that $q' > \frac{1}{2} > q''$.

Considering the experience of LI- and MI-players in this experiment, we expect that LI-players will choose the $(0, q'')$ equilibria more often than unexperienced CO-players since the strategy ws/lc creates "cycle moves" in FCBC (see Rabinowitz et al. 1966), which excludes the mutual desirable cell (see Figure 1):[5]

$$\begin{array}{ccc} 1,1 & & 0,1 \\ & \nearrow & \downarrow \\ 0,0 & \longleftarrow & 1,0 \end{array}$$

Figure 1: Ws/lc-"cycle moves" in the FCBC

In contrast to MFC, one subject in the group of MI-players (in the setting of Table 4 it is player 1) receives no clear-cut correlation. This (missing) information should be sufficient information in order to motivate a review of the strategy ws/lc in FCBC. In the context of the adaptive learning models, MI-players have to choose a parametrization of EWA, where foregone payoffs are discounted with less than one. MI-players should accept a loss for one period in order to reach the mutual desirable cell and should choose the $(1, q')$ equilibria more often than CO-players.

However, unlike the MFC, there is no single, dominant strategy in FCBC. There is no "common ground" on which actions can be interpreted in this game. As soon as players start to develop individual different strategies, coordination becomes increasingly difficult. For that reason we expect that in the second phase of FCBC players (especially inexperienced CO-players) will decrease their performance, so that on an aggregate level data looks like coordination on the $(\frac{1}{2}, \frac{1}{2})$ equilibrium.

3.2 Experimental results

Experiments took place in the same sessions in June and August, 2003, after subjects had played MFC for two phases of 30 periods each (except for the control treatment). Players stayed in the same groups as in the MFC phases. It was announced that the structure of the game changed. CO-players earned on average 5.30 euros. CO-players received the same amount of information as LI-players during the game. Consequently, the mean payoff for each LI-player in phase one was 0.67 points per round and for each player in the control treatment 0.68 points, while MI-players earned 0.74 points. Note that following the exact "ws/lc cycle" described in the previous section would give an expected payoff of 0.66 points per period. Table 5 shows the average rates for the equilibrium $(0, q'')$.

[5] Unless one starts with this cell by pure chance.

$0, q''$	MI	LI	CO
$avrg$ period $1-10$	17.5	22.3	10.0
$avrg$ period $21-30$	14.1	19.2	9.1
$avrg$ phase 1	16.3	19.4	13.0
$avrg$ period $31-40$	40.0	30.7	35.0
$avrg$ period $51-60$	20.0	26.9	29.1
$avrg$ phase 2	28.0	27.1	31.3

Table 5: Average percentage $(0, q'')$ equilibria choices

The comparison of choice rates reveals that CO-players choose the $(0, q'')$ equilibrium significantly less (on average 6 percent) than LI-players,[6] while there is no significant difference between LI- and MI-players and between CO-players and MI-players in the first phase. In contrast, there is no significant difference between all rates in the second phase. Note that all rates increase significantly from the first phase to the second phase in a way that they reach on average a rate of approximately 25 percent, which equals the outcome of chance moves. This corresponds to the payoff of CO-players (0.49 points), which also reflects chance moves. But, also the payoff for LI- and MI-players decreases (0.61 and 0.66 points, respectively). For the second type of equilibria $(1, q')$ the data also supports our expectations. Indeed, MI-players react in the first (second) phase of FCBC to losses in 34.9% (38.9%) with a change of the action. Hence, in the parametrization of EWA foregone payoffs are discounted with 0.35 (0.39, respectively). The average rate for the first phase of choices for the $(1, q')$ equilibria is significantly higher for MI-players than for LI- and CO-players,[7] but there is no significant difference between LI- and CO-players. Table 6 reports the basic findings for rates of choices according to the $(1, q')$ equilibria.

$1, q'$	MI	LI	CO
$avrg$ period $1-10$	50.8	40.7	44.1
$avrg$ period $21-30$	75.8	55.3	66.6
$avrg$ phase 1	61.6	51.0	54.4
$avrg$ period $31-40$	29.1	30.7	19.1
$avrg$ period $51-60$	65.0	46.1	30.8
$avrg$ phase 2	48.3	41.2	26.6

Table 6: Average percentage $(1, q')$ equilibria choices

[6] $p - value = 0.02$ for the equivalence.

[7] The rate for MI-players is on average 7 / 10% higher than the rate for CO- / LI-players (both $p - values < 0.01$).

As expected, the rate for the rate for the $(1, q')$ equilibria decreases for all types of players significantly in the second phase.[8] Though, the average reduction is not the same for the different types of players (CO-players -28%, MI-players -13%, LI-players -10%). Earlier, we showed evidence that the strategy formation process for medium informed players is faster than for low informed. We observe the same for the strategy transformation process. LI-players react in the first (second) phase to losses in 53.9% (46.1%) with a change in their action (compared to 34.9%, 38.9%, respectively, for MI-players). The strategy adaptation of LI-players, which is malevolent in FCBC, is less pronounced. Our findings are in accordance to the psychological literature that says partial reinforcement is more robust against "delearning" (e.g., Bereby-Meyer & Roth, 2003). More information does not enhance players' performance per se, but makes strategies more sensitive for changes in the game structure. Consequently, their rate of reduction for the $(1, q')$ equilibria is smaller for LI- than for MI-players. On the other hand, CO-players, who start without any prior strategy to be transferred, develop individual different strategies, so that in the aggregated data, it seems that they coordinate on the third equilibrium, $(\frac{1}{2}, \frac{1}{2})$. However, we assume that with more repetitions of the FCBC also the performance of LI- and MI-players moves towards chance moves as the common strategy ws/lc erodes.

4 Discussion

The aim of our study is to explore basic facts on the process of strategy formation and development. Therefore, subjects faced a game with a unknown structure in the first part of the experiment. Traditional learning theories as RI and EWA fail in the setting of MFC since subjects do not learn in the action space, but in the strategy space. Key elements in this puzzle are perceived correlations. In the low information treatment, players form their strategy on the basis of the relation between the first six to seven actions and the outcomes. Clear-cut correlations as in the medium information treatment increase the speed of formation. On the other hand, this speed in the formation process hinders MI-players to explore the game structure entirely. Consequently, the strategy of MI-players is more "fragile" towards minor changes in the game structure. In this sense, the fastness in the formation causes a lack of exploration, while more careful explorations take more periods for the formation, but lead to more robust strategies with respect to a failure in coordination in the MFC.

[8] $p - value < 0.01$

In the second part of the experiment, we focus on the process of strategy transformation. For low informed players experience of the earlier games and the resulting strategy led to a focus on one specific type of equilibria. The process of strategy adaptation is slow compared to medium informed players, who review their strategies, which enable them to focus on different, more profitable equilibria. But as the game structure of FCBC does not promote one dominant strategy, more information and a faster adaptation process lead to a faster development of individual different strategies and, thus, to an erosion of the dominance of one type of equilibria. On an aggregate level, this looks like coordination on the mixed equilibrium.

In summary this analysis shows that more structural information and clear-cut correlations lead to a faster strategy formation. This, however, results in less robust strategies with respect to minor changes in the game structure. In case of strategy transformation, more information decreases the time needed for adaptation, but increases the variety of individual strategies, which can hinder coordination.

References

[1] Baddeley, A.D. 1986. *Working memory.* Oxford (U.K.): Clarendon Press.

[2] Bereby-Meyer, Y. & Roth, A. 2003. Partial reinforcement and the sustainability of cooperation. Mimeo.

[3] Cameron, C. & Ho, T.-H., 1999. Experience-weighted attraction learning in normal-form games. *Econometrica, 67*, 827-874.

[4] Cameron, C. , Ho, T.-H. & Chong, K. 2003. Models of thinking, learning, and teaching in games. *American Econnomic Review, 93*, 192-195.

[5] Cox, J.C., Shachat, J. & Walker, M. 1997. An experiment to evaluate Bayesian learning of Nash equilibrium play. mimeo.

[6] Fischbacher, U. 1999. zTree: A toolbox for readymade economic experiments. *Working Paper No. 21*, University of Zurich.

[7] Kelley, H.H., Thibaut, J.W., Radloff, R. & Mundy, D. 1962. The development of cooperation in the 'Minimal Social Situation'. *Psychological Monographs, 76* (19).

[8] Mitropoulos, A. 2001. Learning under minimal information: An experiment on mutual fate control. *Journal of Economic Psychology, 22*, 523-557.

[9] Nyarko, Y. & Schotter, A. 2000. An experimental study of belief learning using elicited beliefs. Mimeo.

[10] Oechssler, J. & Schipper, B. 2003. Can you guess the game you're playing? *Games and Economic Behavior, 43*, 137-152

[11] Pruitt, D.G. 1970. Reward structure of cooperation: The decomposed prisoner's dilemma game. *Journal of Personality and Social Psychology, 7*, 21-27.

[12] Rabinowitz, L., Kelley, H.H. & Rosenblatt, R.M. 1966. Effects of different types of interdependence and response conditions in the minimal social situations. *Journal of Experimental Social Psychology, 2*, 169-197.

[13] Roth, A. & Erev, I. 1995. Learning in extensive form games. *Games and Economic Behavior, 8*, 164-212.

[14] Selten, R. & Stöcker, R. 1986. End Behavior in sequences of finite prisoner's dilemma supergames. *Journal of Economic Behavior and Organization, 7*, 47-70.

[15] Sidowski, J.B., Wyckhoff, L.B. & Tabory, L. 1956. The influence of reinforcement and punishment in a minimal social situation. *Journal of Abnormal and Social Psychology, 52*, 115-119.

[16] Thibaut, J.W. & Kelley, H.H. 1959. *The social psychology in groups.* New York: Wiley.

Appendix

A. Instructions

Thank you for participating in our experiment. We kindly ask you to refrain from any public announcements and attempts to communicate directly with other participants. In case you violate this rule we have to exclude you from this experiment. If you do have any questions, please rise your hand and one of the persons who run the experiment, will come to your place and clarify your questions.

In this experiment you will be assigned randomly and anonymously in groups of two participants. In total you interact within 120 periods with the same person. All participants received the same instructions as you did. The 120 periods are partitioned into four phases of 30 periods each. In each period you (and also the other person) have to choose between two actions, a or b, and receive a payoff of either 0 or 1 ECU per period. You will not be told the mechanism which determines the payoff, but we assure you that only your and the action of the other person will decide on the payoff. In particular, there is no random process controlling the payoff. The four phases of the game are different as the mechanism which determines the payoff for you and the other person, is identical for both of you in the first phase and changes in the second, third and fourth phase. Within each phase the mechanism remains unchanged.

After each round you will receive the information on which payoff you earned in that round as well as your accumulated payoff so far. At the end of each phase you will be asked to describe the mechanism which controls your payoff. At the end of the experiment we will exchange all ECUs earned at a rate of 1 ECU = 0.08 euros and pay off the participants.

B. The pay-off matrices for LI-/MI-players

$Player2, Player1$	a	b	a	b	a	b	a	b
a	$0,0$	$1,0$	$1,0$	$0,0$	$1,0$	$0,0$	$1,0$	$1,1$
b	$0,1$	$1,1$	$1,1$	$0,1$	$0,1$	$1,1$	$0,1$	$0,0$
$Periods$	$1-30$		$31-60$		$61-90$		$91-120$	
$Game$	MFC		MFC		$FCBC$		$FCBC$	

C. The pay-off matrices for the FCBC Control group

$Player2, Player1$	a	b	a	b
a	$1,0$	$0,0$	$1,0$	$1,1$
b	$0,1$	$1,1$	$0,1$	$0,0$
$Periods$	$1-30$		$31-60$	
$Game$	$FCBC$		$FCBC$	