The Economic Journal, **115** (*January*), 200–223. © Royal Economic Society 2005. Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

THE DECISION MAKER MATTERS: INDIVIDUAL VERSUS GROUP BEHAVIOUR IN EXPERIMENTAL BEAUTY-CONTEST GAMES*

Martin G. Kocher and Matthias Sutter

Economics has devoted little attention to whether the type of decision maker matters for economic decisions. However, many important decisions like those on monetary policy or a company's business strategy are made by (small) groups rather than an individual. We compare behaviour of individuals and small groups in an experimental beauty-contest game. Our findings suggest that groups are not smarter decision makers *per se* but that they learn faster than individuals. When individuals compete against groups, the latter significantly outperform the former in terms of payoff.

In economics a decision maker is usually modelled as an individual. However, in many real-life situations the decision makers are, in fact, groups rather than individuals, such as families, boards of directors, legislatures or committees. Households and firms, the main decision-making agents in economic theory, are typically not individuals but groups of people with a joint stake in economic decisions. Similarly, political or military decisions as well as decisions on monetary policy, for instance, are often taken by groups rather than individuals.

Traditional economic theory remains silent on the influence of the type of decision maker on actual decisions. Public choice theory, of course, primarily deals with group decision making but its focus is on the aggregation of individual preferences and the rules of decision making within groups. It therefore widely neglects behavioural differences between individuals and groups, which are at the heart of social psychology, but also highly relevant for economics, given the importance of group decisions.

One straightforward explanation for the irrelevance of the decision maker in economic theory is the structure of many models. For instance, if a Nash equilibrium or a maximising choice exists, economic theory predicts the optimal strategy to be chosen, irrespective of the type of decision maker. If, however, decision-making agents do not act according to equilibrium predictions, behavioural explanations for economic decisions gain importance. Thus, a growing literature has concentrated on the impact of different characteristics of the decision maker, among which the differences between male and female decision makers have been most thoroughly studied (Eckel and Grossman, 1998, 2001;

* We would like to thank Colin Camerer, David de Meza, Rosemarie Nagel, Rupert Sausgruber, four anonymous referees, conference participants at the annual meetings of the Royal Economic Society 2001, European Economic Association 2001, Canadian Economic Association 2001, Economic Science Association 2001 and European Public Choice Society 2001, as well as seminar participants at CentER, Tilburg University for many helpful suggestions and comments. All errors are ours, of course. Financial assistance by the Royal Economic Society's Small Budget Scheme, the Gesellschaft fuer Experimentelle Wirtschaftsforschung and the Center for Experimental Economics at the University of Innsbruck (sponsored by Raiffeisen Landesbank Tirol) is gratefully acknowledged.

Andreoni and Vesterlund, 2001; Gneezy *et al.*, 2003). Only very recently, has the relevance of whether the decision maker is an individual or a small group caught considerable attention.

In this paper, we contribute to this growing literature by addressing the influence of the type of decision maker on the rationality of decision making in an interactive context. In particular, we focus on two main research questions:

(1) Are groups more rational decision makers in the sense that their decisions are closer or converge faster to the equilibrium prediction of the strategic situation?
(2) Do groups outperform individuals in terms of payoff when competing against individuals?

As our vehicle of research we use an experimental beauty-contest game, which will be introduced in Section 1, where we will also explain in detail why we think that the beauty-contest game is suitable to answer these research questions.

Our first research question is related to a growing body of economic literature examining whether groups behave and decide differently from individuals. Almost all of the relevant studies in economics use the toolbox of experimental economics to analyse the behaviour of individuals and groups under controlled laboratory conditions. An interesting general result is the fact that differences in decision making between groups and individuals can be explained neither by simple aggregation of individual preferences or choices nor by simple theories of group decision making.[1]

Bornstein and Yaniv (1998) have studied individual versus group behaviour in a standard, *one-shot* ultimatum game, where a fixed amount of money c is split between a proposer and a responder. If the responder accepts the proposer's offer x, she gets x and the proposer keeps $c - x$. However, if the responder rejects the offer, both get nothing. Bornstein and Yaniv compare two treatments, one with individuals playing against individuals and one with groups (of three subjects each) playing against groups. Their main result is that groups are more (game-theoretically) rational players than individuals by demanding more than individuals in the role of proposer and accepting relatively lower offers in the role of responder. Cox (2002) has examined individual and group decisions in an investment game (Berg *et al.*, 1995), where a trustor can send an amount $x \leq c$ to a trustee. The amount x is tripled and the trustee can send back any amount $y \leq 3x$. Final payoffs are determined by $c - x + y$ for the trustor and $3x - y + c$ for the trustee, who also receives an initial endowment c. Cox finds no significant difference between groups and individuals in amounts sent but groups return significantly smaller amounts, which indicates that groups behave more in line with payoff maximisation in the role of trustees. Cason and Mui (1997) have studied individual and group (of two subjects) behaviour in dictator games, where an individual (group) dictates the allocation of c ($2c$) dollars. Their results indicate that group choices are more other-regarding than individual choices, which is in contrast to the findings of Bornstein and Yaniv or Cox.

[1] For a survey of theories of group decision making in psychology, see Levine and Moreland (1998).

Cox and Hayne (2002) have explored decision making of groups and individuals in common value auctions, characterised by risky outcomes. Although both groups and individuals deviate from rational bidding when they have more information, groups are more affected by the 'disadvantage' of information, leading to the conclusion that groups are less rational decision makers than individuals. The studies of Bone *et al.* (1999) and Rockenbach *et al.* (2001) are somewhat related to the paper by Cox and Hayne in that they investigate group decision making under risk. Neither paper can provide evidence that group decisions comply significantly better with expected utility theory. However, Rockenbach *et al.* also find that groups take the better risks, meaning that they accumulate significantly more expected value at a significantly lower total risk.

Finally, a recent paper by Blinder and Morgan (forthcoming) studies group versus individual decision making in an urn problem and in a monetary policy experiment. Blinder and Morgan are particularly interested in whether groups are slower in decision making than individuals. Measuring the speed of decision making by the amount of information needed before reaching a decision, they find no support for the widely held belief that it takes longer for groups to reach a decision. Furthermore, group decisions are on average superior to individual decisions by about 4% in terms of payoff. Note, however, that Blinder and Morgan use a within-subjects design. Subjects act subsequently both as individuals and as members of a group (of five persons each). There is no independent comparison of individual and group decision making and there is no interaction between decision-making units.

When examining differences in the convergence to equilibrium predictions between individuals and groups – the second part of our first research question – we enter largely unexplored territory in economics. We are not aware of any contribution which addresses this question in an interactive setting.[2] However, analysing differences in learning, sophistication and strategic teaching seems to be a promising avenue to explain why group decisions cannot adequately be accounted for by simple aggregation of individual choices.

Our second research question concentrates on the effects of the type of decision maker in a direct interaction between individuals and groups. To the best of our knowledge, this question has not been assessed so far in the economics literature. Nevertheless, we think it is important to analyse from an economic point of view whether groups can outperform individuals or *vice versa* in direct interaction, because on financial markets, for instance, there are always both types of investors, single individuals and investor groups.[3] Given the resemblance between financial markets and the beauty-contest game, finding one type of decision maker

[2] After finishing this paper, Cooper and Kagel (2003) was brought to our attention. It addresses individual and group behaviour in a signalling game and shows that groups learn much faster to play strategically and converge faster to equilibrium predictions.

[3] It does not seem to be a coincidence that, as far as we know, the only paper with field data on the differences in performance of groups and individuals is on mutual fund management. Prather and Middleton (2002) find that there is no appreciable difference between the outcomes of team-managed and individually-managed funds. Note, however, that their study had considerable difficulties in distinguishing between team and individually-managed funds.

outperforming the other would have important implications for financial decision making.

The remainder of the paper is organised as follows. Section 1 describes the beauty-contest game and argues in detail why it is most suitable for addressing our research questions. Hypotheses on individual versus group behaviour are derived in Section 2. Section 3 reports on a first experimental series where individuals interact with individuals and groups interact with groups. Section 4 analyses the data of Section 3 with respect to some prominent learning theories and examines the learning characteristics of the two types of decision makers. Section 5 is devoted to our second experimental series where we let individuals compete against groups in order to address our second research question. Section 6 concludes the paper.

1. The Beauty-Contest Game

In a beauty-contest game – which was already likened to professional investment activity by Keynes (1936) – N decision makers, be it individuals or groups, simultaneously choose a real number from the interval $I \equiv [0, 100]$. The mean of all choices for round t is denoted $\bar{x}_t$. The winner is the decision maker whose number is closest to a number x^*, being defined as $p \cdot \bar{x}_t$, where $p \in (0, 1)$ is fixed for all rounds and announced at the beginning of the game. This game is dominance solvable. The process of iterated elimination of dominated strategies leads to the game's unique equilibrium at which all players choose zero.[4] The game-theoretic structure makes the beauty-contest game an ideal tool for studying how many iterations of eliminating dominated strategies a decision maker actually applies (Camerer, 1997). Textbook rationality would require the decision maker to iterate infinitely.

Previous studies (Nagel, 1995; Duffy and Nagel, 1997; Ho *et al.*, 1998; Bosch-Domènech *et al.*, 2002; Güth *et al.*, 2002) found that decision makers iterate only a few steps, i.e. the depth of reasoning is rather limited. Surveys of experimental results of beauty-contest games by Camerer (1997, 2003), Nagel (1999) or Camerer *et al.* (2003) show that behaviour is characterised by very similar patterns across very different subject pools and (session) sample sizes. First round guessing averages are typically in the 20s and 30s, and they decline in the course of repetition towards the game theoretic equilibrium, without reaching it. Choosing the game-theoretic equilibrium is rare and would never have led to winning the game.

Interestingly, the beauty-contest game, although extensively studied, has been performed exclusively with individuals as decision makers. By extending the experimental setting to a group against group treatment and a group against individuals treatment, where groups always consist of three subjects each, we are able to explore whether the decision maker matters in an interactive situation.

[4] Rational players will exclude the interval $[100p, 100]$ because any number in this interval is dominated by $100p$. If a rational player believes all others to be rational as well (by also excluding the interval $[100p, 100]$), she will exclude $[100p^2, 100]$, and so on. Choosing zero remains the only non-excluded strategy, given common knowledge of rationality.

The beauty-contest game has several attractive characteristics for studying reasoning processes, rationality and learning of different types of decision makers (individuals or groups). In contrast to most of the previous games that have been applied to analyse differences between group and individual decision making, it does not confound effects of rationality and learning, in which we are primarily interested, with the effects of social preferences, like inequality aversion, fairness or reciprocity that always play a role in simple bargaining games. Furthermore, loss or risk aversion cannot occur in the beauty-contest game.[5] Despite these advantages the beauty-contest game is still interactive (which is not the case in the experiment of Blinder and Morgan, forthcoming, for instance), has a clear economic interpretation and is very simple to explain. It is, therefore, an excellent framework to answer our research questions.

2. Individual versus Group Behaviour

2.1. *Individual versus Group Reasoning*

The widely held belief that groups reach more rational or 'better' decisions than individuals is far from being confirmed by psychological literature. In the idealised form of the group superiority argument groups are, e.g., considered to balance biases, catch errors and stimulate thoughtful work (Davis, 1992). Since the 1950s the conventional wisdom of group superiority has been challenged by numerous experiments (among other classics, see Asch, 1956), leading to the conclusion that group discussion 'can attenuate, amplify, or simply reproduce the judgmental bias of individuals' (Kerr *et al.*, 1996, p. 693). There are two prominent approaches to explain why groups might actually fail to make better decisions than individuals.

First, group conformity and self-censorship may lead to so-called 'groupthink', which can result in symptoms like stereotyping outside people, putting pressure on inside people who disagree, closed-mindedness or incomplete survey of available options and failure to assess the risks of preferred options (Sniezek, 1992; Kleindorfer *et al.*, 1993; Mullen *et al.*, 1994). Groups with designated leaders are especially prone to groupthink's adverse effects, resulting in inefficient decision making.

Second, groups tend to polarise individual attitudinal judgment in many circumstances. This effect is also known under the label 'risky shift' (Stoner, 1961). It contrasts the intuitive conjecture that groups tend to moderate extreme positions and has been found in many different settings (Davis, 1992; Kerr *et al.*, 1996). In economics it is sometimes referred to as the 'group polarisation hypothesis' (Cason and Mui, 1997). But note that there is also considerable evidence that groups may shift individual attitudes towards moderate positions (Moscovici, 1985).

Hence, the psychological literature fails to detect any *general* difference in decision making between individuals and groups. However, distinguishing between distinct types of decision tasks can reveal some *specific* differences between individuals and groups. Basically, one can distinguish between intellective tasks and judgmental tasks, where the former have a clear *ex-post* evaluation criterion for the quality of

[5] Compared to the experiments of Bone *et al.* (1999) or Rockenbach *et al.* (2001), the beauty-contest game is also not concerned with meeting or deviating from the axioms of expected utility theory.

performance but the latter lack such a criterion. Intellective tasks can be further differentiated with respect to their 'demonstrability', i.e., to which degree the knowledge of the solution to the task is shared by group members once it is voiced.

Typically, groups perform better than individuals on (non-interactive) intellective tasks (which are often tasks against nature), meaning that groups more often guess correctly than individuals, or groups are, on average, closer to the correct solution than individuals (Hastie, 1986; Levine and Moreland, 1998). This is particularly the case for decision tasks which are highly demonstrable, because in such situations 'truth wins', i.e. the group is most likely to adopt a correct solution advocated by a single group member. On judgmental tasks, there does not seem to be a systematic difference between individuals and groups in terms of performance.

The beauty-contest game exhibits characteristics of both an intellective task and a judgmental task. The judgmental aspect arises from the interactive structure of the beauty-contest game. Correct expectations on other decision makers' guesses are crucial for one's own guess. Obviously, the intellective task consists in the iterated elimination of dominated strategies and the correct adaptation of one's own guess to one's expectations of guesses submitted by other participants. The intellective part of the beauty-contest game can be considered as demonstrable, since the rationale of the game and the downward dynamics of guesses can be explained relatively easily.

Provided that groups are better in solving intellective tasks and taking into account that groups are neither systematically superior nor inferior to individuals with respect to judgmental tasks, we arrive at the following predictions:

(1) Groups apply deeper levels of reasoning than individuals, which implies that group guesses are closer to the game-theoretic equilibrium.
(2) If groups compete directly against individuals in a beauty-contest game, groups should win the contest more often than individuals.

2.2. *Individual versus Group Learning*

The beauty-contest game is repeated in our experiments in order to obtain some evidence on learning of individuals and groups. Due to the fact that previous experimental studies in economics have heavily relied on individuals as decision makers, it is a hitherto unresolved issue whether there are differences in learning between individuals and groups.

According to information load theory, based on the work by Chalos and Pickard (1985), groups have higher decision consistency and are able to process high information load in intellectual tasks better than individuals.[6] When repeating a beauty-contest game, decision makers have to process the information from the previous rounds (distribution of numbers, mean, winning number etc.) in order to form expectations about guesses in the current round. Information load theory would predict that groups are better in processing this information than

[6] See also Hill (1982, pp. 520–2) for a survey of group versus individual performance in learning tasks.

individuals. Hence, groups should be faster in realising the strategic nature of the game and, as a consequence, we expect the process of convergence towards the equilibrium to be faster for groups than for individuals.

3. Experiment I: Competition among Homogeneous Decision Makers

3.1. *Experimental Design*

In our first series of experiments individuals interact with individuals, and groups (of three subjects each) interact with groups. The parameter p was set to 2/3.[7] The experiment was run as a classroom experiment. There were two parallel sessions on May 11th (with 17 individuals and 17 groups, respectively) and on June 6th, 2000 (with 18 decision makers in each treatment).[8] In total, 140 first-year-students from undergraduate economics courses at the University of Innsbruck participated in this series of experiments, providing us with 35 observations per treatment and round. Our subjects had not been confronted with game theory in any class before participating in the experiment. Assignment to an experimental treatment (individual or group) and to a specific group in the group treatment was random.

The winner of each round in the individual treatment was paid 140 Austrian Schillings (about 10.5€), whereas winning groups were paid three times the individual amount (420 Austrian Schillings). Hence, we kept the per-subject monetary incentives constant across the individual and group treatments. In case of a tie, the amount was split equally between individuals or the respective groups. Winners were paid privately in cash at the end of the experiment, all others received nothing.

Sessions lasted at most 40 minutes and were conducted as follows. Subjects received written instructions, which were read aloud, offering subjects the opportunity to ask private questions. Sessions consisted of four rounds. In each round subjects wrote their guesses on a separate response card. These cards were collected after each round, chosen numbers were read aloud and written on an overhead projector. Then, we calculated and announced the total sum, the average, two-thirds of the average and the winning number. Once this information had been revealed, the next round was started.

Subjects in the individual treatment were isolated from each other and were not allowed to communicate. They were given up to five minutes time per round to decide on their number.

Groups in the group treatment gathered in the large Aula of the faculty. Each group sat at a separate table. The minimum distance to the next group (table) was

[7] Nagel (1995) has shown that players are systematically influenced by the parameter p. Since we are interested in differences between individuals and groups but not in the influence of p, we restrict ourselves to a single parameter choice.

[8] Like in Nagel (1995), we have chosen a relatively large number of decision makers. Convergence of choices towards the game-theoretic equilibrium is supposed to be faster in relatively large sessions, since single players have less influence on the outcome. It has been shown that a small number of decision makers can result in some erratic behaviour (Ho *et al.*, 1998), who had a session size of three and seven, respectively), which could potentially blur the picture of our comparison between individuals and groups. Note however that $N = 2$ is an exception where relatively fast convergence is observed, because the lower number always wins if $p < 1$ (Grosskopf and Nagel, 2001).

about 5 metres. Groups had five minutes time to discuss face-to-face and agree on a single number to be written down on the 'group card' for a given round.[9] Group members were requested to speak with a low voice and were strictly forbidden to speak to members of other groups.

3.2. *Experimental Results*

3.2.1. *First round behaviour*

The mean and median of first round chosen numbers in the experimental sessions are 34.9 and 32, respectively, for individuals as well as 30.8 and 29.1 for groups. As in previous studies of the beauty-contest game, first-round choices are generally far away from equilibrium. Numbers below 10 are quite infrequent (6% of groups and 12% of individuals). Dominated choices (those larger than $100p$) rarely occur. We have only one observation (out of 35) for groups and four (out of 35) observations for individuals, which indicates that this intellectual aspect of the game has been understood by the vast majority of participants.

The cumulative frequencies of guesses in round 1 are plotted in Figure 1(*a*), showing that individual and group data are rather close to each other. In fact, there is no significant difference in the mean of individuals and groups (Mann-Whitney U-test). Given this evidence, we have to reject our hypothesis that groups guess significantly smaller numbers than individuals, at least for the first round. Obviously, groups do not systematically reason deeper than individuals with regard to the iterated elimination of dominated strategies. There is weakly significantly evidence, however, that the variance of guesses is smaller for groups than for individuals ($p < 0.1$, Levene-test). This is an indication that decision making in groups tends to average out outliers in our setting.

One reason for the apparent lack of differences between groups and individuals, regarding the mean of chosen numbers, could be the fact that in the first round neither decision maker has any information on behaviour in this sort of game. Consequently, the group advantage in information processing – as claimed by information load theory – cannot arise in the first round. Furthermore, there is some evidence (Mennecke and Valacich, 1998) that groups, which are put together *ad hoc*, as is the case in our experiments, need some time to coordinate their activities and to share their understanding of the game. Five minutes of discussion might have been too short to substantiate into significantly lower guesses of groups. This latter idea is corroborated by the explanations of choices, which we asked subjects to write down on a sheet of paper after the decisions. Reasoning and elaboration of choices do not seem to be much different between individuals and groups in the first round according to these explanations.

3.2.2. *Behaviour in rounds 2, 3, and 4*

In Figure 2 we plot the transitions of guessed numbers from round t to round $t + 1$ for $t = 1$, 2, and 3. Observations below the diagonal indicate that the chosen

[9] Group cards with more than one number on it would have been invalid. Yet, there was no such case.

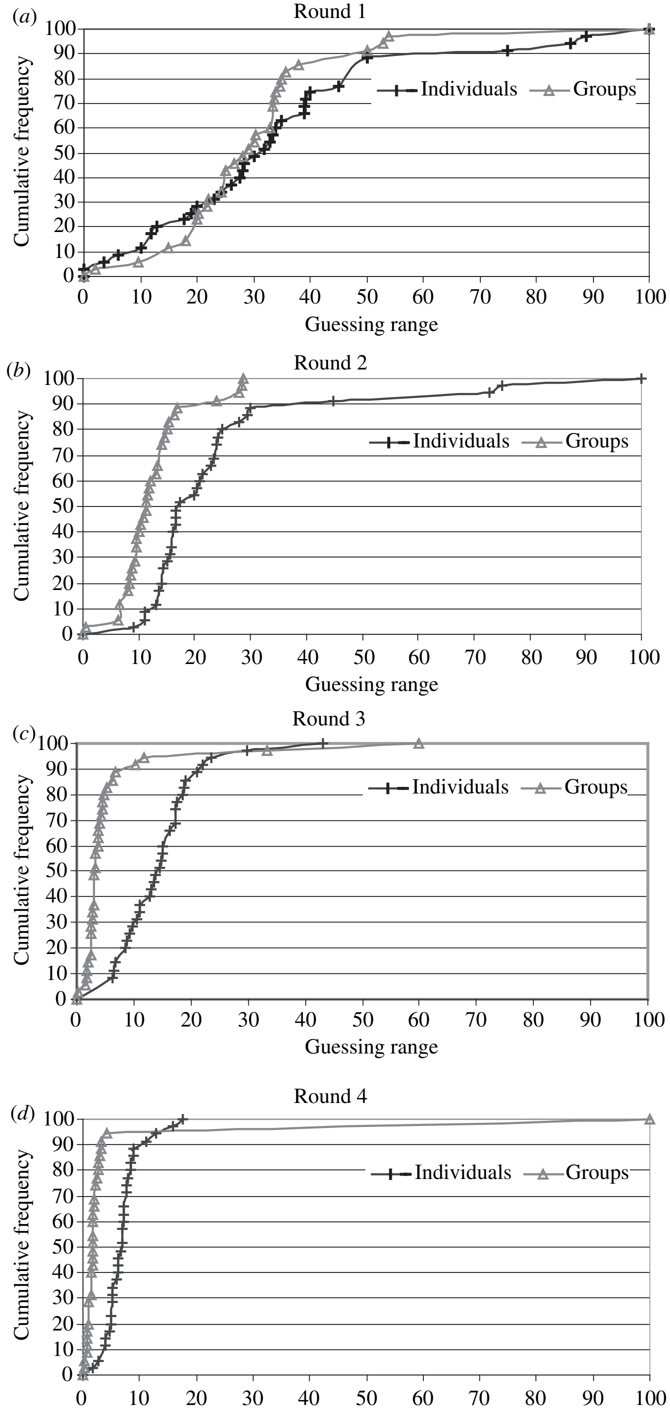

Fig. 1. *Cumulative Frequencies of Guesses*

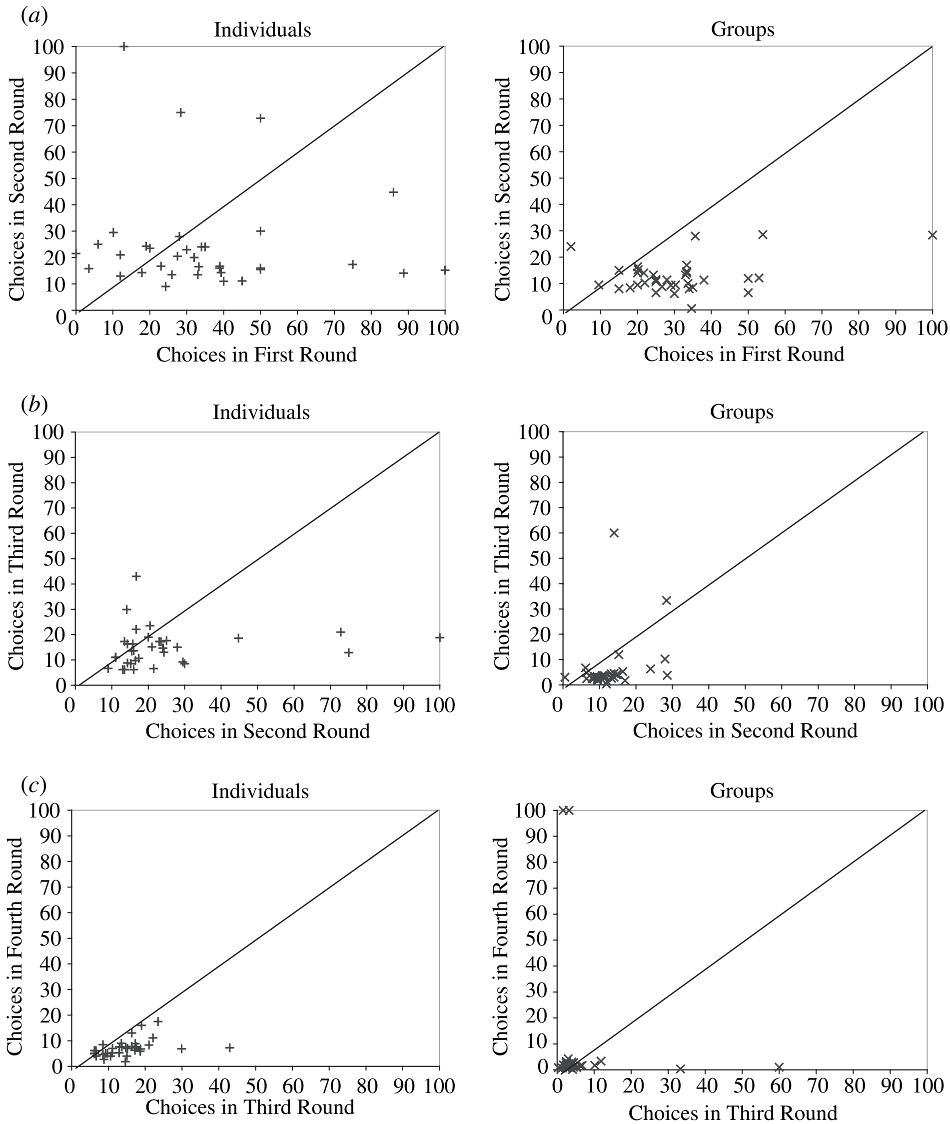

Fig. 2. *Transitions from Round t to Round t + 1*

number in round $t + 1$ is smaller than the number in round t. As can be seen, chosen numbers decline significantly over time in both the individual and the group treatment ($p < 0.025$ and $p < 0.001$ for any transition from round t to round $t + 1$ in the individual, respectively group treatment; Wilcoxon signed-ranks test). Only 18, respectively 11, out of 105 observations for each treatment lie above the diagonal.

Note that 11 out of 35 individuals, but only 1 out of 35 groups, choose a higher number in round 2 than in round 1. Individual and group behaviour is significantly different in this respect ($\chi^2 = 10.05$; $df = 1$; $p < 0.01$), indicating that groups understand the intellectual aspect of the beauty-contest game significantly faster than individuals.

Comparing both decision makers with respect to chosen numbers in rounds 2 to 4, we find that groups choose systematically lower numbers ($p < 0.001$ in any round; Mann-Whitney U-test). This is also immediately clear from looking at panels (*b*), (c) and (*d*) in Figure 1, where cumulative frequencies of group guesses are systematically to the left of individual guesses. Given the fact that there was no statistical difference in chosen numbers in the first round, the results seem to be a first indication for group learning to be faster than individual learning.

However, the mean as well as the median of chosen numbers are already slightly lower for groups than for individuals in the first round. Therefore, the reason for systematic differences in chosen numbers in rounds 2 to 4 might be due to the lower reference point (mean of round 1) in the group treatment. To check for this possibility, we re-calculated chosen numbers in rounds 2 to 4 as a fraction of the previous round's mean and tested whether fractions are different between individuals and groups. Actually, groups choose systematically lower fractions ($p < 0.001$ in any round; Mann-Whitney U-test), corroborating our hypothesis that groups converge much faster towards the equilibrium level than individuals do. This result is also confirmed when we use a single decision maker as its own control by comparing the ratio of her guesses in round $t + 1$ and round t. This resulting ratio is significantly lower for groups than for individuals in any round ($p < 0.01$; Mann-Whitney U-test).

Likewise, the percentage changes of medians are also considerably larger for groups than for individuals until round 3 (see Table 1). From round 3 to round 4 there does not seem to be a difference in the percentage changes of the median guess between individuals and groups. However, this is mainly due to the fact that medians in both group sessions were already very low in round 3 (2.63 in session 1, and 3.74 in session 2, respectively, but 9.71 and 15.74 for the individual sessions).

Table 1

Means and Medians of Rounds 1–4

	Session 1			Session 2		
Round	Mean	Median	Median(t)/Median($t-1$)	Mean	Median	Median(t)/Median($t-1$)
(*a*) *Individuals*						
1	39.66	28.40		30.32	33.00	
2	21.86	16.50	0.58	27.50	22.00	0.67
3	12.59	9.71	0.59	16.99	15.74	0.72
4	6.34	5.30	0.55	7.83	7.60	0.48
(*b*) *Groups*						
1	30.71	30.32		30.86	28.52	
2	11.39	9.51	0.31	13.94	12.35	0.43
3	6.13	2.63	0.28	6.24	3.74	0.30
4	7.56	1.70	0.65	7.18	1.74	0.47

3.3. *Depths of Reasoning*

As we have seen so far, subjects do not choose the equilibrium solution of the game, neither in the first round nor in any of the repetitions. It has already been shown by Nagel (1995), Stahl (1996), Duffy and Nagel (1997) and Ho *et al.* (1998) that a model of iterated best reply describes subjects' behaviour better than the equilibrium prediction obtained by iterated elimination of dominated strategies.[10] Classifying subjects according to the number of steps of their reasoning, in the first round we have level-0 players choosing arbitrarily in the given interval I, with the mean being 50, whereas level-1 players give a best reply to level-0 players by choosing $50p = 33{\cdot}3$. A level-2 player chooses $50p^2$ and so on. Only players with infinite steps of reasoning will choose the equilibrium number zero. In the context of repetition, step reasoning means that after round 1 level-0 players will choose, on average, the mean of the previous round, level-1 players will choose pm_{t-1}, with m_{t-1} as the previous round's mean, and so on.

Denoting player i's guess in round t by $x_{i,t}$, then player i's depth of reasoning in round t (indicating her iterated best reply) is defined as the value of d that solves $x_{i,t} = p^d m_{t-1}$. We group the continuous d values into discrete categories ($d = 0$, 1, 2, 3, 4) by defining neighbourhood intervals for guesses in round t with boundaries $[p^{d+1/2}m_{t-1}, p^{d-1/2}m_{t-1}]$, the right-hand boundary for $d = 0$ being m_{t-1}. All guesses $x_{i,t} > m_{t-1}$ are aggregated into a single category with $d < 0$. For $t = 0$, we set $m_0 = 50$, which has been shown to be a reasonable assumption in this type of beauty-contest games (Duffy and Nagel, 1997; Ho *et al.*, 1998).[11]

Table 2 reports relative frequencies of individuals' or groups' depths of reasoning. The italicised figures represent the modal values of d. For individuals we have either $d = 1$ or $d = 2$ as modal values. This is in marked contrast to groups, for which we also get $d = 3$ (round 4) and even $d = 4$ (round 3) as modal values. A Mann-Whitney U-test confirms the impression arising from Table 2 that groups apply deeper levels of reasoning (have higher ds) than individuals for rounds 2 to 4 ($p < 0.003$ in any round), but not for the first round ($p > 0.6$).

Previous experiments with individuals as decision makers have not found any significant evidence that subjects employ increasing depths of reasoning within the first three or four rounds of a beauty-contest game (Nagel, 1995; Duffy and Nagel, 1997). We can, however, provide clear evidence here that groups show increasing depths of reasoning in the transitions from round 1 to round 2 ($p < 0.01$, sign test on whether group i's depth of reasoning increased, decreased or remained unchanged from round t to round $t + 1$) and from round 2 to round 3 ($p < 0.01$, sign test). There is no further increase in depth of reasoning between rounds

[10] Recently, Camerer *et al.* (2004) have proposed a cognitive hierarchy theory of one-shot games which can explain behaviour in beauty-contest games very well.

[11] Note that the classification of subjects according to their depths of reasoning has some problems. Somebody choosing 33, for instance, can both be a level-0 player ($d = 0$), since level-0 players choose randomly from the interval I, but also a level-1 player ($d = 1$) playing best response to the average level-0 player. As a consequence, especially the categorisation of our first round data in Table 2 should be treated with caution. We are, however, quite confident that the step reasoning model is useful in explaining the data from the second round on. Written remarks of participants on how they arrived at their chosen numbers confirm that impression. We would like to thank an anonymous referee for drawing our attention to the problems of the step reasoning model.

Table 2
Relative Frequencies of Depths of Reasoning in Rounds 1–4

Depth	Round 1	Round 2	Round 3	Round 4
(*a*) Individuals*				
$d < 0$	0.11	0.11	0.11	0.03
$d = 0$	0.14	0.09	0.03	0.06
$d = 1$	*0.37*	0.26	*0.34*	0.09
$d = 2$	0.14	*0.34*	*0.34*	*0.66*
$d = 3$	0.06	0.17	0.17	0.09
$d = 4$	0.09	0.03	0.00	0.06
$d > 4$	0.09	0.00	0.00	0.03
(*b*) Groups*				
$d < 0$	0.09	0.00	0.06	0.06
$d = 0$	0.06	0.09	0.03	0.00
$d = 1$	*0.40*	0.06	0.03	0.03
$d = 2$	0.31	*0.40*	0.17	0.23
$d = 3$	0.09	0.34	0.23	*0.29*
$d = 4$	0.03	0.09	*0.37*	0.20
$d > 4$	0.03	0.03	0.11	0.20

*Data from both sessions in the respective treatment are pooled.

3 and 4. For individuals we do not find a statistically significant increase in depth of reasoning until round 3, but in the transition from round 3 to round 4 a statistically significant increase can be observed ($p = 0.043$, sign test).

Bosch-Domènech *et al.* (2002) provide evidence that once subjects reach the second or third reasoning level, they often jump all the (infinite) steps towards the Nash equilibrium: 'one, two, (three), infinity'. In our sessions we never had any individual or group choosing zero. This is a quite reasonable behaviour since zero could pay off only in case all other competitors would also choose the Nash equilibrium. In other words, the Nash solution is not trembling-hand perfect. What is usually taken for rational behaviour (choosing Nash) represents, in fact, an ignorance of other players' bounded rationality.[12] In our experiments, groups – as well as individuals – correctly anticipate that other participants do not go all the way to infinite levels of reasoning. However, groups proceed systematically further to the theoretical rationality threshold than individuals, immediately after the first round of experience with the game.

Comparing patterns of learning of individuals with the ones of groups leads to the conclusion that groups learn faster and adapt faster to a newly introduced task than individuals do. One explanation for this conclusion could be the possibility of discussing the structure and the dynamics of the beauty-contest game within groups. Yet, our result that even individuals increase their depths of reasoning in the final round might be an indication that more experience with the game can serve as a substitute for group discussion and groups' higher capacity in

[12] This is related to the false consensus phenomenon in psychology, which implies that subjects attach more importance to their own decisions or information than to other subjects' decisions or information when forming expectations on other people's decisions (Dawes, 1990). Engelmann and Strobel (2000) have shown that this false consensus effect vanishes if subjects are given appropriate monetary incentives and representative information. Given that participants in our experiment receive full information about others' decisions in previous rounds, it is not surprising that our data do not indicate a false consensus effect.

information processing. Summarising the evidence from rounds 2 to 4, we can, therefore, confirm our hypothesis that groups are smarter at responding to past evidence, meaning that they are superior in processing feedback in the course of repetition, even though they are not genuinely better in the pure deduction of the dynamics of the game in the very first round.

There are however several other possible explanations which deserve some attention. Groups may choose lower numbers from the second round on, because they expect other groups to reason deeper compared to the expectations that individuals form about other individuals. After having understood the basic dynamics of the game or the step reasoning underlying the unravelling process, i.e. before choosing a number in round 2, groups might take into account that other groups are also able to grasp the consequences of the dynamics. Hence, they decide to choose lower numbers than they would have done if they had interacted with individuals.[13] Section 5 will provide more evidence for this line of reasoning.

Camerer *et al.*'s (2001, 2002) notion of sophistication in repeated games provides another explanation for the group results in our beauty-contest experiments. Sophisticated players know that others are going to learn and, thus, they 'jump ahead', by iterating more steps of reasoning and choosing lower numbers in the beauty contest game than a best response model would probably predict. One might, for instance, assume that the proportion of sophisticated players in a population is q and that they are independently drawn and assigned to the individual and group treatments. Then in any group of three persons the chance of having at least one sophisticated player is $1-(1-q)^3$, which is larger than q for any $q > 0$. If 'truth wins' (meaning that a group as a whole acts sophisticatedly if there is at least one sophisticated person in the group), it is reasonable to expect groups to choose lower numbers than individuals. Of course, *a priori* there is no 'true' number in the beauty contest game but sophisticated group members might prevent groups from choosing dominated numbers or might track the development of means in preceding rounds in order to derive a projection of what is going to happen in the next round. In any case, if we have a proportion of players that are sophisticated and we assume that 'truth wins' within groups, the probability of sophisticated play of groups is, of course, higher than the one of individuals.

4. Econometric Estimation of Learning in the Beauty-Contest Game[14]

Previous studies have already applied learning theories to account for the dynamics of (individual) decision making behaviour in repeated beauty-contest games. Nagel (1995) finds support for a simple 'directional' learning model, which

[13] In line with this rationale, Bornstein *et al.* (2004) provide evidence that groups and individuals react differently to the type of decision maker they are interacting with. Bornstein *et al.* find in a trust game that individual behaviour is independent of the other side being an individual or a group. Contrary to that, group behaviour is strikingly contingent on the type of decision maker in the other role. Behaviour is much more in line with payoff-maximisation if a group is paired with a group than if a group is paired with an individual. Judging from the experimental questionnaires, Bornstein *et al.* conclude that groups expect other groups to be more rational and, consequently, less generous than individuals. Individuals do not show similar differences in expectations.

[14] We owe the estimation to Colin Camerer, Kuan Chong, Teck-Hua Ho and Xin Wang.

is based on Selten and Stoecker (1986). Learning direction theory predicts that subjects change unsuccessful behaviour in the direction of behaviour which would have been successful in the past. Using Nagel's (1995) data, Stahl (1996) reports that a combination of reinforcement and directional learning produces the best fit to Nagel's data. Reinforcement learning (see Roth and Erev, 1995) captures the basic insights from psychology that choices leading to good outcomes in the past are more frequently repeated in the future (law of effect: Thorndike, 1898) and that learning curves are relatively steep in early periods, but flattening out afterwards (power law of practice: Blackburn, 1936). Contrary to reinforcement learning models, where only actually chosen strategies are reinforced, belief learning models, such as weighted fictitious play (Fudenberg and Levine, 1998), use actual *and* foregone payoffs (in case a strategy different from the actual strategy would have been chosen) to enforce a strategy's probability of being chosen.

Camerer and Ho (1999) have integrated reinforcement learning models and belief-based models into a single experience-weighted attraction (EWA) learning model. As laid out in subsequent papers (Camerer *et al.*, 2001; Ho *et al.*, 2002), they find that belief-based learning and EWA fit repeated beauty-contest data substantially better than reinforcement learning or quantal-response equilibrium.

In EWA, attractions determine the probabilities of choosing different strategies through a logistic response function. Player i's strategy j and all other players' strategies in period t are denoted respectively by $s_i^j\,(t)$ and $s_{-i}\,(t)$. Player i's payoff is then $\pi_i\left[s_i^j,\, s_{-i}(t)\right]$. Strategies have initial attractions $A_i^j(0)$. To arrive at the EWA attraction updating (1) one has to define an indicator function $I(x,y)$ to be zero if $x \neq y$ and one if $x = y$:

$$A_i^j(t) = \frac{\phi N(t-1)A_i^j(t-1) + \left\{\delta + (1-\delta)I\left[s_i^j, s_i(t)\right]\right\}\pi_i\left[s_i^j, s_{-i}(t)\right]}{N(t-1)\phi(1-\kappa)+1}. \tag{1}$$

Note that the experience weight is updated according to $N(t) = N(t-1)\phi$ $(1-\kappa) + 1$ and the logit response function (2) with λ as response sensitivity maps attractions into probabilities:

$$P_i^j(t+1) = \frac{e^{\lambda A_i^j(t)}}{\sum_{k=1}^{m} e^{\lambda A_i^k(t)}}. \tag{2}$$

Table 3 reports in panel (a) the total log likelihood (in and out-of-sample) and hit rates for different learning models (as specified in Ho *et al.*, 2002) and in panel (b) the estimates of the parameters δ, ϕ and κ, which will be interpreted below. In-sample estimates are calibrated on about 70% of the decisions. The rest of the observations is used to test for out-of-sample validation, which gives an important clue of the predictability of a model. As can be seen both for the overall data (individual and group numbers) as well as for individuals and groups separately, the EWA model and the belief-based model fit best.[15]

[15] Estimation based upon the functional EWA model (fEWA) of Ho *et al.* is not reported, because it fits beauty-contest data worse than the original EWA model due to some restrictions in the parameters of the fEWA model (Camerer *et al.*, 2001; Ho *et al.*, 2002).

Table 3

Learning – Groups versus Individuals

(*a*) *Fit of learning models*

Model	*EWA*		*Belief-based*		*Reinforcement*‡	
Sample	%Hit†	LL	%Hit†	LL	%Hit†	LL
Overall ($N = 280$)	4.50	−964.00	4.50	−968.00	3.54	−1217.00
Overall in-sample (200)*	5.21	−700.06§	5.21	−697.47§	3.33	−878.70§
Overall out-of-sample (80)*	2.72	−277.13	2.72	−278.90	4.09	−343.56
Individuals (140)	6.13	−489.77	6.13	−491.52	4.11	−610.13
Individuals in-sample (100)*	7.41	−365.43§	7.41	−361.63§	4.54	−444.62§
Individuals out-of-sample (40)*	2.94	−135.86	2.94	−136.79	3.03	−170.11
Groups (140)	2.86	−473.50	2.86	−473.50	3.01	−613.90
Groups in-sample (100)*	3.00	−343.43§	3.00	338.84§	2.16	−436.87§
Groups out-of-sample (40)*	2.50	−141.59	2.50	141.57	5.14	−181.63

(*b*) *Parameter estimates for EWA and belief-based learning model* (averaged across subjects and time)

Model	*EWA*				*Belief-based*			
Parameter‖	ϕ	κ	δ	λ	ϕ	κ	δ	λ
Individuals	0.71	0.00	0.88	0.59	0.70	*0.00*	*1.00*	1.06
Groups	0.64	0.09	1.00	0.53	0.64	*0.00*	*1.00*	0.52

*Calibrated on 70% of the subjects. For both individual and group experimental data, 25 out of 35 subjects are used for calibration and 10 are used for model validation ($t = 4$).

†Number of hits counts the occasions when prob(chosen strategy) = maximum(predicted probabilities). Each count is adjusted by the number of strategies sharing the maximum.

‡Reinforcement learning with payoff variability (Erev *et al.*, 1999).

§Figures are not LL, but BIC (Bayesian Information Criterion), given by LL $-$ $(k/2)\ \log(NT)$, where k is the number of parameters, N of subjects and T of periods.

‖Fixed parameters in italic.

Panel (*b*) of Table 3 gives the parameter estimates of the EWA and belief-based models for groups and individuals separately.[16] The parameter ϕ reflects the decay of previous attractions (i.e., the probability of choosing a certain strategy) due to forgetting or deliberate ignorance of old experience in case the learning environment is changing and subjects are aware of that. A lower ϕ puts a higher decay on old experience. The parameter κ controls for the growth rate of attractions. A low κ, as it is typical for beauty-contest games, implies that attractions are weighted averages of lagged attractions and payoffs. The weight on foregone payoffs is captured by δ, where $\delta = 0$ means that foregone payoffs play no role at all, whereas $\delta = 1$ implies that foregone payoffs have the same weight as actual payoffs (as assumed in the belief-based model). Finally, λ is the response sensitivity for mapping attractions into choice probabilities. Initial attractions in the estimations are based on empirical frequencies of first round guesses. That means that relative frequencies in the actual data are used to compute first-round attractions, with the

[16] Pooling would also work well in our case, since it does not degrade the fit of any learning model significantly (likelihood ratio test). However, since we are interested in the differences between groups and individuals, we present separate models for groups and individuals.

parameter λ chosen in such a way that attractions, when multiplied by λ, exponentiated and normalised, match actual frequencies.

There are two noteworthy differences between groups and individuals notable in Table 3. Groups have a significantly lower ϕ, meaning that groups adapt faster to a changing environment, i.e. to the dynamics of chosen numbers and averages.[17] Furthermore, when considering the EWA model, which is the more general learning model than the belief-based model, the δ-estimate for groups equals one, meaning that they weigh foregone payoffs (which would have resulted from other strategies) exactly like actual payoffs when updating choice probabilities. Individuals, however, have a significantly smaller δ-estimate of 0.88, implying that individuals weigh foregone payoffs less than actual payoffs and, hence, process less information when updating choice probabilities. A smaller δ-estimate leads to a slower convergence to equilibrium in a winner-takes-all beauty-contest game.

5. Experiment II: Competition among Heterogeneous Decision Makers

Even though groups learn the characteristics of the beauty-contest game faster, and, therefore, converge faster to the equilibrium than individuals, we cannot conclude from our first series of experiments whether one type of decision maker outperforms the other in terms of payoff. In order to assess the comparative performance of different types of decision makers we have run a second series of beauty-contest experiments, where groups compete directly against individuals.

5.1. *Experimental Design*

The second experimental series was conducted on January 15th, 2001 at the University of Innsbruck. As in the first series, participants were first-year students of undergraduate economics courses, and groups each consisted of three subjects. Contrary to the first series, we had less decision makers per unit of observation in order to get more data on performance and payoffs. We ran three sessions with 20 participants each, where two individuals and one group (with three subjects) composed one independent unit of observation. Thus, we gathered a total of 12 independent observations, with 24 individuals and 12 groups.

Subjects received written instructions, which were read aloud. Again we always answered questions privately. Then, a random assignment to the type of decision maker and to specific groups took place. After the assignment procedure, individuals and groups were separated. Groups were led to separate rooms where they could discuss their decision in private and transmit their decision via a computer network, using the software z-Tree of Fischbacher (1999). Individuals remained in the lab. No decision makers could figure out with which group/individual(s) they were paired.

[17] Our groups seem to behave similar to experienced individuals. Camerer *et al.* (2002) estimate a EWA model with data from experienced subjects who participated twice in a beauty-contest game. They find that experienced players are more likely to be sophisticated, i.e. to anticipate how others learn. In the learning model they have higher decay rates of previous attractions.

The beauty-contest game was played for four rounds. In each round, a winning individual (group) received 80 (240) Austrian Schillings. After having made their decisions and before they were informed about the choices of the other decision makers in their unit, decision makers were asked to explain briefly the reasons for their decision and their reasoning process on a sheet of paper. Then, decision makers were informed of the numbers chosen by the two individuals and by the group,[18] the average of all chosen numbers, two-thirds of the average, and the winning number. After that, the next round was started.

5.2. *Experimental Results*

Table 4 summarises average chosen numbers and the number of winners for each type of decision maker. Note, first, that groups choose on average higher numbers than individuals in the first round, which is another indication for our conclusion from the first experimental series that groups are not better decision makers *per se*.[19]

In rounds 2 to 4, on average groups choose lower numbers than individuals. However, averaging over all independent observations, chosen numbers are not significantly different between groups and individuals in any of the four rounds. Since we have 12 independent units of observation, the insignificance with respect to chosen numbers might be caused by different time paths of chosen numbers in single units. Indeed, if we check for the order of chosen numbers within independent units (of two individuals and one group each), we observe the following: groups choose significantly more often the minimum number within a unit (26 times in all four rounds, individuals only 22 times; $p < 0.01$, binomial test with 2/3 probability for individuals), whereas individuals choose significantly more often the maximum number (40 times versus 8 times for groups; $p < 0.05$, binomial test).

Table 4

Average Numbers and Winners – Groups versus Individuals

	Round 1	Round 2	Round 3	Round 4	Overall	Average profit
(*a*) Overall averages						
Individuals ($N = 24$)	32.12	30.25	22.89	14.48		
Groups ($N = 12$)	40.70	26.94	16.27	10.78		
(*b*) Number of Winners*						
Individuals ($N = 24$)	9 (0.38)	4 (0.17)	6 (0.25)	7 (0.29)	26 (0.27)	86.7
Groups ($N = 12$)	3 (0.25)	8 (0.67)	6 (0.50)	5 (0.42)	22 (0.46)	146.7

*The relative frequencies of winning individuals, respectively groups, is given in parenthesis.

[18] We indicated whether a number originated from a group or from an individual (either individual 1, or individual 2).

[19] Analysing the reasons for choosing a certain number, it becomes obvious that there is no apparent difference between the reasoning processes of individuals and groups in the first round of the game. Both types of decision makers state quite frequently that they had no idea of what would happen and, thus, had just picked some random number. About the same number of individuals and groups (two or three) indicated that a reasonable average in the first round would be 50. Therefore, they had chosen a number around 33 (two-thirds of the expected average).

Interestingly, we find no difference in the first round, which resembles our results from the first experimental series with homogeneous decision makers and all our explanations from above apply again. In the first round the existing information may have either been too scarce, or groups may not have been able to process the existing information properly, possibly because group cohesion has not yet been established to a sufficient extent. Judging from the written explanations for their decisions and from the number of dominated choices (which was even higher for groups than for individuals), it seems very unlikely that groups expected individuals to choose relatively higher numbers, which could have rationalised groups' high first round choices.

Looking at winners, we find 9 (out of 24) individuals and 3 (out of 12) groups winning in the first round. If winning was randomly distributed among decision makers, we would have expected 8 individuals and 4 groups to win. Yet, in the second round we observe just the reverse. We have 8 (out of 12) winning groups, but only 4 (out of 24) winning individuals. Applying a binomial test (with 2/3 probability for individuals), groups win significantly more often than individuals ($p < 0.05$). In rounds 3 and 4, groups win more often than expected but not significantly so. This result indicates that individuals are able to catch up a bit in rounds 3 and 4 and it is in line with evidence from Section 3, where we have found individuals to increase the depths of reasoning from round 3 to 4. The crucial advantage of groups stems from their more successful choices in round 2, indicating again that groups are able to adapt much faster to the environment of the beauty-contest game and to discuss the information available after round 1.[20]

One way to look at the differences in information processing of individuals and groups is to relate a decision maker's number to the mean of the previous round. For that purpose, we take the average of the two numbers of the individuals within a unit and the number of the group and relate these figures to the previous round's mean. This yields 12 independent observations for each type of decision maker for any transition from round t to $t + 1$ (for $t = 1, 2, 3$). In the transitions from round 1 to 2, and from round 2 to 3, groups choose a significantly lower fraction of the previous round's mean than individuals ($p < 0.05$ for each transition, Mann-Whitney U-test).

However, choosing ever-lower numbers need not always be a successful strategy, of course. In 18 out of 48 winning cases, the median number within the unit was the winning number. In the remaining 30 cases, choosing the minimum number led to winning the contest. Interestingly, winning groups and winning individuals differ with respect to the order of the winning number ($\chi^2 = 2.71$, $df = 1$, $p = 0.1$). Groups' winning numbers are 17 times the minimum number within the unit but only 5 times the median number. Individuals win 13 times with the minimum number and 13 times with the median number. In 9 out of 13 cases where individuals won with the median number, groups had chosen the minimum number.

[20] Note that groups even have the opportunity to simulate the game within the group to grasp a better understanding of the dynamics of the game.

Aggregating over all four rounds, groups win 22 times (out of 48 cases; with 16 expected wins if winning was random), meaning that each group wins on average in 1.8 out of 4 rounds. Individuals, on the contrary, win 26 times (with 32 expected wins), with each individual winning in only 1.1 out of 4 rounds. The frequency of group wins is significantly larger than the average frequency of individual wins within units ($p < 0.05$; Mann-Whitney U-test). Obviously, the higher frequency of group wins transforms into higher payoffs. Members of groups earn on average 146.7 Schillings, whereas individuals receive only 86.7 Schillings, which is about 40% less than group members.

Hence, we are now able to give a clear answer to the second research question motivating our research project: Groups outperform individuals in the beauty-contest game. The explanation for this finding has already been established from the first experimental series: Groups learn faster, most probably due to a better processing of feedback information.

6. Conclusion

This paper has addressed whether the type of decision maker has any influence on actual decisions. Even though the psychological literature has dealt with this topic for quite a long time, the main focus of psychological research has typically been put on non-interactive tasks. Economics, on the contrary, is mainly interested in interactive decisions. Whereas in psychology the literature is ambiguous whether groups are better or worse in reasoning or learning than individuals, recent results from economic experiments suggest that groups might behave relatively more in line with game-theoretic predictions than individuals do. However, stronger interest in group versus individual decision making is a rather recent phenomenon in economics and results are not clear-cut so far. None of the studies we are aware of has investigated the capabilities of groups with respect to reasoning processes and learning. The beauty-contest game provides a suitable framework to compare individuals and groups with respect to reasoning and learning and with respect to relative performance. Given the large number of economic decisions taken by small groups, a thorough experimental analysis of these concepts seems to be of the utmost importance.

Our results do not lend support to the view that groups are superior decision makers *per se*. In both experimental series, first round choices do not differ significantly between groups and individuals with respect to the mean. Groups may not be able to process the existing information properly, possibly because group cohesion and a group discussion structure have not been established to a sufficient extent, yet. Although a difference in expectations may partly drive this result, it is difficult to assess its impact and there are several indications that expectations do not differ between groups and individuals in the first round of the beauty-contest game.

However, when the game is repeated, groups learn the dynamics of the game significantly faster, as information load theory would predict, and increase their depths of reasoning. When directly interacting with individuals, groups significantly outperform individuals in terms of payoff. In the course of repetition, however, individuals can partly compensate their limited information processing capacity by experience.

Given our evidence, it seems reasonable that many important and recurrent decisions in societies are entrusted to groups. In a world of specialists, it may be beneficial to use groups as decision makers instead of individuals. However, as economists we still know too little on such important questions as, for instance, which tasks should actually be entrusted to small groups as decision makers[21] or which internal structure of small groups contributes best to reach 'optimal decisions' and prevents adverse effects like groupthink or overconfidence biases. There is clearly a need for more economic research with respect to group decision making. Such research could ultimately lead to an improved economic theory of decision making that is able to account for the differences between different types of decision makers.

In particular, we think that future research in this field should address the following questions:

(1) How do groups make their decisions, i.e. how do groups aggregate the choices preferred by single group members into a single group decision?[22] Blinder and Morgan (forthcoming) discuss three simple, intuitive models of group decision making. None of them ('the whole is equal to the sum of its parts'; 'the median voter theory'; 'may the best man or woman win') gets them very far. Even though public choice theory has a lot to say on group decision making, it cannot resolve the question of differences between group and individual decision making and it has barely addressed the question of how decisions that require unanimity are actually taken within groups.

(2) Which is the optimal group size? The quality of group decisions in committees, executive boards etc. might depend on the size of a group. Certainly, an increasing group size not only increases human capital but also transaction costs for reaching decisions and the mere costs of decision making since more heads are more expensive. Research on public good provision has not found a significant effect of group size on the provision of public goods (Isaac and Walker, 1994) but this might have been caused by the fact that group members had to make an individual decision whether to contribute. Therefore, it would be interesting to see whether group size has a significant influence in unitary groups, where group members have to agree on a single joint decision. Preliminary evidence by Sutter (2004) suggests that larger groups are better in information processing than smaller groups and, therefore, more successful in a beauty-contest game.

[21] A recent paper by Cox and Hayne (2002), for instance, shows that groups may even do worse than individuals in common value auctions. However, the issue does not seem to be resolved yet, because the prevailing information conditions seem to be crucial for the judgment whether groups are more or less rational than individuals in their context.

[22] Like one of the referees, who asked for more attention to be devoted to this question, we think that an important agenda for future research on group decision-making is to open up the black box of the internal decision-making processes in groups and to gain insights into the dynamics leading to a single group decision. A first step in this direction is to use video-experiments like Bosman *et al.* (2002) have applied in a bargaining game.

University of Innsbruck
University of Innsbruck and Max Planck Institute for Research into Economic Systems Jena

Date of receipt of first submission: August 2002
Date of receipt of final typescript: February 2004

References

Andreoni, J. and Vesterlund, L. (2001). 'Which is the fair sex? Gender differences in altruism', *Quarterly Journal of Economics*, vol. 116, pp. 293–312.

Asch, S. E. (1956). 'Studies of independence and conformity: a minority of one against a unanimous majority', *Psychological Monographs*, vol. 70 (416), all pages.

Berg, J., Dickhaut, J. and McCabe, K. (1995). 'Trust, reciprocity, and social history', *Games and Economic Behaviour*, vol. 10, pp. 122–42.

Blackburn, J. M. (1936). 'Acquisition of skill: an analysis of learning curves', IHRB Report No. 73.

Blinder, A. S. and Morgan, J. (forthcoming). 'Are two heads better than one? An experimental analysis of group vs. individual decision making', *Journal of Money, Credit and Banking*, (forthcoming).

Bone, J., Hey, J. and Suckling, J. (1999). 'Are groups more (or less) consistent than individuals?', *Journal of Risk and Uncertainty*, vol. 8, pp. 63–81.

Bornstein, G. and Yaniv, I. (1998). 'Individual and group behaviour in the ultimatum game: are groups more 'rational' players?', *Experimental Economics*, vol. 1, pp. 101–8.

Bornstein, G., Kocher, M., Kugler, T. and Sutter, M. (2004). 'Whom do you trust more in bargaining – groups or individuals?', University of Innsbruck, mimeo.

Bosch-Domènech, A., García-Montalvo, J., Nagel, R. and Satorra, A. (2002). 'One, two, (three), infinity, …: newspaper and lab beauty-contest experiments', *American Economic Review*, vol. 92, pp. 1687–701.

Bosman, R., Hennig-Schmidt, H. and van Winden, F. (2002). 'Exploring group behaviour in a power-to-take video experiment', University of Bonn, Working Paper.

Camerer, C. F. (1997). 'Progress in behavioural game theory', *Journal of Economic Perspectives*, vol. 11, pp. 167–88.

Camerer, C. F. (2003). *Behavioural Game Theory. Experiments in Strategic Interaction*, Princeton: Princeton University Press.

Camerer, C. F. and Ho, T.-H. (1999). 'Experience-weighted attraction learning in normal form games', *Econometrica*, vol. 67, pp. 827–74.

Camerer, C. F., Ho, T.-H. and Chong, J.-K. (2001). 'Behavioural game theory: thinking, learning, and teaching', paper presented at the Nobel symposium, November 14.

Camerer, C. F., Ho, T.-H. and Chong, J.-K. (2002). 'Sophisticated EWA learning and strategic teaching in repeated games', *Journal of Economic Theory*, vol. 104, pp. 137–88.

Camerer, C. F., Ho, T.-H. and Chong, J.-K. (2003). 'Models of thinking, learning, and teaching in games', *American Economic Review, Papers and Proceedings*, vol. 93, pp. 192–5.

Camerer, C. F., Ho, T.-H. and Chong, J.-K. (2004). 'A cognitive hierarchy theory of one-shot games and experimental analysis', *Quarterly Journal of Economics*, vol. 119, pp. 861–98.

Cason, T. N. and Mui, V.-L. (1997). 'A laboratory study of group polarisation in the team dictator game', Economic Journal, vol. 107, pp. 1465–83.

Chalos, P. and Pickard, S. (1985). 'Information choice and cue use: an experiment in group information processing', *Journal of Applied Psychology*, vol. 70, pp. 634–41.

Cooper, D. J. and Kagel, J. H. (2003). 'Are two heads better than one? Team versus individual play in signalling games', mimeo, Case Western Reserve University.

Cox, J. C. (2002). 'Trust, reciprocity, and other-regarding preferences: groups vs. individuals and males vs. females', in (R. Zwick, and A. Rapoport, eds.), *Advances in Experimental Business Research*, pp. 331–50, Dordrecht: Kluwer Academic Publishers.

Cox, J. C. and Hayne, S. C. (2002). 'Barking up the right tree: are small groups rational agents?', University of Arizona, Working Paper.

Davis, J. H. (1992). 'Some compelling intuitions about group consensus decisions, theoretical and empirical research, and interpersonal aggregation phenomena: selected examples, 1950–1990', *Organizational Behaviour and Human Decision Processes*, vol. 52, pp. 3–38.

Dawes, R. M. (1990). 'The potential nonfalsity of the false consensus effect', in (R. M. Hogarth, ed.), *Insights in Decision Making: A Tribute to Hillel J. Einhorn*, pp. 179–99. Chicago and London: University of Chicago Press.

Duffy, J. and Nagel, R. (1997). 'On the robustness of behaviour in experimental ''beauty contest'' games', Economic Journal, vol. 107, pp. 1684–700.

Eckel, C. C. and Grossman, P. J. (1998). 'Are women less selfish than men? Evidence from dictator experiments', ECONOMIC JOURNAL, vol. 108, pp. 726–35.

Eckel, C. C. and Grossman, P. J. (2001). 'Chivalry and solidarity in ultimatum games', *Economic Inquiry*, vol. 39, pp. 171–88.

Engelmann, D. and Strobel, M. (2000). 'The false consensus effect disappears if representative information and monetary incentives are given', *Experimental Economics*, vol. 3, pp. 241–60.

Erev, I., Bereby-Meyer, Y. and Roth, A. E. (1999). 'The effects of adding a constant to all payoffs: experimental investigation, and a reinforcement learning model with self-adjusting speed of learning', *Journal of Economic Behaviour and Organization*, vol. 39, pp. 111–28.

Fischbacher, U. (1999). 'Z-tree: Zurich toolbox for readymade economic experiments', University of Zurich, Working Paper No. 21, Institute for Empirical Research in Economics.

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

Gneezy, U., Niederle, M. and Rustichini, A. (2003). 'Performance in competitive environments: gender differences', *Quarterly Journal of Economics*, vol. 118, pp. 1049–74.

Grosskopf, B. and Nagel, R. (2001). 'Rational reasoning or adaptive behaviour? Evidence from two-person beauty contest games', Harvard University, Harvard NOM Research Paper No. 01–09.

Güth, W., Kocher, M. and Sutter, M. (2002). 'Experimental 'beauty contests' with homogeneous and heterogeneous players and with interior and boundary equilibria', *Economics Letters*, vol. 74, pp. 219–28.

Hastie, R. (1986). 'Review essay: experimental evidence on group accuracy', in (G. Owen and B. Grofman, eds.), *Information Pooling and Group Decision Making*, pp. 129–57, Westport, CT: JAI Press.

Hill, G. W. (1982). 'Group versus individual performance: ''Are $N+1$ heads better than one?''', *Psychological Bulletin*, vol. 91, pp. 517–39.

Ho, T.-H., Camerer, C. F. and Chong, J.-K. (2002). 'Functional EWA: a one-parameter theory of learning in games', University of California Berkeley, Working Paper.

Ho, T.-H., Camerer, C. F. and Weigelt, K. (1998). 'Iterated dominance and iterated best response in experimental ''p-beauty contests''', *American Economic Review*, vol. 88, pp. 947–69.

Isaac, R. M. and Walker, J. M. (1994). 'Group size and the voluntary provision of public goods: experimental evidence utilizing large groups', *Journal of Public Economics*, vol. 54, pp. 1–36.

Kerr, N. L., MacCoun, R. J. and Kramer, G. P. (1996). 'Bias in judgment: comparing individuals and groups', *Psychological Review*, vol. 103, pp. 687–719.

Keynes, J. M. (1936). *The General Theory of Interest, Employment and Money*, London: Macmillan.

Kleindorfer, P. R., Kunreuther, H. C. and Schoemaker, P. J. H. (1993). *Decision Science. An Integrative Perspective*, New York: Cambridge University Press.

Levine, J. M. and Moreland, R. L. (1998). 'Small groups', in (D. T. Gilbert *et al.*, eds.), *The Handbook of Social Psychology*, 4th edition, vol. 2, pp. 415–69.

Mennecke, B. E. and Valacich, J. S. (1998). 'Information is what you make of it: the influence of group history and computer support on information sharing, decision quality, and member perceptions', *Journal of Management Information Systems*, vol. 15, pp. 173–97.

Moscovici, S. (1985). 'Social influence and conformity', in (G. Lindzey, and E. Aronson, eds.), *Handbook of Social Psychology*, 3rd edition, vol. 2, pp. 347–412.

Mullen, B., Anthony, T., Salas, E. and Driskell, J. E. (1994). 'Group cohesiveness and quality of decision making: an integration of tests of the Groupthink hypothesis', *Small Group Research*, vol. 25, pp. 189–204.

Nagel, R. (1995). 'Unravelling in guessing games: an experimental study', *American Economic Review*, vol. 85, pp. 1313–26.

Nagel, R. (1999). 'A survey of experimental beauty contest games: bounded rationality and learning', in (E. Budescu *et al.*, eds.), *Games and Human Behaviour: Essays in Honour of Amnon Rapoport*, pp. 105–42. New Jersey: Lawrence Erlbaum.

Prather, L. J. and Middleton K. L. (2002). 'Are $N+1$ heads better than one? The case of mutual fund managers', *Journal of Economic Behaviour and Organization*, vol. 47, pp. 103–20.

Rockenbach, B., Sadrieh, A. and Matauschek, B. (2001). 'Teams take the better risks', University of Erfurt, Working Paper.

Roth, A. E. and Erev, I. (1995). 'Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term', *Games and Economic Behaviour*, vol. 8, pp. 164–212.

Selten, R. and Stoecker, R. (1986). 'End behaviour in sequences of finite prisoner's dilemma supergames: a learning theory approach', *Journal of Economic Behaviour and Organization*, vol. 7, pp. 47–70.

Sniezek, J. A. (1992). 'Groups under uncertainty: an examination of confidence in group decision making', *Organizational Behaviour and Human Decision Processes*, vol. 52, pp. 124–55.

Stahl, D. O. (1996). 'Boundedly rational rule learning in a guessing game', *Games and Economic Behaviour*, vol. 16, pp. 303–30.

Stoner, J. A. F. (1961). 'A comparison of individuals and group decisions involving risk', Cambridge, MA: Unpublished Master's Thesis (MIT).

Sutter, M. (2004). 'Are four heads better than two? An experimental beauty contest game with teams of different size', Max Planck Institute for Research into Economic Systems, Working Paper.

Thorndike, E. L. (1898). 'Animal intelligence: an experimental study of the associative processes in animals', *Psychological Review*, vol. 2, Monograph Supplement, no. 8.