Journal of Risk and Uncertainty, 8, 63–81 (1999)

Are Groups More (or Less) Consistent Than Individuals?*

JOHN BONE jdb1@york.ac.uk
Department of Economics and Related Studies, University of York, York YO1 5DD, UK

JOHN HEY jdh1@york.ac.uk
Department of Economics and Related Studies, University of York, York YO1 5DD, UK

JOHN SUCKLING jrs1@york.ac.uk
Department of Economics and Related Studies, University of York, York YO1 5DD, UK

Abstract

There is now overwhelming experimental evidence that individuals systematically violate the axioms of Expected Utility theory. In reality, however, many economic decisions are taken by, or on behalf of, groups whose members have a joint stake in those decisions. This paper reports on an experiment in which pairs of individuals are tested for Common-Ratio inconsistencies. We find that the agreed choices of subject-pairs follow a pattern of inconsistency very close to that of individuals' choices. We also look for evidence that group participation increases the consistency of the individuals themselves. With one solitary exception, we find none.

Key words: Expected Utility theory, Common-Ratio Effect, experiments, groups

JEL Classification: D81, C91, C92

1. Introduction

The experimental evidence on the violation of Expected Utility (EU) theory, by individuals in properly motivated experiments, is now overwhelming.[1] Of course, there is always room for debate about whether such experimental results are representative of behaviour in real economic settings. The central issue here is the extent to which experimental conditions reproduce the essential features of the real environment in which agents operate. But one related issue that is largely overlooked, in the context of these results, concerns the agents themselves. In reality, many economic decisions are taken by, or at least on behalf of, groups rather than isolated individuals.

This paper reports on an experimental investigation into collective decision-making under uncertainty. We asked pairs of individuals to agree choices between financial prospects in which they had a joint stake. Our primary aim was to see whether, in experimental conditions, groups violate EU theory in the same manner, and to the same extent, as individuals appear to do.

But we also had a secondary, parallel aim. One obvious way in which groups differ from individuals is that group members are able, and indeed have an incentive, to argue and discuss with each other. Whatever influence this may have on the group's agreed choices, there are some grounds for supposing that it might increase the EU-consistency of the individual members concerned. Our experiment provided an opportunity of testing this ideas as well.

The specific form of EU-violation which we examined was the Common-Ratio Effect. In Section 2 we define and explain this. Section 3 then elaborates our main empirical concerns, and discusses some of the conceptual background. The experiment itself is detailed in Section 4, and the results examined in Section 5. Section 6 provides a summary of the results, and some concluding observations.

2. The Common-Ratio Effect

The Common-Ratio Effect is a simple form of violation of EU theory. A typical construction involves a positive scalar $\alpha < 1$, three monetary (£) prizes with values $X > Y > Z$, and m distinct non-zero probability values $p_1 > p_2 > , \ldots, > p_m$. These define a set of m prospect-pairs {**R*i***, **S*i***}, each pair corresponding to some p_i, and comprising one "Risky" and one "Safe" prospect:

R*i* £X with probability αp_i; £Z otherwise
S*i* £Y with probability p_i; £Z otherwise

By construction, α is the common ratio of the winning probabilities in each prospect-pair. We describe as a *CR-set* any set of prospect-pairs {{**R1**,**S1**}, ...,(**R*m***,**S*m***}} with this structure. In our experiment each CR-set comprised three pairs ($m = 3$), and is therefore called a *CR-triple*.

The standard EU axioms require that, within a given CR-set, the preference between each prospect-pair be the same. That is, **R1** is preferred to **S1** only if **R*i*** is preferred to **S*i*** for every pair {**R*i***,**S*i***} in that CR-set, and likewise **S1** is preferred to **R1** only if **S*i*** is preferred to **R*i***.

Firstly, the Reduction axiom implies that {**R*i***,**S*i***} may be equivalently described as:

R*i* **R1** with probability p_i/p_1; £Z otherwise
S*i* **S1** with probability p_i/p_1; £Z otherwise

These two (compound) prospects differ only in the first outcome, which occurs in each case with probability p_i/p_1. So, secondly, the Independence axiom requires that the preference over these two prospects {**R*i***,**S*i***} correspond to that over {**R1**,**S1**}.

Nevertheless, in experimental conditions individuals are often found to have preferences which differ across pairs in a given CR-set. Specifically, there is an observed tendency to prefer **S*i*** for high values of p_i and **R*i*** for low values of p_i. This is known as the *Common-Ratio Effect*.

3. Two empirical questions

In this study, we are interested in two broad empirical questions:

1 How does the EU-consistency of groups compare with that of individuals?
2 Do individuals learn, through group interaction, to be more EU-consistent?

Economic theory customarily classifies the main decision-making agents as *households* and *firms*. In reality, most households and firms are not single individuals but groups of people with a joint stake in economic decisions. This is in addition to groups, such as investment or gambling syndicates, formed explicitly to undertake uncertain financial ventures. Although groups may delegate decision-making to individual representatives, their choices must ultimately be agreed, in some way, by group members. So it is of empirical interest to know whether choices agreed by groups are more, or less, EU-consistent than are those made by individuals.

This is also a theoretical question, of course. In theory, the simplest situation is where group members are individually EU-consistent, and have identical preferences over the prospects in which they share a joint stake. Then, so long as their agreed choices are efficient, these too will conform to EU theory. But group members, however EU-consistent as individuals, will in general have some divergent preferences. Agreement will then require compromises by individual members, and group choices will depend on the bargaining principles by which such compromises are reached.

Economists commonly view such principles in one of two ways.[2] The *strategic* approach is to analyse the bargaining process as a non-cooperative game. On this view, there is no reason to suppose that agreed decisions will be in any way collectively rational. More promising, in this sense, is the *axiomatic* approach of Nash and others, which explicitly applies axioms of rationality to group choices. But typically these are just the basic axioms (e.g., Transitivity) of sure choices. Indeed, we know from Harsanyi (1955) that the only arbitration principle guaranteed to produce fully EU-consistent choices is sum-utilitarianism, i.e., maximises a linear function of individual vNM utilities.

In the context of financial prospects in which group members have a joint stake, there is a specific reason why group choices may be EU-inconsistent. This relates to the (re-)allocation of joint prizes between group members, and is explained in Section 4.

So there seems to be a clear theoretical presumption that groups will be EU-inconsistent. But then empirically, of course, so too are individuals. Given this, the question remains as to whether groups will be more, or less, inconsistent than are individuals. Theory does not seem to help here, most obviously because it does not give any guidance as to the degree of group inconsistency, but also because it generally pre-supposes individuals to be consistent in the first place. Almost all theoretical work on bargaining, whether strategic or axiomatic, assumes each group member to have a cardinal utility function defined over final (sure) outcomes, and Harsanyi's theorem explicitly assumes individuals to be EU-consistent. We therefore view this as an open empirical question, of which, we distinguish two aspects. The first is the relative incidence of EO-inconsistency, as between groups and individuals. The second is the type of inconsistency involved.

In order to investigate this question, we constructed an experiment in which pairs of subjects were asked to agree choices from CR-triples, given a joint stake in the prospective prizes. Of courses, for a controlled comparison of groups and individuals, we had separately to measure the EU-consistency of our subjects as individuals. One potential ambiguity here is that the group encounter itself may have some influence on members' individual preferences. Indeed, it is not implausible that, through group discussion, members might become more EU-consistent as individuals.

The idea that people might learn to be more EU-consistent is sometimes associated with Savage, following his remarks on the Allais paradox (Savage, 1972, pp. 101–104). Savage was here defending EU theory on normative grounds. But since his position entails that EU-violations are errors of some kind, it leads naturally to a positive hypothesis that individuals can recognise, and rectify, any such errors.

Even if true, this leaves open the question of what conditions might facilitate this. Discussion, however, would seem to be a promising candidate, especially where the individuals concerned each have a mutual interest in its resolution, as in our experiment. The most obvious scenario involves a teacher/learner relationship, where one partner who is already EU-consistent persuades the other, possibly along the lines of the argument in Section 2. An alternative scenario has both partners starting from a position of EU-inconsistency, but then discovering their mutual error through constructive discussion. Perhaps this would be most likely, if at all, in cases where their individual preferences are (initially) in conflict, this being the trigger for discussion.

We decided, therefore, to elicit our subjects' individual preferences both before and after their group encounters. This would give us not only two standards of comparison for the group choices, but also information relevant to the Savage hypothesis. Evidence of increased EU-consistency would provide suggestive support for the hypothesis. Of course, the lack of such evidence would not falsify it; in this sense our experiment did not constitute a scientific test.

So the experiment comprised three consecutive stages, in each of which subjects were given the same set of choice problems. It might be thought that this repetition could itself induce EU-learning, independently of any group discussion. It is certainly plausible that repetition can induce adaptive learning, where the experience of (suboptimal) outcomes prompts subjects to modify their behaviour. Indeed, there have been many experimental investigations of this idea.[3] But, as described in the following section, our experiment provided no outcomes until all choices had been completed. So adaptive learning of this type could not have occurred.[4] Of course, it may be that repeated experience merely of the choice task, regardless of outcome, provides some opportunity for learning. This idea seems rather more speculative, but in any case the above 'model' of learning through discussion provides a rudimentary framework for the separate measurement of any such effects. We return to this in Section 5.

4. The experiment

Our subjects registered choices from four CR-triples, giving a total of 12 prospect-pairs in all. Following the notation given in section 2, let $\{\mathbf{R}i_t, \mathbf{S}i_t\}$ denote the ith prospect-pair (i = 1, 2, 3) of triple t (t = 1, 2, 3, 4).

The triples all had in common the probability parameters:

$$\alpha = 0.5;\ p_1 = 1,\ p_2 = 0.5,\ p_3 = 0.2$$

and were distinguished only by their prospective prizes, as shown in Table 1.

The final column of Table 1 gives the ratio (ρ) of the expected monetary values of $\mathbf{R}i_t$ and $\mathbf{S}i_t$ in a given triple. It is a measure of the attractiveness of $\mathbf{R}i_t$ relative to the corresponding $\mathbf{S}i_t$. Of course, it is not the definitive measure; a risk-averse individual may nevertheless prefer $\mathbf{S}i_t$ to $\mathbf{R}i_t$ in any of these triples, and will do so in the case of triple 1. But (ρ) nevertheless provides a convenient and natural ordering of the four triples.

The experiment consisted of three consecutive stages. At each stage, the twelve prospect-pairs were given to each subject, each time in an independently randomised order. Stage 2 differed from stages 1 and 3 in that (immediately prior to it) the subjects were paired-up, at random. Each partnership was given the twelve prospect-pairs, as above except that the prizes were all doubled, and asked to agree a choice from each.

Table 1. Prospective (£) prizes for the four triples

	X	Y	Z	ρ
triple 1	30	15	0	1
triple 2	35	15	0	1.17
triple 3	30	12	0	1.25
triple 4	35	12	5	1.67

At stage 2, therefore, each prospect-pair took the form:

$\mathbf{R}i_t$ £2X with probability αp_i; £2Z otherwise
$\mathbf{S}i_t$ £2Y with probability p_i; £2Z otherwise

If these prizes were divided equally, then for each partner the prospect-pairs would correspond exactly to those in stages 1 and 3. However, we did not attempt to impose any specific division of the prizes, instead leaving this for the partners to decide for themselves. In fact, we required them explicitly to register not only their agreed choice of prospect, from each prospect-pair, but also their agreed division of its prospective prizes.

To illustrate why we did this, consider triple 2. Given an equal division of prizes, the second prospect-pair of this triple can be depicted as:

$\mathbf{R2}_2$	J	K
0.25	£35	£35
0.75	£0	£0
v	13.65	3.03

$\mathbf{S2}_2$	J	K
0.5	£15	£15
0.5	£0	£0
v	13.03	3.07

where the partners are denoted {J,K}, and their expected utility values (v) are based on the illustrative vNM utility functions:

$$u_j = (10 + x_j)^{0.9} \qquad u_k = (10 + x_k)^{0.4}$$

In this example, K is the more risk-averse of the two individuals, as reflected in the fact that he prefers the “safe” prospect $\mathbf{S2}_2$ while J prefer $\mathbf{R2}_2$.

However, J might be persuaded to agree to $\mathbf{S2}_2$, given suitable accompanying side-payments. For example, suppose that K agrees to pay J £6 in the event that they win, and J to pay K £3.50 otherwise. This, in effect, gives them the joint prospect:

$\mathbf{S2}'_2$	J	K
0.5	£21	£9
0.5	−£3.5	£3.5
v	13.69	3.04

This is a prospect which, *ex ante*, Pareto-dominates $\mathbf{R2}_2$. So, although she prefers $\mathbf{R2}_2$ to $\mathbf{S2}_2$, J might nevertheless agree to the latter, given contingent side-payments of this type.

But now consider the third prospect-pair {$\mathbf{R3_2}$,$\mathbf{S3_2}$}. Evidently, the correspondingly re-allocated version of $\mathbf{S3_2}$ does *not* Pareto-dominate $\mathbf{R3_2}$:

$\mathbf{R3_2}$	J	K
0.1	£35	£35
0.9	£0	£0
v	10.22	2.72

$\mathbf{S3'_2}$	J	K
0.2	£21	£9
0.8	−£3.5	£3.5
v	8.71	2.92

This illustrates an important proposition: the structure of Pareto-dominance relations between correspondingly-allocated joint prospects is not identical across different prospect-pairs within a given CR-set. It follows that if partners are able freely to (re-)allocate prospective prizes, then their choices among joint prospects may appear as inconsistent. Thus, the agreed choice of $\mathbf{S2_2}$ from {$\mathbf{R2_2}$,$\mathbf{S2_2}$}, but of $\mathbf{R3_2}$ from {$\mathbf{R3_2}$,$\mathbf{S3_2}$}, could actually reflect the Pareto-dominance of $\mathbf{S2'_2}$ over $\mathbf{R2_2}$, there being no corresponding Pareto-dominance of $\mathbf{S3'_2}$ over $\mathbf{R3_2}$. Let us call this an *allocation effect*.[5]

Since we felt unable to prevent covert side-payments, we decided instead to ask partnerships explicitly to record their agreement on how their prizes were to be divided. This would at least give us some information on the extent of any allocation effects.

There was another reason for allowing partners freely to distribute the prizes between themselves. The situation thereby corresponds more closely to that facing groups in the real world, where there is no externally-imposed division of financial prospects in which members have a joint stake.

The experiment proceeded as follows. On recruitment, each subject received a copy of the instructions (see Appendix) together with a booklet containing answer forms for one set of 12 prospect-pairs. In each booklet both the order of the prospect-pairs, and also the (left/right) order of the two prospects in each pair, were individually randomised. Subjects completed these (stage 1) booklets unsupervised, and in their own time, before reporting to an appointed session for stages 2 and 3. Here, having first submitted their completed booklets, subjects were paired-up at random, by drawing coloured counters from an opaque bag.

Each partnership was given a new booklet for stage 2, containing randomised answer forms for the 12 joint prospect-pairs. The session took place in a large hall, within which partnerships were dispersed so that they could each discuss their choices in privacy. When completed, they submitted the booklet to the experimenter, who issued each partner with a new (stage 3) booklet, containing randomised answer forms for the 12 individual prospect-pairs. Subjects then completed and submitted these booklets, again in privacy.

The payment mechanism, specified in the advance instructions, was as follows. Over the three stages of the experiment, each subject was presented with 36 prospect-pairs, either as an individual (stages 1 and 3) or in a partnership (stage 2).

On completion, one of these 36 prospect-pairs was selected at random, independently for each partnership, for playing-out. Both the selection of the prospect-pair, and the playing-out of the chosen prospect, were done by the subject(s) concerned, again by drawing counters from a bag.

If the selected prospect-pair was from stage 1 or 3, then each partner had his chosen prospect independently played out, his payment then being the corresponding prize. If it was from stage 2, then the jointly chosen prospect was played out as a single event, each partner then receiving his agreed division of the joint prize. Had the partnership failed to register an agreed choice (or prize division) from the selected prospect-pair, then each partner would be paid £Z, the lowest prize corresponding to that prospect-pair.

Likewise, of course, an individual making choices at stage 1 or 3 might also have been curious to know the default outcome, i.e., the consequence of failing to register a choice from any given prospect-pair. Unlike for partnerships, however, we would not expect this knowledge to have any bearing on his choice (unless he prefers the default outcome to each prospect in that pair). So, following convention, we did not specify such defaults for stages 1 and 3. Neither did we specify time limits on registering choices at stages 2 and 3. As anticipated, this did not appear to cause any problems; all individuals and partnerships completed their choices in good time.

5. The results

Subjects were undergraduate and graduate students at York University, being largely but not exclusively economists. There were 46 individual subjects, and thus 23 partnerships. So there was a total of 276 responses (choices) at stage 2, and 552 at each of stages 1 and 3.

Let us look first at individuals' choices. Table 2 records the number of individuals who gave EU-consistent responses in n of the four triples, at each of stages 1

Table 2. Number of individuals EU-consistent in n triples

		stage 3					
l	n	0	1	2	3	4	total
stage 1	0	4	2	0	1	0	7
	1	2	9	2	2	0	15
	2	1	3	1	1	0	6
	3	0	5	4	5	1	15
	4	0	2	0	0	1	3
	total	7	21	7	9	2	46

and 3. Only one individual, whom we shall call Leonard, was fully EU-consistent ($n = 4$) at each stage. Leonard was one of only seven individuals who were consistent in at least three triples at each stage.

As measured by n, the overall level of EU-consistency actually fell between stages 1 and 3. There were 9 individuals whose responses improved (above the diagonal in Table 2), but 17 whose responses deteriorated (below the diagonal). The median individual was consistent in two triples at stage 1, but in only one at stage 3. The corresponding mean values were, respectively, 1.83 and 1.52 triples. As a proportion of full-consistency ($n = 4$), this mean score was 0.46 at stage 1, and 0.38 at stage 3.

This overall deterioration in EU-consistency might conceal some instances of improvement through group interaction. Following the discussion in Section 3, let us identify as potential teachers those seven individuals scoring $n \geq 3$ throughout, of whom Leonard was uniquely the best-qualified. Sure enough, his partner was among the improvers, registering one EU-consistent triple at stage 1, and three at stage 3. However, none of the other six potential teachers had partners who improved. Alternatively, we can look for common improvement in partnerships without a teacher. There was none. In fact, in none of our 23 partnerships did both partners improve. So Leonard (or rather his partner) is our sole piece of evidence in support of the Savage hypothesis.

Our primary empirical question concerns the EU-consistency of the partnerships themselves. Before examining this we should report on one very clear result. As discussed in Section 4, the allocation effect provides a possible source of EU-inconsistency for partnerships, and in an attempt to monitor this we asked them to record their agreed division of joint prizes. Of the 23 partnerships in our experiment, all but one agreed to divide all prizes equally. Furthermore, the pattern of this one partnership's prospect choices was very typical. So we can conclude that allocation effects generally played no part in our partnerships' choices.

The prevalence of equal-division agreements may be somewhat surprising. Partners seemed to be overlooking even the opportunity of efficiency gains available through differential risk-sharing. This was not, of itself, a direct concern of the experiment. But it does call for some comment, which we offer in the concluding section.

As regards their prospect choices, only one of the partnerships was fully EU-consistent ($n = 4$). This was Leonard and his partner who, interestingly, agreed choices which coincided exactly with Leonard's own. Throughout the experiment Leonard, accompanied or otherwise, chose $\mathbf{S}\boldsymbol{i}_t$ from each prospect-pair in triples 1–3, and $\mathbf{R}\boldsymbol{i}_t$ in triple 4. We should note in passing that his was not the unequal-division partnership.

Of the remaining 22 partnerships, 13 registered $n = 1$, which was therefore the median value. The mean score was 1.48 triples, which as a proportion of full-consistency is 0.37. So the average incidence of EU-consistency was remarkably similar to that for individuals at stage 3.

We found, therefore, that groups and individuals were closely comparable in the extent of their EU-inconsistency. But, as suggested in Section 3, we might also ask whether the type of inconsistency differs between groups and individuals.

For a given triple t, let **RRS** (for example) denote a respondent's choice of $\mathbf{R1}_t$, $\mathbf{R2}_t$, and $\mathbf{S3}_t$. There are eight possible response patterns, which may be categorised as follows:

EU: either **SSS** or **RRR**
CR: either **SSR** or **SRR** (the Common-Ratio Effect)
other: any of {**SRS, RSS, RSR, RRS**}

Table 3 records the proportion of responses falling into each of these three categories, by triple and in total. The breakdown by triple is informative. Firstly, the incidence of EU-consistency, at all stages, was markedly greater in triple 4 than in other triples. This perhaps not surprising, given the relative attractiveness (ρ) here of **RRR**. Indeed, only one respondent (an individual at stage 1) chose **SSS** in triple 4. Secondly, partnerships were somewhat more EU-consistent than were individuals in triple 4, and somewhat less so in triples 1–3. These facts are connected, as will shortly be explained.

As regards the type of inconsistency, of the non-EU responses (individual or partnership) almost all were CR. Across all triples and respondents, around 90% of responses were either EU or CR, although there are some variations in this by triple. It therefore seems useful to identify a class of response pattern of which EU and CR are (exhaustive) special cases. We call this a General Common-Ratio (GCR) pattern, defined by the absence of any (re-)switching from $\mathbf{R}i_t$ to $\mathbf{S}i_t$, as p_i

Table 3. EU and CR responses by triple

		stage 1	stage 2	stage 3
triple 1	EU	0.48	0.22	0.28
	CR	0.52	0.74	0.61
	other	0.00	0.04	0.11
triple 2	EU	0.41	0.22	0.33
	CR	0.54	0.61	0.57
	other	0.04	0.17	0.11
triple 3	EU	0.30	0.17	0.37
	CR	0.61	0.61	0.54
	other	0.09	0.22	0.09
triple 4	EU	0.63	0.87	0.54
	CR	0.28	0.13	0.37
	other	0.09	0.00	0.09
total	EU	0.46	0.37	0.38
	CR	0.49	0.52	0.52
	other	0.05	0.11	0.10

falls. Within a CR-triple there are four such patterns, and we can order them as follows:

SSS SSR SRR RRR

i.e., according to the point, within the triple, at which the **S** → **R** switch (if any) occurs. In terms of the Machina-Marschak triangle, these are the four response patterns consistent with the fanning-out hypothesis (Machina, 1982).

The distribution of responses across the four GCR patterns, as ordered above, was unimodal for most triple:stages. The exceptions were 4:2, 2:3, 3:3, only the last of which was markedly bi-modal. Given this, and given that overall around 90% of responses were GCR, we can perhaps think in terms of a 'representative respondent'. For each triple, the median response pattern (of those conforming to GCR) was the same at each stage. For triples 1 and 2 it was **SSR**, for triple 3 it was **SRR** and for triple 4 it was **RRR**. Thus, the median (i.e., representative) GCR switch point was monotonically related to the value of ρ, and was the same for partnerships and individuals.

If most responses are GCR then we would expect that, within each triple, the proportion of respondents choosing $\mathbf{R}i_t$ increases with i, i.e., as p_i falls. Table 4 confirms this. It records this proportion for each prospect-pair, at each stage. In addition to confirming the GCR response pattern, the data shows that for any given i there was a general tendency for the proportion choosing $\mathbf{R}i_t$ to increase with t, i.e., as ρ increased. There were a few exceptions to this, however.

Table 4 shows also that partnerships were more inclined to choose $\mathbf{R}i_t$, both in total and for most prospect-pairs, than were individuals. This is consistent with the evidence (Table 3) that on average groups were more EU-consistent than individuals in triple 4, and less so in other triples. For both individuals and partnerships, as ρ increases the distribution of GCR responses shifted away from **SSS** and towards

Table 4. Proportional aggregate responses ($\mathrm{R}i_t$) by prospect-pair

	stage 1	stage 2	stage 3	1 and 3
$\mathbf{R1_1}$	0.07	0.00	0.04	0.02
$\mathbf{R2_2}$	0.22	0.22	0.39	0.11
$\mathbf{R3_3}$	0.59	0.74	0.65	0.48
$\mathbf{R1_2}$	0.13	0.26	0.07	0.02
$\mathbf{R2_2}$	0.33	0.52	0.41	0.17
$\mathbf{R3_2}$	0.67	0.83	0.59	0.48
$\mathbf{R1_3}$	0.17	0.22	0.13	0.07
$\mathbf{R2_3}$	0.54	0.52	0.52	0.41
$\mathbf{R3_3}$	0.76	0.78	0.67	0.59
$\mathbf{R1_4}$	0.65	0.87	0.63	0.52
$\mathbf{R2_4}$	0.89	0.91	0.91	0.83
$\mathbf{R3_4}$	0.91	1.00	0.93	0.87
total	0.49	0.57	0.50	0.38

RRR, the EU-consistent endpoints of the GCR range. But for any given triple the distribution of partnerships' responses was positioned more towards **RRR** than was that of individuals. This difference is only slight since, as already noted, the median GCR response for any given triple was the same for individuals and partnerships. But for each of triples 1−3 individuals' responses were sufficiently closer to **SSS** for their overall EU-consistency to be higher, while for triple 4 the partnerships' responses were sufficiently closer to **RRR** for the reverse.

We can therefore summarise our findings on group choices as follows, with reference to Table 3. The average level of EU-consistency, at 0.37, was close to that for individuals, especially at stage 3. As regards the type of inconsistency, the similarity was also very close. Responses exhibiting the Common-Ratio effect accounted, on average, for a further 0.52 of partnerships' overall responses, again just as for individuals. Both for partnerships and for individuals, the representative respondent had a GCR response pattern which shifted rightwards (away from **SSS** towards **RRR**) as the value of ρ increased. Together with the general tendency for partnerships to favour $\mathbf{R}\boldsymbol{i}_t$ more than did individuals, this largely accounts for the differences, by triple, in EU-consistency between partnerships and individuals.

These are our main findings. However, our data contains further information which may be of interest, in respect of both individual and group choices. Firstly, the final column in Table 4 records the proportion of individual respondents who chose $\mathbf{R}\boldsymbol{i}_t$ at each of stages 1 and 3. This gives some indication of the degree of 'repetition-consistency' between the two individual stages. For example, while 0.67 of individuals chose $\mathbf{R3}_2$ at stage 1, and 0.59 at stage 3, only 0.48 chose $\mathbf{R3}_2$ at both stages. Simple calculation reveals that 0.19 must have changed from $\mathbf{R3}_2$ to $\mathbf{S3}_2$, and 0.11 changed in the reverse direction. So 0.30 of responses, for this prospect-pair, changed between the two stages. It follows that 0.70 of responses remained unchanged (and therefore that 0.22 of individuals repeatedly chose $\mathbf{S3}_2$).

Table 4 shows that in total, i.e., over all prospect-pairs, 0.38 of responses were $\mathbf{R}\boldsymbol{i}_t$ at both stages. So by a similar calculation just over three-quarters (0.77) of all responses were unchanged. Aggregated by triple, the repetition-consistency rate ranged from 0.71 in triple 2 to 0.84 in triple 4.

Secondly, for any prospect-pair in which both partners registered repeated and identical choices, we would expect their joint choice to coincide with this. Table 5 records the instances of Pareto-inefficiency, where this was not the case. It does so

Table 5. Pareto-inefficiencies

	both **R**	both **S**
triple 1	0^{7}	0^{26}
triple 2	0^{7}	5^{22}
triple 3	1^{13}	2^{18}
triple 4	1^{39}	1^{2}
total	2^{66}	8^{68}

according to the triple containing the prospect-pair in question, and according to whether each partner's individual preference was for $\mathbf{R}i_t$ or $\mathbf{S}i_t$. The superscript figure, in each cell, shows the total number of instances (within that triple) in which partners registered such preferences repeatedly and unanimously. For example, within triple 3 there were 13 instances where, for some given i, both partners repeatedly chose $\mathbf{R}i_3$. In one of these, the partnership nevertheless chose $\mathbf{S}i_3$.

The general incidence of Pareto-inefficiency, defined thus, appears to be quite low. The obvious exception is in triple 2, where there were five instances of partnerships choosing $\mathbf{R}i_2$ despite their partners' repeated preference for $\mathbf{S}i_2$. This is consistent with the finding that, overall, partnerships tended to favour $\mathbf{R}i_t$ more than did individuals, although of course we would not have expected this to over-ride unanimous individual preferences.

Thirdly, in section 2 it was noted that the Common-Ratio Effect involves a violation of at least one of the Reduction and Independence axioms of EU theory. Generalisations of EU theory, to accommodate such deviant preferences, normally involve some weakening of the Independence axiom. An example of this is the fanning-out hypothesis mentioned above, which rationalises GCR responses. But other approaches are possible.[6]

Although such questions were not our central concern, our data does contain some indirectly relevant information. Assuming that a larger prize is preferable to a smaller one, then there are some response patterns, across prospect-pairs in different triples, which would specifically violate Independence. For example, if £35 is preferable to £30 then according to Independence $\mathbf{R3}_2$ is preferable to $\mathbf{R3}_1$. Since $\mathbf{S3}_2$ is the same as $\mathbf{S3}_1$, it would therefore be inconsistent (given Transitivity) to choose both $\mathbf{R3}_1$ and $\mathbf{S3}_2$. Indeed, there is one response pattern which would simply violate the assumption that a larger prize is preferable. This is the choice of $\mathbf{R1}_1$ and $\mathbf{S1}_3$.

Table 6 shows the set of similarly inconsistent response patterns, and records their incidence. Each respondent (individual or partnership) could have committed up to 6 such violations. Thus, up to 276 violations could have occurred at each of stages 1 and 3, and up to 138 at stage 2. Given this, the overall incidence seems to

Table 6. Cross-triple violations of Independence

		stage 1	stage 2	stage 3
$\mathbf{R1}_1$ &	$\mathbf{S1}_2$	1	0	1
	$\mathbf{S1}_3$	1	0	0
$\mathbf{R2}_1$ &	$\mathbf{S2}_2$	0	1	3
	$\mathbf{S2}_3$	2	0	2
$\mathbf{R3}_1$ &	$\mathbf{S3}_2$	1	1	5
	$\mathbf{S3}_3$	2	1	3
$\mathbf{S1}_4$ &	$\mathbf{R1}_2$	1	1	0
	$\mathbf{R1}_3$	1	0	2
total		9	4	16

be quite low. It is at its highest at stage 3, in respect of $\mathbf{R3}_1$, as described above. Interestingly, one of the culprits here was Leonard's erstwhile and impressionable partner. In fact, his selection of the prospect $\mathbf{R3}_1$ constituted the only difference between his stage 3 choices and Leonard's.

6. Summary and conclusions

Leonard apart, we found no support for the hypothesis that individuals learn to be more EU-consistent either through discussion or through repetition. Indeed, between stages 1 and 3 there is a fall in the overall incidence of EU responses. However, the date does suggest that individuals' choices are far from arbitrary. There is strong evidence of a systematic GCR response pattern in all triples, although again the overall incidence of this falls slightly between stages 1 and 3.

Partnership choices, likewise, appear not to be arbitrary. The incidence of Pareto-inefficiency, and of cross-triple Independence violations, is low. As with individuals, there is a strong GCR response pattern in all triples. For partnerships, the overall incidence of both EU and GCR response patterns is very similar to that for individuals, although there are some differences by triple.

Finally, we note that non-EU responses, in the case of partnerships, cannot be attributed to allocation effects, since in all cases except one the partners agreed to divide all prizes equally. This requires further comment. Partners may have been explicitly concerned with fairness, an idea which has been explored in other experimental work. However, the focus of such work is typically on zero-sum distributional issues, such as in bargaining or ultimatum games.[7] If our subjects were guided solely by a fairness principle, it was strong enough to steer them not only away from distributionally unfair shares, but also towards shares which we can presume to be allocatively inefficient.

This can be seen by pursuing the example in section 4. Consider the joint prospect $\mathbf{S2}_2$, with prizes divided as follows:

$\mathbf{S2}_2^+$	J	K
0.5	£21	£9
0.5	−£5	£5
v	13.12	3.10

Although unequal, this is an *ex ante* envy-free allocation, which J and K each prefer to $\mathbf{S2}_2$ with equal division. It illustrates the possibility of efficient risk-sharing under which, roughly, the less risk-averse individual (here J) bears proportionately more of the risk. Of course, without knowing subjects' individual degrees of risk-aversion, we could not be sure in any given case that equal division was Pareto-inefficient. But we can be fairly confident that it was so in more than 1/23 of our partnerships.

Our own conjecture is that partners would have agreed to efficient (unequal) divisions, had they perceived this possibility for mutual gain, but that they simply failed to do so. Perhaps this was because the prospect, or its representation, was sufficiently complex to hide the possibility, or perhaps because the number of choices diverted attention from it. We are currently engaged in a new series of experiments designed specifically to explore this conjecture. Preliminary results appear to support it.

Acknowledgments

Our thanks to an anonymous referee, whose detailed comments have immeasurably improved this paper. Thanks also to participants in the Italian Experimental Economics conference in Trento, June 1997, and the FUR conference in Mons, July 1997, for helpful discussion and comments. We are grateful to the Economics Department at York, for finding this experiment, and also to Jill Barnard for providing the equipment.

Notes

1. Camerer (1995) provides an excellent overview.
2. A succinct and accessible treatment can be found in chapter 11 of Rasmusen (1994).
3. Experiment work in this area has, to our knowledge, largely concentrated on strategic interaction, although for a recent exception see Friedman (1998). In the strategic context, repetition also permits dynamic strategies through which, in particular, collusion may be facilitated. It is important to distinguish this from learning as such. Ledyard (1995) provides a discussion of these issues in the context of public good contributions. For an experimental investigation of strategic learning in the proper sense, see Erev and Roth (1995).
4. Even had we provided interim outcomes, any suboptimality experienced as a result of Common-Ratio inconsistency is, to say the least, of an ethereal type.
5. This is only a suggestive analysis. For a more complete account see Bone (1998).
6. Carlin (1992) offers some experimental evidence that the problem lies with Reduction rather than with Independence as such. Cubitt, Starmer and Sugden (1998) explore, experimentally, an analogous decomposition in terms of dynamic choice axioms.
7. A seminal reference is Ochs and Roth (1989). Roth (1995) provides a survey of strategic bargaining experiments, including an extended discussion of this issue.

Appendix: A Group Pairwise Choice Experiment

Welcome to this experiment. The instructions are simple, and if you follow them carefully, you may make a considerable amount of money which will be paid to you in cash immediately after the experiment.

The experiment is in three parts:

1. the first you will do in private before attending the session into which you have been booked;
2. the second you will do in conjuction with one other participant when you attend for this session;
3. the third you will do in private at the end of the session.

You must complete all three parts of the experiment. We anticipate that the session into which you have been booked will last no more than one hour in total.

All three parts of the experiment involve a number of Choice Problems. On each of these you will be asked to choose one or other of the two Options on offer. In the first part of the experiment, you will be choosing as an individual, as indeed in the third part, so you should choose which of the two you personally prefer. In the second part you and the other participant have to jointly choose one or other of the two Options on offer. This makes this second part of the experiment somewhat different from the first and third parts—we will have more to say on this shortly.

In order to provide an incentive for you to consider your choices carefully, we will be using the following payment scheme for your participation in the experiment. After you have completed the experiment, we will pick *one* of the Choice Problems at random: that is *one* picked randomly from the Choice Problems in the first, second and third parts of the experiment. If the randomly picked Choice Problem is from the first or third part of the experiment, we will play out for real the Option you chose on that Problem—and pay you accordingly.

If the randomly picked Choice Problem comes from the second part of the experiment, the procedure we will use will be slightly different—though the basic idea is the same. Once again, we will play out for real the Option you chose on that question; to then decide how to pay you and the other participant off we will follow decisions you made earlier about the division of the possible payoffs. Thus, in the second part of the experiment, we will not only ask you and the other participant to come to a joint decision as to which Option you are choosing but also to make a statement about how the various possible payoffs would be divided up between you.

Option A:
£20 with probability 1.0

outcome	1's payoff	2's payoff	total payoff
£20			£20

Option B:
£40 with probability 0.75
£2 with probability 0.25

outcome	1's payoff	2's payoff	total payoff
£40			£40
£2			£2

Let us give an example. Consider the following Choice Problem:

There are various possibilities for you and the other participant. For example:

1. Choose (A) and split the proceeds £10 to you and £10 to him/her
2. Choose (A) and split the proceeds £12 to you and £8 to him/her
3. Choose (B) and if £2 is the outcome, both of you get £1; whilst if £40 is the outcome, you get £19 and he/she gets £21
4. Choose (B) and you get £2 whatever happens, while he/she gets £0 if £0 is the outcome, and gets £38 if £40 is the outcome
5. Choose (B) and he/she gets £12 whatever is the outcome, while you lose £10 if £0 is the outcome (i.e. you pay him/her £10) and get £28 if £40 is the outcome

More generally, your choices will be one of the following:

6. Choose (A), and you get £x while he/she gets £$(20 - x)$
7. Choose (B), and you get £y and he/she gets £$(2 - y)$ if £0 is the outcome, while you get £z and he/she gets £$(40 - z)$ is £40 is the outcome.

You are free to choose either 6. and any value of x that you agree or 7. and any values of y and z that you agree.

If you have not come to an agreement on a Choice Problem and that Problem is chosen at the end to be played out for real, then we will take the smallest amount in that Choice Problem and split it equally between you.

In the second part of the experiment, you will be asked to compete a box like the one above for each of the Choice Problems. Note that the total of the second and third columns must be the same as the entry in the first and last columns. So, for example, if you chose 1. above, you would do the following:

Option A:	Option B:
£20 with probability 1.0 CHOSEN	£40 with probability 0.75 £2 with probability 0.25

outcome	1's payoff	2's payoff	total payoff
£20	£10	£10	£20

outcome	1's payoff	2's payoff	total payoff
£40			£40
£2			£2

If, however, you chose 4. above you would do as illustrated below:

Option A:	Option B:
£20 with probability 1.0	£40 with probability 0.75 £2 with probability 0.25 CHOSEN

outcome	1's payoff	2's payoff	total payoff
£20			£20

outcome	1's payoff	2's payoff	total payoff
£40	£2	£38	£40
£2	£2	£0	£2

To summarise, the experiment involves the following:

Part 1: Before you come to attend the experimental session, you will do the following. On each of the Choice Problems (appended to these instructions) you must choose either Option A or Option B.

Part 2: You will attend the experimental session, hand in your answers to Part 1, and be paired with some other participant. You will be presented with a sequence of Choice Problems. On each of these Choice Problems you and the other participant with whom you are paired must choose either Option A or Option B as well as indicating how the various possible payoffs would be divided up between you.

Part 3: After the experimental session, you—again as an individual—will be presented with a sequence of Choice Problems. Once again on each of the Choice Problems you must choose either Option A or Option B.

At the end of the experiment, *one* of the Choice Problems from Parts 1, 2 and 3 will be picked at random. If the randomly picked Choice Problem is from Parts 1 or 3, then your chosen Option will be played out for real and you will be paid accordingly. If the randomly picked Choice Problem is from Part 2 then the Option you (jointly) chose earlier will be played out for real—and the two of you paid accordingly—according to the division of the payoffs that you specified earlier. If you had not reached any agreement then the smallest amount of money in that Choice Problem will be split equally between you.

If there is anything in these instructions that you do not understand please ask one of the experimenters.

References

Bone, John. (1998). "Risk-sharing CARA individuals are collectively EU," *Economics Letters* 58, 311–317.

Camerer, Colin F. (1995). "Individual Decision Making," in John H Kagel and Alvin E Roth (eds.), *The Handbook of Experimental Economics*. Princeton U.P.

Carlin, Paul S. (1992). "Violations of the reduction and independence axioms in Allais-type and common-ratio effect experiments," *Journal of Economic Behavior and Organization* 19, 213–235.

Cubitt, Robin P., Chris Starmer and Robert Sugden. (1998). "Dynamic choice and the Common-Ratio Effect: an experimental investigation," *Economic Journal* 108, 1362–1380.

Erev, Ido, and Alvin E. Roth. (1995). "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term," *Games and Economic Behavior*, 8(1), 164–212.

Friedman, Daniel. (1998). "Monty Hall's Three Doors: Construction and Deconstruction of a Choice Anomaly," *American Economic Review* 88(4), 933–946.

Harsanyi, John. (1955). "Cardinal utility, individualistic ethics, and interpersonal comparisons of utility," *Journal of Political Economy* 63, 309–321.

Ledyard, John O. (1995). "Public Goods: A Survey of Experimental Evidence," in John H Kagel and Alvin E Roth (eds.), *op cit*.

Machina, Mark. (1982). "Expected Utility Analysis Without the Independence Axiom," *Econometrica*, 50, 277–323.

Ochs, Jack, and Alvin E. Roth (1989). "An experimental study of sequential bargaining," *American Economic Review* 79, 355–384.

Rasmusen, Eric. (1994). *Games and Information*, 2nd edn. Oxford: Blackwell.

Roth, Alvin E. (1995). "Bargaining Experiments," in John H Kagel and Alvin E Roth (eds.), *op. cit*.

Savage, Leonard J. (1972). *The Foundations of Statistics*, 2nd. edn. New York: Dover Publications.

The Economic Journal, **107** (*September*), 1465–1483. Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK and 350 Main Street, Malden, MA 02148, USA.

A LABORATORY STUDY OF GROUP POLARISATION IN THE TEAM DICTATOR GAME*

Timothy N. Cason and Vai-Lam Mui

This paper introduces the team dictator game to study whether social dynamics within a group can cause groups' decisions to differ systematically from individuals' decisions. In the individual dictator game, a subject dictates the allocation of y dollars; in the team dictator game, a team of two subjects dictates the allocation of $2y$ dollars. We derive and test competing predictions for the two dominant psychological theories of group polarisation in this context. The data indicate that team choices tend to be dominated by the more other-regarding member. This result is more consistent with Social Comparison Theory than Persuasive Argument Theory.

This paper introduces the team dictator game to study whether social dynamics within a group can cause decisions made by groups to differ systematically from decisions made by individuals for a given problem. Most studies of economic decisions assume that the decision is made by a single individual. For example, in studying a firm's behaviour, economists usually assume that an individual – the owner or the manager – decides what to produce, how much to produce, or whether to invest in a particular project. It is also customary when studying consumer behaviour to assume that decisions are made by an individual consumer.

However, many economic decisions are made by a group of individuals instead of a single person. For example, in large corporations, important investment and production decisions are usually made by a top management team. For married couples, the decision regarding whether to buy an expensive consumer durable – like a car or a house – is often the result of intense discussion or even debate between the wife and husband.

Although most economic studies assume that decisions are made by individuals, economists have studied extensively how individuals' preferences or decisions – such as individual choices in voting decisions – are aggregated into a collective choice.[1] However, these studies tend to focus on how strategic interactions among individuals may affect the mapping from individual preferences to the collective choice under different 'rules of the game', and do not consider whether social dynamics matter in such settings.[2]

Contrary to the economic literature, the importance of social influence in

* We thank Nicholas Argyres, Catherine Eckel, Daniel Friedman, Timur Kuran, Matthew Rabin, and seminar participants at Santa Clara University, University of California at Irvine, and the 1994 Public Choice Society meeting for helpful comments. We are also grateful to the editor and three anonymous referees for extremely helpful comments that have led to substantial improvement of the paper. Financial support from the USC Zumberge Faculty Research and Innovation Fund is gratefully acknowledged.

[1] The seminal theoretical works are Arrow (1951) and Black (1958). Fiorina and Plott (1978) is an important early study of group decision-making in committee experiments.

[2] A notable exception is Akerlof (1991). Citing the social psychology literature on obedience and groupthink, Akerlof demonstrates how individual members' undue obedience to authority can cause groups to make suboptimal decisions.

decision-making has long been a central concern of social psychologists. In a classic study, Stoner (1961) found that following group discussion, groups made riskier decisions compared to decisions made by individual members prior to the discussion. His findings stimulated an extensive literature that studies the extent to which group decisions differ systematically from individual decisions. A major finding in this literature is the group polarisation hypothesis: group discussion moves decisions to more extreme points in the same direction as the initial tendencies of the group members' individual choices.

Notice that even if one believes that social dynamics can affect group decision making, it is possible that such social factors are insignificant in determining economic decisions. This could be true if the decision-makers – whether they are groups or single individuals – are constantly under the discipline of market competition. For example, when the decision-makers are managers of firms in competitive markets, the force of market competition may be sufficient to ensure that the firm will end up making the same kind of investment decisions, regardless of whether the decision is made by a single individual or a team of managers. However, the situation can be substantially different when the decision-maker/makers are not facing strong discipline from market competition. In such a situation, it is possible that group dynamics may create a systematic difference between group and individual decisions.

This paper reports a team dictator experiment to investigate this issue. In the individual dictator game, each subject is paired anonymously with another subject, and one is selected to 'dictate' the allocation of y dollars. In the team dictator game, subjects are anonymously placed into groups of four, and two subjects form a team to dictate the allocation of $2y$ dollars among the four subjects. The individual dictator game has been studied extensively by economists (for example, Forsythe *et al.* 1994; Hoffman *et al.* 1994; Bolton *et al.* 1995). By extending the dictator game to this team setting, we are able to incorporate some key insights from the psychological literature of group polarisation to study whether group dynamics matter in economic decisions. To the best of our knowledge, this is the first study that attempts to incorporate this psychological literature into economics.

In the individual dictator game, subjects chosen to be 'dictators' have complete control over the allocation of surplus between themselves and one other subject. It is obvious that if the dictator's only goal is to maximise her earnings, then she should allocate the entire surplus to herself. However, in several independent experiments researchers have found that many dictators give a substantial share of the surplus to other subjects (see Roth (1995) for a survey). While a consensus explanation for this other-regarding behaviour is yet to emerge, most explanations in the literature incorporate subjects' concerns for certain social goals to explain these deviations from income-maximising behaviour. Moreover, strategic considerations are absent in the dictator game, and in this extremely simple setting, the decision-maker/makers are completely free from any competitive pressure. For these reasons, the dictator game is an ideal environment to test whether group dynamics matter in economic decisions: if group dynamics matter at all, they should matter here.[3]

The rest of the paper is organised as follows. Section I briefly summarises the psychological literature on group polarisation and discusses two competing explanations of the phenomenon – *Social Comparison Theory and Persuasive Argument Theory*. Section II describes the experimental design and procedure, and derives the hypotheses to be tested. Section III presents the results. The most interesting finding is that when teams consist of members who have made different individual choices, the team choices tend to be dominated by the more other-regarding member; i.e. the team member who took less of the y dollars when making her individual dictator decision. This tends to make team choices less self-regarding than individual choices for these teams, and the results are more consistent with Social Comparison Theory than Persuasive Argument Theory. Section IV concludes.

I. SOCIAL COMPARISON, PERSUASIVE ARGUMENT, AND GROUP POLARISATION

Psychologists have found that decisions made by groups differ systematically from decisions made by individuals. Initially, researchers conjectured that group discussion would generally lead to compromise between more extreme preferences of individual members and produce more moderate decisions. However, many laboratory studies in psychology suggest otherwise.

Stoner (1961) reported the first experimental demonstration that group decisions led to riskier choices than individual decisions. Stoner used the Choice Dilemma Questionnaire that was devised by Wallach and Kogan (1959) to investigate individual risk-taking behaviour. Each Choice Dilemma describes a situation in which a person has to choose between two actions that may lead to two different outcomes with uncertainty. The subject's task was to advise the decision-maker in the story regarding how much risk he should take in facing the choice problem. The following is a sample problem used by Stoner:

> Mr. A, an electrical engineer who is married and has one child, has been working for a large electronics corporation since graduating from college five years ago. He is assured of a lifetime job with a modest, though adequate, salary, and liberal pension benefits upon retirement. On the other hand, it is very unlikely that his salary will increase much before he retires. While attending a convention, Mr. A is offered a job with a small, newly founded company with a highly uncertain future. The new job would pay more to start and would offer the possibility of a share in the ownership if the company survived the competition of the larger firms.
>
> Imagine that you are advising Mr. A. Listed below are several probabilities or odds of the company's proving financially sound. Please check the *lowest* probability that you would consider it worthwhile to make Mr. A to take the new job.

[3] We are grateful to an anonymous referee for helping us highlight this point.

[ ] The chances are 1 in 10 that the company will prove financially sound.
[ ] The chances are 3 in 10 that the company will prove financially sound.
[ ] The chances are 5 in 10 that the company will prove financially sound.
[ ] The chances are 7 in 10 that the company will prove financially sound.
[ ] The chances are 9 in 10 that the company will prove financially sound.
[ ] Place a check here if you think that Mr. A should not take the new job, no matter what the probabilities.

In his original experiments, Stoner asked the subjects to first make recommendations for twelve Choice Dilemma problems. Having made their choices individually, the subjects were then asked to make a recommendation as a group to the decision-maker in each problem. They assembled together, and discussed each item until they reached consensus on a group recommendation. Stoner found that for most problems, group choices on the whole reflected greater willingness to take risk than the average individual choices of the group members. This finding was referred to as the 'risky-shift' phenomenon, and it immediately initiated a wave of experimental studies.

In many independent experiments, researchers were able to replicate this risky-shift phenomenon. However, they also found that for certain choice problems, the group became more cautious than the average individual members. This was referred to as the 'cautious-shift' phenomenon. In fact, in Stoner's original experiment, while the subjects exhibited risky-shift in ten of the twelve choice problems, they exhibited cautious shift in the other two problems.

Teger and Pruitt (1967) was the first to recognise that in these Choice Dilemma experiments, there was a systematic correlation between the mean initial response of the individuals and the direction of the shift of the group's decision. When the initial mean of the individual choices is relatively risky, group discussion generally causes further shift toward the risky extreme. In contrast, problems that elicit relatively cautious individual means tend to result in further shifts in the cautious direction after group discussion.

In view of this observation and further evidence, Moscovici and Zavalloni (1969) proposed the group polarisation hypothesis: group discussion moves decisions to more extreme points in the same direction as the average of the group members' initial individual choices. Following many studies there is now a consensus among psychologists that group polarisation is a well-established empirical phenomenon (Baron *et al.* 1992; Brown, 1986; Isenberg, 1986; Myers and Lamm, 1976).

While researchers have proposed many explanations to account for this phenomenon, two hypotheses – Social Comparison Theory (SCT) and Persuasive Argument Theory (PAT) – emerged as the dominant explanations. According to PAT, people are influenced by the number and persuasiveness of pro and con arguments that they recall from memory when making decisions. Therefore, group discussion will cause an individual to change her position in a given direction to the extent that the discussion exposes her to persuasive arguments favouring that direction. When subjects engage in group discussion, they are engaging in a process of pooling their arguments together. The final decision of the group is affected by this pool of arguments. If the initial mean response of the individuals exhibits a preference toward a particular position, then it is likely that the subjects will be exposed to more persuasive arguments in favour of this position during the discussion. Hence, group discussion is likely to produce a shift in choices in favour of the initial pre-discussion tendency (Burnstein *et al.* 1973; Bishop and Myers, 1974; Brown, 1974, 1986; Vinokur and Burnstein, 1978).

SCT provides an alternative perspective. According to this theory, people are motivated both to perceive and to present themselves in a socially desirable way. This theory further suggests that people desire to perceive themselves as more favourable than what they believe to be the average tendency. To accomplish this, a person observes how other people behave, and she then adjusts her behaviour to present herself in a socially more favourable way. Since group interaction induces subjects to engage in such a social comparison process, it will elicit a shift of choices in a direction of greater perceived social value (Brown, 1986; Levinger and Schneider, 1969; Myers *et al.* 1980).

Notice that both theories maintain that group interaction can affect decision-making because it enables a member of a group to obtain new information that can change her behaviour. However, the mechanisms through which this occurs are different. SCT emphasises that when interacting with others, subjects will concentrate on gathering information to determine what is socially desirable. PAT, on the other hand, postulates that people are influenced by the number and persuasiveness of pro and con arguments. If subjects are exposed to convincing arguments that appeal to considerations other than social desirability, then group discussion can cause a shift of choices in the direction that is contrary to what they perceive to be socially desirable.[4]

For example, if subjects in the dictator game perceive that other-regarding behaviour is socially desirable, then SCT predicts that group choices will be more other-regarding than individual choices, regardless whether the group's initial tendency is self-regarding or not. However, PAT predicts that for groups in which individual choices exhibit an initial tendency towards self-regarding behaviour, the group choices will tend to be more self-regarding since group discussion will generate more arguments in support of self-regarding behaviour. We explain these empirical hypotheses more precisely in the next section.

II. EXPERIMENTAL PROCEDURE AND HYPOTHESES

We adopt the exchange framing for the dictator decisions, and our instructions follow as closely as possible to the exchange treatment in Hoffman *et al.* (1994). In the exchange framing of the (individual) dictator game, one subject was randomly assigned the role of the seller and the other the role of the buyer. The seller chose a price that ranged between 0 and 5 dollars. The buyer was

[4] Also notice that contrary to PAT – which emphasises the importance of group discussion – SCT implies that a subject's behaviour can change if there are ways through which she can gather information about others' behaviour even in the absence of discussion. For example, merely observing what others have chosen can induce behavioural change according to SCT.

Table 1
Summary of Sessions

Session number	Number of subjects	First choice	Second choice
Session 1	16	Individual	Team
Session 2	20	Team	Individual
Session 3	20	Team	Individual
Session 4	16	Individual	Team
Session 5	24	Team	Individual
Session 6	16	Individual	Team
Session 7	24*	Team	Individual
Session 8	24	Individual	Team
Session 9	28	Individual	Team

* One subject participated in an earlier session, so his choices in this later session (and those of his teammate) were removed from the data before analysis.

required to buy at that price. If the seller chose a price of P dollars, the buyer received $5-P$ dollars while the seller received P dollars, so the seller dictated the allocation of the $5 surplus.

In our experiment, each seller made two decisions – an individual decision and a team decision. The decisions were made sequentially within a twenty-minute session. In the team decision a pair of two seller subjects chose a price, and a pair of two buyer subjects were paired with each seller team. The seller team also chose a price that ranged between 0 and 5 dollars. If the seller team chose a price of P dollars, each buyer received $5-P$ dollars while each seller received P dollars, so the seller team dictated the allocation of the $10 surplus. Thus, the design kept per-subject monetary incentives constant across the team and individual decisions.[5]

The primary treatment variable in our experiment is the ordering of the two decisions within a session. We refer to the sessions that began with the individual decision as the I–T (Individual–Team) treatment, and the sessions that began with the team decision as the T–I (Team–Individual) treatment. The different sequencing of decisions permits us to test both the group polarisation hypothesis and an alternative 'observer effect' hypothesis proposed by Hoffman *et al.* (1994).

Instructions are available by request from the authors. Table 1 summarises the nine sessions.[6] Subjects were recruited from three large microeconomics principles classes. All sessions were conducted in a large classroom, and all

[5] In our design, the monetary stakes are $5 per pair of subjects for each decision. This differs from the $10 stakes used in Hoffman *et al.* (1994). Forsythe *et al.* (1994) find that there is no significant difference in offer distributions when the total surplus changes from $5 to $10. However, they find that paying subjects in individual dictator games increases the amount kept by dictators compared to hypothetical choices.

[6] The number of subjects was allowed to vary across sessions for practical reasons, in order to maximise the number of observations for each session. We believe a range of 16 to 28 all constitutes a large enough group so that each subject assigns a sufficiently small probability that any other subject is the one (of 15 to 27) with whom they are paired. This belief is supported empirically by our inability to find significant behavioural differences across group sizes; for example, average individual offers on the first decision were $1.15, $1.25 and $1.21 for groups of size 16, 24 and 28, respectively.

participants immediately received a \$3 cash appearance fee. Subjects first randomly drew an identification number that determined their role (either buyer or seller) and the subject with whom they were paired for the team decision. The identification numbers were designed so that subjects could not determine which were buyers and which were sellers. They then filled out subject payment sheets that included their address for mailing their cash experiment payment.[7]

The experimenter then passed out decision forms to all subjects. Sellers allocated the surplus by circling payment divisions, which were available in discrete intervals of \$0.50. For the team decision, every team was called to the front of the room (by identification numbers) and excused to the hallway to discuss their decision and fill out the form in private. Both buyers and sellers received decision forms to preserve anonymity. The buyer decision forms elicited expectations about the seller offer, but the instructions emphasised that these choices did not affect buyer payments from the experiment. Teams were not given an explicit time limit to reach a consensus, but all finished within five minutes.

What do the two psychological theories summarised in Section I predict for the team dictator experiments studied here? To answer this question, we need to operationalise the concept of group polarisation in our setting. The two theories require us to classify offers as self-regarding or other-regarding. Therefore, for an empirical test of these theories, it is necessary to first establish a central or neutral point for the offer distributions. Unfortunately, as Myers and Lamm (1976) point out (p. 607), it is difficult (they say 'impossible') to define a neutral point on the altruism-selfishness continuum.

One possibility is to use the 'fair' offer of \$2.50 as the neutral point because it gives equal payment to each subject. Offers below \$2.50 are self-regarding, and offers above \$2.50 are other-regarding. On the other hand, an economist who believes that subjects only care about pecuniary rewards might suggest the 'rational' offer of \$0 as the neutral point. However, the data do not seem to suggest that either of these are appropriate neutral points. In fact, nearly all offers lie in the interval [\$0, \$2.50].

In what follows we consider an empirically-defined neutral point, similar to the approach taken by Abelson (1973), Butler and Crino (1992), and Knox and Safford (1976). In particular, we employ \$1.50 as the neutral point because it is the overall median of the offers. Our results are qualitatively unchanged if we use \$1.00 (the median of the I-T treatment) or \$1.35 (the overall mean) as the neutral point.

Let $\bar{y}_k$ denote the mean of the individual dictator game seller offers for the two members of the kth team. If $\bar{y}_k$ is less than the neutral point – that is, on average, the team members offered less than the sample median when making

[7] Although paying subjects through the mail after the session is a departure from traditional procedure, we do not believe this affects the results. We mailed the payments in order to complete the sessions in less than 20 minutes, and we emphasised in the experiment that all payments would be mailed in cash within 48 hours of the session. Since results in our individual dictator decisions replicate the Hoffman *et al.* (1994) results that employ immediate cash payment (see Section III), the empirical evidence indicates that this procedural departure is innocuous.

their individual offers – we classify this team as a 'self-regarding team'. If $\bar{y}_k$ is more than the median, we classify this team as an 'other-regarding team'.[8]

In most previous studies, SCT and PAT predict the same qualitative shifts after group discussion. Therefore, researchers have studied the content of group discussion ('content analysis') or manipulated the information available through discussion to differentiate between the two explanations (Isenberg 1986).[9] To keep our experimental design similar to existing studies of the individual dictator game, we made no attempt to observe team negotiations, so the procedure did not permit a content analysis of the team bargaining. However, an advantage of the present design is that the value-laden context of the dictator game implies that SCT and PAT make different predictions in the I–T treatment.

In particular, PAT predicts that following group discussion, compared to the mean of their individual offers, the self-regarding teams will make team offers that are more self-regarding, and the other-regarding teams will make team offers that are more other-regarding. In contrast, SCT predicts that both the self-regarding teams and the other-regarding teams will make team offers that are more other-regarding than the mean of their individual offers.

We now explain how we arrive at these predictions. Recall that PAT predicts that following group discussion, team offers will shift and become more extreme than the initial pre-discussion tendency. This occurs because if the initial mean individual response of the team members exhibits a preference toward a particular position, then it is likely that the members of this team will be exposed to more persuasive arguments favouring this position during discussion. In the I–T treatment, PAT predicts this shift in choices because discussion takes place after the individuals make their decisions and before the team makes its decision. However, PAT predicts no shift in the T–I treatment because subjects are not exposed to any additional arguments after the team makes its decision and before the individuals make their decisions.

Persuasive Argument Hypothesis for the Team Dictator Game: In the I–T treatment, compared to the mean of their individual offers, the self-regarding teams will make team offers that are more self-regarding, and the other-regarding teams will make team offers that are more other-regarding.

[8] As pointed out by a referee, if we were to extend the dictator game to allow for teams to consist of more than two members, the mean of the members' individual offers may not be an appropriate measure of the central tendency of these individual offers. For example, suppose a six person team chose to offer the following vector (3, 3, 3, 3, 3, 10) in their individual decisions when the surplus to be allocated in the individual decision is \$20. Suppose further that the overall sample median offer was 4. This team had a mean of 4·17, so it will be considered an other-regarding team if the classification is established by comparing the group mean to the overall sample median. This seems inappropriate because five out of six members were more self-regarding than the median. However, in our current experiment, a team only consists of two members. Since the median and the mean are identical for a two person team, this complication does not arise.

[9] For example, consider a design in which in one treatment, subjects are only allowed to communicate the amount that they propose the team allocate to itself, and are not allowed to offer supporting arguments. This treatment would eliminate the role that persuasive argument may play in affecting decisions. In another treatment, subjects are required to offer supporting arguments. In the current experiment, subjects have complete freedom in deciding whether or not to offer any supporting arguments when making their proposals.

SCT makes a different prediction. Recall that according to SCT, subjects are motivated both to perceive and to present themselves in a socially desirable way. This theory further suggests that people desire to perceive themselves as more favourable than what they believe to be the average tendency. An individual observes how other people present themselves, and adjusts her presentation accordingly. Since group discussion induces the agents to engage in such a comparing process, it will elicit an average shift in a direction of greater perceived social value.

If subjects perceive that other-regarding behaviour is socially desirable, social comparison will cause subjects to shift their offers in that direction after group discussion. In the I–T treatment, discussion takes place after the individuals make their decisions and before the team makes its decision. SCT thus predicts that in the I–T treatment, both the self-regarding teams and the other-regarding teams will make team offers that are more other-regarding than the mean of their individual offers. On the other hand, SCT predicts no systematic shift in the T–I treatment because there are no additional interactions which will induce subjects to engage in social comparison after the team makes its decision and before the individuals make their decisions. To summarise:

Social Comparison Hypothesis for the Team Dictator Game: In the I–T treatment, both the self-regarding teams and the other-regarding teams will make team offers that are more other-regarding than the mean of their individual offers.

As pointed out by a referee, it is possible that subjects may perceive self-regarding behaviour as socially desirable in this setting. Because our subjects are recruited from economics classes (as is typical for economics experiments), there are at least two reasons for this possibility (Marwell and Ames, 1981). First, there may be a selection bias among these subjects. Students who are attracted to economics classes may be more concerned about pecuniary benefits and therefore may consider taking most of the surplus to be the socially desirable action, at least among this particular group. Secondly, if learning economics has led subjects to think that the economically 'correct' decision is to take the entire surplus, then they may perceive that self-regarding behaviour is socially desirable. According to this operationalisation, SCT predicts that in the I–T treatment both the self-regarding and other-regarding teams will make a more self-regarding team offer compared to the mean of their individual offers.

Our conjecture was that this self-regarding version of SCT is unlikely to be confirmed by the data, because in most dictator experiments subjects exhibit a significant degree of other-regarding behaviour. Moreover, our data provide no evidence that increased economics training affects decisions in this setting.[10] As we report below, the data are inconsistent with this self-regarding version of SCT but support the other-regarding version of SCT, since the vast majority of teams that shift make more other-regarding offers on the team decision.

[10] The mean individual offer is identical ($1.28) for the sellers with no previous economics class ($N = 71$) and for the sellers with one previous economics class ($N = 20$).

Finally, the team dictator game also enables us to test an explanation of subjects' behaviour, which extends an idea proposed by Hoffman *et al.* (1994). They argued that subjects in the (individual) dictator game may be concerned about whether their offers are judged unfair by the experimenter, so the experimenter as observer could be increasing other-regarding behaviour. To test this observer effect hypothesis, they conducted a clever double blind dictator experiment, in which individual subjects' decisions could not be known either by the experimenter or by anyone else. They show that the offer distributions in the double blind dictator game are significantly more self-regarding than offer distributions in other experimental treatments that do not ensure subject anonymity to the experimenter. However, Bolton *et al.* (1995) fail to detect an observer effect in the dictator game when keeping the instruction frame constant. Moreover, Laury *et al.* (1995) also fail to detect an observer effect in a laboratory public goods environment, and Bolton and Zwick (1995) only find a very modest observer effect in the ultimatum game.[11] In his review of the laboratory bargaining literature, Roth (1995) concludes that the observer effect, if it exists at all, is insignificant.

The observer effect we test is different from the previous studies, however, because the previous studies manipulate the observability of the subjects' choices to the experimenter. In our team dictator game all subjects' choices are observed by the experimenter, but the *number* of observers has been increased when moving from the individual decisions to the team decisions. In the team decision, a seller's final choice is observable to an additional person other than the experimenter – namely, his teammate. Importantly, this explanation does not include any role for information gathering in group discussion and therefore applies with equal force in the T–I and I–T treatments.

Observer Effect Hypothesis of the Team Dictator Game: In both the T–I and the I–T treatments, teams will make team offers that are more other-regarding than the mean of their individual offers.

Note that both the social comparison and the observer effect hypotheses predict that team offers should be more other-regarding than the mean individual offers in the I–T treatment. However, the explanations for this shift are different in the two hypotheses, and they generate different predictions in the T–I treatment. According to the observer effect hypothesis, individuals behave in a less self-regarding way when the number of observers increase because they do not want to be perceived by others as selfish. Therefore, it predicts that team offers will be more other-regarding than the individual offers in both the T–I and the I–T treatments.

On the other hand, SCT maintains that individuals desire to perceive themselves as better than the social average (Brown, 1974). Once group discussion induces subjects to engage in social comparison and causes them to behave in a more other-regarding way, because subjects are concerned about their self-perception, they will continue to behave in the same way even if the

[11] In an ultimatum game, the buyer has the opportunity to accept or reject the seller offer. In the case of rejection, both subjects receive zero payments.

number of people observing their behaviour is reduced. Hence, SCT predicts shifts in choice in the I–T but not in the T–I treatment.

III. RESULTS

For all 46 teams, the Appendix tables present (a) the individual decision by the two team members separately (columns 1 and 2); (b) the average of these two individual offers (denoted $\bar{y}_k$ in column 3); (c) the offer made by this team on the team decision (denoted y_{kt} in column 4); (d) the classification of the mean individual offer as other-regarding or self-regarding (column 5); and (e) the direction of the shift on the team offer relative to the mean individual offer of the team (column 6).

Consistent with previous dictator game experiments, offers range widely from \$0 to equal split allocations. The pooled median is \$1.50. The individual offer distributions in both the I–T and T–I treatments are not significantly different from the dictator share distributions in Hoffman *et al.* (1994) that employ exchange wording and are not double-blind; Characteristic Function (CF) test (χ^2 (D.F. = 4)) statistics range between 5 and 7.[12] This provides evidence that the different procedures and subject pool we employ do not change subjects' behaviour compared to previous studies of the individual dictator game.

Recall that according to PAT, groups with a mean individual offer less than \$1.50 would polarise toward \$0 on their team offer, while groups with a mean individual offer greater than \$1.50 would polarise toward \$5 on their team offer. The data do not support this hypothesis. Table 2*a* shows that 23 of the 25 teams in the I–T treatment have a mean individual offer different from \$1.50, and only two shift toward the predicted poles (i.e. align in the indicated corners of Table 2*a*). Eleven teams shift in the direction that contradicts PAT, and 10 teams do not change. A three-by-three contingency table test rejects the null hypothesis that the direction of shift is independent of whether the mean individual offer is self-regarding or other-regarding, in favour of choice shifts toward the neutral point [χ^2 (D.F. = 4) = 12·52, p = 0·03]. This is the direct opposite of PAT.[13] This leads to our first conclusion.[14]

[12] Epps and Singleton (1986). The 10% critical value for a one-tailed test is 7·78 for the CF test. For tests that fail to reject the null hypothesis (that is, those with p-value > 0·1), we usually suppress the p-value. Forsythe *et al.* (1994) find that the CF and the Anderson–Darling test have the most power in these dictator game settings, and we prefer the CF test because it does not require continuous distribution functions. We also conducted all of the distributional tests using the more familiar Kolmogorov–Smirnov test. Conclusions are unchanged, so we report only the CF test.

[13] The middle row of Table 2*a* shows the teams with a mean individual offer equal to the neutral point, and PAT makes no prediction for such teams. The middle column of Table 2*a* shows the teams whose team offers do not differ from their mean individual offers, and most of these teams consist of members who make identical offers in their individual decisions. Group discussions in these teams are not likely to generate many new persuasive arguments, so they are not very useful for testing PAT. These observations suggest a two-by-two test that eliminates the middle column and row of Table 2*a*. This test also indicates that teams shift toward the neutral point rather than toward the predicted poles [χ^2 (D.F. = 1) = 4·17, p = 0·05].

[14] Daniel Friedman has pointed out to us that this conclusion might be sensitive to alternative operationalisations of the null hypothesis. In particular, under a null hypothesis of random behaviour and independent errors, the likelihood of observing shifts toward the predicted poles is less than shifts away from

Table 2

Frequency of Shifts Toward Other-Regarding and Self-Regarding Team Offers and Classification of Mean Individual Offer

	Direction of shift (team)			
	Self-regarding (team < mean individual)	No change (team = mean individual)	Other-regarding (team > mean individual)	Total (row)
(a) Individual–Team (I–T) treatment Mean of the individual offer is				
Self-regarding (< $1.50)	1*	6	9†‡	16
Neutral (= $1.50)	2	0	0†‡	2
Other-regarding (> $1.50)	2	4	1*†‡	7
Total (column)	5	10	10	25
(b) Team–Individual (T–I) treatment Mean of the individual offer is				
Self-regarding (< $1·50)	1	3	5‡	9
Neutral (= $1.50)	0	2	2‡	4
Other-regarding (> $1.50)	3	2	3‡	8
Total (column)	4	7	10	21

* Observations in these cells are consistent with Persuasive Argument Theory.
† Observations in these cells are consistent with Social Comparison Theory.
‡ Observations in these cells are consistent with the Observer Effect.

Conclusion 1: The data in this team dictator experiment provide no support for Persuasive Argument Theory.

Next consider SCT, which predicts that team offers will move in the other-regarding direction for both self-regarding and other-regarding teams. The data are more favourable to SCT. Ten of the 25 teams in the I–T treatment shift in the other-regarding direction, five teams shift in the self-regarding direction, while 10 teams do not shift. However, shifts toward other-regarding offers are not significantly more frequent than shifts toward self-regarding

the poles due to the possibility of a regression toward the mean. We conducted a simulation to determine the strength of this possible bias, and found that our conclusions remain unchanged and that we can reject this null hypothesis of random behaviour. For the simulation we drew 1,000 samples of 50 independent individual offers and 25 team offers from the empirical distribution of offers. (The sample size of 25 teams is the same as in the I–T treatment.) We then classified the mean individual offers and the direction of the team offer shift as in Table 2. As expected, shifts toward the poles are less common than shifts away from the poles, with an average (over the 1,000 samples) of 6·31 teams shifting toward the poles and 13·17 teams shifting away from the poles. However, we observe only two teams shift toward the poles in the I–T treatment (Table 2*a*), which is less than the fifth percentile of the pole-shift rate in the simulation (three teams). Therefore, we reject this null hypothesis of random behaviour and independent errors at 5% using a Monte Carlo test, in the direction of more shifts inconsistent with PAT.

offers using a binomial test ($p = 0{\cdot}15$).[15] The other-regarding shift is present primarily for the teams who are individually self-regarding. Nine of these 16 teams shift in the other-regarding direction, while only 1 shifts in the self-regarding direction. For these self-regarding teams, the data reject the null hypothesis that other-regarding and self-regarding shifts are equally likely in favour of the shift predicted by SCT (binomial test $p = 0{\cdot}01$).[16]

The contingency table and binomial tests only capture the direction of shifts but do not account for their magnitude. To analyse shift magnitudes, we explicitly consider the 'bargains' struck by team members. Consider the following team bargaining equation, where y_{kt} denotes the offer made by the kth team, and individual offers are ordered $y_{k1} \geqslant y_{k2}$ so that the subject with the 1 index is more other-regarding:

$$y_{kt} = \alpha_0 + \alpha_1 y_{k1} + \alpha_2 y_{k2} + \epsilon_\kappa, \tag{1}$$

where ϵ_k is an error term. The hypothesis that the team members have an equal impact on the team offer implies $\alpha_1 = \alpha_2$. If, as predicted by SCT, team offers shift in the other-regarding direction compared to the individual offers, the team offers will tend to be dominated by the individual offer of the more other-regarding team member. This will imply that $\alpha_1 > \alpha_2$.

Estimates of (1) are shown in Table 3 for the two treatments, in all cases excluding the 13 teams comprised of sellers with equal individual offers.[17] Column (1) presents the I–T treatment and Column (2) presents the T–I treatment. For the I–T treatment, the team offer responds positively to the individual offer of the more other-regarding team member (y_{k1}), and the coefficient on the more self-regarding team member's individual offer (y_{k2}) is negative but not significantly different from zero. This indicates that when a team consists of members who have made different individual offers, the team offer tends to be dominated by the more other-regarding member. The F-statistic in the bottom of Table 3 indicates that for the I–T treatment, the data reject the null hypothesis that each team member has equal influence on the team offer (p-value $= 0{\cdot}03$).

Conclusion 2: In the I–T treatment, team offers tend to shift in the other-regarding direction, especially for the teams who make self-regarding individual offers. Our estimate of a team bargaining equation reveals that team offers are dominated by the more other-regarding member. Combined with conclusion 1, these findings provide better support for SCT than PAT.

[15] The binomial test we employ calculates the likelihood that the outcome x will be observed m times or less in a sample of size n, under the null hypothesis that outcomes x and not-x are equally likely. For this test we exclude the teams which do not shift on the team choice.

[16] The simulation based on the null hypothesis of random behaviour and independent errors discussed in footnote 14 indicates that our conclusions regarding SCT are also robust. Other-regarding and self-regarding shifts are equally likely for the teams overall, so the null hypothesis in the binomial test that each shift is equally likely is still appropriate. For the sample of teams who are individually self-regarding, the simulation indicates that other-regarding shifts occur more frequently than self-regarding shifts. Nevertheless, using either a Monte Carlo test or binomial test employing the simulation probability for other-regarding shifts, the data still reject this null hypothesis of random behaviour in favour of the shift predicted by SCT.

[17] With equal individual offers there is no bargaining disagreement to explain. Results are similar when using all 46 teams.

Table 3

Estimates of Team Bargaining Equation

(Dependent variable: Team offer (surplus per buyer))

Variable or statistic	I–T treatment (1)	T–I treatment (2)
More other-regarding individual offer (α_1)	0·654**	0·040
	(0·255)	(0·217)
More self-regarding individual offer (α_2)	−0·460	0·294
	(0·274)	(0·280)
Intercept (α_0)	0·416	1·189**
	(0·393)	(0·461)
Observations	18	15
Adjusted R^2	0·22	−0·04
F-statistic testing $\alpha_1 = \alpha_2$	6·064†	0·387

Notes: Standard errors are in parentheses; * denotes significantly different from 0 at 10% level; ** denotes significantly different from 0 at 5% level; † denotes reject $\alpha_1 = \alpha_2$ at the 5% level. All estimates exclude the 13 teams with equal individual choices.

The fact that team offers shift in the other-regarding direction in the I–T treatment is consistent with both SCT and the observer effect hypothesis. However, the two hypotheses make different predictions in the T–I treatment. In particular, the observer effect predicts that team offers will be more other-regarding than the mean of the team members' individual offers in the T–I treatment, while SCT does not predict any systematic shift.

The bargaining equation estimate in Column (2) of Table 3 indicates that the team offer is not dominated by either the more other-regarding or the more self-regarding member for the T–I treatment. A comparison between the adjusted R^2 in the I–T and T–I equation estimates reveals that although this simple bargaining equation performs reasonably well in the I–T treatment, it performs poorly in the T–I treatment. This provides some evidence against the observer effect hypothesis. However, Table 2*b* indicates that 10 of the 21 teams in the T–I treatment shift in the other-regarding direction, while only four of these teams shift in the self-regarding direction (Binomial test $p = 0{\cdot}09$). The rate of shifts toward other-regarding behaviour is very similar in both the I–T and the T–I treatments, and a two-by-two contingency table test fails to reject the null hypothesis that the shift rate toward other-regarding offers on the team decision is independent of the (I–T versus T–I) decision order (χ^2 (D.F. = 1) = 0·08). The magnitude of this shift in offers is modest in the T–I treatment, however. As shown in the appendix, the mean team offer is only six cents larger than the mean individual offer.[18]

[18] Also note that the corresponding difference in the I–T treatment is only nine cents. But if we exclude those teams that consist of members who make identical individual offers, this difference increases to eighteen cents in the I–T treatment but only increases to eight cents in the T–I treatment. None of these differences in mean offers nor any differences in offer distributions are statistically significant according to conventional tests, however.

Conclusion 3: Consistent with the observer effect, more teams shift in the other-regarding direction than in the self-regarding direction in the T–I treatment. The frequency of this shift is not significantly different in the I–T and T–I treatments. However, estimates of a bargaining equation and a comparison of the mean offers indicate that the magnitude of this shift is smaller in the T–I treatment. These mixed results therefore do not provide strong evidence to distinguish between SCT and the observer effect.

Finally, we consider the role of demographic factors in explaining choices.[19] Gender is the only demographic factor that significantly affects seller offers. Individual offers are about equal for females (mean offer = \$1.37) and males (mean offer = \$1.26) and are not significantly different (t-value = 0·51, $p = 0{\cdot}61$). This lack of a gender effect in the dictator game was reported previously by Bolton and Katok (1995). However, an analysis of team offers suggests a possible role for gender in affecting team offers. Twenty of the 46 teams were comprised of one male and one female. Eight of these 20 teams either had initially equal preferences (four teams) or made a team offer that was equidistant from the two individual offers (four teams). Of the remaining 12 teams with a team offer dominated by one team member, nine of the team offers are closer to the individual offers made by females. The estimate for a gender bargaining equation analogous to equation (1) using all 20 teams of differing gender is

$$\underset{}{y_{kt} =} \underset{(0{\cdot}197)}{0{\cdot}244} \underset{(0{\cdot}105)}{+0{\cdot}702 y_{kf}} \underset{(0{\cdot}082)}{+0{\cdot}230 y_{km}}, \quad N = 20, \bar{\mathrm{R}}^2 = 0{\cdot}72 \tag{2}$$

where y_{kf} is the individual offer of the female member and y_{km} is the individual offer of the male member of the kth team (standard errors in parentheses). Although the coefficients on both individual offers are highly significant, the female team member appears to dominate the team bargaining. Furthermore, we can strongly reject the hypothesis that the team members have an equal impact on the team offer. ($F(\text{D.F.} = 1, 17) = 13{\cdot}51$, $p < 0{\cdot}01$).[20] This result is summarised by our final conclusion:

Conclusion 4: In teams comprised of one female and one male subject, the team offers are dominated by the female team member.

We find this result puzzling, especially since most experimental economic studies fail to find a significant gender effect. We hesitate to draw any strong conclusions because of the small sample size of teams comprised of one female and one male subject, and also because unlike other hypotheses tested above, we do not have any a priori intuitive explanation for a gender effect. However, we note that Bolton and Katok (1995) observe that in studies which find gender effects (Brown-Kruse and Hummels, 1993; Eckel and Grossman, 1994),

[19] After all the decisions were made, subjects filled out a brief demographic questionnaire. The demographic data were used to determine if dictator decisions differed systematically by gender, the amount of economics training, age, cultural background and other factors. (The University of Southern California has a very diverse undergraduate student population, and more than one third of our subjects were born outside the United States.)

[20] This significant gender effect in team bargaining suggests that (1) might be misspecified because it does not include gender. We therefore reestimated a version of (1) that includes gender interaction effects for both treatments, but these interaction effects were insignificant and our conclusions were unchanged.

subjects learn the gender of others in their group. In contrast, studies which fail to observe significant gender effects usually provide no such information to subjects. Our study falls into the first category, and our finding adds some strength to this observation regarding the differences between studies that find significant gender effects and those that do not. Additional laboratory studies may help us to determine the potential generality and significance of this observation better.

IV. CONCLUSION

This paper introduces the team dictator game and uses the psychological theories of group polarisation to study whether social dynamics within a group can cause groups' decisions to differ systematically from individuals' decisions. We operationalise the two dominant theories of group polarisation – Persuasive Argument Theory and Social Comparison Theory – in the context of the dictator game, and show that they generate different predictions. Perhaps the most interesting finding of our experiment is that when teams consist of members who have made different individual choices in the I–T treatment, the team choices tend to be dominated by the more other-regarding member. This makes team choices more other-regarding than individual choices for these teams, although the difference is modest.

Overall, in this setting, our data provide strong evidence against PAT, and are more consistent with SCT than PAT. The data do not provide strong evidence to distinguish between the SCT and the observer effect hypotheses, but our results are consistent with findings in previous studies which indicate that the observer effect is probably not very significant.

Beyond these empirical findings, a contribution of this paper is the incorporation of some insights from social psychology regarding group behaviour to study economic decision-making. We believe that this exercise provides a reasonably good case for the need of future research along these lines.

For example, a natural extension is to study whether the empirical findings reported here generalise or perhaps become sharper when the size of teams increases. By varying the size of teams systematically, one can determine whether any differences between group and individual choices are magnified or diminished as group size increases. Another experiment could solicit an individual offer, followed by a team offer, then followed by an additional individual offer (i.e. an I–T–I design). This would provide another test to distinguish SCT from the observer effect, because it would allow us to compare individual choices by the same subject before and after group discussion.

The hypotheses discussed in this paper generate implications for the study of organisations that can be tested in field studies. For example, when choosing between alternative production processes, manufacturing firms are making economic decisions that often have important social consequences, such as how much or how little harm to inflict on the environment. If managers consider protecting the environment to be socially desirable, SCT implies that decisions by a management team will be systematically biased toward processes that are more protective of the environment compared to individual decisions. In contrast, PAT implies that decisions by a management team will exhibit a bias that magnifies the individual team members' average preferences toward the environment. It predicts that 'non-protective' management teams will adopt a production process that is less protective of the environment compared to individual decisions. Additional laboratory and field studies along these lines may provide useful insight in how organisations make decisions that have social consequences.

Finally, a limitation of the current study is that we have not addressed the question of what determines the persuasiveness of an argument; nor have we provided any explicit theory that specifies in detail the process through which an individual uses the information conveyed by others' choices to update his belief regarding what constitutes socially desirable behaviour in a particular context. For example, to further evaluate the Persuasive Argument Theory, it would be useful to determine if a person's perception of the relative persuasiveness of different arguments depends mainly on their content, or whether their ranking is substantially affected by the order in which the arguments are presented. Similarly, Social Comparison Theory postulates that an individual uses the information conveyed by the choices of others to update his belief regarding what constitutes socially desirable behaviour. However, the theory does not specify whether an individual uses the simple average of the observed choices of others to do so, or whether he puts a disproportional weight on the majority opinion in a group composed of more than two individuals. Further studies of these issues are essential for a more complete understanding of how group dynamics affect economic decision making.

University of Southern California

Date of receipt of final typescript: October 1996

References

Abelson, R. (1973). 'Comments on "group shift" to caution at the race track.' *Journal of Experimental Social Psychology*, vol. 9, pp. 517–21.

Akerlof, G. A. (1991). 'Procrastination and obedience.' *American Economic Review*, vol. 81, pp. 1–19.

Arrow, K. J. (1951). *Social Choice and Individual Values*. New York: Wiley.

Baron, R. S., Kerr, N. L. and Miller, N. (1992). *Group Process, Group Decision, Group Action*, Pacific Grove, California: Brooks/Cole Pub. Co.

Bishop, G. D. and Myers, D. G. (1974). 'Informational influences in group discussion.' *Organizational Behavior and Human Performance*, vol. 12, pp. 92–104.

Black, R. D. (1958). *The Theory of Committees and Elections*. Cambridge: Cambridge University Press.

Bolton, G. E. and Katok, E. (1995). 'An experimental test for gender differences in beneficent behavior.' *Economics Letters*, vol. 48, pp. 287–92.

Bolton, G. E. and Zwick, R. (1995). 'Anonymity versus punishment in ultimatum bargaining.' *Games and Economic Behavior*, vol. 10, pp. 95–121.

Bolton, G. E., Katok, E. and Zwick, R. (1995). 'Dictator game giving: rules of fairness versus acts of kindness.' *International Journal of Game Theory*, forthcoming.

Brown, R. (1974). 'Further comment on the risky shift.' *American Psychologist*, vol. 29, pp. 468–70.

Brown, R. (1986). *Social Psychology*. Second Edition. New York: the Free Press.

Brown-Kruse, J. and Hummels, D. (1993). 'Gender effects in laboratory public goods contribution.' *Journal of Economic Behavior and Organization*, vol. 22, pp. 255–67.

Burnstein, E., Vinokur, A. and Trope, Y. (1973). 'Interpersonal comparison versus persuasive argument: a more direct test of alternative explanations for group-induced shifts in individual choices.' *Journal of Experimental Social Psychology*, vol. 9, pp. 236–45.

Butler, J. K. and Crino, M. D. (1992). 'Effects of initial tendency and real risk on choice shift.' *Organizational Behavior and Human Decision Processes*, vol. 53, pp. 14–34.

Eckel, C. and Grossman, P. (1994). 'Chivalry and solidarity in ultimatum games.' Working Paper, Department of Economics, Wayne State University.

Epps, T. W. and Singleton, K. (1986). 'An omnibus test for the two-sample problem using the empirical characteristic function.' *Journal of Statistics and Computer Simulation*, vol. 26, pp. 177–203.

Fiorina, M. P. and Plott, C. R. (1978). 'Committee decisions under majority rule: an experimental study.' *American Political Science Review*, vol. 72, pp. 577–98.

Forsythe, R., Horowitz, J., Savin, N. E. and Sefton, M. (1994). 'Fairness in simple bargaining experiments.' *Games and Economic Behavior*, vol. 6, pp. 347–69.

Hoffman, E., McCabe, K., Shachat, K. and Smith, V. (1994). 'Preferences, property rights and anonymity in bargaining games.' *Games and Economic Behavior*, vol. 7, pp. 346–80.

Isenberg, D. J. (1986). 'Group polarization: a critical review and meta-analysis.' *Journal of Personality and Social Psychology*, vol. 50, pp. 1141–51.

Knox, R. and Safford, R. K. (1976). 'Group caution at the race track.' *Journal of Experimental Social Psychology*, vol. 12, pp. 317–24.

Laury, S. K., Walker, J. M. and Williams, A. W. (1995). 'Anonymity and the voluntary provision of public goods.' *Journal of Economic Behavior and Organization*, vol. 27, pp. 365–80.

Levinger, G. and Schneider, D. J. (1969). 'Test of the "risk is a value" hypothesis.' *Journal of Personality and Social Psychology*, vol. 11, pp. 165–9.

Marwell, G. and Ames, R. (1981). 'Economists free ride, does anyone else? Experiments on the provision of public goods.' *Journal of Public Economics*, vol. 15, pp. 295–310.

Moscovici, S. and Zavalloni, M. (1969). 'The group as a polarizer of attitudes.' *Journal of Personality and Social Psychology*, vol. 12, pp. 125–35.

Myers, D. G. and Lamm, H. (1976). 'The group polarization phenomenon.' *Psychological Bulletin*, vol. 83, pp. 602–27.

Myers, D. G., Bruggink, J. B., Kersting, R. C. and Schlosser, B. A. (1980). 'Does learning others' opinion change one's opinion?' *Personality and Social Psychology Bulletin*, vol. 6, pp. 253–60.

Roth, A. E. (1995). 'Bargaining experiments.' In *Handbook of Experimental Economics* (ed. J. Kagal and A. E. Roth), pp. 253–348. Princeton, NJ: Princeton University Press.

Stoner, J. A. F. (1961). 'A comparison of individual and group decisions under risk.' Unpublished Master's Thesis, Massachusetts Institute of Technology, School of Management.

Teger, A. I. and Pruitt, D. G. (1967). 'Components of group risk taking.' *Journal of Experimental Social Psychology*, vol. 3, pp. 189–205.

Vinokur, A. and Burnstein, E. (1978). 'Novel argumentation and attitude change: the case of polarization following group discussion.' *European Journal of Social Psychology*, vol. 8, pp. 335–48.

Wallach, M. and Kogan, N. (1959). 'Sex differences and judgement process.' *Journal of Personality*, vol. 27, pp. 555–64.

Table A1

Raw Data and Team Classifications for I–T Treatment

More other-regarding individual offer (y_{k1}) (1)	More self-regarding individual offer (y_{k2}) (2)	Mean individual offer ($\bar{y}_k$) (3)	Team offer (y_{kt}) (4)	Classification of mean individual offer (5)	Direction of shift on team offer (6)
0.5	0	0.25	0	SR	SR
0.5	0	0.25	1.5	SR	OR
1	0	0.5	0.5	SR	NC
1	0	0.5	0.5	SR	NC
1	0.5	0.75	1	SR	OR
1	0.5	0.75	1	SR	OR
1	0.5	0.75	1	SR	OR
1	0.5	0.75	1	SR	OR
1.5	0	0.75	1.5	SR	OR
2	0	1	2	SR	OR
2	0	1	2.5	SR	OR
1	1	1	1	SR	NC
2	0	1	1	SR	NC

Table A1 continued opposite

Table A1 (*cont.*)

More other-regarding individual offer (y_{k1}) (1)	More self-regarding individual offer (y_{k2}) (2)	Mean individual offer ($\bar{y}_k$) (3)	Team offer (y_{kt}) (4)	Classification of mean individual offer (5)	Direction of shift on team offer (6)
1	1	1	1	SR	NC
1	1	1	1	SR	NC
2·5	0	1·25	2·5	SR	OR
2·5	0·5	1·5	1	N	SR
1·5	1·5	1·5	0·5	N	SR
2	1·5	1·75	1	OR	SR
2	1·5	1·75	2	OR	OR
2	2	2	2	OR	NC
2·5	1·5	2	2	OR	NC
2	2	2	2	OR	NC
2·5	2	2·25	0	OR	SR
2·5	2·5	2·5	2·5	OR	NC
Mean = 1·58 s.e. = 0·13 Median = 1·50	Mean = 0·80 s.e. = 0·16 Median = 0·50	Mean = 1·19 s.e. = 0·13 Median = 1·00	Mean = 1·28 s.e. = 0·15 Median = 1·00		

Notes: SR, Self-Regarding; OR, Other-Regarding; N, Neutral; NC, No Change.

Table A2

Raw Data and Team Classifications for T–I Treatment

More other-regarding individual offer (y_{k1}) (1)	More self-regarding individual offer (y_{k2}) (2)	Mean individual offer ($\bar{y}_k$) (3)	Team offer (y_{kt}) (4)	Classification of mean individual offer (5)	Direction of shift on team offer (6)
0	0	0	0	SR	NC
0	0	0	0	SR	NC
1	0	0·5	2	SR	OR
1	0·5	0·75	1·5	SR	OR
1	0·5	0·75	1	SR	OR
1·5	0	0·75	1	SR	OR
1·5	0·5	1	1	SR	NC
1·5	1	1·25	1·5	SR	OR
1·5	1	1·25	0	SR	SR
2	1	1·5	2	N	OR
1·5	1·5	1·5	1·5	N	NC
2·5	0·5	1·5	2·5	N	OR
1·5	1·5	1·5	1·5	N	NC
2	1·5	1·75	2	OR	OR
2	1·5	1·75	2	OR	OR
2·5	1·5	2	1·5	OR	SR
4	0	2	1	OR	SR
2·5	2	2·25	2·5	OR	OR
2·5	2·5	2·5	2·5	OR	NC
3·5	2	2·75	1·5	OR	SR
3	3	3	3	OR	NC
Mean = 1·83 s.e. = 0·22 Median = 1·50	Mean = 1·05 s.e. = 0·19 Median = 1·00	Mean = 1·44 s.e. = 0·18 Median = 1·50	Mean = 1·50 s.e. = 0·17 Median = 1·50		

Notes: SR, Self-Regarding; OR, Other-Regarding; N, Neutral; NC, No Change.

J Risk Uncertainty (2007) 35:129–148
DOI 10.1007/s11166-007-9020-y

Individual and group decision making under risk: An experimental study of Bayesian updating and violations of first-order stochastic dominance

Gary Charness · Edi Karni · Dan Levin

Published online: 11 September 2007

Abstract This paper reports the results of experiments designed to test whether individuals and groups abide by monotonicity with respect to first-order stochastic dominance and Bayesian updating when making decisions under risk. The results indicate a significant number of violations of both principles. The violation rate when groups make decisions is substantially lower, and decreasing with group size, suggesting that social interaction improves the decision-making process. Greater transparency of the decision task reduces the violation rate, suggesting that these violations are due to judgment errors rather than the preference structure. In one treatment, however, less complex decisions result in a higher error rates.

Keywords Decision making under risk · Group decisions · Bayesian updating · First-order stochastic dominance

JEL Classification D80 · C91 · C92

G. Charness
Department of Economics, University of California, Santa Barbara, CA 93106, USA

E. Karni (✉)
Department of Economics, Johns Hopkins University, Baltimore, MD 21218, USA
e-mail: karni@jhu.edu

D. Levin
Department of Economics, The Ohio State University, Columbus, OH 43210, USA

Among the tenets of rational decision making under risk, monotonicity with respect to first-order stochastic dominance (that is, preference for a better chance of winning a larger sum of money) is the most compelling.[1] Violations of monotonicity with respect to first-order stochastic dominance may, therefore, be regarded as decision-making errors rather than as genuine expressions of preferences.[2] In situations involving decision making under risk, the application of Bayes' rule incorporates the impact of new information on the risks associated with alternative courses of action seems compelling, but it is not always intuitive. Consequently, uninitiated decision makers are more likely to err when faced with situations requiring a re-evaluation of risks in the light of new information.

Most experimental studies of individual decision making in the face of risk examine the patterns of choice displayed by subjects acting in isolation. The detection and classification of systematic errors committed by individuals *acting in isolation* in an unfamiliar laboratory environment might improve the understanding of the psychology of individual decision makers. However, the significance of such errors for understanding economic behavior is far from obvious. Seldom, in real life, does an individual find himself in situations in which he must make an important decision in isolation. Typically, when facing decisions with important consequences, individuals seek advice from family members, friends, and colleagues and, if necessary, consult experts. In other words, most decisions are made in a "social context." It is natural to suppose that, ceteris paribus, the less familiar and/or more complicated the decision problem, the more inclined the decision maker is to consult others before acting. We hypothesize that seeking advice and/or merely deliberating the implications of choosing alternative courses of action helps improve the understanding of the decision problem and makes errors less likely. Consequently, the study of group decision making should enhance the relevance of the experimental results for the analysis of economic behavior.

This paper presents an experimental study of the hypothesis that the simpler the problem and the less likely it is to involve emotions, the more likely a decision maker acting in isolation is to reach the correct decision when one exists. In addition, ceteris paribus, decisions made by groups of decision makers are more likely to be correct than decisions made by individual decision makers in isolation, and the larger the group, the more likely the decision is to be correct.

The paper contains two main findings. First, when making decisions in isolation, a substantial number of individuals choose first-order stochastically dominated alternatives. Second, when social interaction is allowed, the number of such violations tends to decrease with the size of the group.

The experimental design, which builds on Charness and Levin (2005), requires subjects to choose between risky prospects under four distinct

[1] Preference relations having expected utility representation are monotonic with respect to first-order stochastic dominance if and only if the utility function is increasing monotonically in income or in wealth.

[2] Kahneman and Tversky (1972) report such violations. However, their subjects changed their minds once the nature of the alternative involved was made apparent.

treatments of different levels of complexity in terms of the potential presence of affect, Bayesian updating, and compounding lotteries. To discern the effect of social interaction, we replicate some of the individual treatments with groups comprising two and three individuals.

Our study differs from Charness and Levin (2005) in at least three important respects. First, we examine choices involving alternatives that may be ranked according to first-order stochastic dominance. Second, to identify the factors responsible for errors, we consider a series of simplifications of the decision problem that allow us to test the hypothesis that greater transparency tends to reduce the observed number of violations. Third, and perhaps most importantly, we investigate the effect of group decision making on the prevalence of violations of first-order stochastic dominance.

Whether groups (or individuals receiving advice) are better at avoiding making systematic errors is an empirical question. While it might seem intuitive that a broader base of opinions would tend to mitigate the effects of misunderstanding and incompetence, thereby leaving less room for error, the verdict of the existing literature is not clear cut. Kerr, MacCoun and Kramer (1996) survey the psychological literature on the relative susceptibility of individuals and groups to systematic judgmental biases, finding that there is no simple answer to the question of performance. In fact, one determinant appears to be whether a decision problem requires a certain flash of insight that, once provided, makes matters transparent; this is known as "truth wins."[3] In this case, it seems intuitive that a team should perform at least as well as any individual member. However, some judgmental decisions do not have this flavor, so that group dynamics may either improve or impair the decision-making process.

In all of the situations considered here, there is a demonstrably correct course of action. It is, therefore, tempting to consider these decisions to be clear enough so that an individual who understands the problem can readily convince the other group members to choose the correct alternative. However, while applying Bayes' rule or correctly analyzing compound lotteries may seem simple to some, subjects in the laboratory and decision makers in the field may have difficulty with these tasks and may well not understand the underlying mechanics and principles. By varying the degree of complexity in our design, we offer evidence of how simple a decision must be to generate relative error rates consistent with a truth-wins paradigm.

Recent experimental studies indicate the tendency of groups to outperform individuals in situations involving decisions and games. Blinder and Morgan (2005) ran two laboratory experiments—one a statistical urn problem, the other a monetary experiment—to test the hypothesis that groups make decisions more slowly than individuals. In both experiments, groups were just as quick as individuals to reach decisions and outperformed individuals on average. Kocher and Sutter (2005) find that groups learn more quickly than individuals in a beauty contest (guessing game). Cooper and Kagel (2005)

[3]See Cooper and Kagel (2005) for a discussion of this issue.

study the behavior of individuals and two-player teams in signaling games; teams consistently play more strategically than individuals do and in some cases, the improvement exceeds the truth-wins (order-statistic) norm. Our study indicates that larger groups tend to commit fewer errors. This is true for every variation in the decision task, whether simple or complex, that we test.

With one exception, we find that reducing complexity tends to reduce errors. Only the elimination of the need to perform Bayesian updating (while still requiring subjects to contend with a compound lottery) results in a higher error rate for both individuals and pairs. We offer a tentative explanation for this anomaly, based on the *representativeness bias* (see Kahneman and Tversky 1972). This finding needs to be replicated before concluding that this is indeed a phenomenon to contend with and not merely a fluke.

The remainder of the paper is organized as follows. In the next section we describe the hypotheses and experimental design. In Section 2 we summarize the findings. We offer some discussion of our results in Section 3, and the conclusions appear in Section 4.

1 The hypotheses and the experiment

1.1 Hypotheses

Our basic premise is that all the tasks presented to the participants in our experiment are choice problems that have right and wrong answers. Specifically, choices violating monotonicity with respect to first-order stochastic dominance are bad, representing decision-making errors. Furthermore, these errors are more likely to occur when the choice is accompanied by affect, when the stochastic order is less transparent and when the choices are made in isolation.

To test these hypotheses, we began by confronting the subjects with a two-stage process. In the first stage, the subjects acquired information, and in the second stage they were asked to choose between two lotteries (risky prospects). If utilized correctly—that is, by the application of Bayes' rule—the information acquired in the first stage permits the subject to order the lotteries that figure in the second stage, by first-order stochastic dominance.

The presence of the information-acquisition stage simultaneously introduces several factors that may contribute to decision making errors. First, if the information acquired is considered to be "good news" or "bad news" it might be accompanied by an emotional response (affect) blurring the difference between the subsequent alternatives and biasing the choice in a predictable direction, resulting in violations of monotonicity with respect to first-order stochastic dominance. Second, to the uninitiated, the need to update the beliefs using Bayes' rule is confusing and blurs the stochastic order of the risky prospects encountered in the second stage.

In the experiment described below we start by confronting the subjects with tasks that confound all these issues. We then disentangle the issues by first removing affect and then, successively, increasing the transparency of the task. Finally, to test for the effect of social interaction, we introduce group decisions and observe and compare the choices made by groups consisting of two and three individuals to those made by solitary decision makers.

1.2 Experimental design

We conducted Web-based sessions on the UCSB campus, with students recruited by e-mail from the same general student population (but with different students) as in Charness and Levin (2005); approximately 65% of the participants were female. Sessions lasted about 45 minutes on average. Participants met in the lab and were given a handout explaining the experimental set-up; detailed, hands-on instructions were provided on the Web site, and participants were required to correctly answer questions testing comprehension.

In our design, there are two equally likely states of the world, Up and Down, and two lotteries (Left and Right) consisting of two different sides of the screen, from which the individual can draw face-down 'cards' that may be either black or white. There is always a mix of colors on the Left side, while the Right side has cards of only one color. Throughout the experiment,[4] only black cards have value, that is, black cards are assigned positive payoff, while white cards pay nothing. In the state Up, the Left side has four black cards and two white cards, while the Right side has six black cards. In the state Down, the Left urn has two black cards and four white cards, while the Right urn has six white cards.

	Left Side	Right Side
Up ($p = .5$)	●●●●○○	●●●●●●
Down ($p = .5$)	●●○○○○	○○○○○○

Our design includes four different treatments. We refer to the first, most complex, treatment as $ABCD$, where A signifies the potential presence of *affect*, B the need for *Bayesian updating*, C the presence of *compounding* effect, and D the ranking by first-order stochastic *dominance*. In this treatment, subjects are asked to select one of six cards on either the left or right side of the screen. The card is then revealed and replaced, with the face-down cards on the screen then 'shuffled'. The subject then selects a second card, knowing that the state (Up or Down) remains the same as for the first card, and this card is then revealed; people were paid one unit of experimental currency for each black card drawn.

We are primarily interested in the choice behavior when the first draw is from the Left urn, as previous evidence indicates that the error rate is very low in all cases when the first draw is from the Right urn. To ensure receiving relevant observations for each individual, during the first 20 (out of 60) periods

[4] Except in the BCD treatment, described below, where the paying color is not determined until after the first draw.

in the session we required the first draw to be from the Left side in the odd-numbered periods and from the Right side in the even-numbered periods.

After drawing a black card from the Left side, a second draw from the Right side stochastically dominates a second draw from the left side. Therefore, the subject should switch to the Right side for the second and final draw of the period. (Updating his prior according to Bayes' rule, the subject's probability of success, that is, winning a unit of the experimental currency, from the second draw are: 5/9 if he draws again from the Left side, and 6/9 if his second draw is from the Right side). By the same logic, after having drawn a white card from the Left side, the decision maker should stay with the Left side for the second draw of the period. (Updating his prior by Bayes' rule, the subject's probability of success in the second draw are 4/9 if he draws again from the Left side, and 3/9 if he draws from the Right side).[5]

A subject's choice for the second draw may be biased by his emotional response to the observed outcome. A black card in the first draw from the Left represents success (and an award), and may induce a sensation of optimism reinforcing the Left side choice. In other words, having just won a prize drawing from the Left, the subject may consider this to be his lucky side, and be reluctant to switch. On the other hand, drawing a white card first may instill a sensation of disappointment (failure) with the first Left draw and "push" the subject to switch sides for the second draw. To the extent that such heuristics exist and influence decisions, they may exacerbate the violations of monotonicity with respect to first-order stochastic dominance.

In order to mitigate the potential influence of such sensations, we implemented a treatment, dubbed *BCD* (the *Affect* being removed), in which the first draw from the Left was not associated with a payoff. Moreover, to counteract the possible presence of a more subtle emotional response that might accompany success or failure when observing the first draw, we did not tell the subject whether it is the black or white card that has a positive payoff in the second draw until after the first card was drawn. This was feasible due to the symmetry of the distribution of black and white cards across the Up and Down states. People were required to make their initial draws from the Left side, observing the color of the card before replacement; there was no payment for the first draw. At the bottom of the screen, the subject was then informed as to the color that would pay on the second draw.

In our third treatment, *CD*, we eliminate the need for Bayesian updating while preserving the compound-lottery structure involving combining continuation probabilities after different initial outcomes. In this treatment prizes were awarded for drawing black cards during 80 periods. As before, after a first black draw from the Left the correct Bayesian posterior probability of

[5] In the case of an initial black draw, there is a 2/3 chance that the state is Up. If the state is indeed Up, then by drawing from the Left side, the subject has a 2/3 chance of drawing another black card. If the state is Down (probability 1/3), then the subject has a 1/3 chance of drawing a black card. Thus, the expected value is 2/3*2/3 + 1/3*1/3 = 5/9. However, by choosing from the Right side for the second draw, the subject has a 2/3 (6/9) chance of drawing a black ball. The calculations are similar for the case of an initial draw of a white card.

Up is 2/3 and that of Down is 1/3. In order to eliminate the need for Bayesian updating, we simplified the procedure so that there is only one draw in a period. We offer subjects a choice between the Left and Right side, informing them that the probability of state Up is 2/3 (the same as in the BCD treatment after a would-be successful first draw); this matches the case when a would-be-successful color is drawn first in the BCD treatment. Note that to compute overall probabilities a subject must analyze, for the Left side, a compound lottery involving different states of Nature.

The fourth treatment, D, focuses on dominance. In this treatment, done on paper, subjects were offered a reduced lottery instead of the compound lottery mentioned in the preceding paragraph. Subjects were presented with a choice of choosing the Left or Right side after being informed that we would roll a nine-sided die (actually ten-sided, but if 0 came up we rolled again); if a subject chose Left, a prize of $2 would be awarded if numbers 1-5 came up, while if the decision maker chose Right, this prize would be awarded if numbers 1-6 came up. We handed out a sheet of paper with this choice at the end of BCD or CD sessions, rather than repeating the choice multiple times. This procedure is equivalent to the case (in the more complex treatments) where the first draw has been made from the Left side and was successful, as there is the same 5/9 probability of success by drawing again from the Left side and a 6/9 chance of success by instead drawing from the Right side.

The procedure that we chose to follow in these experiments, that is, starting with the $ABCD$ treatment rather than the simpler ones, reflects our view that decision making under risk and under uncertainty, typically involves all, or some, of these factors. One makes a choice and experiences an emotional response to the outcome; this outcome also generates information that necessitates updating the priors, and the calculations may also involve compound lotteries.

1.3 Individual and group decisions

The second part of our study examines the social-interaction effects on decisions. Our interest in this issue stems from the obvious fact that most real-life decisions take place in environments in which individuals can, and do, consult and seek advice before making decisions, thereby benefiting from the experience and expertise of others. We hypothesize that the opportunity to exchange ideas and opinions improves the decision making ability of individuals. To test this hypothesis, we administered some of the treatments with groups instead of individuals acting as decision making units. Specifically, we repeated treatments $BCD/CD/D$ and CD/D with pairs of subjects making joint decisions, and treatments $BCD/CD/D$ with groups of three subjects making decisions.[6] The subjects in each group were permitted to speak to each other (quietly) and could reach decisions in any manner, while experimenter

[6] We have not investigated group effects for the ABCD treatment, as we are more interested in simpler environments and trying to identify where the truth-wins paradigm breaks down.

intervention (e.g., a coin flip) was available if needed. In practice, no group ever failed to reach a decision on its own.

To grasp our hypotheses, consider the case of the *BCD* treatment. In a population of subjects an unknown proportion, say x, of the subjects are Bayesians (that is, subjects who know and are able to apply Bayes' rule). We hypothesize that, without fail, Bayesians choose the dominating option. The rest of the subjects, whose proportion in the population is $(1-x)$, employ other means to assess the alternative risks and make their choices. Among these subjects, a certain proportion, say $(1-\alpha_1)$, may actually "guess" the right choice and pick the dominating alternative, without using Bayesian formula. Thus, $(1-\alpha_1)(1-x)$ is the proportion of non-Bayesian subjects in the population that, by chance or by intuition, made the right choice. Clearly, $\alpha_1(1-x)$ is the proportion of non-Bayesian subjects that made the wrong selection.

We are reluctant to speculate on what might be reasonable values for x or α_1. In particular, if the non-Bayesian participants are as likely to make the right decision as to make the wrong decision, then $\alpha_1 = 0.5$. However, it is conceivable that, even if they do not know how to apply Bayesian rule, subjects may nevertheless intuitively choose correctly more often than not. Thus, we hypothesize that $\alpha_1 < 0.5$.

We propose here an hypothesis that would allow us to estimate the values of x and α_1, and test them. To illustrate how this is done, suppose that Bayesian decision makers do not err. Then observing the proportion of wrong choices, say k_1, in the individual experiments, we learned that:

$$\alpha_1 (1-x) = k_1 \tag{1}$$

Turning next to decisions by groups, we hypothesize that whenever matched with non-Bayesians, the Bayesians are able to convince them of the right way of looking at the (posterior) alternatives and, consequently, choose the right alternative—this would be the case if the truth-wins norm applies. We also hypothesize that, even if Bayesians are not represented in a group, the mere deliberation of the alternatives among members of a group, tends to increase the number of right choices. Formally, let k_2 be the proportion of wrong choices by groups consisting of pairs of individuals. Then,

$$\alpha_2 (1-x)^2 = k_2, \tag{2}$$

where it might seem a natural presumption that $\alpha_2 \leq \alpha_1$; we test this presumption later. That is, k_2 is the proportion of non-Bayesians that were randomly matched with other non-Bayesians and who jointly chose the wrong answer.

If we alternatively insist that $\alpha_2 = \alpha_1 = \alpha$, these two equations allow us to estimate (solve) simultaneously for x and α. Call these estimates x^* and α^*. But if in fact, $\alpha_2 < \alpha_1 \leq 0.5$, then x^* overestimates the true value of x, and

Table 1 Treatment summary—individuals

Treatment	First draw restriction	Payment	Participants	Periods
$ABCD$[a]	In first 20 periods	Black cards from both draws	57	60
BCD	Left draw only	Second draw only, color TBD	57[b]	80
CD	None	Black cards	48	80
D	None	Black card	104	1

[a] As mentioned earlier, we only consider $ABCD$ decisions made after a first draw from the Left urn.

[b] One person played solo for the 30 periods of the three-person BCD treatment.

α^* overestimates α_1 and α_2, where the overestimation of α_2 is relatively larger than that of α_1.[7]

Observe next that in groups of three subjects, the proportion of wrong choices should satisfy $k_3 = \alpha_3 (1-x)^3$, where x is the true proportion of Bayesians in the population and $\alpha_3 \leq \alpha^* \leq \alpha_2$. Our hypothesis does not indicate unambiguously the relationship between the magnitude of k_3 and $\alpha^* (1-x^*)^3$. To see why, note that

$$\alpha_2 (1-x)^2 = k_2 = \alpha^* \left(1-x^*\right)^2 . \tag{3}$$

Thus, solving α^* from Eq. 3, gives

$$\alpha^* \left(1-x^*\right)^3 = \frac{\alpha_2 (1-x)^2}{(1-x^*)^2} \left(1-x^*\right)^3 = \alpha_2 (1-x)^2 \left(1-x^*\right) \gtreqless \alpha_3 (1-x)^3 = k_3, \tag{4}$$

where we used the inequality $\alpha_3 \leq \alpha_2$ and that x^* overestimates x. However, $k_3 \leq \alpha^* (1-x^*)^3$ implies that $\alpha_3 < \alpha_2$. Hence, the last inequality, that non-Bayesian trios make better guesses than non-Bayesian pairs, is a testable hypothesis.

In the CD and D treatments, the need for Bayesian updating is removed. That is, it can be modeled along the same lines as above assuming that $x = 0$. Note, however, that because the treatments are less complex, it is not necessarily true that the proportions of subjects that make the wrong selection remain the same across treatments. It seems reasonable to suppose that this proportion declines when the stochastic order becomes more transparent. Formally, let α_i^{CD} and α_i^{D}, $i = 1, 2, 3$ denote the proportion of subjects that

[7] To see this, note that $\alpha_i = \alpha$, $i = 1, 2$ implies that

$$\frac{k_2}{(1-x)^2} = \frac{k_1}{1-x}$$

thus $x^* = 1 - \frac{k_2}{k_1}$. If $\alpha_1 > \alpha_2$ then, by the same reasoning, $\hat{x} < 1 - \frac{k_2}{k_1}$, where $\hat{x}$ is the true value of x. Thus, $\alpha_i^* = k_i / (1-x^*)^i > k_i / \left(1-\hat{x}\right)^i = \alpha_i$, $i = 1, 2$. Hence,

$$\frac{\alpha_2^*}{\alpha_1^*} = \frac{k_2}{k_1 (1-x^*)} > \frac{k_2}{k_1 \left(1-\hat{x}\right)} = \frac{\alpha_2}{\alpha_1}.$$

Table 2 Treatment summary—groups

Treatment	First draw restriction	Payment	Groups	Periods
BCD-2 person	Left draw only	Second draw only, color TBD	30	30
BCD-3 person	Left draw only	Second draw only, color TBD	21	30
CD-2 person	None	Black cards	36	30
D-2 person	None	Black card	66	1
D-3 person	None	Black card	21	1

make the wrong selection in the CD and D treatments, respectively, rather than in the BCD treatment. We hypothesize that $\alpha_i > \alpha_i^{CD} > \alpha_i^{D}$, $i = 1, 2, 3$.

We summarize our treatments with individual decision makers in Table 1, and those with groups in Table 2.

Participants were paid \$0.30 for each successful draw in the $ABCD$, BCD, and CD treatments. In order to pay reasonably similar amounts across treatments, we had 80 periods in the individual BCD and CD treatments rather than the 60 periods in the $ABCD$ treatment (80 decisions counted, compared to 120) and also included the D treatment at the end of the session. No person could participate in more than one session or treatment (with the exception of the BCD/D and CD/D sessions). Group decisions took longer, so we reduced the number of periods to 30 and increased the show-up fee (to \$8 from \$5) in these sessions.[8] On average, our participants earned approximately \$16–17 for a 1-h session.

2 Results and analysis

2.1 Individual decisions

We present our findings regarding individual decision making according to the treatments. The aggregate results are summarized in Table 3 below.

Individual treatment ABCD: For this treatment we document two types of first-order stochastic dominance violations: (1) Violations that follow success (namely, a draw of a black card) and receiving a reward in the first round and (2) violations that follow failure (namely, a draw of a white card) and not receiving a reward in the first round. Recall that monotonicity with respect to first-order stochastic dominance implies that success should have resulted in the subject switching to the Right lottery for the second draw. Violations in this instance mean that the subject didn't switch. Similarly, after a first draw of a white card (representing a failure), subjects should have stayed with the Left

[8] We chose to maintain the same marginal benefit for choosing the right color, rather than keeping the same show-up fee and changing the marginal incentives. While it is possible (although unlikely) that the slightly larger show-up fee caused participants to take the task less seriously, this would only serve to lessen the difference between the error rates in individual and group decisions.

Table 3 The violation rates of first-order stochastic dominance, individual decisions

Treatment	Following success	Following failure	Aggregate error %
ABCD	37.5% (225/600)	40.0% (234/585)	38.7% (459/1,185)
BCD	18.8% (420/2,234)	39.9% (908/2,276)	29.4% (1,328/4,510)
CD	30.2% (1,158/3,840)	–	–
D	8.7% (9/104)	–	–

lottery for the second draw. Violations in this case mean that the subjects did switch. The aggregate violation rate of the first type, that is, following success is 37.5% and the aggregate violation rate of the second type, that is, following failure is 40.0%; there is no significant difference between these rates. The combined aggregate error rate is 38.7%.[9]

Recall that the implementation of the *ABCD* treatment involved the forced and voluntary choice of the side of the first draw. Specifically, the first 20 periods are forced first draws (alternating Left and Right) and the last 40 periods are voluntary first draws. The results of the forced and voluntary treatments displayed separately, are as follows: The error rate with the forced-first-draw is 34.5%, (101/293), following success and 40.4% (111/277) following failure; the error rate with the voluntary-first-draw is 40.4% (124/307) following success and 39.9% (123/308) following failure.

Perhaps the most important, and striking, aspect of these findings is the high rate of violations of monotonicity with respect to first-order stochastic dominance. A second aspect of these findings is the remarkable similarity of the error rates following success and failure. This suggests that success induces a tendency to erroneously "stay the course" while failure induces a tendency to, again erroneously, change course. These tendencies may reflect a bias due to affect. To test for the effect of affect, we remove it in the next treatment.

Individual treatment BCD: Charness and Levin (2005) observed a large reduction in aggregate error rates, from 47.0 to 28.2%, when the affect is eliminated; the reduction was particularly dramatic in the case of a successful (same color as the one paying on the second draw) first draw. There the reduction is from 36.8 to 13.5%. In both of these cases the reductions were strongly statistically significant.

In our experiment, the overall reduction in error rates after removing the affect is smaller, yet still large (from 38.7 to 29.4%). The difference between these two rates is marginally significant in a conservative ranksum test using each individual's overall rate as one observation ($Z = 1.34$, $p = 0.090$, one-tailed test). Remarkably, this reduction is due entirely to the very large

[9] This is considerably lower than the aggregate error rate in the *ABCD* treatment of Charness and Levin (2005). In that experiment, however, the amount paid for a successful draw differed for the two urns, rendering the decision somewhat more complex. Due to this differential payment, not all Bayesian errors in Charness and Levin (2005) reflected violations of first-order stochastic dominance.

reduction following success (from 37.5 to 18.8%).[10] There is no noticeable reduction following failure. These distinct tendencies are puzzling and seem to merit further study.[11] Moreover, this rather large drop in the error rate following success is responsible for some other anomalies in the findings, to be discussed below. Finally, the binomial test shows that the error rate following "failure" is significantly higher than the error rate following "success" ($Z = 4.81$, $p = 0.000$), since this was higher for 44 individuals and lower for 9 individuals (and the same for the other four people in this treatment).

Individual treatment CD: In going from the BCD to the CD treatment we eliminated the need for Bayesian updating. That is, subjects are told that the probability of Up is 2/3. It's "as if" they made a successful first draw and now face the second decision. However, in this treatment there is no actual first draw, and the one draw that remains is equivalent to the second draw in treatments $ABCD$ and BCD.

It is well documented in the psychological and experimental economics literature that subjects are not good Bayesians.[12] This suggests that eliminating the need for Bayesian updating would have reduced the error rate. In fact, however, the error rate *increased* significantly from that in the BCD treatment (from 18.8 to 30.2%; $Z = -2.40$, $p = 0.017$, in a two-tailed ranksum test).

Individual treatment D: The last treatment simplifies the task further by eliminating the complication due to compounding. This enables us to observe, directly, how decision makers chose between two lotteries with probabilities of 5/9 and 6/9, respectively, of winning the same prize. Only nine individuals out of 104 (8.7%) chose Left.[13] This number is small enough to suggest that the violations are due to errors and misunderstandings, and don't actually reflect the participants' preferences. Checking each individual who made choices in D (as well as decisions in BCD or CD), the D error rate is lower than the

[10] The reduction is statistically significant, with $Z = -2.92$, and $p = 0.004$, two-tailed test.

[11] While we cannot be certain why the effect of affect was smaller here than in Charness and Levin (2005), we note that the corresponding ABCD and BCD treatments in that experiment were slightly different, in that people received 7/6 unit for successful draws from the Right side (compared to 1 unit for a successful draw from the Left side). This added a layer of complexity that appears to have affected behavior (note the reduction in the overall ABCD error rate from 48.3 to 38.7%). Perhaps affect is less determinative with simpler problems. In any case, the reduction after success is similar in both cases; the puzzle is the lack of reduction in the error rate after failure, compared to the modest reduction found in Charness and Levin (2005).

[12] See, for example Tversky and Kahneman (1971, 1973), Kahneman and Tversky (1972), and Grether (1980, 1992). More recent studies (e.g., Ouwersloot, Nijkamp and Rietveld 1998; Zizzo et al. 2000) also provide strong evidence that experimental subjects are often not even close to being 'perfect Bayesians'.

[13] One person admitted he hadn't read the instructions. A colleague also suggests that some subjects may simply have been "jerking our chain" by deliberately choosing the dominated strategy.

Table 4 The violation rates of first-order stochastic dominance, group decisions

Treatment	Following success	Following failure	Aggregate error %
BCD-1	18.8% (420/2,234)	39.9% (908/2,276)	29.4% (1,328/4,510)
BCD-2	15.4% (67/434)	32.6% (152/466)	24.3% (219/900)
BCD-3	7.5% (23/307)	10.8% (35/323)	9.2% (58/630)
CD-1	30.2% (1,158/3,840)	–	–
CD-2	23.0% (248/1,080)	–	–
D-1	8.7% (9/104)	–	–
D-2	3.0% (2/66)	–	–
D-3	0.0% (0/21)	–	–

BCD error rate for 35 subjects and is higher for nine subjects; the D error rate is lower than the CD error rate for 35 subjects and is never higher for anyone. The binomial test finds $Z = 3.18$ ($p = 0.001$) and $Z = 5.92$ ($p = 0.000$) for the respective differences, both significant.

2.2 Group decisions

We turn next to the evidence concerning group decisions. As mentioned earlier, the members of each randomly-formed group consulted with each other before making a group choice. Each group was left to its own devices as to the decision process.

The main findings are summarized in Table 4. Our findings indicate a general and significant drift towards a reduction of the error rates as the group size increases.

Group treatment BCD: With groups of two subjects, the error rate under the BCD treatment is lower than with individual decision makers after both success and failure, and the aggregate error rate is reduced by over five percentage points, to 24.3%. However, this difference is not statistically significant.

The reduction is considerably more dramatic with groups of three subjects. In this case, the error rates following both success and failure, are less than half the corresponding rates with two-subject groups, and the aggregate error rate is only 9.2%, less than that for either individual or paired decision makers; ranksum tests give $Z = 3.57$ ($p = 0.000$) and $Z = 2.96$ ($p = 0.003$) for the respective differences, both statistically significant. While the improvement is considerable for decisions following failure and following success, the effect is strongest following an unsuccessful first draw.

We note that the error rate after failure in the first draw is higher than that after a success in each of the four cases in Tables 3 and 4. A simple binomial test indicates that this will happen by chance with $p = 0.062$, on the margin of statistical significance.

Group treatment CD: This treatment indicates that groups are less prone to violations of first-order stochastic dominance than are individuals; in this case, the error rate for two-subject groups is 23.0%. As with all comparisons across individual and group decision-making entities, this rate is lower than

Table 5 Regression of individual error rates across conditions

	Coefficient	Std. error	Z-statistic	$P > \|Z\|$
CD	0.1228	0.0343	3.58	0.000
D	−0.0984	0.0267	−3.68	0.000
Pair	−0.0637	0.0277	−2.30	0.021
Trio	−0.0975	0.0428	−2.28	0.023
Constant	0.1832	0.0254	7.20	0.000

Number of obs. = 384; Adjusted $R^2 = 0.1416$

the corresponding 30.2% rate for individuals. The difference between these two rates is marginally significant in a ranksum test using each individual's overall rate as one observation ($Z = 1.37$, $p = 0.085$, one-tailed test). We have no evidence concerning the possible effect of increasing the group size from two to three subjects, although the evidence from the BCD treatment suggests the hypothesis that such an increase will reduce the error rate significantly.

As with individual decision makers, the error rate for pairs is lower in the BCD treatment than in the CD treatment (15.4 versus 23.0%). However, the difference between these two rates is at most marginally significant in a ranksum test using each pair's overall rate as one observation ($Z = 1.42$, $p = 0.156$, two-tailed test[14]); of course, this is a rather conservative test.

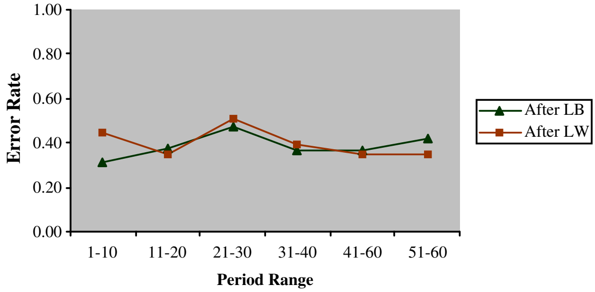
Group treatment D: In this treatment, group decision making nearly eliminated the errors found in the individual case. With groups consisting of two subjects, the error rate was down to 3.0%; more telling, perhaps, is the 0% error rate observed for groups consisting of three subjects. Checking each pair who made choices in D as well as decisions in BCD or CD, the D error rate is lower than the BCD error rate for 12 pairs and is higher for one pair, while the D error rate is lower than the CD error rate for 24 pairs and is higher for one pair. The binomial test finds $Z = 3.05$ ($p = 0.001$) and $Z = 4.60$ ($p = 0.000$) for the respective differences from randomness. Checking each trio who made choices in D as well as decisions in BCD, the D error rate is significantly lower (at $p = 0.001$) than the BCD error rate for 13 groups and was never higher. The binomial test finds this one-sidedness is significant ($Z = 3.61$, $p = 0.000$).

These very low error rates for the groups tend to confirm our claim that the observed errors are due to mistakes rather than genuine expressions of the participants' preferences.

We perform a simple regression to illustrate in a more comprehensive manner the significance of the difference in error rates across treatments (to keep the comparison fair, we only include successful draws in BCD), as well as across group size. We use each decision-making entity's overall error rate in

[14]We cannot use a one-tailed test, as we have no *ex ante* prediction that error rates will increase going from *BCD* to *CD*.

Fig. 1 Error rate over time in ABCD



the BCD or CD treatment (whichever applies) as one observation and count the D choice (if any) as another observation, with standard errors clustered by individual. We regress this error rate against various dummies (Table 5).

We see that in relation to the BCD baseline, the error rate in the CD treatment is significantly higher and the error rate in the D treatment is significantly lower. The error rate with larger groups (compared to the individual baseline) is significantly lower, with the negative coefficient for the trio dummy higher than the coefficient for the pair dummy.

Since we have differing number of periods for group and individual decisions, this could potentially cause a problem in our comparison. However, choices are rather stable over time, as can be seen in Figures 1, 2, 3, 4, 5 and 6. Furthermore, if the error rate is decreasing over time, having more periods in the group treatments would only serve to increase the difference between the error rates for individual and group decisions.

3 Discussion

We turn next to an interpretation of these findings based on the model of Section 1.3. Consider the treatment BCD. Our findings indicate that $k_1 = 0.294$ and $k_2 = 0.243$. If we presume that $\alpha_2 = \alpha_1 = \alpha$, calculations show

Fig. 2 Error rate over time in BCD-1

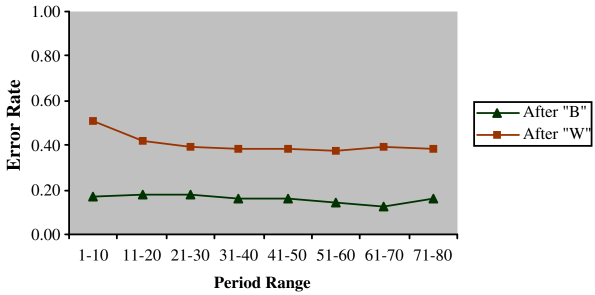


Fig. 3 Error rate over time in BCD-2

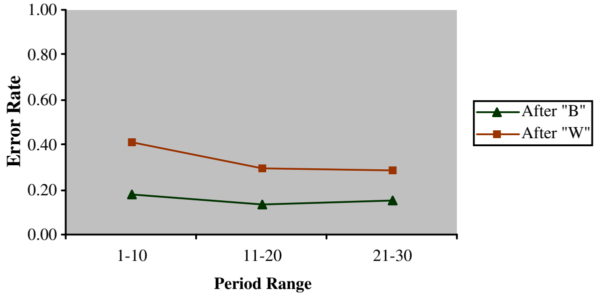


Fig. 4 Error rate over time in BCD-3

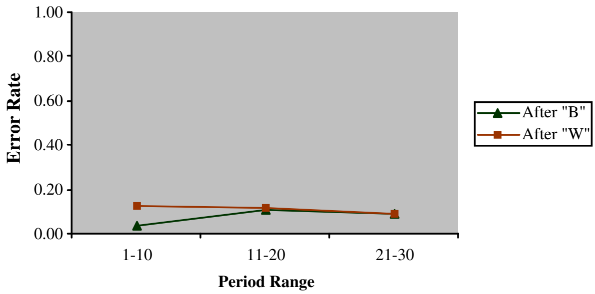


Fig. 5 Error rate over time in CD-1

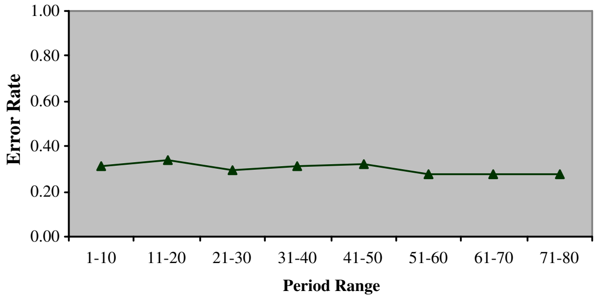
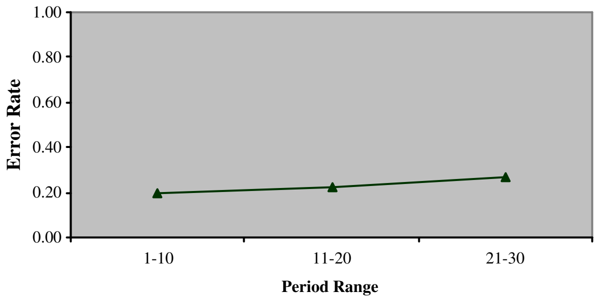


Fig. 6 Error rate over time in Cd-2



that $x^* = 0.173$ and $\alpha^* = 0.356$.[15] Note, however, that x^* overestimates the value of x, the proportion of Bayesian decision makers in the subject population, and α^* overestimates α_1 and α_2, where the overestimate of α_2 is relatively larger than that of α_1. This lends support to the hypothesis that $\alpha_1 < 0.5$. Moreover, we argued that $\alpha^* (1 - x^*)^3 \geq k_3$ implies that $\alpha_3 < \alpha_2$. But,

$$k_3 = 0.092 < 0.201 = 0.356\,(1 - 0.173)^3\,.$$

Hence, our hypothesis that $\alpha_3 < \alpha_2$ is also supported by our findings. There are a number of different statistical tests that could be used to test for the degree of significance. First, we note that this is equivalent to testing how certain we are that our observed value of 0.092 for k_3 (in Table 4) indicates a true value smaller than 0.201. One approach considers the error rate for each trio making decisions. Using a standard t-test on the 21 trios, we can reject a true mean of 0.201 ($p = 0.001$, one-tailed test). However, the distributional assumption of this test may be questionable; in fact, while the mean error rate was 0.092, the median error rate for trios in BCD was only 0.033 (one error out of 30 chances), with seven groups with error rates above 0.092 and 14 groups with error rates below it. Thus, we instead use bootstrapping to construct a 95-percent confidence interval around the observed mean (0.041 to 0.149). Clearly our calculated value for k_3, 0.201 falls well outside this range.

Consider next the CD and D treatments. Observe that $\alpha_1^{CD} = 0.302 > \alpha_2^{CD} = 0.230$ and $\alpha_1^{D} = 0.087 > \alpha_2^{D} = 0.030 > \alpha_3^{D} = 0$. Moreover, $\alpha_i^{CD} > \alpha_i^{D}$, $i = 1, 2$. All of these inequalities are consistent with our hypothesis. Since these respective differences consist of five directional tests all going in the predicted direction, the likelihood that this would occur by chance is $p = 0.031$.

Finally, calculating α_1 and α_2 using the error rate following success in the BCD treatment, we obtain $\alpha_1 = 0.228$ and $\alpha_2 = 0.226$.[16] Since x^* is an overestimate, the estimates of α_1 and α_2 also overestimates the true values. Thus, $\alpha_i < \alpha_i^{CD}$, $i = 1, 2$. This is not what we expected and it goes against the hypothesis stated earlier. It is related to the puzzle, pointed out before, that the error rates following success under the BCD treatment exceed those under the CD treatment.

The increased aggregate error rates, for both individuals and pairs, associated with the simplification from BCD to CD warrants further examination. If this turns out to be a real phenomenon rather than a fluke, one possible explanation might be the influence of the *representativeness bias*, which occurs when subjects pay too much attention to the sample (recent) information, wrongly believing that it represents the true distribution and so effectively neglecting the (given) base rate. For example, after a success (failure) the probability of Up may seem really close to 1 (0). If this bias is sufficiently

[15] This is the solution to the equations: $\alpha\,(1 - x) = 0.294$ and $\alpha\,(1 - x)^2 = 0.243$.

[16] The restriction to error rates following success is for the purpose of comparison with the other treatments. The estimated values are based on the solution of the equations $\alpha_1\,(1 - 0.174) = 0.188$ and $\alpha_2\,(1 - 0.174)^2 = 0.154$, where $x^* = 0.174$ and the error rates, 0.188 and 0.154 are from Tables 3 and 4, respectively.

strong and dominates the effect of the simplification of the task that occurs when moving from BCD to CD, eliminating it could explain the higher error rate in the CD compared to that of the BCD treatment, as giving the correct posterior may help to silence the bias.[17] This argument also assumes that the representativeness bias dominates a "conservatism" bias (see Edwards 1982) that occurs when the subject ignores, or "under-updates," the prior distribution given new evidence, instead holding on to the base rate. There is considerable evidence that subjects underweight base rates in this way,[18] and our results after success are consistent with this notion. We noted earlier that the representativeness bias should also reduce the error rate after an initial failure (since the probability of Up seems like 0 there is no point in trying the Right urn); in fact, the only comparison we have across treatments, $ABCD$ versus BCD in Table 3, shows dramatic improvement from removing the affect after an initial success, but no improvement after an initial failure. Perhaps people react asymmetrically to good news and bad news.[19]

Note that the beneficial effect of increasing group size is present, with improvement from 39.9 to 32.6 to 10.8% for individuals, pairs, and trios, respectively (Tables 3 and 4). In fact, we observe, in every case, a monotonic relationship between group size and the violation rate. Yet our results are not completely consistent with a truth-wins norm, as it appears that α does vary across conditions of complexity and group size. In principle, this norm predicts that we should see more improvement in the error rates when going from individuals to pairs than when going from pairs to trios; while this is true in the D treatment, it is not at all true in the BCD treatment. Thus, perhaps the less complex tasks are more like flashes of insight and are less deliberative, while deliberation plays a greater role in more complex environments.

4 Conclusions

The two most striking findings of our study are: First, there is a substantial number of individuals who, when making decisions in isolation, violate the most basic tenet of rational decision making by choosing first-order stochastically dominated alternatives. Second, the incidence of choice of dominated alternatives is negatively correlated with the size of the group; in fact, every comparison across group size favors the larger group. Thus, social

[17] Using the methodology of modern experimental economics, Grether (1980) finds evidence of the representativeness bias in a task involving identifying the urn from which a ball was drawn. El-Gamal and Grether (1995) show that people are more prone to the representativeness bias than the conservatism bias. Ganguly, Kagel and Moser (2000) discuss a similar effect. This may help explain why we don't see a big drop going from treatment BCD to treatment CD, but the observed *increase* is still surprising.

[18] See for example Kahneman and Tversky (1972), Tversky and Kahneman (1973) and Ganguly, Kagel and Moser (2000). Interestingly, the latter study finds that a more abstract context (such as ours) tends to *reduce* this bias.

[19] Charness and Levin (2005) also find this asymmetric improvement from removing affect.

interaction, whether through the presence of experts or consultation, improves the decision-making process in the laboratory.

Typically, important individual decisions are made in social environments, enabling the decision makers to discuss their options and benefit from the experience and expertise of others. Therefore, the two conclusions above lend support to the claim that the experimental evidence indicating *systematic deviations from the courses of action prescribed by normative models of decision making under risk and uncertainty, such as expected utility theory, are due, in part, to the artificial isolation imposed by the experimental setting*. These violations tend to be less pronounced when social interaction is allowed. Our findings also suggest that, in addition to the presence of experts (Bayesian decision makers, in our particular context) the mere opportunity to discuss the alternatives and to clarify their meaning is important. To the extent that this also prevails in the field, perhaps a group decision-making environment may be more representative of non-laboratory conditions than one where individuals operate on their own.

Third, the presence of affect, the need to apply Bayesian updating and the complexity of the alternatives tend to contribute to erroneous judgments and poor decisions. This is in line with earlier results obtained by Charness and Levin (2005). Other evidence consistent with this finding is reported in Kagel and Levin (1986) and Charness and Levin (2006).

The results of this paper also raise some issues that merit further investigation. First, the systematic higher rates of violations of first-order stochastic dominance following a failure compared to those following success is puzzling and further examination of this tendency is warranted. Second, the significant increase in the violation rate following a successful draw, when we have dispensed with the need to perform Bayesian updating, is troubling; further examination of this issue seems worthwhile. Third, we examine joint decision making by requiring that the group jointly reach a decision. An alternative test of the social-interaction effect should allow subjects to discuss their decisions and than make their choice individually. Finally, our study focused on violations of monotonicity with respect to first-order stochastic dominance. It would be interesting to see if a similar tendency for reductions in violations of the independence axiom or the sure thing principle obtain when social interaction is allowed.

Acknowledgements We thank two anonymous referees, Nick Burger, Asen Ivanov, Jim Peck, Tim Salmon, and in particular John Kagel for valuable comments. Charness thanks the Department of Economics at UCSB for funding, Karni thanks the Department of Economics at Johns Hopkins University for funding, and Levin thanks the National Science Foundation for funding.

References

Blinder, Alan and John Morgan. (2005). "Are Two Heads Better than One? An Experimental Analysis of Group versus Individual Decision Making," *Journal of Money, Credit, and Banking* 37, 789–811.

Charness, Gary and Dan Levin. (2005). “When Optimal Choices Feel Wrong: A Laboratory Study of Bayesian Updating, Complexity, and Psychological Affect,” *American Economic Review* 95, 1300–1309.

Charness, Gary and Dan Levin. (2006). “The Origin of the Winner’s Curse: A Laboratory Study,” WP UCSB.

Cooper, David and John Kagel. (2005). “Are Two Heads Better Than One? Team versus Individual Play in Signaling Games,” *American Economic Review* 95, 477–509.

Edwards, Ward. (1982). “Conservatism in Human Information Processing.” In Daniel Kahneman, Paul Slovic, and Amos Tversky (eds), *Judgement Under Uncertainty: Heuristic and Biases* 359–369, Cambridge, UK: Cambridge University Press.

El-Gamal, Mahmoud and David Gretcher. (1995). “Are People Bayesian? Uncovering Behavioral Strategies,” *Journal of the American Statistical Association* 90, 1137–1145.

Ganguly, Ananda, John Kagel, and Donald Moser. (2000). “Do Asset Market Prices Reflect Traders’ Judgment Biases?” *Journal of Risk and Uncertainty* 20, 219–245.

Grether, David. (1980). “Bayes’ Rule and the Representativeness Heuristic: Some Experimental Evidence,” *Journal of Economic Behavior and Organization* 17, 31–57.

Grether, David. (1992). “Testing Bayes’ Rule as a Descriptive Model: The Representativeness Heuristic,” *Quarterly Journal of Economics* 95, 537–557.

Kagel, John and Dan Levin. (1986). “The Winner’s Curse and Public Information in Common Value Auctions,” *American Economic Review* 76, 894–920.

Kahneman, Daniel and Amos Tversky. (1972). “Subjective Probability: A Judgment of Representativeness,” *Cognitive Psychology* 3, 430–454.

Kerr, Norbert, Robert MacCoun, and Geoffrey Kramer. (1996). “Bias in Judgment: Comparing Individuals and Groups,” *Psychological Review* 103, 687–719.

Kocher, Matthias and Matthias Sutter. (2005). “The Decision Maker Matters: Individual Versus Team Behavior in Experimental Beauty-contest Games,” *Economic Journal* 115, 200–223.

Ouwersloot, Hans, Peter Nijkamp, and Piet Rietveld. (1998). “Errors in Probability Updating Behaviour: Measurement and Impact Analysis,” *Journal of Economic Psychology* 19, 535–563.

Tversky, Amos and Daniel Kahneman. (1971). “Belief in the Law of Small Numbers,” *Psychological Bulletin* 76, 105–110.

Tversky, Amos and Daniel Kahneman. (1973). “Availability: A Heuristic for Judging Frequency and Probability,” *Cognitive Psychology* 5, 207–232.

Zizzo, Daniel, Stephanie Stolarz-Fantino, Julie Wen, and Edmund Fantino. (2000). “A Violation of the Monotonicity Axiom: Experimental Evidence on the Conjunction Fallacy,” *Journal of Economic Behavior and Organization* 41, 263–276.

Journal of Economic Perspectives—Volume 26, Number 3—Summer 2012—Pages 157–176

Groups Make Better Self-Interested Decisions†

Gary Charness and Matthias Sutter

A decision maker in an economics textbook is usually modeled as an individual whose decisions are not influenced by any other people, but of course, human decision-making in the real world is typically embedded in a social environment. Households and firms, common decision-making agents in economic theory, are typically not individuals either, but groups of people—in the case of firms, often interacting and overlapping groups. Similarly, important political or military decisions as well as resolutions on monetary and economic policy are often made by configurations of groups and committees rather than by individuals. Economic research has developed an interest regarding group decision-making—and its possible differences with individual decision-making—only rather recently. Camerer (2003) concludes his book on *Behavioral Game Theory* with a section on the top ten open research questions for future research, listing as number eight "how do teams, groups, and firms play games?" Potential differences between individual and group decision-making have been studied over the past ten to 15 years in a large set of games in the experimental economics literature.

In this paper, we describe what economists have learned about differences between group and individual decision-making. This literature is still young, and in this paper, we will mostly draw on experimental work (mainly in the laboratory) that has compared individual decision-making to group decision-making, and to individual

■ *Gary Charness is Professor of Economics, Department of Economics, University of California at Santa Barbara, Santa Barbara, California. Matthias Sutter is Professor of Experimental Economics, Department of Public Finance, University of Innsbruck, Innsbruck, Austria, and Professor of Economics, Department of Economics, University of Gothenburg, Göteborg, Sweden. Their email addresses are ⟨charness@econ.ucsb.edu⟩ and ⟨matthias.sutter@uibk.ac.at⟩.*

† To access the Appendix, visit http://dx.doi.org/10.1257/jep.26.3.157.

doi=10.1257/jep.26.3.157

decision-making in situations with salient group membership.[1] In a nutshell, the bottom line emerging from economic research on group decision-making is that groups are more likely to make choices that follow standard game-theoretic predictions, while individuals are more likely to be influenced by biases, cognitive limitations, and social considerations. In this sense, groups are generally less "behavioral" than individuals. An immediate implication of this result is that individual decisions in isolation cannot necessarily be assumed to be good predictors of the decisions made by groups. More broadly, the evidence casts doubts on traditional approaches that model economic behavior as if individuals were making decisions in isolation.

We focus on three main lessons in this paper. First, the use of rationality as a useful assumption for studying real-world economic behavior may not be as problematic as some have argued. In this context, what we mean by rationality is that cognitive limitations (in the sense of bounded rationality) apply less to groups and that groups engage in more self-interested behavior than do individuals. In fact, we find that such rationality applies pretty well to group decisions, and we argue that groups are at least an element in most decisions. People always belong to some groups (for instance, males or left-handed people, and the like) and their behavior may well be affected when a sense of group membership is present. In addition, many important economic decisions—including decisions where consequences affect individual decision units, such as buying a home or choosing a health insurance plan—are made after some consultations with others, even if they are not explicitly part of a group decision-making process. Thus, while the behavioralist critique of deviations from the rational paradigm is important and has many applications, we should be careful about how we describe economic agents in our models.[2] If we were to specify that most of these agents are acting in social or group contexts, then the claim that they are rational actors would be strengthened.

A second lesson is that, from a social point of view, group decision-making may be a method for individuals to try to protect themselves from the consequences of their own behavioral irrationalities or limitations. Suppose an individual is very present oriented and so has great difficulties in saving for retirement. Perhaps through participation in groups at work or in one's social, political, recreational, or religious life, one can achieve at least a modicum of success in assuring a retirement income. As another example, perhaps one does not have the self-discipline to exercise on one's own but will do so with regularity if one forms or joins a group of people who jog together or meet to play tennis. In a business environment, one might find it personally nearly impossible ever to fire anyone, even if the result is that one's business goes bankrupt. But it might be possible to achieve this end by being part of a committee that makes such decisions. In short, group membership

[1] The evidence from laboratory experiments has the advantage of allowing for a clean and controlled analysis of group decision-making and group membership effects because subjects are randomly assigned to making a choice individually or as a group member. This is more difficult with field data, but not impossible, as Lesson 2 below will confirm.

[2] See Levitt and List (2007) for an account of the behavioralist critique.

and group participation can facilitate people doing things that they wish (on some important level) to do, but might be unable to do without the support of a group.

The third lesson is that in some environments—for example, in cases where trust and cooperation lead to improved social welfare—it might make sense to have individuals making decisions, and in other cases—for example, when deeper levels of insight or analytical problem-solving and coordination are especially valuable—it might make sense to have groups making decisions. For example, a considerable body of experimental literature suggests that, perhaps because individuals are unselfish or socially oriented, they are able to reach welfare, socially efficient outcomes in situations like the prisoner's dilemma or a "trust game."[3] In these settings, group decision-making/membership presumably leads to lower social welfare (in the sense of the total social material payoffs), because the element of trust or cooperation is sharply reduced. However, we will explore a number of other settings where group decision-making is more sophisticated and effective. Thus, researchers can start groping toward a provisional taxonomy concerning where and when it is optimal to have a group process or an individual one.

We discuss these three lessons in the following sections. We intersperse the discussions with evidence, primarily experimental, for the story being told.[4] Building on this evidence, we then discuss the major sources for differences in decisions made by individuals and groups before we conclude with an outlook on promising avenues for future research on group decision-making.

Lesson One: Groups are More Cognitively Sophisticated

We look first at experiments that compare individual and group decisions where each player is only concerned with making the best selfish decision without regard to social considerations.[5] This category includes investment or portfolio decisions, tournaments, and tasks where the ability to reason through the problem is important due to some cognitive limitation or psychological bias that typically affects the outcome.

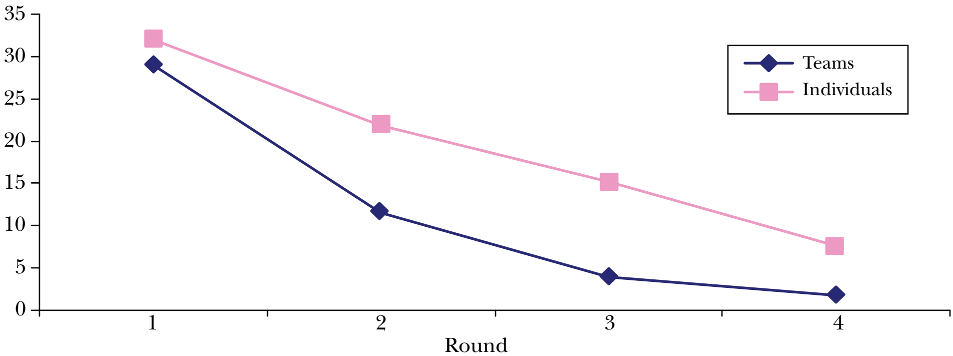
One well-known example is the beauty-contest game (also known as "the guessing game"). In this simultaneous move game, a set of n decision makers chooses a number from the interval $[0, 100]$, and the winner is the decision maker whose

[3] It may also be possible that an individual's social concerns are directed at one's group in the case of group membership. We shall take up this point later.

[4] While we only discuss a few studies in each section, we present a brief description of the most important other studies supporting our conclusions in an online Appendix available with this paper at ⟨http://e-jep.org⟩.

[5] We focus here on results from experimental economics, with less emphasis on psychological research: see Levine and Moreland (1998) for an account of small-group research in psychology. From our perspective, the research in experimental economics has two particular advantages: 1) the ubiquitous use of financial incentives, a condition that is often not met in experimental research in the field of psychology; and 2) the use of simpler paradigms that allow for benchmarking behavior to standard game-theoretic predictions. Psychological paradigms are often much more complex, thereby making it more difficult to characterize general patterns of behavioral differences between individuals and groups.

Figure 1
Median Number Chosen by Groups and Individuals in a Beauty-Contest Game



Source: Kocher and Sutter (2005).
Note: In this simultaneous move game, a set of n decision makers chooses a number from the interval $[0, 100]$, and the winner is the decision maker whose number is closest to p times the average chosen number, with p being some fraction less than 1.

number is closest to p times the average chosen number, with p being some fraction less than 1. The name of the beauty-contest game comes from the Keynes (1936) analogy between beauty contests and financial investing in the *General Theory*: "It is not a case of choosing those which, to the best of one's judgment, are really the prettiest, nor even those which average opinion genuinely thinks the prettiest. We have reached the third degree where we devote our intelligences to anticipating what average opinion expects the average opinion to be. And there are some, I believe, who practice the fourth, fifth and higher degrees." Similarly, in a beauty-contest game, the choice requires anticipating what average opinion will be.

However, since $p < 1$, the rational equilibrium choice will be zero. For example, a player might begin by asking what the right choice will be if all other players choose randomly over the interval between 0 and 100 with $p = 2/3$, a standard value in the experimental literature. In this case, the expected value of the average of a random choice would be 50. If one anticipates that people are guessing randomly, the best response (assuming one's own guess does not distort matters) is 33.3. However, if one anticipates that everyone else will anticipate and also best-respond to random choice, the best response is 22.2. Continuing this pattern of inference through multiple iterations, the equilibrium choice is zero.

Several studies show that in the beauty-contest game, groups choose systematically lower numbers, thus suggesting that they are reasoning more deeply about the strategy of the game and are expecting the other parties to reason more deeply as well (Kocher and Sutter 2005; Kocher, Strauss, and Sutter 2006; Sutter 2005). Kocher and Sutter (2005) find that groups think one step ahead of individuals, leading them to quicker convergence towards equilibrium play, as is shown in Figure 1, which presents the median number chosen by groups (of three subjects

Figure 2
An Urn Experiment

	Left Urn	Right Urn
Up ($p = .5$)	●●●● ○○	●●● ●●●
Down ($p = .5$)	●● ○○○○	○○○ ○○○

Source: The experiment is from Charness, Karni, and Levin (2007).

each) and individuals across four rounds. When groups and individuals compete against each other (rather than groups competing against groups, or individuals against individuals), groups outperform individuals significantly by earning under the rules of the game roughly 70 percent more than individuals (Kocher and Sutter 2005; Kocher, Strauss, and Sutter 2006).[6] One possible explanation why groups choose lower numbers is that the groups, in thinking through the situation, also expect other groups to think more deeply than individuals.

Two papers by Charness, Karni, and Levin (2007, 2010) specifically examine deviations from rational behavior (by looking at error) rates in tasks involving violations of first-order stochastic dominance, and in a task involving the well-known conjunction fallacy described in Tversky and Kahneman (1983). In these studies, comparisons are made among the error rates for different group sizes.

Charness, Karni, and Levin (2007) set up a situation (see Figure 2) with a left urn and a right urn, where the state of the world is "up" or "down" with equal probability; this state is fixed for two periods. A person draws a ball, observes the color, and the ball is replaced. In the "up" state, there are four black balls and two white balls in the left urn, and in the "down" state there are two black balls and four white balls in the left urn. The right urn contains six black balls in the "up" state and six white balls in the "down" state. The most interesting case is when the first draw is from the left urn, as is required in some periods. In the original set-up, black balls pay and white balls don't. With a "good" draw (black ball) one should switch to drawing from the right urn, while with a "bad" draw (white ball) one should stay with the left urn.[7] Of course, this violates the common "win-stay, lose-shift" heuristic and is thus counterintuitive. In another treatment, subjects do

[6] In all comparisons, the per-capita incentives were kept constant across conditions, meaning that for an identical set of decisions in a particular game, the payoffs per head were identical for individuals and each single group member.
[7] To see this, note that given the draw of a black ball, the probability that the state is "up" is 2/3. If it is "up", then the probability of drawing a black ball is 2/3; if it is "down", the probability of drawing a black ball is 1/3. Since $(2/3 \times 2/3) + (1/3 \times 1/3) = 5/9$ and the probability of drawing a black ball from the right urn is 2/3, one should switch. By the same token, the probability of drawing a black

Table 1
Error Rates in an Urn Experiment in Which One Choice Stochastically Dominates the Other
(ABCD refers to the treatment with affect, Bayesian updating, a compound lottery, and dominance, while BCD, CD, and D drop one condition in turn)

Group size	*ABCD*	*BCD*	*CD*	*D*
1	.375	.188	.302	.087
2	—	.154	.230	.030
3	—	.075	—	.000

Source: Charness, Karni, and Levin (2007).
Notes: The table shows error rates in an experiment in which the choice to draw from one urn first-order stochastically dominates the choice to draw from the other. (See text for a description of the experiment.) We only consider choices after a successful first draw, as we do not have observations for the CD and D cases after unsuccessful first draws.

not know before drawing which color will pay off, with the first draw (unpaid, informational only) made automatically from the left urn. In this way, there is no sense of success or failure (and corresponding emotions) upon observing the color of the ball drawn. Removing the psychological affect in this way was found to substantially reduce the error rate in Charness and Levin (2005). A third treatment performs the Bayesian updating for the subjects, a fourth treatment eliminates the compound lottery, and a fifth treatment only considers dominance (drawing from an urn with six good balls out of nine or an urn with five good balls out of nine). Table 1 shows the corresponding error rates.

Since first-order stochastic dominance is a very basic principle, it is clear that these refusals to switch are violations of rationality. In all cases, the error rate goes down as the number of people in the decision-making group increases. In the case of dominance, the rate goes to a flat zero.

Charness, Karni, and Levin (2010) consider the Linda paradox, where this question is asked:

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which is more probable:
>
> (*a*) Linda is a bank teller.
> (*b*) Linda is a bank teller and is active in the feminist movement.

ball from the left urn given that the first draw was a *white* ball is $(1/3 \times 2/3) + (2/3 \times 1/3) = 4/9$, while the probability of drawing a black ball from the right urn is only $1/3$.

Table 2

Violations of the Conjunction Rule in an Experiment Undertaken with Individuals, Pairs, and Trios

Study	*Details*	*Incorrect answers/ total sample*	*Error rate (percent)*
Individuals			
T&K, 1983	UBC undergrads, no incentives	121/142	85.2
CKL, 2010	UCSB students, singles, no incentives	50/86	58.1
CKL, 2010	UCSB students, singles, incentives	31/94	33.0
CKL, 2010	UCSB students, total singles	**81/180**	**45.0**
Pairs			
CKL, 2010	UCSB students, in pairs, no incentives	27/56	48.2
CKL, 2010	UCSB students, in pairs, incentives	5/38	13.2
CKL, 2010	UCSB students, total in pairs	**32/94**	**34.0**
Trios			
CKL, 2010	UCSB students, in trios, no incentives	10/39	25.6
CKL, 2010	UCSB students, in trios, incentives	5/48	10.4
CKL, 2010	UCSB students, total in trios	**15/87**	**17.2**

Source: Charness, Karni, and Levin (2010); Tversky and Kahneman (1983).
Notes: This question was asked in the experiment: Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which is more probable: (a) Linda is a bank teller. (b) Linda is a bank teller and is active in the feminist movement. (Since condition *b* imposes an extra restriction, it quite clearly cannot be more probable than *a*.) UBC is the University of British Columbia; UCSB is the University of California, Santa Barbara.

Since condition *b* imposes an extra restriction, it quite clearly cannot be more probable than *a*. And yet Tversky and Kahneman (1983) report that 85 percent of respondents answer *b*. This seems a shocking violation of rational choice, no doubt due to cognitive limitations. The question was asked with and without incentives for a correct answer; people in groups consulted with each other, but then made individual decisions. Table 2 presents the data from the study for singles, pairs, and trios.

Once again, we see a clear pattern of reductions in the error rate as the number of people in the group grows. For example, without incentives the error rate drops from 58.1 percent with singles to 48.2 percent with pairs to 25.6 percent with trios. We also note that people do far better when they are provided with financial incentives, perhaps the more realistic case.

We close this section with two experimental results in games where the issue is cognitive ability. Cooper and Kagel (2005) study the "limit-pricing game" where one player acting as a market incumbent with either high or low costs of production has to decide on an output level before another player acting as a potential entrant makes a decision about market entry. In this setting, game-theoretic considerations suggest that the incumbent should choose the "limit-pricing" output with higher quantities and thus lower prices than would otherwise prevail, in order to deter market entry of the potential entrant, which could lead to still-lower prices. Indeed, Cooper and Kagel find that groups (of two persons each) play strategically far more often and thus are more successful in deterring market entry. This is particularly true in situations where the market parameters (through cost functions) change, in which cases groups are faster in learning the new "limit-pricing" output to deter market entry.

Finally, another example of how groups often see more deeply into a strategic situation is the two-person "company takeover game." In this game, a seller has a single item to sell. The item has a specific value to the seller, which the seller knows. However, the item will be worth 50 percent more than that to the buyer, but the buyer knows only a distribution of potential values for the seller. If the bid is at least as large as the seller's value, the buyer acquires the company after paying the bid. The optimal bid is zero, yet the vast majority of buyers fail to condition their bids on winning, and so select a positive bid (say, the expected value of that distribution).[8] An insightful bidder will recognize that potential values (seller values) above the bid are irrelevant, and so will condition her bid appropriately. This set-up is effectively a form of the "winner's curse," where the winner of an auction loses money. Casari, Zhang, and Jackson (2010) analyze group and individual behavior in this game. They find that groups fall prey to the "winner's curse" of overbidding significantly less often than individuals do, by a margin of about 10 percentage points. A similar finding of less overbidding by groups (by reducing their bids in a contest by about 25 percent) is reported in Sheremeta and Zhang (2010). In both papers, groups learn to reduce their bids from communication inside the group, indicating that groups are better in learning rational bidding strategies than individuals.

These examples (and others in the online Appendix) are rather compelling in illustrating that group choices in decision-making environments characterized by cognitive limitations (bounded rationality) are closer to the predictions of standard theory than are individual choices. These findings let us conclude that groups are more rational decision makers in the sense that economists have defined.

Lesson Two: Groups Can Help with Self-Control and Productivity Problems

Nearly everyone has self-control problems, such as procrastination, not exercising despite the lasting benefits of doing so, and being unable to control one's spending to save money. A lack of self-control or even motivation is also often found in the workplace, so that productivity is far from optimal. People engage in

[8] It is easy to show that the optimal bid is zero. Suppose one bids x from the interval $[0, 100]$. Assuming a uniform distribution, the average relevant seller value is not 50, but is instead $x/2$, since values above x lead to no sale. Thus the expected value to the buyer conditional on acquiring the company is 50 percent more or $3x/4$, so one loses $x/4$ on average, and choosing $x = 0$ is best.

a wide variety of commitment mechanisms to cope with these issues. For example, researchers quite often employ the commitment device known as co-authorship. One does not wish to let down a co-author (who presumably produces!), so one works harder. In a sense, this form of production is enhanced by being in a group. In this section, we present evidence from experimental and empirical studies that suggest that group decision-making and group membership can help to alleviate these self-control problems.

The evidence in this embryonic area is limited. It is difficult to observe self-control problems in the laboratory, so the experimental evidence on this topic comes from field experiments.[9] One such experiment was conducted by Falk and Ichino (2006). They let subjects perform a real-effort task, which was to put letters into envelopes for a mass mailing. In one condition, subjects had to perform the task alone in a room, while in another condition there were two subjects in the room, and both could easily watch the performance of the other. Falk and Ichino find that in the condition where groups of subjects were working, average productivity was 16 percent higher than in the isolated condition, indicating that peer effects in the group had a positive impact on productivity. Mas and Moretti (2009) also report such positive spillovers in a supermarket chain where the introduction of high-productivity workers into shifts increased the average individual productivity. While in the previous two examples, the wages of subjects were independent of their coworkers, Hamilton, Nickerson, and Owan (2003) examined how productivity in a garment factory in California changed when the plant shifted from an individual piece-rate to a group piece-rate production system (where a group member's wage did depend on the other group members' performance). While the problem of free-riding in groups (Holmstrom 1982) might decrease average productivity, Hamilton, Nickerson, and Owan (2003) find that the adoption of a group payment scheme at the plant improved worker productivity by 14 percent on average, even after controlling for systematic selection of high-ability workers into work groups. Interestingly, their data also reveal that an increase in a group's heterogeneity in ability levels increases productivity.

Babcock and Hartman (2011) investigate peer effects at the level of individual connections, and leverage the approach to shed light on peer mechanisms. In a field experiment with college freshmen, they elicited friendship networks and offered monetary incentives in some treatments for using the recreation center. Their main findings are that treated subjects with treated best friends put forth significantly more effort toward the incentivized task than do treated subjects with control best friends. The peer effect is about 20 percent as large as the direct individual effect of the incentive. There is also clear evidence of a mechanism: subjects coordinate with

[9] List (2011) provides a taxonomy of field experiments. Broadly speaking, they can be categorized into artefactual experiments (real-world participants, perhaps from business or the public sector, brought into the laboratory setting), framed field experiments (real-world participants knowingly participating in experiments in a natural setting), and natural field experiments (real-world participants unknowingly participating in a real-world experimental setting).

best friends to overcome pre-commitment problems or reduce effort costs. Their results highlight subtle peer effects and other mechanisms that often go undetected.

In a related paper, Babcock, Bedard, Charness, Hartman, and Royer (2012) find evidence that pairing people helps to overcome problems with exercising and studying. In a field experiment involving studying and a field experiment involving exercise, large team effects operate through social channels. These experiments feature exogenous team formation and opportunities for repeated social interactions over time; one suspects that the effects would be substantially larger with endogenous group formation. In any case, in the pay-for-study intervention, people assigned to the team treatment frequented the study room considerably more often than people assigned to the individual treatment. The team-compensation system induced agents to choose their effort as if they valued a marginal dollar of compensation for their teammate from two-thirds as much to twice as much as they valued a dollar of own compensation. The paper concludes that the social effects of monetary team incentives can be used to induce effort at significantly lower cost than through direct individual payment.

Recent evidence from microfinance suggests that the frequency of meeting with others to discuss micro-loans is positively associated with repayment rates, thus helping to avoid self-control problems due to a wish for immediate gratification (Laibson 1997), which increases default risks. While the effects of group liability—where borrowers are organized in groups in which they are the guarantors of each other's loans—on default rates have been diverse (Armendariz de Aghion and Morduch 2005),[10] Feigenberg, Field, and Pande (2011) show that more frequent meetings of Indian microfinance borrowers lead to substantially lower default rates. People in a group that met once per month were 3.5 times more likely to default on a second loan than people in a group that met once per week. While this study does not provide direct evidence that people who met in groups default less frequently than people who did not (although extrapolation suggests that this is the case), it does appear that these meetings generated a form of economically valuable social capital that promoted more trustworthy behavior. In fact, there was considerably more external social interaction amongst members of the weekly group than amongst members of the monthly group. In this sense, organizing people into groups that meet frequently can enhance responsible behavior.

Lesson Three: Groups May Decrease Welfare Because of Stronger Self-interested Preferences

In the first two lessons, we have argued that decision making in groups leads to choices that are closer to predicted choices under the standard assumptions of

[10] In a carefully controlled natural field experiment on group versus individual liability in microfinance credits in the Philippines, Giné and Karlan (2011) do not find a difference in repayment rates between group and individual liability contracts.

Table 3
Social Welfare in a Trust Game
(as a fraction of the maximum possible payoff)

		Second-mover	
		Individual	*Group*
First-mover	Individual	0.77	0.84
	Group	0.69	0.62

Source: This is a trust game described in Kugler, Bornstein, Kocher, and Sutter (2007).
Note: Social welfare is the actual payoff per person divided by the maximum possible payoff.

rationality and that help individuals to overcome or at least contain their behavioral biases. While all of this seems like a desirable influence of group decision-making, we have not yet addressed how group decision-making may affect social welfare as we have defined it above (as total social material payoffs). We attend to this issue here, showing that decision making in groups may, in fact, be detrimental for social welfare in specific situations, whereas it is good for social welfare in others. Because the evidence in this relatively young field of research is still emerging, we are not yet able to provide a definitive taxonomy of when group decision-making is good for welfare and when it is bad, but we can lay some cornerstones upon which such a taxonomy could be built in the future.

We start with evidence from a game originally termed "the investment game" but now more commonly known as "the trust game." In this game, the first player can send an amount $x \leq c$ to a second player. The second player receives $3x$, and can send back any (non-tripled) amount $y \leq 3x$, which finishes the game. In this setting, the standard game-theoretic prediction is that the first player won't expect to get anything back, and so will send nothing. Given that an increase in the amount x is associated with higher social welfare (as the sum of payoffs for both players), the standard prediction is associated with the least efficient outcome.

Kugler, Bornstein, Kocher, and Sutter (2007) have run a trust game where either individuals, or groups of three subjects each, were in the role of first- or second-mover. They find that groups send significantly smaller amounts (by about 20 percentage points) as first-movers, and also return on average smaller amounts (although this second result was statistically insignificant). Hence, group choices are closer to the standard rationality paradigm. Table 3 shows social welfare in the four different conditions in the experiment as a fraction of the maximum possible payoff per subject. If first-movers are groups, social welfare is significantly smaller. Since second-movers are only making redistributive choices, they do not affect social welfare.[11]

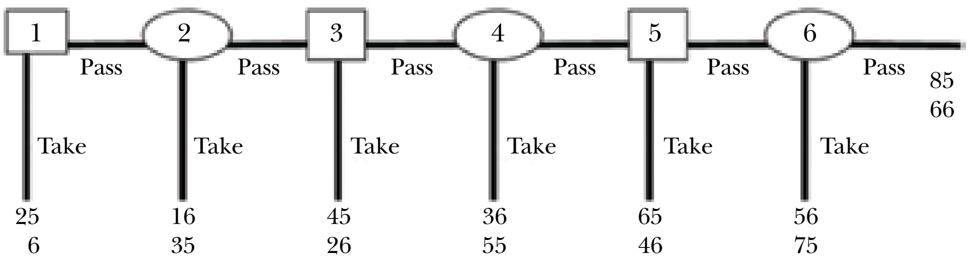
[11] Cox (2002) finds that groups as second-movers return significantly smaller amounts than individuals do. Again, this does not affect total social payoffs, since second-movers only redistribute money.

Instead of using group decision-making, Song (2008) has studied how group representatives make decisions on behalf of their group in a trust game. This means that the representative had to make a decision that determined the outcome of a three-person group. Song finds that group representatives send about 20 percent less as first-movers and return about 40 percent less as second-movers than individuals who decide only for themselves. These results support the earlier work of Kugler, Bornstein, Kocher, and Sutter (2007) on the negative effect of group decision-making on social welfare when trust is crucial to increase social welfare.

The "centipede game" can be viewed as a multistage version of the trust game. There are two stakes on the table: one large and one small. Players must decide either to pass the stakes to the other player, at which point both stakes increase in size, or end the game by taking the larger stake for themselves and giving the smaller stake to the other player. The payoffs are arranged such that if one passes the stakes in a particular stage and the opponent immediately ends the game in the next stage, one receives less than if one had taken the payoff and not passed the stakes.The centipede game is played for a limited number of rounds. Thus, backward induction suggests that players should end the game earlier, rather than run the risk of getting a lower payoff in the event that the other player "takes" at the next move. Figure 3 displays the centipede game used in a study by Bornstein, Kugler, and Ziegelmeyer (2004) in which they let individuals play against individuals, and groups (of three subjects each) against groups. They find that individuals' median action is to "take" at node 5, while the median action of groups is to "take" at node 4. The difference is statistically significant and yields also significantly smaller payoffs for group members (50 on average) than for individuals (58 on average). Hence, the evidence shows a similar pattern as in the trust game: group play is more likely to conform to the rationality standard of game theory, but as a result, group play is also less likely to reap the potential efficiency gains.

As a final piece of evidence that group behavior may be bad for social welfare, we refer to a classic prisoner's dilemma. Of course, a prisoner's dilemma game is the familiar setting in which each of two players will find it a dominant strategy to defect, but if they can coordinate on cooperation, their combined payoff will be larger. Charness, Rigotti, and Rustichini (2007) study how individuals play this game on behalf of groups: that is, when they are making (individual) choices in front of their group members and when their actions influence the other group members' payoffs (referred to as "payoff commonality"). They find that cooperation rates go down considerably and significantly when individuals play this game against an out-group member in front of their in-group and when payoff commonality applies. Hence, while defection is the self-interested choice here, group membership makes this choice more frequent, but as a consequence social welfare is reduced. In sum, the evidence summarized so far suggests that in trust games, centipede games, and prisoner's dilemma games (all of which share the characteristic that they have a unique and socially inefficient, pure-strategy Nash equilibrium) group

Figure 3
A Centipede Game



Source: Bornstein, Kugler, and Ziegelmeyer (2004).
Notes: Player 1's decision nodes are denoted by squares, and Player 2's by circles. At the start of this game, the large stake is 25 and the small stake is 6. Each time a player passes, both stakes are increased by 10. At each terminal node, the top number shows the payoff for Player 1 and the bottom for Player 2 if the game ends at that stage.

decision-making and group membership decrease social welfare, because groups show too little trust regarding cooperation from their interaction partners.

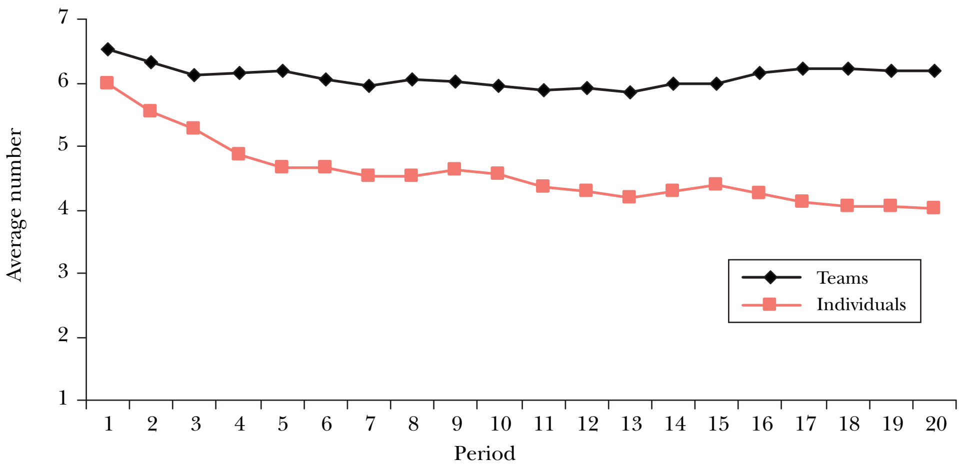
This negative effect of groups on social welfare does not generalize to all games, however. In particular, there is strong evidence that in games with multiple pure-strategy equilibria—commonly referred to as coordination games—group decision-making helps achieve efficient coordination, thus increasing social welfare.

Charness, Rigotti, and Rustichini (2007) consider a battle-of-the-sexes game. This is a 2 × 2 game often described with a story like this one: A couple agrees to get together, but they cannot remember where they agreed to meet. Both parties know that the husband preferred to attend a certain sports event and the wife preferred to attend a certain play. Both parties receive higher benefits if they coordinate on a location, yet they cannot communicate with each other. This setting has two pure-strategy equilibria where both parties attend the same location, either the sports event or the play.[12] Efficiency in this game requires successful coordination (avoiding the outcomes in which the couple ends up in different places). Charness, Rigotti, and Rustichini (2007) show that *salient* group membership (one person in the pair plays in front of an audience of one's group members) significantly increases the rate of successful coordination compared to the rate in a situation without salient group membership. In this case, salient group membership leads to better social outcomes.

Some coordination games have multiple equilibria that are Pareto-ranked—that is, some equilibria are more efficient than others. For example, the "weakest link" game studied in Feri, Irlenbusch, and Sutter (2010) shares this feature, and it works like this: There are five players, which can be either individuals or groups with

[12] There is also a mixed-strategy equilibrium.

Figure 4
Effort Levels of Individuals and Groups in a Weakest Link Game



Source: Feri, Irlenbusch, and Sutter (2010).
Notes: This game, denoted *WL-BASE*, is described in Feri, Irlenbusch, and Sutter (2010). There are five players, which can be either individuals or groups with three members each. Each player (either an individual or a three-person group) chooses an effort level between 1 and 7. The payoff each player receives gets higher if they all choose to exert more effort, but it also gets lower—at a faster rate—the lower the minimum choice (or "weakest link") of all players.

three members each. Each player (either an individual or a three-person group) chooses an effort level between 1 and 7 (group members may communicate briefly first). The payoff each player receives gets higher if they all choose to exert more effort, but it also gets lower—at a faster rate—the lower the minimum choice (or "weakest link") of all players. In this setting, it turns out that any setting where all the players choose the same level of effort will be an equilibrium. The biggest payoffs for all players together will arise if everyone coordinates on a high level of effort. But the weakest-link dynamic tends to push toward coordinating on a lower level of effort. Feri, Irlenbusch, and Sutter find that the three-player groups not only play more efficient high-effort equilibria more often than individuals, but also are more successful in avoiding miscoordination (which in this case means picking different effort levels). Figure 4 shows the average effort levels across 20 periods for individuals and groups, indicating a large and significant difference in the ability to coordinate on more efficient outcomes. Social welfare is on average 24 percent higher when groups play this coordination game than when individuals make decisions.

In short, the effect of group decision-making on social welfare can go in either direction. The pattern emerging from the evidence seems to indicate that more rational choices of groups decrease social welfare when games have a unique pure-strategy equilibrium (with a dominant strategy, in fact), but that groups are more successful in coordinating on more efficient equilibria when a multiplicity of equilibria exist. The common denominator for these seemingly divergent effects of group decision-making may be that groups put more weight on own payoffs than do individuals (something discussed also in the next section). Studying the learning of groups and individuals, Feri, Irlenbusch, and Sutter (2010) find that groups are more sensitive to the attractions of different strategies and take into account more strongly the potential payoffs of previously not-chosen strategies. These learning characteristics of groups imply that payoffs play a significantly larger role in determining their choice probabilities than they do for individuals, leading to a higher frequency of choosing dominant strategies in trust games ("do not trust"), centipede games ("take"), or prisoner's dilemma games ("defect"), but also to a higher frequency of choosing more efficient equilibria in coordination games.

Sources of Differences in Individual and Group Decisions

Why might groups behave in a more rational manner than individuals? We explore three possible reasons: 1) multiple brains are better at seeking answers; 2) multiple brains are better at anticipating the actions of other parties and thus better at coordinating behavior with what other parties are likely to do; and 3) groups may be more likely than individuals to emphasize monetary payoffs over alternative concerns, such as fairness or reciprocity towards another player.

Our first possible explanation for differences between individuals and groups is that groups can potentially benefit from having multiple brains. In some cases, this may lead to better decisions in the sense of avoiding errors. In addition to the examples given in Lesson One, consider an information cascade game. Here players receive a private signal and then announce a public belief in sequential order: for example, players might look at one marble drawn from a bag, and then announce their belief as to whether the bag is two-thirds white marbles or two-thirds black marbles. Later players must then compare their own private signal to the public beliefs of others. In an information cascade, players ought to disregard their private information and instead follow the belief being expressed by many others at some stage of the game. Fahr and Irlenbusch (2011) find that groups make fewer mistakes in an information cascade experiment than individuals (and thus earn more money).[13] Evidence from psychology supports the argument that social interaction improves the decision-making process. For instance, in letters-to-numbers problems, where a random coding of the letters A–J to the numbers 0–9 needs to be solved, groups do much better than individuals by taking about 30 percent fewer trials to solve the problems (Laughlin, Bonner, and Miner 2002). Likewise, in the "Wason selection task," developed to test whether individuals employ the rules of formal logic when

[13] Also in information cascade experiments by Alevy, Haigh, and List (2007), professional traders were shown to be better able to discern the quality of public signals. One possible explanation for the superiority of professional traders over college students might be that professional traders are more used to being in a group, so they make better decisions, an interpretation that would be consistent with the findings by Fahr and Irlenbusch (2011).

testing conditional statements of the form "if p, then q," groups have solution rates of 50 percent while individuals have solution rates of 11 percent (Maciejovsky and Budescu 2007). The Wason selection task is an example of a "truth wins" problem: that is, a problem where the solution is difficult to reach without grasping a specific insight, but then the solution is easily explained to another individual. In such cases, groups can be expected to solve the problem with higher probability. Consider that a fraction p of all individuals has the specific insight to solve the problem, then the likelihood that a group with n members solves the problem is $1 - (1 - p)^n$, which is larger than p (if $p < 1$). The likelihood $1 - (1 - p)^n$ is often referred to as the "truth-wins benchmark." While groups typically do better than individuals in such insight problems, they rarely meet or exceed the truth-wins benchmark.[14]

In an interesting experiment from the psychology literature, groups actually beat this benchmark. Michaelson, Watson, and Black (1989) grouped together students in a class (average group size of six) and asked them to answer questions based on assigned reading, with the scores counting towards the course grade. These tasks ranged from recalling specific concepts from the reading to ones requiring higher cognitive ability and a deeper understanding to being able to synthesize concepts. The key comparison was between the highest score of any individual in a group and the average score of the group on the task; the notion behind this comparison is to test the view that, in an organizational context, group decisions will be better than the decisions of the most knowledgeable group member. In fact, a remarkable 97 percent of all groups outperformed their best member. Each person first completed the task individually and then retook the test as a member of a group that could have internal discussions. Group scores were compared with the highest score for any individual in the group. In the economics literature, choices made in the Cooper and Kagel (2005) limit-pricing game and in the Maciejovsky and Budescu (2007) Wason selection task provide examples where groups do better than the truth-wins benchmark.

A second possible reason why groups make more rational decisions than individuals, especially in interactive games, is that group members are better able to put themselves into the shoes of their competitors when discussing their own strategy. It seems that the need to discuss the game with another group member often leads to a discussion regarding how the group members would play the game, making it a salient feature then to consider the other player's available strategies and payoffs more extensively than individuals would do (Cooper and Kagel 2005). For this reason, groups can be better prepared to anticipate the actions of other players. From there, it is only a short step to think about the best reply to one's own expectation about the opponent's most likely strategy. As a consequence, group behavior is pushed towards the standard game-theoretic predictions. This insight is consistent with what has been observed in the limit-pricing game of Cooper and Kagel (2005). Further support is presented in Sutter, Czermak, and Feri (2010).

[14] Meaning that their solution rates stay below $1 - (1 - p)^n$ but remain above p.

They let individuals and groups make choices in simple two-player games (with unique pure-strategy, Pareto-inefficient Nash equilibria). Groups play the Nash equilibrium in these games about 10 percentage points more frequently than individuals, and the main reason is that they expect their opponent to play the Nash equilibrium more frequently than individuals expect this from individuals. Accordingly, groups more often play the equilibrium as a best response to their own beliefs.

A third reason why groups may behave "less behaviorally" than individuals is that groups may be more concerned with their own group's monetary payoffs and thus disregard more frequently the payoffs of the other player. Communication within groups may change an individual's reference point for optimization. Instead of maximizing own payoffs, individuals may consider the joint payoff (or welfare) of those engaged in the discussion as the appropriate target for optimization. Psychologists have long been emphasizing such an effect of communication: Elster (1986, pp. 112–113), for instance, has suggested that it is "pragmatically impossible to argue that a given solution should be chosen just because it is good for oneself. By the very act of engaging in a public debate . . . one has ruled out the possibility of invoking such reasons. To engage in discussion can in fact be seen as one kind of self-censorship, a pre-commitment to the idea of rational decision." By rational decision, however, Elster (1986) refers to decisions which are advantageous for the *group* of communicating subjects as a whole, but not necessarily aligned with (and sometimes even contrary to) the interests of other players in the opponent group. Such an argument links our discussion to the long-standing literature on in-group/out-group effects. (For an overview from an economic perspective, interested readers might start with Chen and Li 2009.) By design, group decision-making creates an in-group—one's own group—and an out-group—with whom the own-group is interacting. Social psychology has coined the term "discontinuity effect" (for example, Schopler et al. 2001) to describe the fact that, typically, groups act more competitively and more selfishly when interacting with other groups than when individuals interact with individuals.

Conclusion

The existing literature that compares group and individual decision-making provides considerable evidence that groups make choices that are more rational in a standard game-theoretic sense than those of individuals. As a result, group decision-making and being a member of a group can overcome cognitive biases and limitations. However, making decisions in groups does not always lead to increases in social welfare, which raises the question: Under which conditions is individual or group decision-making better for society as a whole? We have identified several games (with unique equilibria) where individual decision-making yields higher welfare, while in coordination games (with multiple equilibria), groups achieve more efficient outcomes.

Since group decision-making is present in a wide variety of economic environments, this issue has considerable practical relevance. Generally, decision making in groups seems to be most effective when there is a good degree of diversity in the group and when the environment is a participatory one in which diverse ideas can be expressed (rather than an environment with a dominant and intimidating personality). For example, any single individual group member could have an insight that sheds light on what would otherwise be a blind spot for the group; it pays to broaden the base. Still, it seems best to have groups of modest size, so that interior coordination problems and "social loafing"—in this case, reduced effort—are manageable. As Surowiecki (2004, pp. 190–91) wrote: "If small groups are included in the decision-making process, then they should be allowed to make decisions. If an organization sets up teams and then uses them for purely advisory purposes, it loses the true advantage that a team has: namely, collective wisdom." It is noteworthy, however, that it remains to be determined what constitutes an ideal group size. A useful starting point here is Forsyth's (2006) work on group size and performance. We suspect that the optimal size of the group will depend on factors such as the complexity of the decision, but more research is clearly needed here.

Some other open issues for future research include the influence of different communication media on group decisions. Do group dynamics change when video calls substitute for face-to-face communication? Another relatively unexplored area is the effect of internal conflicts on the rationality and character of group decisions: that is, what happens when the payoffs to members of a group are not identical? Groups can be a way of diffusing decision-making and avoiding responsibility, but they can also be a powerful force for more careful and productive decisions. Ultimately, the goal of comparing individual and group decision-making is to identify the contexts and types of decisions where each is likely to work best.

References

Alevy, Jonathan E., Michael S. Haigh, and John A. List. 2007. "Information Cascades: Evidence from a Field Experiment with Financial Market Professionals." *Journal of Finance* 62(1): 151–80.

Armendariz de Aghion, Beatriz, and Jonathan Morduch. 2005. *The Economics of Microfinance.* MIT Press.

Babcock, Philip, Kelly Bedard, Gary Charness, John Hartman, and Heather Royer. 2012. "Letting Down the Team: Social Effects of Team Incentives." Unpublished paper.

Babcock, Philip, and John Hartman. 2011. "Coordination and Contagion: Peer Effects and Mechanisms in a Randomized Field Experiment." Unpublished paper.

Bornstein, Gary, Tamar Kugler, and Anthony Ziegelmeyer. 2004. "Individual and Group Decisions in the Centipede Game: Are Groups More 'Rational' Players?" *Journal of Experimental Social Psychology* 40(5): 599–605.

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

Casari, Marco, Jingjing Zhang, and Christine

Jackson. 2010. “Do Groups Fall Prey to the Winner’s Curse?” IEW Working Paper 504, Institute for Empirical Research in Economics, University of Zurich.

Charness, Gary, Edi Karni, and Dan Levin. 2007. “Individual and Group Decision Making under Risk: An Experimental Study of Bayesian Updating and Violations of First-Order Stochastic Dominance.” *Journal of Risk and Uncertainty* 35(2): 129–48.

Charness, Gary, Edi Karni, and Dan Levin. 2010. “On the Conjunction Fallacy in Probability Judgment: New Experimental Evidence Regarding Linda.” *Games and Economic Behavior* 68(2): 551–56.

Charness, Gary, and Dan Levin. 2005. “When Optimal Choices Feel Wrong: A Laboratory Study of Bayesian Updating, Complexity, and Affect.” *American Economic Review* 95(4): 1300–1309.

Charness, Gary, Luca Rigotti, and Aldo Rustichini. 2007. “Individual Behavior and Group Membership.” *American Economic Review* 97(4): 1340–52.

Chen, Yan, and Xin Li. 2009. “Group Identity and Social Preferences.” *American Economic Review* 99(1): 431–57.

Cooper, David J., and John H. Kagel. 2005. “Are Two Heads Better Than One? Team versus Individual Play in Signaling Games.” *American Economic Review* 95(3): 477–509.

Cox, James C. 2002. “Trust, Reciprocity, and Other-Regarding Preferences: Groups vs. Individuals and Males vs. Females.” In *Advances in Experimental Business Research*, edited by Rami Zwick and Amnon Rapoport, 331–50. Dordrecht: Kluwer Academic Publishers.

Elster, Jon. 1986. “The Market and the Forum: Three Varieties of Political Theory.” In *Foundations of Social Choice Theory: Studies in Rationality and Social Change*, edited by J. Elster and A. Hylland, 103–132. Cambridge University Press.

Fahr, René, and Bernd Irlenbusch. 2011. “Who Follows the Crowd—Groups or Individuals?” *Journal of Economic Behavior and Organization* 80(2): 200–209.

Falk, Armin, and Andrea Ichino. 2006. “Clean Evidence on Peer Effects.” *Journal of Labor Economics* 24(1): 39–57.

Feigenberg, Benjamin, Erica Field, and Rohini Pande. 2011. “The Economic Returns to Social Interaction: Experimental Evidence from Microfinance.” http://www.economics.harvard.edu/faculty/field/files/Social_Capital_feb10_ef_rp.pdf.

Feri, Francesco, Bernd Irlenbusch, and Matthias Sutter. 2010. “Efficiency Gains from Team-Based Coordination—Large-Scale Experimental Evidence.” *American Economic Review* 100(4): 1892–1912.

Forsyth, Donelson R. 2006. *Group Dynamics*, 4th edition. Belmont, CA: Thomson Higher Education.

Giné, Xavier, and Dean S. Karlan. 2011. “Group versus Individual Liability: Short and Long Term Evidence from Philippine Microcredit Lending Groups.” June. http://karlan.yale.edu/p/GroupversusIndividualLending.pdf.

Hamilton, Barton H., Jack A. Nickerson, and Hideo Owan. 2003. “Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation.” *Journal of Political Economy* 111(2): 465–97.

Holmstrom, Bengt. 1982. “Moral Hazard in Teams.” *Bell Journal of Economics* 13(2): 324–40.

Keynes, John Maynard. 1936. *The General Theory of Employment, Interest and Money*. Macmillan Cambridge University Press for the Royal Economic Society.

Kocher, Martin G., Sabine Strauss, and Matthias Sutter. 2006. “Individual or Team Decision-Making—Causes and Consequences of Self-Selection.” *Games and Economic Behavior* 56(2): 259–70.

Kocher, Martin G., and Matthias Sutter. 2005. “The Decision Maker Matters: Individual versus Group Behavior in Experimental Beauty-Contest Games.” *Economic Journal* 115(500): 200–223.

Kugler, Tamar, Gary Bornstein, Martin G. Kocher, and Matthias Sutter. 2007. “Trust between Individuals and Groups: Groups are Less Trusting Than Individuals But Just as Trustworthy.” *Journal of Economic Psychology* 28(6): 646–57.

Laibson, David. 1997. “Golden Eggs and Hyperbolic Discounting.” *Quarterly Journal of Economics* 112(2): 443–77.

Laughlin, Patrick R., Bryan L. Bonner, and Andrew G. Miner. 2002. “Groups Perform Better Than the Best Individuals on Letter-to-Numbers Problems.” *Organizational Behavior and Human Decision Processes* 88(2): 606–620.

Levine, John M., and Robert L. Moreland. 1998. “Small Groups.” In *The Handbook of Social Psychology*, 4th edition, vol. 2, edited by Gilbert, D. T., S. T. Fiske, and G. Lindzey, 415–69. McGraw-Hill.

Levitt, Steven, and John A. List. 2007. “What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?” *Journal of Economic Perspectives* 21(2): 153–74.

List, John A. 2011. “Why Economists Should Conduct Field Experiments and 14 Tips for Pulling One Off.” *Journal of Economic Perspectives* 25(3): 3–16.

Maciejovsky, Boris, and David V. Budescu. 2007. “Collective Induction without Cooperation? Learning and Knowledge Transfer in Cooperative

Groups and Competitive Auctions." *Journal of Personality and Social Psychology* 92(5): 854–70.

Mas, Alexandre, and Enrico Moretti. 2009. "Peers at Work." *American Economic Review* 99(1): 112–45.

Michaelson, Larry K., Warren E. Watson, and Robert H. Black. 1989. "A Realistic Test of Individual versus Group Consensus Decision Making." *Journal of Applied Psychology* 74(5): 834–39.

Schopler, John, Chester A. Insko, Jennifer Wieselquist, Michael Pemberton, Betty Witcher, Rob Kozar, Chris Roddenberry, and Tim Wildschut. 2001. "When Groups Are More Competitive Than Individuals: The Domain of the Discontinuity Effect." *Journal of Personality and Social Psychology* 80(4): 632–44.

Sheremeta, Roman M., and Jingjing Zhang. 2010. "Can Groups Solve the Problem of Overbidding in Contests?" *Social Choice and Welfare* 35(2): 175–97.

Song, Fei. 2008. "Trust and Reciprocity Behavior and Behavioral Forecasts: Individuals versus Group-Representatives." *Games and Economic Behavior* 62(2): 675–96.

Surowiecki, James. 2004. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations.* Doubleday.

Sutter, Matthias. 2005. "Are Four Heads Better Than Two? An Experimental Beauty-Contest Game with Teams of Different Size." *Economics Letters* 88(1): 41–46.

Sutter, Matthias, Simon Czermak, and Francesco Feri. 2010. "Strategic Sophistication of Individuals and Teams in Experimental Normal-Form Games." *IZA Discussion Paper 4732.*

Tversky, Amos, and Daniel Kahneman. 1983. "Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment." *Psychological Review* 90(40): 293–315.

Are Two Heads Better Than One? Team versus Individual Play in Signaling Games

By DAVID J. COOPER AND JOHN H. KAGEL*

We compare individuals with two-person teams in signaling game experiments. Teams consistently play more strategically than individuals and generate positive synergies in more difficult games, beating a demanding "truth-wins" norm. The superior performance of teams is most striking following changes in payoffs that change the equilibrium outcome. Individuals play less *strategically following the change in payoffs than inexperienced subjects playing the same game. In contrast, the teams exhibit* positive *learning transfer, playing* more *strategically following the change than inexperienced subjects. Dialogues between teammates are used to identify factors promoting strategic play.* (*JEL* C72, C92, D82, L12)

Many economic decisions are made within a team or group framework with two or more economic agents consulting with each other in deciding what course of action to take. For example, corporate bid teams, often in conjunction with outside consultants, determined bidding strategies in the spectrum (air wave) rights auctions. U.S. monetary policy is determined by the 12-member Open Market Committee of the Federal Reserve Bank. More generally, virtually all significant strategic decisions by corporations are made within a group or team framework. In contrast, much of economic theory and game theory, and most experimental investigations of these theories, make no distinction between strategic decisions made by teams versus individuals. As a result, there is potentially a significant hole in our understanding of large areas of economic behavior. If there are major differences between individuals and teams in important economic settings, direct extrapolation of individual-level research to team performance may be strikingly inconsistent with observed behavior. The results presented here, which find substantial differences between the behavior of individuals and teams in signaling games, indicate these concerns are likely to be well founded.

We report experiments comparing the behavior of individuals versus freely interacting two-person teams in Paul Milgrom and John R. Roberts's (1982) entry limit pricing game. Strategic play in this game revolves around an incumbent monopolist attempting to deter entry by signaling it will be a tough competitor for a potential entrant. Specifically, strategic play takes place through limit pricing, the choice by incumbents of greater quantities (and lower prices by extension) than would prevail in the absence of asymmetric information. Past experiments with signaling games show that equilibrium play emerges only gradually, requiring a number of replications of the game before anything approaching an equilibrium emerges (Jordi Brandts and Charles A. Holt, 1992; Cooper et al., 1997b; Cooper et al., 1999).

We study three different versions of the limit pricing game, which vary in the difficulty of

* Cooper: Department of Economics, Weatherhead School of Management, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106 (e-mail: david.cooper@case.edu); Kagel: Department of Economics, Ohio State University, 410 Arps Hall, Columbus, OH 43210 (e-mail: kagel@economics.sbs.ohio-state.edu). Research support from the National Science Foundation (SBR-9809538) is gratefully acknowledged. Helpful research assistance was provided by Alex Brown, Bethia Cullis, Guillaume Frechette, John Lightle, and Todd Schoellman. We have received helpful comments from Eric Bettinger, John Ham, David Laibson, James Peck, Alvin Roth, seminar participants at Carnegie-Mellon University, Emory University, Harvard University, Ohio State University, University College London, and University of Wyoming, as well as the referees and editor of this journal. Any opinions, findings, and conclusions or recommendations in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

learning to play strategically. We focus on how fast strategic play (e.g., limit pricing) develops for teams versus individuals, as well as identifying possible qualitative differences in the learning/adjustment process between teams and individuals. For all three treatments, the equilibrium toward which play converges is the same for teams and individuals, but teams learn to play strategically faster than individuals. Impressively, in the more difficult games, teams beat a demanding "truth-wins" norm drawn from the psychology literature.

The differences between teams and individuals are most striking following changes in payoffs that affect the equilibrium outcome. In this crossover treatment, subjects first play a game that supports both pure strategy pooling and separating equilibria, with play converging on a pooling equilibrium, and are then switched to a game in which the only pure strategy equilibria are separating. For individuals, experience in the game with pooling equilibria retards adjustment to the separating equilibria following the change in games (compared to individuals who play the same game but with no prior experience), so that there is *negative* learning transfer between games. For teams, however, experience in the game with pooling equilibria enhances adjustment to the separating equilibrium (compared to teams with no previous experience), yielding *positive* learning transfer between games. Thus, teams not only learn to play strategically faster than individuals, but also follow a different learning/adjustment process than individuals.

An important methodological innovation of our procedures is the recording and coding of dialogues between teammates as they coordinate their decisions. These dialogues provide a direct window into teams' learning processes. Analysis of the dialogues indicates that a critical step in monopolists' learning to play strategically is putting themselves in the entrant's shoes, reasoning from the entrant's point of view to infer likely responses to their choice as a monopolist. Statements to this effect are (a) among the most commonly coded statements prior to limit pricing; (b) among the best predictors of whether teams will continue to limit price after their first attempt to do so; and (c) much more common immediately following the change in payoffs in the crossover treatment compared to inexperienced teams playing the same game. Thus, the kind of interactive reasoning that underlies much of game theory has clear empirical validity in our data.

Beyond the importance of our results for economics and game theory, our work also speaks to the large psychology literature on group decision making by consciously "common-purpose" groups seeking consensus on how to solve a specific problem. There are two distinct branches to this literature, one dealing with judgmental questions and another dealing with "eureka-type" problems requiring no special information to solve, but having solutions that tend to be self-confirming if discovered. The latter is closest in spirit to our game. For such problems, psychologists apply a truth-wins norm in judging whether or not teams are superior to individuals. Intuitively, a team should be no less likely to solve a problem than its most able member would be *acting alone*. By comparing the performance of freely interacting teams with this norm, psychologists identify the presence of positive, negative, or zero synergies for teams.

The rich psychology literature on team versus individual play in eureka problems consistently finds that teams typically fall well short of the truth-wins norm (James H. Davis, 1992). In contrast, for two out of our three treatments, teams meet or beat the truth-wins norm, indicating strong positive synergies from the teams treatment. The concluding section of the paper identifies important differences in experimental procedures, and between games with strategic interactions like we study versus games against nature that psychologists typically study, which likely account for the striking differences between our results and those in the psychology literature. We also compare our results to the limited literature on team versus individual play previously reported in the economics literature.

The paper proceeds as follows: Section I characterizes the structure of the limit-pricing game and the equilibria for the games played. Section II defines what we mean by strategic play and relates learning to play strategically in our experiment to the truth-wins norm. Section III outlines our experimental procedures. Results are reported in Section IV. Section V reports on the team dialogues for insights into the learning process and to justify fully our defini-

TABLE 1A—MONOPOLIST PAYOFFS

High-cost monopolist (MH)			Low-cost monopolist (ML)		
Monopolist output	Entrant response		Monopolist output	Entrant response	
	IN	OUT		IN	OUT
1	150	426	1	250	542
2	168	444	2	276	568
3	150	426	3	330	606
4	132	408	4	352	628
5	56	182	5	334	610
6	−188	−38	6	316	592
7	−292	−126	7	213	486

tion of strategic play. The last section of the paper summarizes our main findings and relates our results to the psychology and economics literature on teams versus individuals and cross-game learning.

I. The Limit Pricing Game

The games studied here are based on Milgrom and Roberts's (1982) entry limit pricing model. For our purposes, the industrial organization implications of the model are of secondary importance. We therefore employ a very stylized version of the model, focusing on the signaling aspects of the game.

A. *Structure of the Game*

The limit pricing game is played between an incumbent monopolist (M) and a potential entrant (E). The game proceeds as follows: (a) M observes its type, high cost (MH) or low cost (ML). The two types are realized with equal probability, with this being common knowledge; (b) M chooses one of seven output levels. M's payoff, shown in Table 1A, is contingent on its type, the output level chosen, and E's response; (c) E sees M's output, but not M's type, and either plays IN or OUT. The asymmetric information, in conjunction with the fact that it is profitable to enter against MHs, but not against MLs, provides an incentive for strategic play (limit pricing) by Ms. E's payoff depends on M's type and on E's decision, not on M's choice. As a treatment variable, two different payoff tables, Tables 1B and 1C, were used for

TABLE 1B—ENTRANT PAYOFFS, HIGH-COST ENTRANTS

Entrant's strategy	Monopolist's type	
	High cost	Low cost
IN	300	74
OUT	250	250

TABLE 1C—ENTRANT PAYOFFS, LOW-COST ENTRANTS

Entrant's strategy	Monopolist's type	
	High cost	Low cost
IN	500	200
OUT	250	250

Es. These represent "high-cost" and "low-cost" Es respectively. Only one of these tables was in use at any given time.[1]

Three features of Table 1A capture the main strategic elements confronting Ms. First, all else being equal, Ms are better off if Es choose OUT rather than IN. Second, reflecting their lower marginal costs, MLs generally prefer higher outputs than MHs. This can be seen in Ms' payoffs should they ignore the effect of their choices on E's behavior—MLs would choose 4 as opposed to 2 for MHs. We refer to choice of 2 by MHs or 4 by MLs as the "myopic maxima." Third, 6 and 7 are dominated strategies for MHs, but not for MLs. At these outputs MLs

[1] Payoffs are given in the experimental currency "francs." Francs were converted to dollars with one franc equal to $0.0025. Headings in Tables 1 and 2 have been changed to match the exposition in the text.

can, in theory, perfectly distinguish themselves from MHs.

For either the high- or low-cost entrant payoff table, it always pays for Es to play IN when M is known to be an MH and OUT against an ML. Given the prior probability of the different M types, however, the expected value of OUT is greater than IN in Table 1B (250 versus 187) and the expected value of IN is greater than OUT in Table 1C (350 versus 250).

B. *Equilibrium Predictions*

For Tables 1A and 1B (games with high-cost Es), there exists multiple pure strategy pooling, as well as separating (sequential) equilibria. Pure strategy pooling equilibria occur at outputs 1 to 5. To understand how these work, consider the pooling equilibrium where both MHs and MLs choose 3. Given the prior probabilities over M's type, E's expected value of OUT is greater than IN following a choice of 3. Sequential equilibrium puts no restrictions on E's beliefs following M's choice of an output level that is supposed to be chosen with zero probability in equilibrium. Therefore, it is admissible for E to believe that any deviation from 3 involves an MH type with sufficiently high probability to induce choice of IN. Given these beliefs and the resulting actions by Es, both MHs and MLs achieve higher profits at 3 rather than deviating to some other output level. The other pooling equilibria are constructed in a similar fashion. The pooling equilibria at outputs 3 to 5 involve limit pricing by MHs—choosing higher outputs (and hence lower prices) than would prevail under full information about Ms' type.

Two pure strategy separating equilibria also exist. In both of these, MHs choose 2 and are always entered on; MLs either always choose 6 or 7 and never incur entry. With MLs choosing 6 or 7, MHs cannot profitably imitate them since 2 dominates 6 and 7 for MHs. Out-of-equilibrium beliefs supporting these equilibria are that any deviation involves an MH type with sufficiently high probability to induce entry. These separating equilibria involve limit pricing by MLs.

For Tables 1A and 1C (the limit pricing game with low-cost Es) the expected value of IN is greater than OUT if both types choose the same output level. This destroys any pure strategy pooling equilibrium, with the only pure strategy equilibria being the two separating equilibria just described.

As is typical of signaling games, the limit pricing game suffers from an overabundance of equilibria. To obtain sharper predictions, one must apply equilibrium refinements. The intuitive criterion (In-Koo Cho and David M. Kreps, 1987) reduces the equilibria in games with high-cost Es to pooling at 4 or 5, and the efficient separating equilibrium with MLs choosing 6. For games with low-cost Es, only the efficient separating equilibrium with MLs choosing 6 survives the intuitive criteria.

Our experiments employ three different versions of this limit pricing game: (a) games with only high-cost Es, so that both pure strategy separating and pooling equilibria exist; (b) games with only low-cost Es, for which the only pure strategy equilibria are separating; and (c) a crossover treatment in which subjects play the game with high-cost Es first, in which play reliably converges to the pooling equilibrium at 4, after which they are switched to the game with low-cost Es. This last treatment tests for learning across games.

C. *Defining Strategic Play*

In the analysis of the experimental data that follows, we focus on the development of strategic play for Ms. Strategic play is defined as choice of outputs 3 to 5 by MHs in games with high-cost Es and choice of outputs 5 to 7 by MLs in games with low-cost Es. These are clearly not the only possible definitions of strategic play we could employ since, in terms of the theory, strategic play depends critically on Ms' beliefs about Es' responses to their actions. Based on the dialogues from the team treatment, however, and substantial circumstantial evidence from other experiments, it is clear that Ms' initial choices (which are overwhelmingly at output 2 for MHs and 4 for MLs) involve attempts to maximize their payoffs *ignoring* the effect of their choices on Es' potential responses. Once Ms begin to consider the effect of their choices on Es' responses, their choices almost exclusively involve MHs choosing 3 to 5 in games with high-cost Es and MLs choosing 5 or 6 in games with low-cost Es.

For example, consider games with high-cost Es where MHs' initial choices are focused on 2 for both teams and individuals (see Figures 1 and 2 in Section IV A below). The team dialogues, analyzed in detail in Section V below, make it clear that MHs are essentially not thinking about Es' responses when making these choices. The following dialogue is a prime example of this:

> "I think we should pick 2. What about you? It is the highest for both [meaning the highest payoff for both choices that Es have]." [brackets added]
> "I agree."
> "OK then. 2 it is."
> "It's the highest for both."

In the coding scheme we have developed for quantifying the dialogues between teammates, dialogues like the preceding are coded under the category "Myopic choice as a monopolist." There is nothing in this dialogue (and other dialogues in the "myopic choice" category) where MHs account for the fact that their choice will influence Es' beliefs about their types and/or have an impact on Es' actions. Even though the high entry rate on 2, as compared to 3, 4, or 5, makes it profitable for MHs to limit price from the earliest rounds of the experiments, the myopic choice category is the second-highest coded category prior to a first attempt of MH teams to play strategically (choose 3, 4, or 5).[2] In other words, MHs aren't initially choosing 2 because they (incorrectly) anticipate that it will lead to favorable responses by Es, but rather because they fail to consider that their choice will have an impact on E's response.

Additional circumstantial evidence pointing to the fact that *initially* most Ms fail to consider Es' responses to their choices comes from several sources. Most tellingly, in both earlier experiments with the limit pricing games as well as the current dataset, there are no significant differences between MLs' early plays of the game regardless of whether they faced high or low cost Es, even though entry rates on 4 are far higher when playing against low-cost Es. Further, there is substantial evidence from a number of other experiments that subjects fail to account correctly for their rivals' actions in formulating their initial choices, and only gradually learn to do so.[3]

II. Connections to Team Play in the Psychology and Economics Literatures

A. *Connections to the Psychology Literature on Teams*

The substantial psychology literature on team versus individual subject play distinguishes between judgmental tasks and so-called "eureka" problems.[4] Judgment tasks involve settings where either there isn't a "correct" action, or the correct action is highly unlikely to be discovered by an untrained subject. For example, there are numerous studies comparing the attitudes of teams and individuals toward risk. While the choices of teams systematically differ from those of individuals, they cannot meaningfully be termed more or less correct.[5] By way of contrast, eureka problems have a correct solution (or solutions). While this solution may be difficult to discover, it is self-confirming once discovered and can easily be demonstrated to others. Logic problems like the Tower of Hanoi are good examples of eureka problems. This classic puzzle, while challenging for a novice, is solved by a simple algorithm that can be explained in a few sentences. In other words, once the puzzle has been solved it is easy to show others that it has been successfully solved.

We argue that in signaling games, as subjects actually play the game, the discovery of strategic play corresponds closely to solving a eureka-type problem. The "aha" type insight in

[2] The highest coded category involves taking notice of the feedback information about the population choices of Ms, as well as Es' responses to same.

[3] For example, in describing the failure of the intuitive criterion in signaling games, Brandts and Holt (1992, p. 1357) note that, "a type-L's deviation to decision A might be motivated by the belief that the 'signal,' A or B, will have no effect on the respondent's decision. These beliefs are contradicted by the actual decisions of the respondents ... ."

[4] Much of our discussion of the psychology literature is based on Davis's (1992) review article (see also Gayle W. Hill, 1982).

[5] For a general survey of the psychology literature dealing with group versus individual decisions for which there are no demonstrably correct answers, or difficult statistical decisions, see Norbert L. Kerr et al. (1996).

signaling games is for Ms to realize that their actions affect Es' beliefs and, by extension, choices of Es. Once this insight has been reached, it is relatively straightforward to realize that MHs can gain by imitating MLs (for games with high-cost Es) or MLs can benefit by distinguishing themselves from MHs (for games with low-cost Es). Further, as we will show, these insights are largely self-confirming, since by the time they occur, there is massive evidence that strategic play will have the desired effect of reducing entry.

For many years the prevailing wisdom among psychologists for eureka-type problems was consistent with the folk wisdom that teams outperformed individuals. For example, in a classic experiment, Marjorie Shaw (1932) observed ad hoc, freely interacting, four-person groups working on word puzzles.[6] Shaw found that in most problems both the proportion of solutions and the time to find a solution was superior for groups than for a comparable sample of individuals working privately. This is a typical result when comparing groups and individuals directly, a result attributed to the ability of group members to catch others' errors, reject incorrect solutions, and generally stimulate more thoughtful work (Davis, 1992).

By the mid-1950s, findings of superiority for groups in eureka-type problems came to be viewed with some suspicion. The stimulus for reevaluation was work by Irving Lorge and Herbert Solomon (1955) who proposed the following truth-wins (TW) standard against which to evaluate the superiority of group performance: Assuming that group interactions are neutral, the group should be able to achieve a correct answer if at least one member proposes it. Therefore, if the probability of an individual working alone solving the problem equals p, the probability P of a randomly selected group with r members solving the problem is the probability that this random sample contains at least one individual who can solve the problem: $P = 1 - (1 - p)^r$. The Lorge-Solomon baseline provides a quantifiable measure of synergy. If teams exceed this standard, the inference is that the interaction among group members generates better performance than the individuals could achieve acting independently. Likewise, failure to reach the Lorge-Solomon baseline indicates that interactions actually make teams perform worse than their members would do acting alone.

Lorge and Solomon reanalyzed Shaw's data, with other researchers doing the same for a number of other datasets. The results of this revaluation establish that "... freely interacting groups *very* rarely exceed, sometimes match, and usually fall below the Lorge-Solomon [TW] baseline" (Davis, 1992, p. 7, italics in the original).[7] Psychologists attribute this relative "inefficiency" of groups to social process losses such as reduced effort due to free-riding or coordination problems involved in combining team members' contributions—problems familiar to economists. The first relates to the economics literature on shirking in team production processes (Armen A. Alchian and Harold Demsetz, 1972). The second relates to diminishing marginal productivity and to common coordination issues involved in team production.[8]

Although there is far from a one-to-one mapping between strategic play in the signaling games studied here and the individual choice problems that psychologists use to compare teams with individuals, we argue that learning to play strategically is sufficiently close to a eureka problem as to make comparisons relevant. The comparison is of interest both because beating the TW norm provides evidence for the presence of positive synergies in team play and because, given the frequent evidence from the psychology literature that this norm is *not* satisfied, beating the TW norms is impressive

[6] For example, in one problem there are three cannibals and three missionaries on one side of a river. The puzzle is to get them across to the other side by means of a boat that holds only two people at a time. Further, all missionaries but only one cannibal can row, and never under any circumstances can the cannibals outnumber the missionaries.

[7] Davis (1992) does not discuss the effects of team size in his survey, suggesting that it is at best of second-order importance.

[8] There are, of course, reasons other than process loss as to why teams might fail to meet or beat the TW norm. Communication itself is a costly enterprise that reduces the time subjects have available for solving the problem at hand. Also, if a team randomly selects a dictatorial decision maker who does not consult with team members prior to choosing, average results for these teams would not beat the TW norm. Neither of these factors is likely to have played a role in our experiment.

evidence of the efficacy of team play.[9] This is not to say that the TW norm is the only relevant reference point against which to evaluate the desirability of teams from an economic perspective. That, of course, depends on the context and on the additional costs of using teams as opposed to whatever additional benefits they confer.[10]

B. *Connections to the Economics Literature*

There have been a handful of studies of group versus individual performance in the economics literature. We focus on the subset of these papers that is most relevant to our study, those that study behavior in games.

Gary Bornstein and Ilan Yaniv (1998) study individual versus team behavior in a standard, one-shot ultimatum game experiment. Their main result is that three-member teams are more game-theoretically rational players than individuals, as they demand more than individuals as proposers and are willing to accept less in the role of responders. In contrast, in a dictator game experiment, Timothy N. Cason and Vai-Lam Mui (1997) find that team choices tend to be dominated by the more "other-regarding" member of the team.[11]

James C. Cox and Stephen C. Hayne (2002) explore differences between group and individual bids in common value auctions for once- and twice-experienced bidders. They focus on "rational" bidding, defined as bidding low enough to avoid falling prey to the winner's curse. With a signal sample size of 1, there are no material differences between groups (of size 5) and individuals in bidding. In contrast, with a signal sample size of 5, groups tend to be *less* rational than individuals (reported in two of four treatments, with no differences found in the two other treatments).

Martin G. Kocher and Matthias Sutter (2005) (see also Sutter, 2004) compare individual and team behavior in "beauty-contest" games where decision makers compete for a fixed prize by simultaneously guessing a number in a given interval. The winner is the decision maker whose number is closest to a predetermined fraction of the average of all the numbers that everyone picks. If this fraction is less than one, the unique serially undominated strategy is to guess zero. In practice, optimal guesses depend both on one's own insight into the equilibrium outcome and on the degree to which one believes that others have the same insight. There are no differences between teams (of size 3) and individuals in the first round of the game, but teams learn faster than individuals as they choose lower numbers in subsequent rounds.

To synthesize the results of these experiments, teams do the same or somewhat better than individuals (with the possible exception of the 5 signal treatment in Cox and Hayne). Our experiment differs substantially from the preceding literature on team play in games. In no case do earlier investigators compare teams against individuals using the demanding truth-wins norm employed here. This is appropriate given that the games being studied rely less on a eureka type insight than on subjects' judgments. For example, in the ultimatum game, rejection or acceptance of offers closer to the subgame perfect equilibrium outcome are tied to whether or not own income is the only argument of players' utility functions, a matter of preferences rather than logic.[12] In contrast, we argue that for signaling game experiments there is a eureka insight to be gained, and that the truth-wins norm therefore provides a relevant benchmark for the presence of synergies associated with team performance.

A further key innovation of our study is that we have analyzed the team dialogues to obtain insights into subjects' learning and reasoning

[9] One might also object to the terminology "truth wins" in the context of games. We employ it primarily to connect with the relevant psychology literature.

[10] Further, as one of our referees suggests, outside the lab, team size is likely to be endogenous and may turn out to be optimal for the problem at hand.

[11] James C. Cox (2002), on the other hand, finds that teams return significantly smaller amounts in the trust game than do individuals.

[12] In the beauty contest game, a subject's judgment about how logical other players will be is as important as their own ability to reason about the game. In common value auctions, formulating a strategy to avoid the winner's curse is clearly a task that is demonstrable. It involves avoiding a decision-making bias, however, that virtually all subjects, both students and those presumably practiced in industries subject to the curse, fall prey to (Kagel and Dan Levin, 2002).

processes. This is as much a purpose of the present paper as is making comparisons between teams and individuals.

III. Experimental Design and Procedures

In what follows we refer to individuals who participated in our experiment as "subjects," while "players" refer to agents in the limit pricing game. A player is a single subject in the individual player (1 × 1) sessions, but consists of two subjects in the team (2 × 2) sessions.

A. *General Procedures*

Subjects were recruited through announcements in undergraduate classes, posters placed throughout the Ohio State University campus, advertisements in the campus newspaper, and direct e-mail contact. This resulted in recruiting a broad cross section of mostly undergraduate students and some graduate students. Experienced subject sessions generally took place about a week after the inexperienced subject sessions. Subjects from different inexperienced sessions were mixed in the experienced subject sessions, but subjects were *not* switched between the 1 × 1 and 2 × 2 treatments.[13]

Inexperienced 2 × 2 sessions lasted approximately two hours; inexperienced 1 × 1 sessions lasted approximately one and a half hours. Experienced subject sessions were substantially shorter than this as short, summary instructions were used and subjects were familiar with the game. Subjects were paid $6 for showing up on time with total earnings averaging between $26 and $27 per subject in inexperienced subject sessions. Earnings were higher in experienced subject sessions, averaging a little over $32 person (including the $6 show-up fee), largely as a result of playing more games.

The 1 × 1 sessions were designed to employ between 12 and 16 subjects, with the 2 × 2 sessions employing between 20 and 28 subjects.[14] This results in a somewhat smaller number of players in the team sessions (since each player requires two subjects), but was dictated by the difficulty of assembling larger numbers of subjects, as well the fact that the 2 × 2 treatment must be run in multiples of four subjects (two pairs) and the lab has only 30 work stations. All but two sessions included an even number of subjects so that all subjects participated in every round. The two exceptions were two inexperienced 2 × 2 sessions with 23 subjects, where the experimenter served as the twenty-fourth subject to avoid discarding three subjects. In these cases the solo player was told that her teammate was the experimenter who would agree to all of her choices without any further communication.

Upon arrival, subjects were randomly assigned to computer terminals. A common set of instructions was read aloud, and each subject was given a written copy.[15] All sessions employed abstract terms throughout. For example, Ms were referred to as "A players," with types "A1" and "A2," respectively, and potential Es were described as "B players." Other terms were given similarly meaningless labels. Subjects had copies of both Ms' and Es' payoff tables and were required to fill out a short questionnaire to insure their ability to read them. After reading the instructions, questions were answered out loud and play began with a single practice round followed by more questions.

Before each play of the game the computer randomly determined each M's type and displayed this information on Ms' screens. The screen also showed the payoff tables for both types with the table for that player's type displayed on the left. Ms chose by clicking the output level on the payoff table displayed on their screens. The program automatically highlighted Ms' possible payoffs and required that the choice be confirmed. After *all* Ms had confirmed their choices, each M's choice was sent to the E they were paired with. Es then decided between IN and OUT by clicking the appropri-

[13] Econometric analysis indicates that there are no systematic differences between choices in the inexperienced sessions for subjects who later returned for an experienced subject session and those who did not.

[14] One experienced 1 × 1 session was run with ten subjects and one experienced 2 × 2 session was run with 16 subjects to avoid losing hard to obtain experienced subject data. The first inexperienced 2 × 2 session had 12 subjects. Normally, a session with such a low turnout would have been canceled, but this one was run in part as a final test of the software.

[15] A copy of the instructions is available at http://www.e-aer.org/data/june05_app_kagel.zip.

ate choice on their payoff table. Here, too, possible payoffs were highlighted and subjects were required to confirm their choices.

Following each play of the game, subjects learned their own payoff and Es were told the type of M they were paired with. In addition, the lower-left-hand portion of each player's screen displayed the results of all pairings: M's type, M's output, and E's response ordered by output levels (pooled over *all* M types) from highest to lowest.[16] The screen automatically displayed the three most recent periods of play, with a scroll bar available to see all past periods.

The following rotation procedures were generally employed: subjects switched roles with Ms becoming Es and vice versa every six games for inexperienced sessions, and every four games for experienced sessions. We refer to a block of 12 (8) games in an inexperienced (experienced) session as a "cycle." Within each half-cycle, each M was paired with a *different* E for each play of the game. Inexperienced subject sessions had 24 games divided into two 12-game cycles. Experienced subject sessions had 32 games, divided into four 8-game cycles. The number of games in a session was announced in the instructions. The exceptions to these general procedures occurred in the process of discovering how many replications we could achieve within two hours for the 2×2 treatment.[17] The first two 1×1 inexperienced low-cost E sessions had three 12-game cycles (36 games), as we were still hopeful that we would be able to complete a comparable number of games in the team sessions. The first inexperienced 2×2 session switched roles every four games, with a second inexperienced 2×2 session switching roles every five games. A third inexperienced 2×2 session completed only 18 games (1.5 cycles) due to time constraints (exacerbated by a computer crash). Note that these few deviations from the norm (five out of 33 sessions) all resulted in 2×2 subjects having less experience than 1×1 subjects, which would tend to favor more rapid adjustment to equilibrium in the 1×1 sessions, contrary to the outcomes observed.

B. *Team* (2×2) *Procedures*

Team pairings were determined randomly by the computer at the beginning of each session. Matches could not be preserved between inexperienced and experienced subject sessions due to attrition and mixing of subjects from different sessions.[18] Subjects were not told the identity of the person they were matched with and were asked not to identify themselves.

Teammates were able to communicate and coordinate their decisions using an instant messaging system with full knowledge that these messages would be recorded, but with no other team having access to their messages.[19] In addition to instructing subjects that the instant messaging system was intended to be used for coordinating their decisions, subjects were told to be civil to each other and not to use any profanity. Otherwise, subjects were given no instruction about what messages to send.[20] The message system was open almost continuously, and messages were time stamped with the period of the game being played.

When teams made choices, the relevant payoff table on the screen had a column labeled "partner's choice" on the left and a column labeled "my choice" on the right. When a subject entered a choice, the possible payoffs were

[16] The use of public information in the 1×1 treatment might be expected to crowd out some of the beneficial effect of discussions between team members, as it provides a large amount of group information.

[17] The two-hour constraint on experimental sessions is designed to avoid subject fatigue.

[18] In three of the team sessions, the software had to be restarted, which necessitated new team pairings. Two of these restarts were due to software crashes; the third was due to the session running beyond its advertised time, necessitating the release of some subjects.

[19] The team effect is, of course, inherently confounded with the effect of the particular communication channel the teams use. We have no reason to suppose that written communication is any different from verbal communication, especially for subjects who have grown up with e-mail and instant messaging on the Internet. Two key advantages of instant messaging over face-to-face discussions are (a) transcripts of dialogues are created automatically, and (b) it would be impossible to have team discussions in the lab while preserving confidentiality between team members.

[20] Beyond these instructions, there was no attempt to prevent subjects from sending any message they desired. Virtually all of the discussions were civil. Many teams discussed topics in addition to the experiment and there was some use of profanity. There is little evidence that subjects were inhibited by the knowledge that their messages were being recorded.

TABLE 2—SUMMARY OF EXPERIMENTAL TREATMENTS

	1×1 Treatment	2×2 Treatment
Inexperienced high-cost entrants	5 sessions 70 subjects	6 sessions 128 subjects
Inexperienced low-cost entrants	4 sessions 64 subjects	4 sessions 104 subjects
Experienced low-cost entrants	3 sessions 42 subjects	3 sessions 67 subjects
Crossover sessions[a] High-cost → Low-cost entrants	4 sessions 50 subjects	4 sessions 64 subjects

[a] Data for experienced subjects in games with high-cost Es come from first cycle in this treatment.

highlighted in blue. When a subject's partner entered a choice, the possible payoffs were highlighted in pink. Once choices coincided, possible payoffs were highlighted in red, at which point teammates had four seconds to change their choice before it became binding. Team play started with three minutes to coordinate choices, with a countdown clock shown on the computer screens.[21] If teams failed to coordinate within this time constraint, the dialogue box was closed and one teammate was randomly selected as "leader" with his choice implemented unilaterally. There were virtually no disagreements of this sort.

C. *Experimental Design and Hypotheses*

Table 2 summarizes the four types of experimental sessions conducted: sessions with inexperienced subjects and high-cost Es; sessions with inexperienced subjects and low-cost Es; sessions where subjects who had played in inexperienced subject sessions with low-cost Es were recruited back for more games with low-cost Es; and crossover sessions where subjects who had played in games with high-cost Es were recruited back. Crossover sessions started with a full cycle (eight games) with high-cost Es. This was followed by 24 plays (three cycles) with low-cost Es. When Es' payoffs were changed, subjects were given written copies of the new payoff tables, with a brief set of instructions read out loud indicating that the only change in procedures involved new payoffs for Es.

For each cycle of the games played we examine two related hypotheses:

Hypothesis 1: There will be more strategic play in each of the 2×2 games than in the corresponding 1×1 games.

Hypothesis 2: The level of strategic play in the 2×2 games will meet or beat the TW norm, as described in Section II A, based on the level of play observed in the 1×1 games.[22]

Hypothesis 1 constitutes the minimal requirement to justify team decision making on purely economic grounds.[23] Hypothesis 2, if validated, indicates that interactions between teammates generate positive synergies, giving them a greater chance of success working together than the most able member of the team would have working independently.

Hypothesis 2 is based on the argument that strategic play in signaling games, as it occurs in the lab, has strong similarities to the puzzle-solving branch of the psychology literature dealing with team versus individual play. There are important differences, however, between the sort of individual choice problems studied by psychologists and a strategic environment like the limit pricing game. The most critical of these rests on the inherently stochastic nature of Es' choices. It is well established that subjects' ability to respond in games is sensitive to the payoff premium (Raymond C. Battalio et al., 2001). While playing strategically almost always maximizes expected payoffs in our games, the premium for playing strategically varies quite a bit. To the extent that this premium varies between the 2×2 and 1×1 treatments, it could cause the 2×2 treatments to either overperform or underperform relative to the TW

[21] This time was reduced by 30 seconds between cycles except for cycles following a crossover.

[22] More technically, suppose p_t is the probability that, given the opportunity, an individual plays strategically in cycle t, and P_t is the corresponding probability for a two-person team. Beating the TW norm in cycle t requires $P_t \geq 1 - (1 - p_t)^2$.

[23] Team decision making can, of course, be justified on other grounds including legal, administrative, or political reasons.

norm. We use probit regressions to address this issue since this permits comparisons of the level of strategic play between treatments, while controlling for possible differences in the incentives Ms face for playing strategically.[24] These probits are reported in detail in the Appendix, with the results summarized in the text. They are important to our analysis of whether or not Hypotheses 1 and 2 are supported by the data, but the details can be somewhat tedious and distracting from the overall analysis.

We pose a third hypothesis specific to the crossover sessions. In addition to giving us another venue for comparing the decision making of teams versus individuals, this treatment also allows us to address the question of learning generalizability—the ability to take experience with one game and apply it in a related game. Learning generalizability is an important issue in psychology, and important to economic arguments that rely on learning processes to justify equilibrium outcomes.[25] The usual result from psychology experiments is that there is little or no learning transfer (and sometimes even negative learning transfer) unless subjects are explicitly queued to draw on their previous experiences (see, for example, Gavriel Solomon and David N. Perkins, 1989).[26] One could argue that positive learning transfer is likely to be particularly difficult here since strategic behavior following this crossover requires substantially different actions than before the crossover. Prior to the crossover, MHs learn to imitate MLs, but MLs must learn to distinguish themselves following the crossover.[27]

Hypothesis 3: There will be little or no learning transfer in the crossover treatments. That is, levels of strategic play will be the same, or possibly even lower, following the crossover compared to inexperienced subjects in games with low-cost Es. This will hold for both the 2×2 and the 1×1 treatments.

IV. Experimental Results

A. *Limit Pricing in Games with High-Cost Entrants*

Figure 1 aggregates data from the 1×1 sessions for games with high-cost Es. These data provide a baseline for how strategic play evolves for MHs. In the first cycle of play, choices of each type are concentrated at their myopic maxima; 2 is the modal choice for MHs and 4 is the modal choice for MLs. The lack of strategic play by MHs cannot be attributed to a lack of incentives; even in cycle 1 entry rates for 2 are 46.3 percent higher than for 4. A difference of only 13 percent is needed to make strategic play profitable for MHs.

In the second cycle the difference in entry rates between 2 and 4 becomes even more pronounced, rising to 59.4 percent. Responding to these strong incentives, MHs begin to play strategically with greater frequency, with 4 becoming their modal choice. At the same time, play of MLs becomes even more concentrated at 4. The first cycle of experienced subject play continues these trends: the entry rate differential between 2 and 4 rises slightly, MHs play strategically even more frequently, and MLs choose 4 almost exclusively.

Figure 2 reports the results for the 2×2 sessions for each cycle of play. Comparing Figure 1 with Figure 2, it's clear that the general dynamics of play are similar, as MHs in the $2 \times$

[24] An alternative, and superior, method for dealing with this issue would be to conduct sessions in which both teams and individuals are playing at the same time. Unfortunately our software, which took some time to develop, cannot accommodate such a design. This is one of several issues to be investigated in the future.

[25] Virtually all studies of learning in economics employ an environment in which learning takes place within an essentially stationary environment. Studies of cross-game learning are important since it's unreasonable to expect the exact same game to be repeated over and over, so that if one can justify only convergence to equilibrium in such situations, there would not be much reason to have faith in the widespread applications that are found in the literature. Rather, faith in such applications can be greater if players infer how their opponents will act in one situation as opposed to how they acted in other, related situations.

[26] Here, too, psychology experiments deal primarily with puzzles, or individual skills, and not strategic situations such as those involved in signaling games.

[27] A fictitious play learning model that has worked well in tracking play from previous signaling games (Cooper et al., 1997b) predicts that MLs' strategic play immediately following the change in Es' payoffs will be *less* than in inexperienced control sessions (negative transfer), and will remain so until behavior converges to the equilibrium outcome.

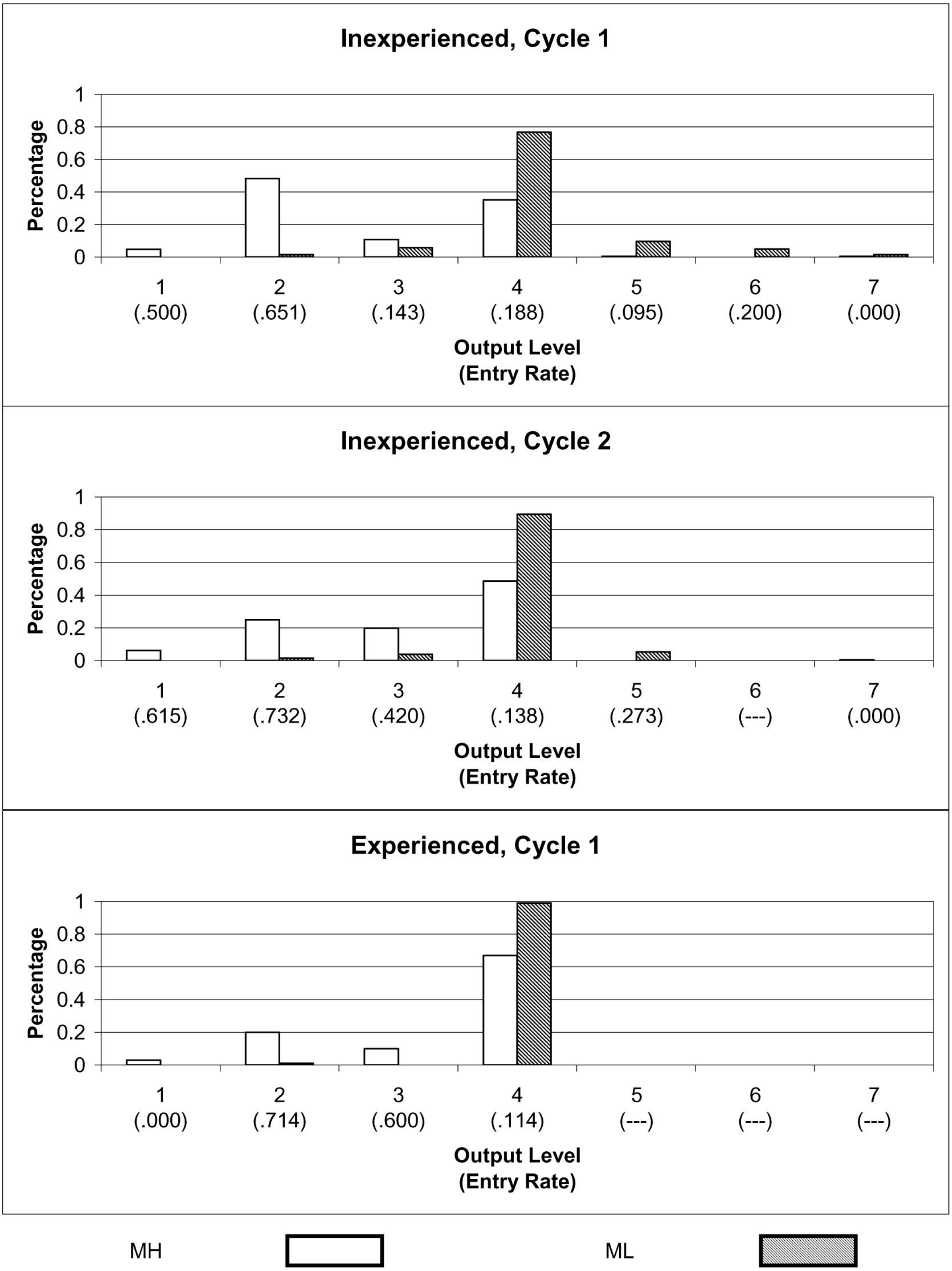


FIGURE 1. POOLED DATA FROM 1 × 1 SESSIONS FOR GAMES WITH HIGH-COST ENTRANTS

Notes: Pure strategy pooling and separating equilibria exist. Bars indicate choice frequencies for MHs and MLs for each possible output level, with entry rates for those output levels shown in parentheses. Cycles consist of 12 (8) plays of the game for inexperienced (experienced) subjects.

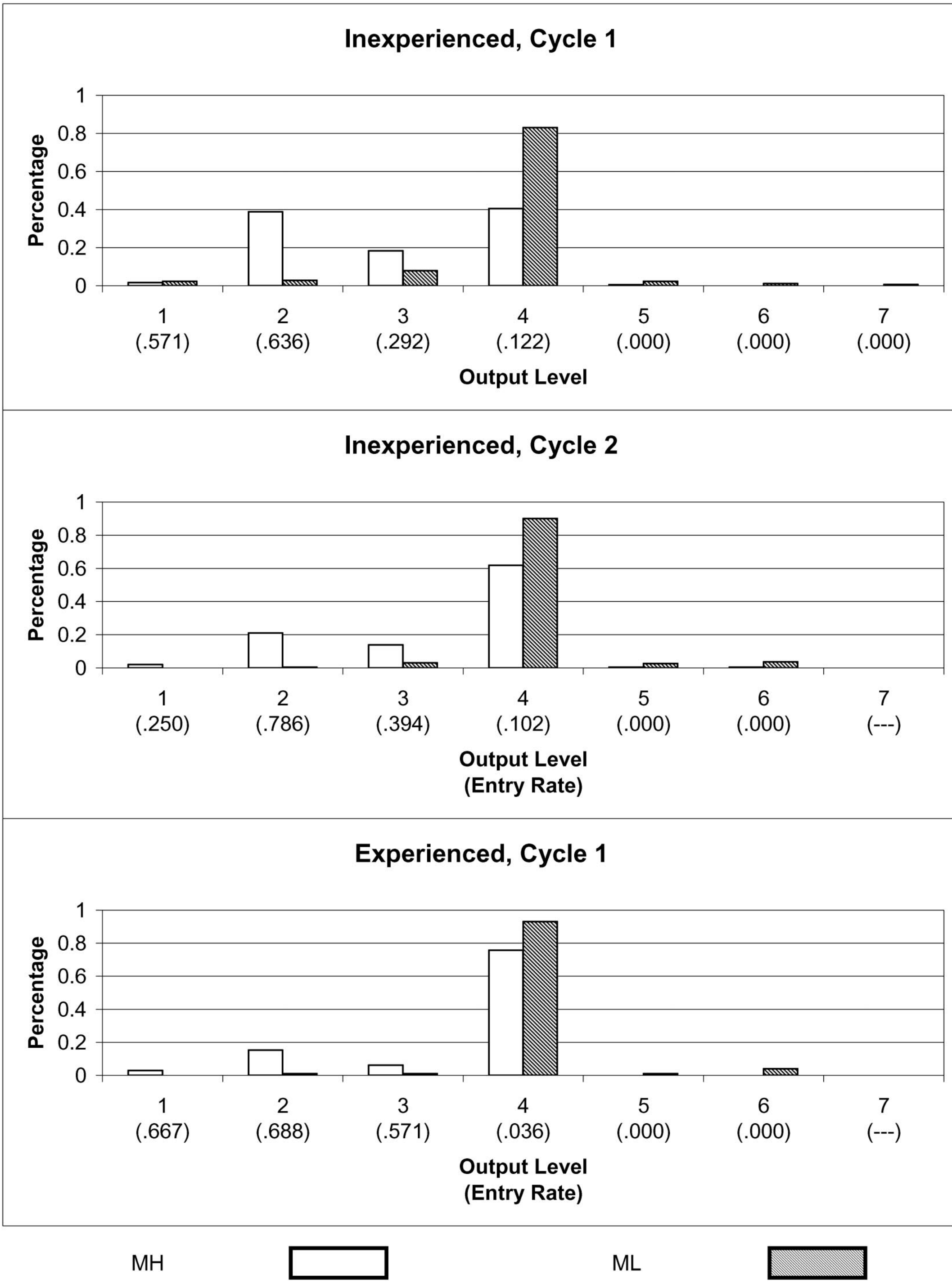


FIGURE 2. POOLED DATA FROM 2 × 2 SESSIONS FOR GAMES WITH HIGH-COST ENTRANTS

Notes: Pure strategy pooling and separating equilibria exist. Bars indicate choice frequencies for MHs and MLs for each possible output level, with entry rates for those output levels shown in parentheses. Cycles consist of 12 (8) plays of the game for inexperienced (experienced) subjects.

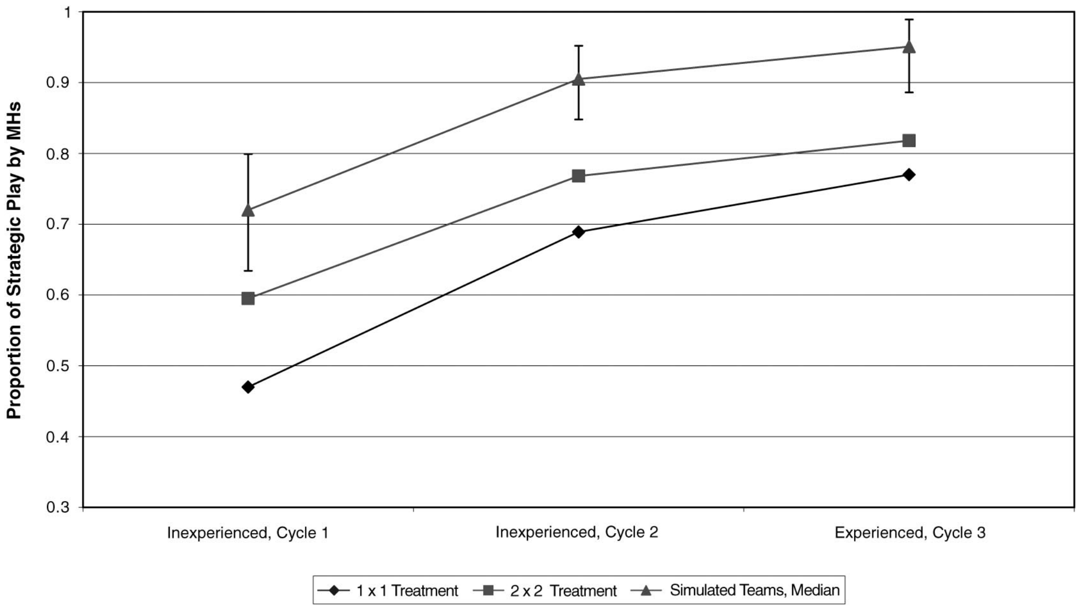


FIGURE 3. COMPARING THE DEVELOPMENT OF STRATEGIC PLAY FOR MHs IN 2 × 2 WITH 1 × 1 SESSIONS IN GAMES WITH HIGH-COST ENTRANTS

Notes: Vertical axis shows frequency of MHs choosing outputs 3–5. Horizontal axis shows cycle of play. Bars give the 90-percent confidence interval for the truth-wins standard.

2 sessions only gradually move from myopic play at 2 to strategic play at 4. Within the first cycle of play there is a distinctly higher frequency of strategic play by MHs in the 2 × 2 treatment, which continues into later cycles as well. Moreover, strategic play is relatively more concentrated on 4 (rather than 3) in the 2 × 2 treatment.

Figure 3 directly compares levels of strategic play (MHs' choices of 3, 4, and 5) between the 1 × 1 sessions and the 2 × 2 sessions, along with the TW norm (shown as filled triangles with error bars for the 90-percent confidence interval).[28] Strategic play emerges somewhat more rapidly for MHs in the 2 × 2 treatment, being some 12.5 percent greater than in the 1 × 1 sessions for the first cycle of inexperienced subject play. The difference between treatments in the proportion of strategic play narrows over time, with 7.9 percent greater strategic play for 2 × 2s in the second cycle of inexperienced subject play, and 4.8 percent greater strategic play in the first cycle of experienced subject play. Quite clearly, Hypothesis 2 fails to be satisfied, as the level of strategic play in the 2 × 2 sessions is well below the 90-percent confidence interval for the TW norm. Thus, differences between the two treatments largely reflect a greater proportion of MHs who immediately

[28] Because of clustering in the data, simulations are needed to calculate the error bars correctly. The simulated 2 × 2 data are based on 100,000 simulated 2 × 2 datasets for each cycle of play, with the same number of teams in each dataset as in the experiment. Simulated 2 × 2 play is based on randomly drawing two individuals (with replacement) from the 1 × 1 sessions. The likelihood of any individual being drawn is proportional to the number of times that individual was an MH in that cycle, with the probability of playing strategically based on the observed frequency of strategic play as an MH in that cycle. A simulated team was considered to have played strategically if either of its members played strategically. The error bars then display the fifth and ninety-fifth of the distribution of percentages of strategic play in a simulated 2 × 2 dataset. If the percentage of strategic play in the actual 2 × 2 data is below the error bar for the simulated data, as in all three cycles of the data for games with high-cost Es, this indicates that over 95 percent of the simulated datasets yield more strategic play than the actual data.

play strategically in the 2 × 2 treatment rather than faster learning, with the teams failing to generate positive synergies as measured by the TW norm.

The results reported in Figure 3 suggest that Hypothesis 1 holds, albeit weakly, for games with high-cost Es. To examine this more formally, we ran probit regressions for MHs' choices, with the dependent variable being whether or not MHs played strategically (i.e., chose output 3, 4, or 5).[29] This analysis finds that (a) without any controls for differential entry rates, there is significantly more strategic play in the 2 × 2 treatment for the first cycle of inexperienced subject play ($p < 0.10$), with no significant differences in later cycles; (b) adding controls for Es' choices shows that Ms respond only weakly to differences in entry rates, as the parameter estimate for this variable is not significant at conventional levels; and (c) the entry rate controls have little impact on the estimated differences between 1 × 1 and 2 × 2 play.

The results for games with high-cost Es may be summarized as follows:

Conclusion 1: In games with high-cost Es, MHs play more strategically with teams than with individuals, with maximal differences observed in the first cycle of inexperienced subject play. Teams do not, however, meet or beat the TW norm.

B. *Limit Pricing in Games with Low-Cost Entrants*

Figure 4 aggregates data from 1 × 1 sessions for games with low-cost Es.[30] Play in the first cycle for inexperienced subjects is similar to games with high-cost Es, with Ms' play clustered at the myopic maxima for both types. There are strong incentives to play strategically as an MH in this first cycle of play, but only weak incentives for MLs to play strategically. One notable difference in early play between these games and those with high-cost Es is that entry rates are much higher here for outputs 2 to 4, consistent with the substantially higher payoffs for IN versus OUT. By the second cycle of inexperienced play there exist strong incentives for both MHs and MLs to play strategically. Steady movement toward strategic play takes place for both types, but this movement is smaller for MLs than for MHs, even though MLs have stronger incentives than MHs to play strategically in the second cycle.

Experienced sessions with low-cost Es are largely a continuation of patterns from inexperienced subject play. MHs increase their play of 4 going from cycle 1 to 2, with little retreat back to 2 thereafter, as it is incentive-compatible for them to choose 4 over 2 throughout (albeit somewhat less so in later cycles of play). MLs' level of strategic play in cycle 1 is roughly what it was in the ending cycle of inexperienced subject play, but steadily increases thereafter, with strategic play largely directed to output 6. This movement is quite slow—only in the third experienced cycle does 6 become the modal choice for MLs. In contrast, in the games with high-cost Es, strategic play by MHs is the modal outcome by the second cycle of the *inexperienced* sessions. This difference is even more striking since the incentives for MLs to play strategically here are roughly the same as for MHs to play strategically in the previous treatment. There is clearly something quite different, and more difficult, about learning to play strategically as an ML than as an MH.

Figure 5 aggregates data from the 2 × 2 sessions for games with low-cost Es. The general pattern of play is the same as in the 1 × 1 sessions, but convergence to the efficient separating equilibrium is much more rapid and complete. Comparing Figure 4 with Figure 5, the more rapid development of strategic play for teams is clear by the second cycle of inexperienced subject play. More striking yet is the near complete convergence to the efficient separating equilibrium in the last two cycles of experienced subject play in the 2 × 2 games, as compared to the 1 × 1 sessions. Not only is there far more strategic play by MLs in the 2 ×

[29] Running random effects probits with the 2 × 2 treatment poses some nonstandard statistical issues. We employ very conservative assumptions regarding the degree of independence between team members, particularly with respect to experienced players which, if anything, bias our results against finding statistical significance between the two treatments. Tests for robustness of the probit results to alternative controls for individual and team effects, definitions of strategic play, and controls for Es' behavior are reported at http://www.e-aer.org/data/june05_app_kagel.zip.

[30] Data from the crossover treatments are *not* included here but are discussed in the next section.

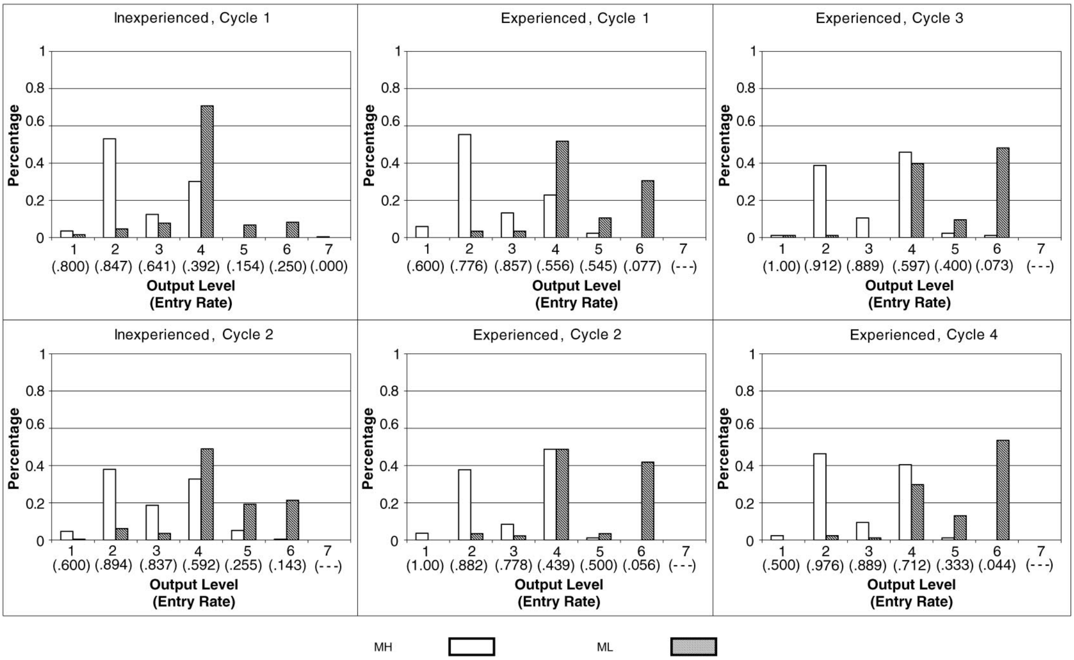


FIGURE 4. POOLED DATA FROM 1 × 1 SESSIONS FOR GAMES WITH LOW-COST ENTRANTS.

Notes: Only pure strategy separating equilibria exist. Bars indicate choice frequencies for MHs and MLs for each possible output level, with entry rates for those output levels shown in parentheses. Cycles consist of 12 (8) plays of the game for inexperienced (experienced) subjects.

2 games for these final cycles, but it is much more heavily concentrated on 6 compared with 5 in the 1 × 1 sessions. Further, there is 100 percent entry on 4, and MHs have almost completely retreated back to choosing 2 in the 2 × 2 games. In contrast, in the corresponding cycles of 1 × 1 sessions, the entry rate on 4 relative to 2 is still sufficiently low that choice of 4 is incentive compatible for MHs.[31]

Figure 6 compares the level of strategic play by MLs between the 2 × 2 and 1 × 1 treatments, as well as simulated outcomes for the TW norm. The results are striking: teams play more strategically than individuals throughout, with the difference growing steadily with experience. They match the TW norm in the two cycles of inexperienced subject play, and then either beat it or fall into the upper tail of the 90-percent confidence interval, for all four cycles of experienced play.[32]

Probits reported in the Appendix show that (a) the differences in the level of strategic play

[31] It might be argued that individuals are likely to keep learning over time so that with enough experience they will catch up with teams. However, we have no evidence for this to date in spite of, in one instance, bringing back twice experienced subjects (see, for example, Cooper et al., 1997b, Figure 5), so the jury is still out on this question. The closest we have seen to anything like the teams' convergence to the efficient separating equilibrium for individuals occurred in games where MHs were prohibited from choosing 6 or 7, and this was announced as part of the instructions (see Cooper et al., 1997a).

[32] While strategic play by MHs (choice of 3, 4, or 5) is clearly an important feature of the data in games with low-cost Es, we do not compare the development of this behavior across treatments. Unlike strategic play by MLs, strategic play by MHs in the low-cost entrant game does not cleanly fit our definition of a eureka-type problem. Recall that a eureka-type problem has a *demonstrably* correct solution. While the entry rate differential between 2 and 4 starts out large, it shrinks steadily over time. So although there is scope, early on, for an "aha" type insight for MHs (in that they can profitably imitate MLs), at some point (unknown to us) individual MHs become increasingly disabused of the profitability of choosing 4 over 2.

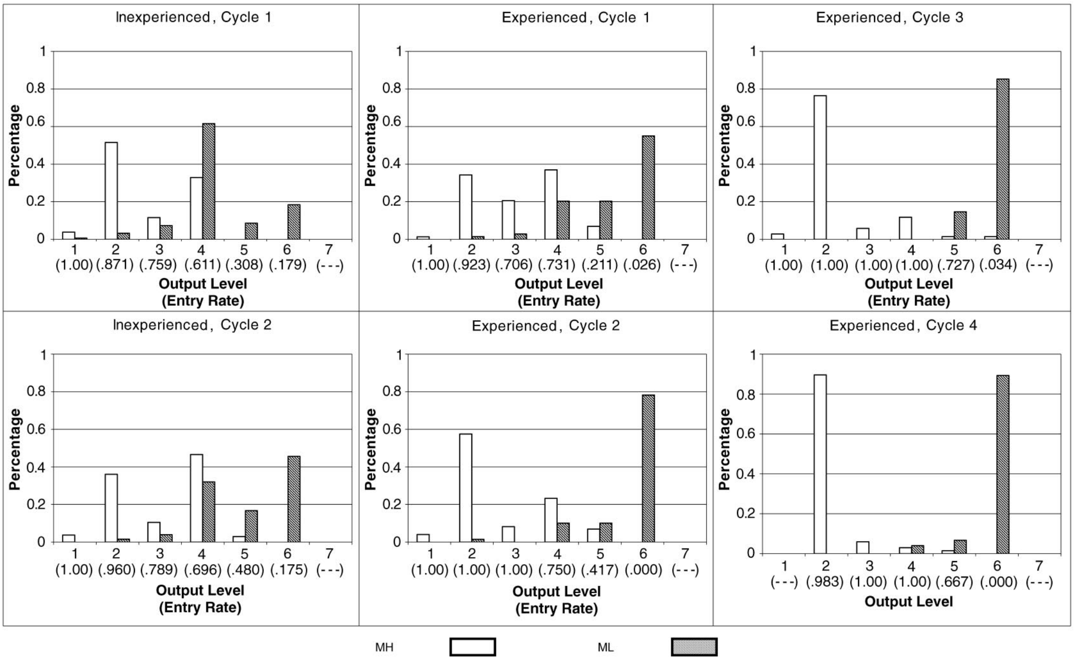


FIGURE 5. POOLED DATA FROM 2 × 2 SESSIONS FOR GAMES WITH LOW-COST ENTRANTS

Notes: Only pure strategy-separating equilibria exist. Bars indicate choice frequencies for MHs and MLs for each possible output level, with entry rates for those output levels shown in parentheses. Cycles consist of 12 (8) plays of the game for inexperienced (experienced) subjects.

for MLs between the team and 1 × 1 treatments are statistically significant in all cycles of play; (b) entry rates have a strong impact on MLs choosing to play strategically; and (c) controlling for entry rate differentials, MLs' level of strategic play is still significantly higher in the teams' treatment for four of the six cycles, although the magnitude of the effect is reduced.

As the preceding suggests, the ability of teams to meet or beat the TW norm may be due to greater entry rate differentials in the 2 × 2 treatment. The results of the probit analysis allow us to explore this issue in more detail. We calculate for each cycle what the proportion of strategic play by MLs would have been in the 2 × 2 sessions, *if the entry rate differential had been identical, in that cycle of play, to the* 1 × 1 *sessions,* and compared these adjusted percentages with the TW norm. These calculations show that after the first cycle of inexperienced subject play, teams meet the TW norm, but beat it only in experienced cycle 3. For example, if teams in the 2 × 2 treatment had faced the entry rates observed for the 1 × 1 treatments in experienced cycle 2, the estimated proportion of strategic play by teams in cycle 2 would have been 63.4 percent rather than the observed frequency of 88.4 percent. This adjusted figure is slightly lower than the TW norm of 70.6 percent, so that teams meet (within the 90-percent confidence interval), but do not exceed, the estimated value of the TW norm.[33] Thus, we cannot reject a null

[33] As an alternative, we have adjusted the 1 × 1 proportions for the difference in entry rate differentials, used the adjusted 1 × 1 proportions to calculate an adjusted TW norm, and compared this adjusted TW norm with the 2 × 2 data. Not surprisingly, this alternative approach yields similar conclusions to adjusting the 2 × 2 data. Specifically, the adjusted TW norm moves up sufficiently to make it difficult to distinguish the adjusted TW norm from the 2 × 2 data in most cycles. More precise statements cannot be made since calculating error bars for this adjusted TW norm is problematic. Because of these difficulties, this alternative approach is inferior to adjusting the 2 × 2 data for entry rate differentials. Note, however, that the probit regressions use an ex-

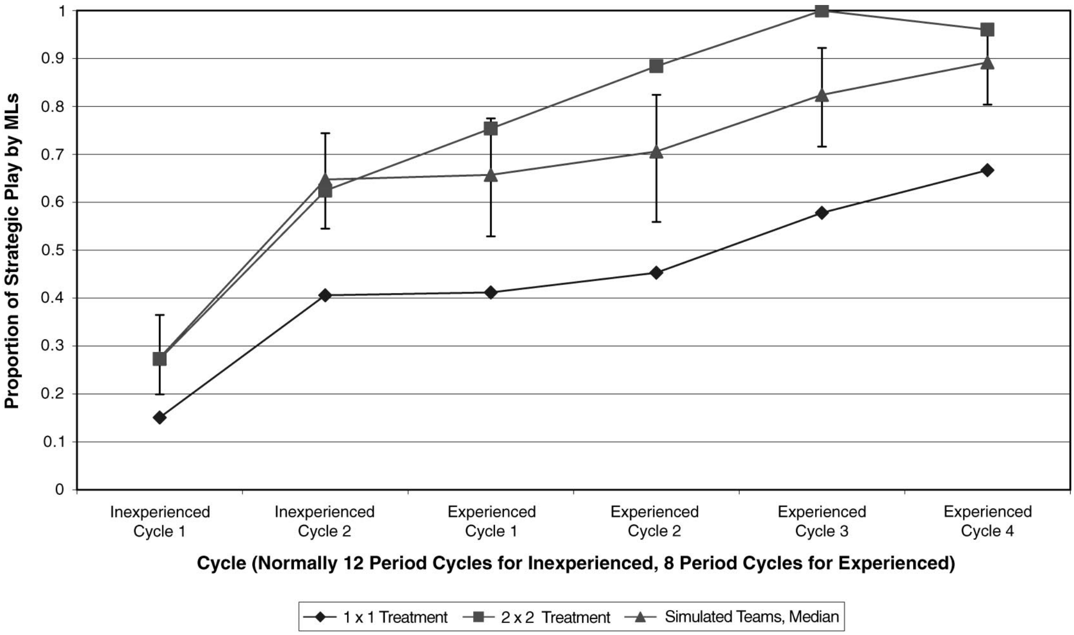


FIGURE 6. COMPARING THE DEVELOPMENT OF STRATEGIC PLAY FOR MLS IN 2 × 2 WITH 1 × 1 SESSIONS IN GAMES WITH LOW-COST ENTRANTS

Notes: Vertical axis shows frequency of MLs choosing outputs 5–7. Horizontal axis shows cycle of play. Bars give the 90-percent confidence interval for the truth-wins standard.

hypothesis that the ability of teams in the low-cost entrant game to *beat* the TW standard is due to greater incentives to play strategically. It is clear, however, that teams *meet* the TW norm after controlling for any differences in MLs' incentives to play strategically between the two treatments.

> *Conclusion 2*: MLs have higher levels of strategic play in the 2 × 2 treatment, meeting or beating the TW norm in all cycles of play. Accounting for entry rate differences between treatments, MLs generally meet, but do not beat, the TW norm. Further, play is much closer to the efficient separating equilibrium in the last two cycles of experienced subject play in the 2 × 2 sessions than in the 1 × 1 case.

C. *Limit Pricing in the Crossover Treatment*

Figure 7 reports data for the 1 × 1 and 2 × 2 treatments (top and bottom panels respectively) by cycle following the crossover. The differences are quite striking on a number of dimensions.

(a) Right from the start, there are substantially higher levels of strategic play for MLs in the 2 × 2 sessions. Figure 8 directly compares MLs' level of strategic play between the two treatments, along with the TW norm. The frequency of strategic play in the 2 × 2 sessions is well above the 90-percent confidence interval for the TW norm for the first two cycles, and slightly exceeds its upper bound in the last cycle. Further, using

tremely conservative approach in controlling for clustering in the data. As described in Section A1 of the on-line Appendix, approaches that make fuller use of the information in the data consistently yield smaller estimates for the marginal effect of entry rate differentials on strategic play by MLs. As such, the adjusted proportions of strategic play for the 2 × 2 data are probably too low, understating how well teams are performing versus the TW norm after controlling for entry rate differentials.

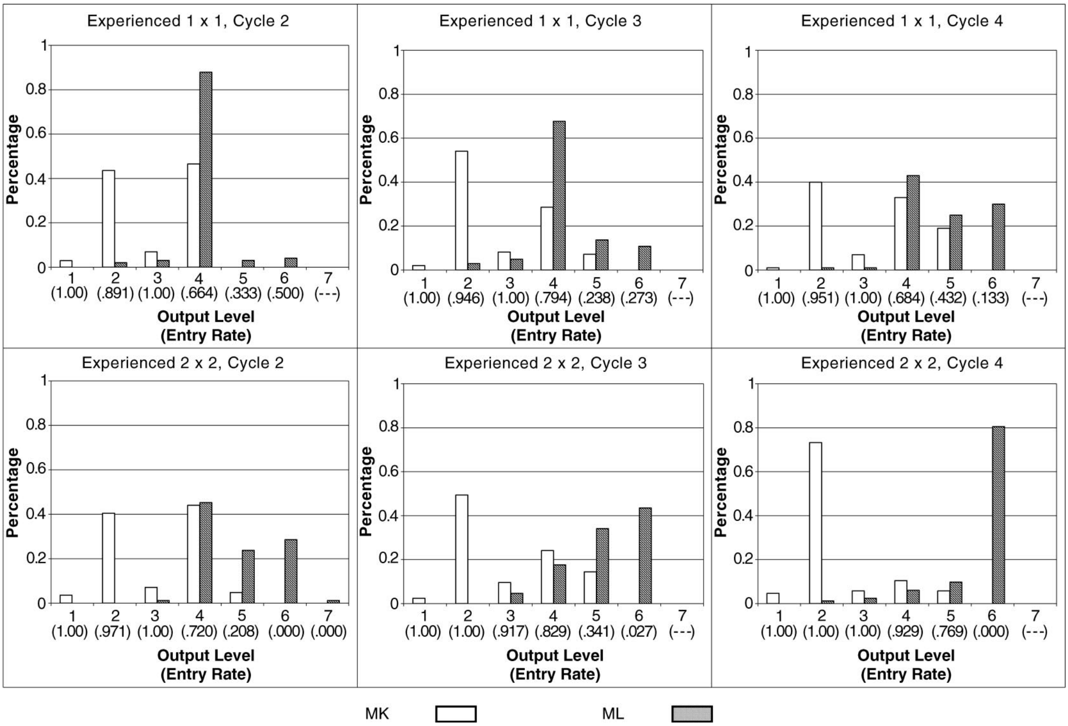


FIGURE 7. POOLED DATA FROM 1 × 1 AND 2 × 2 SESSIONS FOR THE CROSSOVER TREATMENT

Notes: Data are for play following the crossover where only pure strategy-separating equilibria exist. Bars indicate choice frequencies for MHs and MLs for each possible output level, with entry rates for those output levels shown in parentheses. Cycles consist of 8 plays of the game.

the probit regressions to control for entry rate differentials, strategic play for MLs in teams remains significantly above the 90-percent confidence interval of the TW norm in the first two cycles of play. Thus, not only do teams *beat* the TW norm in the crossover treatment, this *cannot* be attributed solely to entry rate differentials between the two treatments.

(b) In Figure 7, it's clear that in the last cycle following the crossover play has fully converged to the efficient separating equilibrium in the 2 × 2 treatment, but is still very much in transition in the 1 × 1 treatment. In the 1 × 1 treatment, MLs' strategic choices are fairly evenly split between 5 and 6 in the final cycle, with 4 still attracting a large number of choices for both MLs and MHs. In contrast, 6 accounts for 89.2 percent of MLs strategic choices for teams and the vast majority of MH choices are at 2.

(c) Comparing the levels of strategic play in Figure 7 with Figures 4 and 5 gets at the issue of cross-game learning. Subjects in Figures 4 and 5 are playing the same game as those in Figure 7, but have no prior experience with the limit pricing game. Subjects in Figure 7 have prior experience with the limit pricing game, but in games with high-cost Es where play converges on the pooling equilibrium at 4. Does prior experience in the game with high-cost Es help with the development of strategic play for MLs in the game with low-cost Es? The answer, shown in Figure 9, is clearly yes for the 2 × 2 treatment, but no for the 1 × 1 treatment. In fact, as probits reported in the Appendix verify, there is significantly *less* strategic play for MLs in the first two cycles

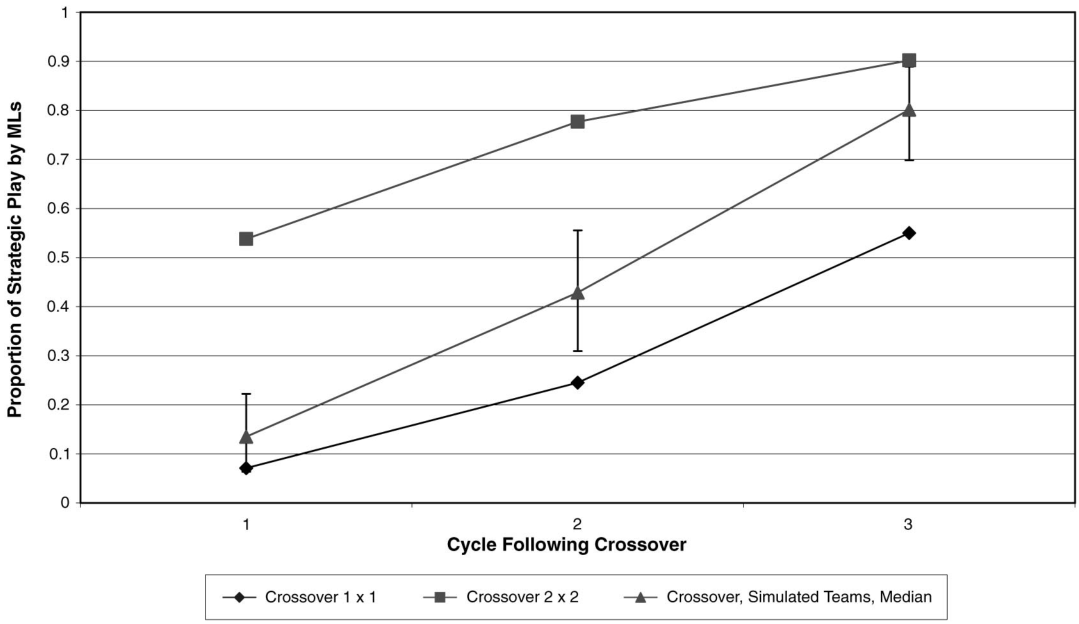


FIGURE 8. COMPARING THE DEVELOPMENT OF STRATEGIC PLAY FOR MLs IN 2 × 2 WITH 1 × 1 SESSIONS *FOLLOWING* THE CROSSOVER TO GAMES WITH LOW-COST ENTRANTS

Notes: Vertical axis shows frequency of MLs choosing outputs 5–7. Horizontal axis shows cycle of play. Bars give the 90-percent confidence interval for the truth-wins standard.

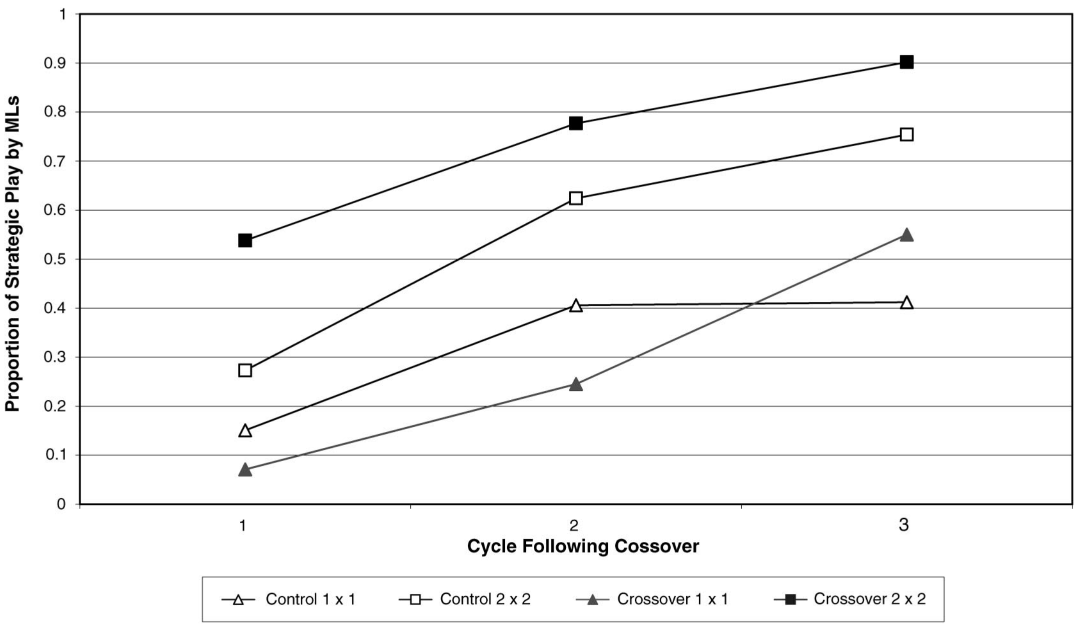


FIGURE 9. CROSS-GAME LEARNING: COMPARING THE DEVELOPMENT OF STRATEGIC PLAY FOR MLs IN THE CROSSOVER TREATMENT (DATA FROM FIGURE 7) WITH PLAY IN GAMES WITH ONLY LOW-COST ENTRANTS (DATA FROM FIGURES 4 AND 5)

Notes: Vertical axis shows frequency of MLs choosing outputs 5–7. Horizontal axis shows cycle of play.

of the 1 × 1 treatment following the crossover than in the first two cycles of inexperienced subject play, both with and without including entry rate measures in the regressions.[34] In contrast, there is significantly *more* strategic play in the first cycle of play following the crossover than in the first cycle of inexperienced subject play in the 2 × 2 sessions. These results are summarized in the following two conclusions.

Conclusion 3: The crossover treatment shows substantially higher levels of strategic play for MLs in the 2 × 2 treatment than in the 1 × 1 case for all cycles following the crossover, beating the TW norm throughout. This conclusion is robust to accounting for different incentives to play strategically between the two treatments in the first two cycles of play. Further, a clean separating equilibrium has emerged in the 2 × 2 treatment by the last cycle of play, while the 1 × 1 treatment is still very much in a state of flux.

Conclusion 4: There is *negative* learning transfer in the 1 × 1 treatment, as prior experience in the game with high-cost Es provided subjects with a slight, but statistically significant, *disadvantage* following the introduction of low-cost Es relative to subjects with no prior experience. However there is *positive* learning transfer under the same conditions in the 2 × 2 treatment, as prior experience in games with high-cost Es *facilitates* the development of strategic play relative to subjects with no prior experience in games with low-cost Es.

There is one final point worth making before analyzing the team dialogues. It is based on the notion that the speed with which strategic play develops in the 1 × 1 treatments is a measure of the difficulty subjects have with learning to play the game. Assuming this is the case, we can rank the games in terms of their degree of difficulty, with the games with high-cost Es and the pooling equilibrium being the easiest, and the crossover treatment being the most difficult. This leads to the following conjecture:

Conjecture: The more difficult it is to learn to play strategically, the greater the advantage of two-person teams over individuals.

The possible reasons for such an inverse relationship are discussed in the concluding section of the paper.

V. Insights into the Learning Process: Analysis of the Team Dialogues

Our original motivation for conducting the teams treatment was to obtain direct insight into the learning process underlying the development of strategic play through analyzing the team dialogues. These dialogues are a natural part of the experimental task and are clearly relevant to the task at hand, thereby providing an unbiased, albeit noisy, window into the underlying learning process.[35]

The primary goal of this section is to provide a brief overview of the major contents of the team dialogues. In doing so we provide further evidence that Ms' initial choices (2 for MHs and 4 for MLs) reflect a basic failure to think strategically, and that the development of strategic play is associated with Ms thinking from the point of view of Es, thereby allowing Ms to anticipate Es' responses to their choices.[36] The

[34] The negative learning transfer reported here for the 1 × 1 treatment differs from the positive learning transfer found in the 1 × 1 treatment reported in Cooper and Kagel (2004). Beyond the use of different subject populations, there are a number of methodological differences between the present experiment and the one reported in Cooper and Kagel (2004). Preliminary results in a follow-up study indicate that the most prominent of these differences has to do with the use of an abstract context here versus the use of meaningful context in the earlier experiment (see Cooper and Kagel, 2003). We plan a full report on the effects of context on learning transfer in games in another paper.

[35] Post-experiment surveys gather data retrospectively, thereby relying on subjects' possibly shaky and/or biased memories of what they were thinking in earlier stages of the experiment. "Talk out loud" techniques allow for the gathering of real-time data, but the dialogues generated are not integral to the task at hand.

[36] Given the differences in team versus individual play documented here, these insights do not necessarily extend to individual play. The latter will require obtaining monologues directly from individual play, a project that plays a prominent role in our future research agenda. Our educated guess is that the development of strategic play by individuals also relies on learning to think from the E's point of view.

TABLE 3—CATEGORIES FOR CODING TEAM DIALOGUES

Category[a]	Description[b]
1	Myopic choice as an M
3i	Pooling as MH type: pure imitation (no rationale given for why others should be imitated)
3ii	Pooling as an MH type: idea is to "fool" E team
3iii	Pooling as an MH type: discussion from point of view of a E team
3iv	Pooling as an MH type: drawing on own actions as a E team
4i	Separating as an ML type: pure imitation (no rationale given for why others should be imitated)
4ii	Separating as an ML type: referring to negative numbers as making it obvious that a 6 or 7 couldn't be an MH
4iii	Separating as an ML type: discussion from point of view of Es
4iv	Separating as an ML type: drawing on own actions as an E
5	Recognizing separating as an E
5i	E team referring to negative numbers as making it obvious that a 6 or 7 couldn't be an MH
6	Following crossover: recognizing that the change in payoff tables will change Es' choices
7	MH types choose 2: recognizing that the high frequency of IN at 3 & 4 make 2 the best choice (not myopic)
8	Reinforcement: explicit reference to making choice due to the past success/failure of past actions
9	Other regarding behavior: must be intent of decision, not an amusing side effect
10	Using the feedback provided about the decisions of others
11	Contamination: talked with someone prior to the experiment about what choices to make
12	Expression of out of equilibrium beliefs
12i	Expression of out of equilibrium beliefs: Es will choose IN for anything other than 4
12ii	Expression of out of equilibrium beliefs: Es will choose OUT for anything other than 2
13	Correcting mistakes of partner (e.g., misreading the payoff table, misreading one's own type, etc.)
14	Convincing a more knowledgeable/sophisticated teammate to deviate from a more profitable strategy
15	Level 1 reasoning by Es: MHs will choose 2, MLs will choose 4
16	Understanding pooling as an E in games with high-cost Es

[a] Category 2 was "advocating myopic choice as a E player." We dropped this category midway through the coding because it was too ambiguous.

[b] For categories where the description starts with a particular action, we are looking for the justification given for taking this action.

dialogues also provide clear evidence that the positive cross-game learning reported *within* the teams' treatment is associated with increased numbers of subjects who anticipate that the change in Es' payoffs will promote increased entry, and who understand the appropriate response to the increased entry. This implies that one of the key things teams have learned from prior experience with the game with high-cost Es is to think strategically, and this thinking readily generalizes to games with low-cost Es.

To analyze the team dialogues, we developed a coding system for types of statements as follows: first, the authors separately read through a common sample of the dialogues, establishing a set of preliminary codings, which were then reconciled, establishing the single list shown in Table 3. This coding scheme is designed to capture all statements that are relevant to subjects' learning to play strategically. Two undergraduate research assistants were then trained (separately) to do the coding. Although there is a fair amount of variance between the two coders (the average cross-coder correlation for the 24 categories is 0.388), much of the variance comes from categories that were not coded very frequently. (The average cross-coder correlation across the five most frequent categories is 0.570, with a minimum of 0.517.)[37] Results reported are based on the average of the two

[37] As an alternative measure of agreement between coders, we calculated the proportion of times that coders agreed on their categorization of a dialogue relative to the total number of observations where at least one coder categorized the dialogue. Averaging over the five most frequent categories, the proportion of agreement is 46.4 percent. Two sources of noise hold down agreements between coders. First, the two coders systematically disagreed on the exact codings for several categories. Second, given the unstructured nature of the dialogues, discussions frequently extend over multiple rounds of the experiment. The agreement between coders is therefore reduced by cases in which the coders agree on what should be coded but disagree on exactly which round it should be coded for.

TABLE 4—SUBJECTS PLAYING STRATEGICALLY FOR THE FIRST TIME (MEANS OF INDEPENDENT VARIABLES)

High-cost entrant game, inexperienced				
Played strategically next opportunity?	All	No	Yes	Fisher exact
M's choice	3.687	3.692	3.698	0.187
E's choice (Entry = 1)	0.180	0.231	0.116	0.370
Cycle 2	0.246	0.077	0.233	0.426
Ever coded category 1	0.418	0.423	0.407	0.829
Ever coded category 3ii	0.123	0.115	0.140	0.291
Ever coded category 3iii	0.369	0.000	0.500	0.000***
Ever coded category 10	0.557	0.346	0.628	0.113
Ever coded category 13	0.107	0.038	0.128	0.258
Ever coded category 15	0.295	0.077	0.326	0.126
# Observations	61	13	43	

Low-cost entrant game, inexperienced				
Played strategically next opportunity?	All	No	Yes	Fisher exact
M's choice	5.571	5.273	5.704	0.028**
E's choice (Entry = 1)	0.286	0.636	0.148	0.005***
Cycle 2	0.429	0.364	0.407	1.000
Ever coded category 1	0.298	0.227	0.352	0.499
Ever coded category 3iii	0.286	0.273	0.315	0.446
Ever coded category 4ii	0.143	0.000	0.167	0.276
Ever coded category 4iii	0.310	0.091	0.352	0.030**
Ever coded category 5 or 5i	0.202	0.182	0.222	1.000
Ever coded category 10	0.595	0.500	0.667	0.644
Ever coded category 15	0.107	0.182	0.093	0.481
# Observations	42	11	27	

* Statistically significant at the 10-percent level.
** Statistically significant at the 5-percent level.
*** Statistically significant at the 1-percent level.

independent codings, unless otherwise stated. In averaging across coders, we are implicitly assuming that errors are independent across coders so that averaging reduces the total error.

A. *What Teams Are Talking about Prior to First Strategic Play*

Table 4 reports coding frequencies for those categories that were coded 10 percent of the time or more for inexperienced teams, up to and including the first time they play strategically.[38] The columns labeled "No" and "Yes" in Table 4 distinguish between teams who played strategically at the next opportunity following their first strategic play (the Yes column) and those who failed to do so (the No column). The latter is not uncommon for inexperienced subjects, occurring for 22.8 percent of the teams that play strategically in games with high-cost Es and 28.9 percent of the time in games with low-cost Es. The "All" column ignores this distinction, averaging across the Yes and No columns.[39] The column labeled "Fisher exact" reports the results of a Fisher exact test for whether there is a statistically significant relationship between continuing to play strategically and the row variable.

Our analysis begins with focusing on the All frequencies. Category 10 refers to teams using the feedback data provided after each game, which record the results of all pairings: M's type, M's output, and E's response. This is by far and away the most frequently coded category

[38] Experienced subject sessions are less informative, as many teams include individuals who have previously figured out how to play strategically, which can lead to very cryptic conversations. Table A 8 in http://www.e-aer.org/data/june05_app_kagel.zip provides a complete set of codings along with the between-coder correlations. We employ all comments up to and including the period a team first played strategically because teams were in constant communication but had the opportunity to play strategically only at randomly determined intervals. Thus, in many cases a team had figured out how to play strategically when they were an E or were playing in the other role as an M.

[39] The frequencies in the All column are not the weighted average of the frequencies in the Yes and No columns, as some teams never got the chance to play strategically again.

prior to strategic play in both the high- and low-cost entrant games.[40] The frequent use of feedback regarding others' choices suggests that pure reinforcement learning, in the spirit of Alvin E. Roth and Ido Erev (1995), is likely to be a poor model, by itself, of team learning.

Category 1 refers to myopic choice as an M player. This is the second most frequently coded category prior to strategic play in games with high-cost Es, and a close third in games with low-cost Es. It corresponds to MHs choosing 2 and MLs choosing 4 because it has the highest payoff *ignoring Es' potential response to these choices.* To quote from a ML team in a session with low-cost Es:

> "I think we should pick 4 because no matter what they pick, the highest payoff will be for either X [In] or Y [Out]." "Cool."

Teams, in their role as Es, make similar statements in looking forward to how they will play as Ms.[41] To quote from a low-cost E team:

> "When we're team A2 [ML] we should always choose #4 because that makes the most money no matter what the B [E] team chooses." "Good point." "And with A1 [MH] I think its option 2."

The high frequency of coding for category 1 supports our contention that early play of the myopic maxima (2 for MHs, 4 for MLs) is nonstrategic in nature.

The categories of greatest interest to us are those that relate to how subjects justify strategic play. Category 3 was designed to capture different types of reasoning underlying strategic play in games with high-cost Es, with category 4 the corresponding category for games with low-cost Es. There are four subcategories in each case. The most frequently coded subcategories are 3iii and 4iii, in both cases consisting of explicit reasoning from the point of view of Es' potential responses to their choices as Ms. For example, to quote from a team deciding how to play as an MH:

> "Let's try 4 again." "OK." "I think the problem with 2 is that Bs [Es] knew we were A1 [MH], so they chose X [IN] to maximize their payoff." "I think you're right."

And to quote from a team in deciding how to play as an ML:

> "If we enter 6 when we are A2 [ML] then everyone will know that we are an A2 [ML] and will guess accordingly, giving us a higher average. ... The 6 pays less but is very clear what we are."

These subcategories are also among the highest of the 24 categories coded, with 3iii the third most frequent coding for games with high-cost Es and 4iii the second most frequent for games with low-cost Es. The second most common subcategories—3ii and 4ii—are close variations on the same theme.[42] From a game theorist's point of view, this strategic empathy—reasoning from the other player's point of view—goes to the heart of thinking and behaving strategically. The codings provide clear evidence that this type of reasoning underlies the development of strategic play.

The Fisher exact tests in Table 4 indicate that being coded for subcategories 3iii and 4iii is a good predictor of whether or not, having played strategically once, teams will continue to play strategically. For games with high-cost Es, the only variable with any significant predictive power regarding continued strategic play is subcategory 3iii. This effect is quite strong; of the 27 teams coded for 3iii by either coder, *none* returned to nonstrategic play. Likewise, subcategory 4iii is a strong predictor for whether strategic play will continue for games with low-cost Es. Only one of the 15 teams coded for 4iii by

[40] A typical dialogue here for an MH team: "Look at the table. Every 4 was sent an X [IN]." "So what then." "Have no clue. You pick a number."

[41] These statements are also included in Category 1.

[42] Subcategory 3ii involves MHs choosing higher outputs in order to "fool" Es into thinking they were MLs. Subcategory 4ii involves explicitly recognizing that the negative payoffs for MHs' choice of 6 or 7 must make it obvious to Es that these choices are from MLs. In games with high-cost Es, subcategories 3ii and 3iii, while related, are not perfectly correlated. In games with low-cost Es, virtually all teams coded for 4ii were also coded for 4iii, but not vice versa. The relatively high frequency of 3iii in games with low-cost Es reflects play passing through a phase where MHs imitate MLs prior to the emergence of a separating equilibrium.

either coder returned to nonstrategic play.[43] Thus, thinking from the point of view of Es plays an important role in whether or not inexperienced Ms backslide into nonstrategic play.

The dialogues provide a few additional insights. First, most experimenters suspect that switching roles in games speeds up the learning process. While it probably isn't worth the time and money to verify directly this methodological conjecture, the relatively frequent codings of categories 5 and 5i (which code for recognizing separating choices by MLs while in the role of an E) in games with low-cost Es provide indirect evidence to this effect.[44] Second, there are lessons to be learned from the less frequently coded categories that did not make it into Table 4. Category 8 codes directly for pure reinforcement learning. It is coded relatively infrequently, particularly for games with low-cost Es (8.2 percent for games with high-cost Es; 3.6 percent for games with low-cost Es). Subcategories 3i or 4i were designed to capture Ms playing strategically purely on the basis of imitation. Neither is coded with any frequency (2.5 percent and 2.4 percent for 3i and 4i, respectively). The infrequent codings for reinforcement learning and pure imitation suggest that teams are not blindly choosing (avoiding) strategies that have done well (poorly) in the past. Rather, the substantially more frequent codings for subcategories 3iii and 4iii suggest that they are trying to figure out *why* these strategies have done well or poorly in the past. Finally, in discussing why teams frequently fail to clear the TW threshold, psychologists often refer to "process loss." One extreme version of process loss is "truth loses"—when a subject who has failed to figure out the strategic aspects of the game convinces a more insightful teammate to play nonstrategically. Category 14, designed to capture this, is not coded often—less than 1 percent in games with high-cost Es and 1.2 percent in games with low-cost Es—indicating that this kind of process loss rarely occurs.

Conclusion 6: Use of feedback about the decisions of others is the most frequently coded category. Categories designed to capture Ms' thinking from the point of view of Es' responses to Ms' choices are the most commonly coded justifications for strategic play, and the second and third most commonly coded categories overall for games with low- and high-cost Es, respectively. This is consistent with the type of strategic anticipation that lies at the heart of game theory.

Conclusion 7: Reverting to nonstrategic play after having played strategically occurs significantly less often for teams that have discussed their strategy as Ms from the point of view of Es. That is, teams recorded as thinking in game theoretic terms are more likely to continue to play strategically than those without such conversations.

B. *Positive Learning Transfer for Teams in the Crossover Treatment*

One of the most striking features of the teams' data is the positive learning transfer in the crossover treatment. The codings provide some insight into the mechanism underlying this. First, most teams almost immediately recognized that the change in payoffs would increase entry rates substantially, particularly for output 4. Category 6 codes for this—recognizing that the change in payoff tables will change Es' choice (following the crossover)—as is coded for 84.5 percent of all teams overall and coded (by at least one coder) for 76 percent of all teams in the *first play of the game* following the crossover.[45] Second, in the first cycle following the crossover subcategory 4iii was coded for 35.7 percent of the teams, compared to 17.3 percent in the first cycle of inexperienced play for games with low-cost Es. This represents a large growth in the frequency of strategic

[43] "The other variables that have significant predictive power are whether M chooses 5 or 6 when first playing strategically, and E's response to M's first strategic play. Teams reverting to nonstrategic play are more likely both to have chosen 5 (rather than 6) and to have been entered on compared with teams that continue to play strategically.

[44] However, subjects almost never explicitly draw on their own experiences as Es in deciding how to play as Ms: subcategories 3iv and 4iv are rarely coded (4.9 percent and 7.1 percent respectively) prior to first strategic play.

[45] For example, from one M team, "Now they are making this interesting. ... Now it'd actually be worth it to pick X [IN] every time."

empathy, as well as an ability to transfer that strategic thinking, in the 2×2 treatment.[46]

We can relate these observations, and the results of the cross-over treatments, to adaptive learning models offered in the literature. Standard adaptive learning models, for example stochastic fictitious play, predict *negative* learning transfer in the crossover treatments, as MLs with prior experience in games with high-cost Es have to unlearn their expectations of high payoffs and low entry rates for choosing 4, beliefs that don't burden inexperienced subjects. One way to overcome this and generate the positive cross-game learning observed in the data is to introduce sophisticated learners into the model who, in this case, (a) anticipate that changing the payoff table for Es will change Es choices; and (b) whose numbers increase as a result of prior experience with the game (see Cooper and Kagel, 2004).[47] From this perspective the negative transfer observed for 1×1 crossover sessions suggests that sophisticated learning develops at a much slower rate in the 1×1 sessions than in the 2×2 sessions. More fundamental differences in the learning processes may exist as well.

> *Conclusion 8*: The strong, positive learning transfer for teams is consistent with an adaptive learning model in which there are growing numbers of sophisticated learners who anticipate that the change in Es' payoffs will promote increased entry and that play of 6 will deter entry. The codings provide supporting evidence for this conjecture. The negative learning transfer found in the 1×1 treatment suggests that sophisticated learning develops at a much slower rate than in the 2×2 sessions, or some fundamental difference exists between the learning processes of teams and individuals.

[46] There is also substantially less backsliding in the crossover treatment than for inexperienced sessions in games with low-cost Es (11.1 percent versus 29.0 percent; $Z = 1.91$, $p < 0.06$, 2-tailed test). This difference is consistent with the importance of subcategory 4iii in preventing backsliding, and the substantially higher frequency with which 4iii is coded in the crossover treatment.

[47] For other examples of this sort of model, see Colin F. Camerer et al. (2002); Milgrom and Roberts (1991); and Dale O. Stahl II (1996).

VI. Conclusions and Discussion

This paper compares team versus individual play in a signaling game based on Milgrom and Roberts's (1982) entry limit pricing game. The focus of the paper is on differences in the learning/adjustment process between two-person teams versus individual subjects. Strategic play develops more rapidly in teams in all three treatments. Further, the superiority of team play increases the more difficult it is for subjects to learn to play strategically, so that in the most challenging games teams meet or surpass the "truth-wins" norm developed by psychologists for "eureka-type" learning problems. Surpassing the truth-wins norm is consistent with positive synergies between teammates and is rarely reported in the psychology literature for similar types of problems. In addition, teams exhibit strong *positive* cross-game learning, whereas individuals exhibit *negative* cross-game learning. The positive learning transfer in the 2×2 treatment also contrasts with results typically reported in the psychology literature, where zero or even negative learning transfer is usually reported (Solomon and Perkins, 1989). It is consistent with adaptive learning models with growing numbers of sophisticated learners who anticipate their opponents' responses to changes in their opponents' payoffs. This interpretation is supported by the team dialogues.

Our results raise two questions. Why do teams perform relatively better, compared to individuals, the more difficult the learning/adjustment process is? Why do teams meet, and even beat, the truth-wins norm in games with low-cost entrants and in the crossover treatment, given that psychology experiments usually report a failure to do so in similar problems? While greater incentives to play strategically can partially explain the strong performance of teams, even after controlling for this factor teams meet the truth-wins standard in games with low-cost entrants and beat it in the crossover treatment. Instead, the answer to both of these questions stems from several related sources.

One possibility is that although the insight underlying the truth-wins model is relevant to our results, the mathematical formula used to generate the truth-wins standard is not appropriate here. This formula assumes that problem solving (playing strategically) involves a *single* "aha" insight. However, the dynamic leading to the separating equilibrium in games with low-cost entrants consists of two distinct stages. Play first passes through a phase where MHs imitate MLs before MLs, under pressure from rising entry rates, begin to separate. Thus, this might be better likened to a multistep problem in which getting to the right answer involves a series of smaller insights rather than a single "aha" insight. In this case, it might well be that the truth-wins norm applies to each step of the learning process, with *both* team members being *equally* likely to provide the crucial insight for each step in the learning process. If this is the case, outcomes are likely to meet or beat the truth-wins norm.[48]

A second possibility is that playing as teams speeds up the development of strategic empathy, generating more "sophisticated" players who can think from the viewpoint of others. Increased numbers of sophisticated learners can account for teams meeting or beating the truth-wins norm in games with low-cost entrants, as well as the positive cross-game learning reported for teams. Comparing team dialogues following the crossover treatment with those of inexperienced teams provides direct evidence for large increases in the number of sophisticated learners as a result of experience in the teams treatment.[49]

The question that remains is why do these factors play a role in our experiment but not in the typical psychology experiment? The answer is that our experimental procedures, typical of those employed in economics, differ in two significant ways from those commonly employed in psychology experiments. First, we are looking at a game, so that strategic interactions between teams with antagonistic goals play a central role. Psychologists typically investigate individual decision problems (puzzles) where each team acts independently to solve a problem. If teams are better able to think from the point of view of others, this insight is relevant for games but not for individual decision problems. Second, our study consists of a number of replications of the same basic problem. In contrast, psychologists typically study one-shot learning problems. If teams differ not in making subjects more sophisticated initially, but instead in how fast subjects learn to become sophisticated, as our experimental results suggest, the ability of teams to meet or beat the truth-wins norm will be realized only once they have gained sufficient experience. Both of these factors can also help explain the positive cross-game learning found with teams here, compared to the absence of positive cross-game learning in the typical psychology experiment.

Given the strong convergence of teams to equilibrium outcomes, particularly in the more difficult games, one might be tempted to conclude that increased reliance on team play is all that is needed to tidy up some of the embarrassing discrepancies between experimental data and economic theory reported in the literature. The limited economics literature on team versus individual play, however, indicates otherwise (see the references cited in Section II). Further, within the structure of the signaling games reported here, we have investigated the ability of teams to overcome violations of equilibrium refinements reported in earlier experiments. In

[48] To be more formal, suppose that playing strategically involves two discrete insights. Assume that the likelihood of obtaining these insights is independent across tasks and across individuals. Let p_1 and p_2 be the probabilities of an individual obtaining each insight. The probability of an individual solving the problem is p_1p_2. The TW norm is therefore $p_1p_2(2 - p_1p_2)$. This, however, is not the probability that a team solves the problem in the absence of any synergies because it ignores the possibility that one team member obtains the first insight and the second team member obtains the second insight. The correct probability that a team solves the problem is $p_1p_2(4 - 2(p_1 + p_2) + p_1p_2)$. Doing some algebra, we can confirm that this probability is always greater than the TW norm. Of course, in such a setting it's possible that teams perform no better, or even worse, than individuals in cases where team members who do not grasp the first insight are not able to work on the problem on any deeper level, and hence are of no use (or even a hindrance) to obtaining further insights.

[49] Yet a third possibility is that there are significant strategic interactions within teams, which spill over into enhanced strategic interactions between teams. The team dialogues provide very little evidence for this, however, as team members seem quite cooperative in their discussions. About the only clear strategic interplay we observe within teams is that at times one team member will give into the partner to play a strategy that he clearly thinks is wrong, in order to convince the partner that it is wrong (as opposed to continued discussions).

particular, we have conducted treatments similar to the one reported in Cooper et al. (1997a) in which play by individuals violates one of the weakest equilibrium selection criteria—single round elimination of dominated strategies. Using experienced subjects, in two out of two sessions team play converges rapidly to an inefficient separating equilibrium that violates this selection criteria.[50] In short, the forward induction arguments underlying even very weak equilibrium selection criteria may be far too subtle for teams (as well as individuals) to adhere to.

Although the present paper answers a number of questions regarding team versus individual play in signaling games, many questions remain to be answered. Primary among these are (a) exploring why we observe negative learning transfer in the 1 × 1 game here unlike the positive learning transfer in the 1 × 1 game reported in Cooper and Kagel (2004); and (b) developing methods to compare directly the thought processes underlying individual subject play with those for teams. Work in progress on the first question suggests that the context used to frame the games plays a critical role in fostering positive transfer (see Cooper and Kagel, 2003). We are only now beginning work on the second question. From a broader perspective, the team procedures employed here, and the instant messaging possibilities of modern laboratory software, open up a number of exciting possibilities for gaining direct insight into behavior in a wide variety of settings.

Appendix

This Appendix contains the details of the regressions reported in the text. Additional analysis examining the robustness of the regression results to alternative specifications and additional details about the codings of the dialogues are posted at http://www.e-aer.org/data/june05_app_kagel.zip.

All of the regressions are probits where the dependent variable is whether Ms have played strategically or not. In games with high-cost (low-cost) Es, the dataset includes only choices of MHs (MLs). For MHs (MLs), output levels 3–5 (5–7) are coded as strategic play.

Because of repeated play by the same subjects, the probits must account for the presence of individual effects. This task is complicated by the structure of the 2 × 2 data, since in the experienced sessions, subjects were matched with a new partner (as was the case when the software malfunctioned and had to be restarted). This requires accounting not just for potential correlation between observations from the same team, but also for potential correlation between observations from different teams that shared a common member. In dealing with this problem we employ the relatively conservative approach suggested by Brent R. Moulton (1986) and Kung-Yee Liang and Scott L. Zeger (1986) to correct the standard errors for clustering.[51] A cluster in defined in terms of "chunks": any two observations that share a common team member must be included in the same chunk.[52] For example, suppose that subjects A and B were a team in an inexperienced subject session. In the following experienced subject session, A is teamed with C, and B is teamed with D. Any observations that include subjects A, B, C, or D are included in a single chunk, even though observations involving C and D as inexperienced subjects have only a tenuous connection. By taking this conservative approach we bias the results against finding statistical significance.

Games with High-Cost Es: Results for games with high-cost Es are reported in Table A1. Explanatory variables include dummies for the cycle,[53] interactions between dummies for the cycle and a dummy for the 2 × 2 treatment, and a measure of Es' choices. The first inexperienced cycle of the 1 × 1 treatment serves as a

[50] These results may be found at http://www.econ.sbs.ohio-state/kagel/violations.intuitive.pdf.

[51] This is the correction used by Stata for clustering. The qualitative results reported are not sensitive to how the individual and team effects are handled. See http://www.e-aer.org/data/june05_app_kagel.zip.

[52] For the 1 × 1 data, all observations from a single subject constitute a chunk.

[53] In regressions not reported here, we added controls for the varying length of a cycle. These additions are statistically significant but do not affect the qualitative results we report.

TABLE A1—PROBIT REGRESSIONS, HIGH-COST ENTRANT SESSIONS
(*Standard errors corrected for clustering at the "chunk" level*)
Dependent variable: Strategic choice by MHs (1003 obs, 174 teams)

Variable	Model 1	Model 2	Model 3
Constant	−0.077	−0.241	−0.154
	(0.129)	(0.240)	(0.263)
Inexperienced cycle 2	0.568***	0.525***	0.548***
	(0.133)	(0.140)	(0.140)
Experienced	0.815***	0.748***	0.784***
	(0.189)	(0.204)	(0.208)
2 × 2 * Inexperienced cycle 1	0.316*	0.299	0.000
	(0.187)	(0.187)	(0.529)
2 × 2 * Inexperienced cycle 2	0.240	0.197	−0.206
	(0.211)	(0.223)	(0.722)
2 × 2 * Experienced	0.170	0.164	−0.232
	(0.274)	(0.269)	(0.809)
Entry rate differential		0.359	
		(0.425)	
1 × 1 * Entry rate differential			0.168
			(0.472)
2 × 2 * Entry rate differential			0.777
			(0.852)
Log likelihood	−609.56	−608.64	−608.07

* Statistically significant at the 10-percent level.
** Statistically significant at the 5-percent level.
*** Statistically significant at the 1-percent level.

base. The 2 × 2-cycle interaction terms capture differences between behavior by teams in that cycle and behavior by individuals from 1 × 1 sessions in the same cycle. The measure of Es' behavior consists of the difference in the current cycle between the entry rates for output levels 2 and 4, which serves as a proxy for the incentives for strategic play. This is calculated across all observations from the same session, since subjects got to see all entry decisions in their session.

Model 1 tests for 2 × 2 effects without controlling for entry rates. A marginally significant 2 × 2 effect is found in the first inexperienced cycle, with no significant effects thereafter. Model 2 adds the control for Es' behavior. Ms are only weakly responsive to the entry rate differential between 2 and 4, as the parameter value is far from statistically significant at conventional levels. The impact on the estimated 2 × 2 effects is minimal. Model 3 checks whether the impact of entry rate differentials differs between the 2 × 2 and 1 × 1 treatments. While the difference is impressive, the parameters are estimated imprecisely so that we cannot reject a null hypothesis of no difference in responsiveness between treatments.

Games with Low-Cost Es: The probits for this treatment are reported in Table A2. The regression specification is similar to Table A1 except that the entry rate differential used is between output levels 4 and 6, reflecting the incentives to play strategically for MLs. Note that the regressions include a dummy for the third cycle of inexperienced play, but no interaction term between this dummy and the 2 × 2 dummy, since none of the inexperienced 2 × 2 sessions included a third cycle.

Model 1 tests for differences in the frequency of strategic play without any controls for entry rates. The results strongly support Hypothesis 1, as the 2 × 2 interactions are strongly significant in all cycles.[54] Model 2 adds the entry rate

[54] The interaction term for the third cycle of the experienced sessions had to be dropped because there was no variation in the data—all MLs in the relevant cell played strategically. This can be interpreted as yielding an arbitrarily large parameter estimate that would be statistically significant at any desired level.

TABLE A2—PROBIT REGRESSIONS, LOW-COST ENTRANT SESSIONS
(*Standard errors corrected for clustering at the "chunk" level*)
Dependent variable: Strategic choice by MLs (1375 obs, 176 teams)

Variable	Model 1	Model 2	Model 3
Constant	−1.032***	−0.451***	−0.385**
	(0.150)	(0.169)	(0.180)
Inexperienced cycle 2	0.795***	0.270	0.212
	(0.157)	(0.177)	(0.188)
Inexperienced cycle 3	1.006***	0.149	0.040
	(0.244)	(0.311)	(0.348)
Experienced cycle 1	0.809***	0.190	0.119
	(0.221)	(0.238)	(0.251)
Experienced cycle 2	0.915***	0.605**	0.585**
	(0.235)	(0.236)	(0.238)
Experienced cycle 3	1.230***	0.638**	0.576**
	(0.227)	(0.248)	(0.258)
Experienced cycle 4	1.463***	0.628**	0.526*
	(0.236)	(0.275)	(0.296)
2×2 * Inexperienced cycle 1	0.424**	−0.046	−0.149
	(0.189)	(0.195)	(0.207)
2×2 * Inexperienced cycle 2	0.553***	0.463**	0.469**
	(0.197)	(0.205)	(0.200)
2×2 * Experienced cycle 1	0.909***	0.709***	0.810***
	(0.260)	(0.225)	(0.232)
2×2 * Experienced cycle 2	1.312***	0.480	0.669**
	(0.336)	(0.309)	(0.330)
2×2 * Experienced cycle 3	Dropped	Dropped	Dropped
	(No variation)	(No variation)	(No variation)
2×2 * Experienced cycle 4	1.320***	0.934***	1.228***
	(0.307)	(0.399)	(0.393)
Entry rate differential		1.791***	
		(0.443)	
1×1 * Entry rate differential			2.076***
			(0.622)
2×2 * Entry rate differential			1.036***
			(0.276)
Log likelihood	−751.45	−703.53	−700.84

* Statistically significant at the 10-percent level.
** Statistically significant at the 5-percent level.
*** Statistically significant at the 1-percent level.

differential as a control for incentives to play strategically as an ML. This is statistically significant at the 1-percent level. With the inclusion of the entry rate differential, the 2×2 interaction terms are weakened across the board with several of them failing to achieve statistical significance. There is still a clear 2×2 effect, but at least part of this effect must be attributed to changes in the incentives to play strategically as an ML. Model 3 permits differing sensitivity to entry-rate differentials between the 2×2 and 1×1 treatments. The results indicate that subjects are roughly twice as sensitive to the entry-rate differential in the 1×1 treatment, but the standard errors of the estimates are sufficiently large that a null hypothesis of no difference cannot be rejected at the 10-percent level. Thus, we regard the estimates as suggestive. With this in mind, the strength of the 2×2 effect is greater overall in Model 3 than in Model 2.

Crossover Sessions: The first half of Table A3 reports probits comparing the 2×2 treatment with the 1×1 treatment for the crossover sessions. The dataset includes all plays by MLs following the crossover. The regression specification is the same as for games with low-cost Es. Since teams are never rematched here, however, clustering at the chunk level is equivalent to clustering at the team level.

TABLE A3—PROBIT REGRESSIONS, CROSSOVER SESSIONS, LOW-COST ENTRANTS
(*Standard errors corrected for clustering at the "chunk" level*)
Dependent variable: Strategic choice by MLs

Test of truth-wins norm

	All crossover data (552 obs, 92 teams)	
Variable	Model 1	Model 2
Constant	−1.471***	−1.068***
	(0.218)	(0.287)
Crossover cycle 2	0.781***	0.413
	(0.240)	(0.298)
Crossover cycle 3	1.596***	1.244***
	(0.238)	(0.265)
2 × 2 * Crossover cycle 1	1.560***	1.023***
	(0.281)	(0.375)
2 × 2 * Crossover cycle 2	1.450***	1.162***
	(0.299)	(0.343)
2 × 2 * Crossover cycle 3	1.170***	0.738*
	(0.308)	(0.404)
Entry rate differential		1.139*
		(0.623)
Log likelihood	−280.30	−274.94

Tests of cross-game learning

	1 × 1 treatment (1119 obs, 116 teams)		2 × 2 treatment (739 obs, 152 teams)	
Variable	Model 1	Model 2	Model 1	Model 2
Constant	−1.032***	−0.626***	−0.603***	−0.327***
	(0.150)	(0.164)	(0.120)	(0.123)
Cycle 2[a]	0.795***	0.281	0.919***	0.827***
	(0.157)	(0.178)	(0.139)	(0.129)
Cycle 3[b]	0.809***	0.163	1.289***	1.079***
	(0.220)	(0.246)	(0.151)	(0.135)
Control, inexperienced cycle 3	1.006***	0.107		
	(0.244)	(0.319)		
Control, experienced cycle 2	0.915***	0.597**	1.798***	1.409***
	(0.234)	(0.237)	(0.176)	(0.232)
Control, experienced cycle 3	1.230***	0.614**	Dropped	
	(0.227)	(0.253)	(No Variation)	
Control, experienced cycle 4	1.463***	0.589**	2.353***	1.863***
	(0.236)	(0.285)	(0.190)	(0.284)
Crossover, cycle 1	−0.439*	−0.648*	0.692***	0.415*
	(0.264)	(0.370)	(0.215)	(0.221)
Crossover, cycle 2	−0.453*	−0.627**	0.444*	0.153
	(0.258)	(0.255)	(0.250)	(0.289)
Crossover, cycle 3	0.349	0.322	0.610**	0.323
	(0.274)	(0.265)	(0.296)	(0.305)
Entry rate differential		1.901***		1.004***
		(0.490)		(0.339)
Log likelihood	−655.43	−601.54	−376.00	−371.91

* Statistically significant at the 10-percent level.
** Statistically significant at the 5-percent level.
*** Statistically significant at the 1-percent level.
[a] Equals 1 for observations in the second inexperienced cycle of the control sessions and the second cycle following the crossover in the crossover sessions.
[b] Equals 1 for observations in the first experienced cycle of the control sessions and the third cycle following the crossover in the crossover sessions.

Model 1 confirms the obvious—MLs play strategically significantly more often following the crossover in 2 × 2 sessions than in 1 × 1 sessions. Model 2 indicates that this difference is robust to controls for differing entry rates.

The second half of Table A3 contains probits for the cross-game learning effects. This includes sessions where subjects have experience only with the low-cost E game, which serve as controls here, as well as with the crossover sessions following the crossover. The variables of interest here are the dummies for the three cycles following the crossover. The structure of the dummies is such that these three parameter estimates capture, respectively, the differences between the first, second, and third cycle following the crossover and the first and second cycle of inexperienced play and first cycle of experienced play in the controls.

For the 1 × 1 data, Model 1 shows a small but statistically significant *negative* crossover effect. Model 2 indicates that this negative effect is robust to controls for the entry-rate differential. Thus, previous experience games with high-cost Es inhibit strategic play in games with low-cost Es.

For the 2 × 2 data, Model 1 shows a statistically significant *positive* crossover effect. As shown in Model 2, the size of this effect is substantially reduced by adding controls for entry-rate differentials, retaining statistical significance only for the first cycle following the crossover. It is difficult to sort out causality here—previous experience with the high-cost entrant game leads both to immediately higher levels of strategic play and to immediately higher entry rate differentials. Based on the dialogues, we are disinclined to attribute the higher rates of strategic play solely to differences in feedback.

Comparing the cross-game learning probits, it is worth noting that once again the responsiveness to entry rates is substantially larger in the 1 × 1 treatment. Even though this result never quite achieves statistical significance, its pervasiveness in games with low-cost Es suggests that teams rely less on the feedback than do individuals. This implies differences in the basic process used by teams and individuals to reason about the limit pricing game.

REFERENCES

Alchian, Armen A. and Demsetz, Harold. "Production, Information Costs, and Economic Organization." *American Economic Review*, 1972, *62*(5), pp. 777–95.

Battalio, Raymond; Samuelson, Larry and Van Huyck, John. "Optimization Incentives and Coordination Failure in Laboratory Stag Hunt Games." *Econometrica*, 2001, *69*(3), pp. 749–64.

Bornstein, Gary and Yaniv, Ilan. "Individual and Group Behavior in the Ultimatum Game: Are Groups More 'Rational' Players?" *Experimental Economics*, 1998, *1*(1), pp. 101–08.

Brandts, Jordi and Holt, Charles A. "An Experimental Test of Equilibrium Dominance in Signaling Games." *American Economic Review*, 1992, *82*(5), pp. 1350–65.

Camerer, Colin F.; Ho, Teck-Hua and Chong, Juin-Kuan. "Sophisticated Experience-Weighted Attraction Learning and Strategic Teaching in Repeated Games." *Journal of Economic Theory*, 2002, *104*(1), pp. 137–88.

Cason, Timothy N. and Mui, Vai-Lam. "A Laboratory Study of Group Polarization in the Team Dictator Game." *Economic Journal*, 1997, *107*(444), pp. 1465–83.

Cho, In-Koo and Kreps, David M. "Signaling Games and Stable Equilibria." *Quarterly Journal of Economics*, 1987, *102*(2), pp. 179–221.

Cooper, David J.; Garvin, Susan and Kagel, John H. "Adaptive Learning vs. Equilibrium Refinements in an Entry Limit Pricing Game." *Economic Journal*, 1997a, *107*(442), pp. 553–75.

Cooper, David J.; Garvin, Susan and Kagel, John H. "Signaling and Adaptive Learning in an Entry Limit Pricing Game." *RAND Journal of Economics*, 1997b, *28*(4), pp. 662–83.

Cooper, David J. and Kagel, John H. "Learning and Transfer in Signaling Games." Case Western Reserve University, 2004.

Cooper, David J. and Kagel, John H. "Lessons Learned: Generalizing Learning across Games." *American Economic Review*, 2003, *93*(2), pp. 202–07.

Cooper, David J.; Kagel, John H.; Lo, Wei and Gu, Qing Liang. "Gaming against Managers in Incentive Systems: Experimental Results with Chinese Students and Chinese Managers." *American Economic Review*, 1999, *89*(4), pp. 781–804.

Cox, James C. "Trust, Reciprocity, and Other-Regarding Preferences: Groups vs. Individuals and Males vs. Females," in Rami Zwick and Amnon Rapoport, eds., *Experimental business research*. Dordrecht: Kluwer Academic, 2002, pp. 331–50.

Cox, James C. and Hayne, Stephen C. "Barking up the Wrong Tree: Are Small Groups Rational Agents?" Unpublished Paper, 2002.

Davis, James H. "Some Compelling Intuitions about Group Consensus Decisions, Theoretical and Empirical Research, and Interpersonal Aggregation Phenomena: Selected Examples, 1950–1990." *Organizational Behavior and Human Decision Processes*, 1992, *52*(1), pp. 3–38.

Hill, Gayle W. "Group versus Individual Performance: Are $N + 1$ Heads Better Than One?" *Psychological Bulletin*, 1982, *91*(3), pp. 517–39.

Kagel, John H. and Levin, Dan. *Common value auctions and the winner's curse*. Princeton: Princeton University Press, 2002.

Kerr, Norbert L.; MacCoun, Robert J. and Kramer, Geoffrey P. "Bias in Judgment: Comparing Individuals and Groups." *Psychological Review*, 1996, *103*(4), pp. 687–719.

Kocher, Martin G. and Sutter, Matthias. "The Decision Maker Matters: Individual versus Group Behaviour in Experimental 'Beauty-Contest' Games." *Economic Journal*, 2005, *115*(500), pp. 200–23.

Liang, Kung-Yee and Zeger, Scott L. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika*, 1986, *73*(1), pp. 13–22.

Lorge, Irving and Solomon, Herbert. "Two Models of Group Behavior in the Solution of Eureka-Type Problems." *Psychometrika*, 1955, *20*(2), pp. 139–48.

Milgrom, Paul and Roberts, John. "Limit Pricing and Entry under Incomplete Information: An Equilibrium Analysis." *Econometrica*, 1982, *50*(2), pp. 443–59.

Milgrom, Paul and Roberts, John. "Adaptive and Sophisticated Learning in Normal Form Games." *Games and Economic Behavior*, 1991, *3*(1), pp. 82–100.

Moulton, Brent R. "Random Group Effects and the Precision of Regression Estimates." *Journal of Econometrics*, 1986, *32*(3), pp. 385–97.

Roth, Alvin E. and Erev, Ido. "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term." *Games and Economic Behavior*, 1995, *8*(1), pp. 164–212.

Shaw, Marjorie. "A Comparison of Individuals and Small Groups in the Rational Solution of Complex Problems." *American Journal of Psychology*, 1932, *44*(3), pp. 491–504.

Solomon, Gavriel and Perkins, David N. "Rocky Roads to Transfer: Rethinking Mechanisms of a Neglected Phenomenon." *Educational Psychologist*, 1989, *24*(2), pp. 113–42.

Stahl, Dale O. "Boundedly Rational Rule Learning in a Guessing Game." *Games and Economic Behavior*, 1996, *16*(2), pp. 303–30.

Sutter, Matthias. "Are Four Heads Better Than Two? An Experimental Beauty-Contest Game with Teams of Different Sizes." Max Planck Institute for Research into Economic Systems, Discussion Papers on Strategic Interaction: Paper No. 15-2004, 2004.

Economics Working Paper Series

2016/018

Individual vs. Group Decision Making: an Experiment on Dynamic Choice under Risk and Ambiguity

Enrica Carbone, Konstantinos Georgalos and Gerardo Infante

The Department of Economics
Lancaster University Management School
Lancaster LA1 4YX
UK

LUMS home page: http://www.lancaster.ac.uk/lums/

Individual vs. Group Decision Making: an Experiment on Dynamic Choice under Risk and Ambiguity

Enrica Carbone*, Konstantinos Georgalos†, Gerardo Infante‡§

November 9, 2016

Abstract

This paper focuses on comparing individual and group decision making, in a stochastic inter-temporal problem in two decision environments, namely risk and ambiguity. Using a consumption/saving laboratory experiment, we investigate behaviour in four treatments: (1) individual choice under risk; (2) group choice under risk; (3) individual choice under ambiguity and (4) group choice under ambiguity. Comparing decisions within and between decision environments, we find an anti-symmetric pattern. While individuals are choosing on average closer to the theoretical optimal predictions, compared to groups in the risk treatments, groups tend to deviate less under ambiguity. Within decision environments, individuals deviate more when they choose under ambiguity, while groups are better planners under ambiguity rather than under risk. We argue that the results might be driven by differences in the levels of ambiguity and risk attitudes between individuals and groups, extending the frequently observed pattern of groups behaving closer to risk and ambiguity neutrality, to its dynamic dimension.

JEL classification: C91, C92, D11, D91, E21

Keywords: Risk, Ambiguity, Inter-temporal Optimisation, Group Decision Making, Learning, Experiment

*Seconda Università di Napoli, 81043, Capua, Italy, ✉ enrica.carbone@gmail.com, ☎ +39 0823274063

†Department of Economics, Lancaster University Management School, LA1 4YX, Lancaster, U.K. ✉ k.georgalos@lancaster.ac.uk, ☎ +44 (0)15245 93170

‡School of Economics, University of East Anglia, NR4 7TJ, Norwich, U.K. ✉ g.infante@uea.ac.uk, ☎ +44 (0)16035 91731

§Financial support from the MIUR- PRIN 2007 "Consumo, Risparmio e Mercati finanziari: Teorie non Convenzionali, Test e Applicazioni", is gratefully acknowledged.

1 Introduction

Most real life economic decisions usually share three main characteristics: (1) the decision environment involves some kind of uncertainty, either objective (risk) or subjective (ambiguity); (2) decisions are made by groups rather than by isolated individuals (e.g. households, executive boards or policy committees) and; (3) decisions involve a sequence of choices over either a long time horizon or after the reception of some relevant information compared of a single choice (e.g. savings, investments or insurance). Standard economic theory relies on the assumption that when an agent is confronted with a stochastic, intertemporal decision problem under uncertainty, she takes into consideration all the possible future states of the world and calculates the optimal solution of this dynamic maximisation problem by applying *backward induction*, satisfying in that way dynamic consistency (what seems to be optimal at time t_2 from the viewpoint of t_1, is still optimal when time t_2 arrives). On top of that, the majority of economic models make no separation between individual and collective decision making assuming that both act in a behaviourally indistinguishable way.

Recently, a vast body of experimental literature has been devoted to the comparison of individual and group decision making. Two recent reviews of this literature (Charness and Sutter (2012) and Kugler et al. (2012)) conclude that groups tend to behave closer to what is defined as *rational* choice by economic theory, comply with the predictions of game theoretical models, as well as to decide in a more self-interested manner. Although there is an affluence of studies on collective choice in static frameworks, there is little empirical evidence of group dynamic decision making.

We present evidence from a consumption/saving laboratory experiment where we study choices from two decision units, namely individuals and groups, in two decision environments, risk and ambiguity. We therefore investigate behaviour in four treatments: (1) individual choice under risk; (2) group choice under risk; (3) individual choice under ambiguity and

(4) group choice under ambiguity, in a stochastic inter-temporal allocation problem. Groups consist of two members and decisions are made after a phase of communication and deliberation. We compare behaviour both within decision units and within decision environments. Within decision units analysis (i.e. individuals (groups) under risk vs. individuals (groups) under ambiguity) allows us to investigate whether the introduction of ambiguity, regarding the future level of income, has any significant impact to the way individuals and groups decide, while within decision environments analysis (i.e. individuals vs. groups under risk (ambiguity)), allows us to explore whether there are fundamental differences between individuals and groups.

Our main results can be summarised as follows. Both groups and individuals substantially deviate from the predicted theoretical optimal level of consumption both under risk and under ambiguity. There are significant treatment effects within a decision environment. We observe an anti-symmetric result, where individuals perform better compared to groups under risk while groups perform better under ambiguity. Likewise, individuals tend to deviate less from the conditional level of consumption when they plan under risk compared to ambiguity, while groups deviate less in an ambiguous environment rather than in a risky one. The majority of the subjects is characterised by considerably myopic (short) planning horizons. We observe a common pattern across all treatments regarding the factors that drive behaviour (e.g. repetition of the task, available wealth) as well as significant gender effects in consumption/saving choices. Finally, we observe precautionary saving behaviour with individuals saving more under ambiguity than under risk and also individuals saving more compared to groups. We argue that the observed differences in behaviour may be the consequence of different risk and ambiguity attitudes, which extends the often observed pattern in static choice experiments of groups being risk and ambiguity neutral, to a dynamic decision framework.

The paper is organised as follows. We start in section 2 by reviewing the related literature on life-cycle experiments, dynamic group choice and group decision making under ambiguity and discuss how our study contributes to this literature. In section 3, we present the decision task as well as the underlying theoretical model that we aim to test. We then move to the experimental design, stimuli and procedures in section 4 while in section 5 we report our results. We then conclude.

2 Related Literature

Many studies in the psychology literature and more recently in the economics discipline, aim to explore differences between individuals and groups in various fields. These studies usually focus on investigating how differently individuals and groups decide, compared to the predictions of some kind of *rational* decision theory. Kugler et al. (2012) and Charness and Sutter (2012) report extensive experimental evidence advocating the superiority of groups regarding decision making that adheres to the game theoretical predictions. When this comparison concentrates on decision making under risk and ambiguity, the main research question that is often explored, is whether individuals and groups are characterised by different attitudes towards risk and ambiguity or if being member of a group alters the individual levels of these attitudes. Baker et al. (2008) find that groups tend to make decisions that are more consistent with risk neutral preferences, Shupp and Williams (2008) using parametric structural estimations find that groups have a lower risk aversion coefficient, Masclet et al. (2009) on the contrary, find that groups opt for the safer choices, Charness et al. (2010) find that groups perform significantly better on a probability reasoning task, Zhang and Casari (2012) report that group choices are more coherent and closer to risk neutrality, Bougheas et al. (2013) find that groups take more risk than individuals, while Baillon et al. (2016) investigate behaviour in Allais paradox and stochastic dominance tasks, reporting that groups violate less often stochastic dominance but they deviated more in the Allais paradox tasks.

More recently, motivated by the extensive experimental evidence of non-neutral ambiguity attitudes (Halevy (2007), Ahn et al. (2014), Hey and Pace (2014) and Stahl (2014) among others[1]) researchers have started investigating group decision making under ambiguity (imprecise probabilities). Early studies concentrated on the effects of social interaction to ambiguity attitudes, rather than on choices by groups (see Curley et al. (1986), Keller et al. (2007), Trautmann et al. (2008) and Muthukrishnan et al. (2009)). Charness et al. (2013) show that ambiguity neutral agents are able to persuade the non-neutral ones to make joint, ambiguity-neutral decisions, Keck et al. (2014) find that groups are inclined to make more ambiguity neutral decisions and that ambiguity averse individuals tend to become ambiguity neutral after they consult with their peers. Brunette et al. (2015) report that groups applying the unanimity rule are less risk averse. They found the same pattern for ambiguity but without significance. Similar work has been done by Levati et al. (2016) who test different voting rules in collective choice under ambiguity and by Lahno (2014) who examines the effects of feedback in decision making under ambiguity.

Almost all the aforementioned studies, investigate decision making in a static framework. Nevertheless, there are a few experiments that investigate collective choice in inter-temporal frameworks. Gillet et al. (2009) find that groups make qualitatively better decisions than individuals in an inter-temporal common pool environment, Charness et al. (2007) report that individuals tend to choose first-order stochastically dominated alternatives more often in a Bayesian updating experiment, Jackson and Yariv (2014) find that social planners exhibited extensive present bias in an inter-temporal common consumption stream experiment, Carbone and Infante (2014) and Carbone and Infante (2015) compared behaviour between individuals and groups in an inter-temporal life-cycle experiment under risk (objective uncertainty) finding significant deviations from the optimal planning strategy as well as significant differences between the treatments, while Denant-Boemont et al. (2016) find that groups are more patient and make more consistent decisions in collective time preferences experiment.

[1]See Etner et al. (2012) for a review of the theoretical models and Trautmann and van de Kuilen (2015) for a review of the experimental evidence.

In order to compare behaviour in a dynamic framework, we use a decision task borrowed from the literature on saving experiments. Literature in incentivised life-cycle experiments is as early as Hey and Dardanoni (1988). A common result of life-cycle experiments is that agents systematically deviate from the theoretically optimal consumption path usually by over-consuming during the early stages of the life-cycle and under-consume later. Several different explanations have been given for this pattern ranging from dynamically inconsistent preferences that include present bias and truncated planning horizons (Ballinger et al. (2003), Carbone and Hey (2004), Carbone (2005), Brown et al. (2009)), cognitive skills (Ballinger et al. (2011)), external habits and social learning (Carbone and Duffy (2014), Feltovich and Ejebu (2014)) to debt aversion (Meissner (2015)). [2] Our study contributes to the literature in the following ways. To the best of our knowledge, this is the first study to compare individual and group decision making in a dynamic framework, under risk and ambiguity. In the field of group decision making under ambiguity, in contrast to Charness et al. (2013), Keck et al. (2014) and Brunette et al. (2015) who compare groups and individual in a static framework, using a life-cycle experimental design, we report the first experiment that studies dynamic group decision making under ambiguity in a task that involves learning and updating of ambiguous beliefs. Generally the experimental literature on dynamic decision making under ambiguity (updating and learning) is very limited. At the individual decision making level there is the work by Cohen et al. (2000) and Dominiak et al. (2012) that test the Ellsberg paradox in a dynamic framework. Similarly, regarding learning under ambiguity, whilst the topic has been quite developed theoretically (see Marinacci (2002), Epstein and Schneider (2007), Epstein et al. (2010), Zimper and Ludwig (2009)) there is lack of both experimental and empirical work. A recent study by Nicholls et al. (2015) tests whether learning helps to reduce the violations of the Ellsberg paradox. Recently, Baillon et al. (2015) study the effect of learning on ambiguity attitudes in an experiment using initial public offerings on the New York Stock Exchange.

[2] For an extensive review of life-cycle experiments see Duffy (2014)

With this study we aim to obtain some preliminary results of dynamic group choice under ambiguity where the participants obtain information during the experiment which allows them to reduce the level of ambiguity.

Finally, regarding the literature on saving experiments, although the modelling advances in the literature of choice under ambiguity have been recently exploited to theoretically analyse life-cycle decisions (Campanale (2011), Peijnenburg (2015)), there is lack of empirical evidence. We add to this literature by reporting an experimental study of life-cycle choice where ambiguity is introduced regarding the future income stream. In addition, unlike previous saving experiments that only test the effects of social influence to consumption decisions, we explicitly test how groups make similar decisions after deliberation. Our work shares similarities with Carbone and Infante (2014) who study individual choice under certainty, risk and ambiguity in a savings experiment and Carbone and Infante (2015) who study individual and group choice under risk. Nevertheless, our study is different from Carbone and Infante (2014) mainly in three points. First, we compare groups and individuals while they focus only on individuals. Second, we study 15-period lifecycles compared to their 5-period. Then, we adopt a Bayesian learning model to represent updating regarding the future states of the world, while they assume uniform distribution of income. Finally, we extend the framework of Carbone and Infante (2015) which we use as a benchmark to introduce ambiguity to the future income.

Following Charness and Sutter (2012), who claim that "Ultimately, the goal of comparing individual and group decision making is to identify the contexts and types of decisions where each is likely to work best", this study provides a framework for understanding differences between individual and group choice in a stochastic inter-temporal consumption-saving problem under ambiguity.

3 Theoretical Framework

We present a simple, discrete-time, finite horizon life-cycle model of consumption and savings decisions without discounting. An agent lives for a finite number of periods T and receives utility $u(c_t)$ from consumption c_t at every period t with $t \in \{1, 2, \cdots, T\}$. At the beginning of each period t, the decision maker is endowed with a stochastic income y_t[3]. At each period, the agent decides how much of her available wealth to consume and how much to save, given that there is a fixed interest rate applied to the savings each period. The wealth at every period (or the cash-on-hand) include the savings up to that period plus the endowed income for the period. There is no borrowing allowed (the consumption choices should be non-negative) and there are no bequest motives (all the available wealth must be consumed by the end of the life-cycle).

The utility that the agent receives from consumption is represented by a concave, additive separable, constant absolute relative risk aversion (CARA) utility function of the following form:

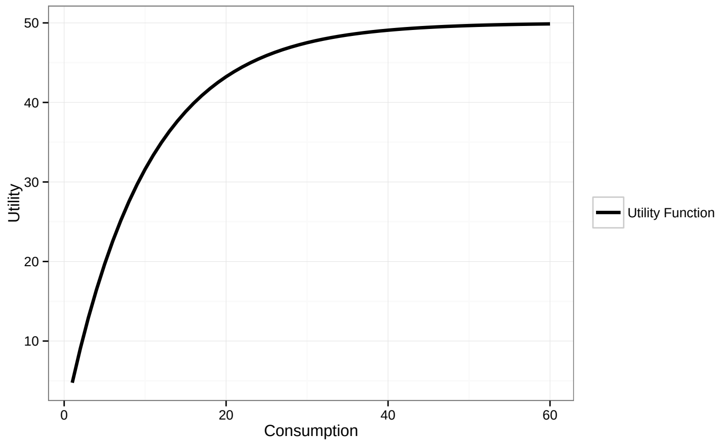
$$U(c_t) = (k - \exp(-\rho c_t))\, \alpha \tag{1}$$

where c_t is the level of the agent's consumption at each period t, ρ is the coefficient of risk aversion and the parameters $\alpha > 0, k > 0$ are scaling factors of the utility function that allow for affine transformations. In order to induce the utility function on the subjects (a conversion function from experimental currency to monetary units) during the experiment we set the following values for the parameters: $\rho = 0.1, \alpha = 50$ and $k = 1$. The shape of the utility function is shown in Figure 1.

The objective of the decision maker is to maximise the utility obtained by the life-cycle consumption. Using the Expected Discounted Utility model, the optimisation program can be

[3]The income generation process is described shortly.

Figure 1: Utility Function



written as:

$$\max_{\{c_t\}} E_t \left[\sum_{t=1}^{T} \beta U(c_t) \right] \tag{2}$$

subject to the inter-temporal budget constraint:

$$w_{t+1} = \alpha_{t+1} + y_t = (1+r)(w_t - c_t) + y_t \tag{3}$$

where w_{t+1} is the wealth of the next period, w_t is the level of wealth or the cash-on-hand at the beginning of period t, α_{t+1} represents the available assets or savings at the beginning of period $t+1$ and r is the rate of return which is known and remained constant during the experiment (at a fixed rate equal to 0.2). The discount rate is assumed to be equal to zero which means that the discount factor β is equal to 1. y_t represents the income of the agent at time t. The income follows a stochastic process which is characterised either by risk or by ambiguity and there are two possible states of the world, a state where the income is High $((\overline{y}_t))$ and a state with Low income $(\underline{y}_t)$. The stochastic process is following a *Bernoulli* distribution which is applied with the aid of a two-colour *Ellsberg*-type[4] urn containing 10 black and white balls in equal proportions[5] representing High and Low income respectively. At each period, a ball is

[4]The Ellsberg-type urns have been introduced in the literature by Daniel Ellsberg seminal paper (Ellsberg (1961)). In this paper he proposed two thoughts experiment with the scope to challenge the "sure thing principle" of the Subjective Expected Utility model (Savage (1954)) and to introduce non-neutral attitudes towards ambiguity. A significant number of experimental studies are making use of either the two-colour or the three-colour urn in order to introduce ambiguity in the lab.

[5]This information was only provided during the risk treatments. During the ambiguity treatments, subjects obtained no information regarding the composition of the urn and thus they were facing ambiguity. During the

randomly drawn from the urn and the colour of the ball defines the state of the world (the income for that period). The sampling method is constituted of draws with replacement so that each draw will not alter the probabilities of the future events. Finally, borrowing is not allowed and therefore the wealth of the agents should be at all times greater or equal to zero. There are no bequest motives and any savings should be consumed before the end of the last period. In addition, there is lack of uncertainty regarding the planning horizon as the agents know the exact length of their life-cycle.

In order to solve for the optimal consumption-savings levels we adopt the value function iteration approach i. The Bellman operator for this problem is given by:

$$V_t(w_t) = u(c_t^*) + E[V_{t+1}(w_{t+1}^*)] \tag{4}$$

where V is the value function and E is the expected operator which is defined as

$$E[V_{t+1}(w_{t+1}^*)] = \mu V_{t+1}(w_{t+1}^{*H}) + (1-\mu)V_{t+1}(w_{t+1}^{*L}) \tag{5}$$

where

$$w_{t+1}^{*s} = (1+r)(w_t - c_t^*) + y^s$$

with $s \in [L, H]$ for Low and High income respectively and μ being the subjective probability (belief) of the agent that the future state of the world will be High[6]. The value function establishes a recursive relation between consumption at every period t and every future period $t+1$. Based on the assumptions above and the constraints imposed by the experimental design, it is possible to calculate an optimal inter-temporal consumption vector $c^* = (c_1^*, \ldots, c_T^*)$ for the agent's life-cycle, for any given level of wealth and for any given level of beliefs regarding the future state of the world. Under these assumptions there is no explicit solution thus we resort to numerical optimisation methods. Using backward induction along with the no bequest constraint (all the wealth must be consumed at the end of the life-cycle), we start from the last

session they had the chance to observe draws from the urn and obtain information regarding the actual distribution.

[6]We elaborate on this issue later.

period, where optimality requires the consumption of all the available wealth, and solve backwards, period by period, for any possible level of wealth. This guarantees that at any period, the Bellman equation is satisfied and the optimal consumption level at period t is a function of the optimal level of consumption at period $t + 1$. Furthermore, similarly to Ballinger et al. (2003), since everything in the experimental design is discrete (the income process, the consumption choices etc.) an exact solution is possible to be calculated and consequently, there is no need for approximation (interpolation).[7] Then, for any given income stream, it is possible to work forward and to recover the optimal levels of consumption and savings, for any corresponding level of wealth. In Figure 2, the optimal life-cycle savings (end-of-period cash balances at the end of each period) path is shown, averaged over 50,000 simulated income streams. As expected[8], the optimal path requires the agent to build a saving profile that is increasing for the first half of the life, reaches a peak at roughly the middle of the life-cycle and then the savings are following a decreasing path till everything is consumed at the last period.

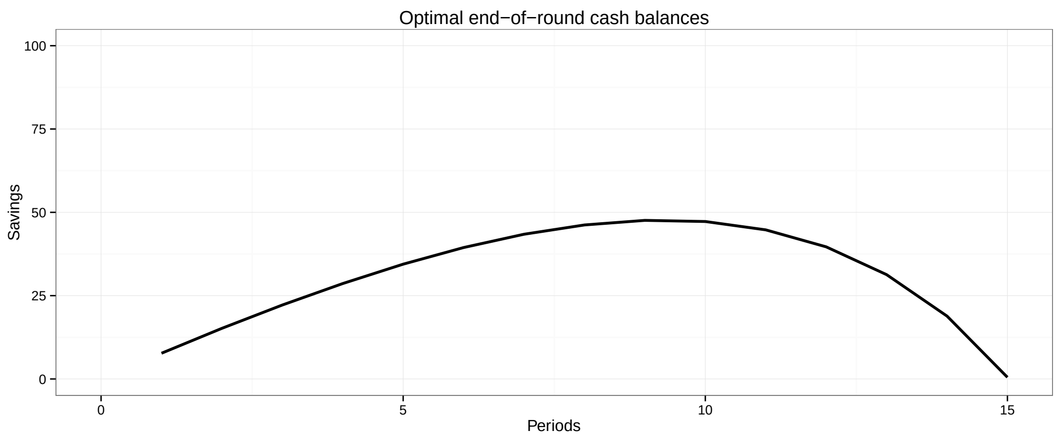


Figure 2: Optimal cash-in-hand holdings (Average of 50000 simulated income streams)

Finally, it is necessary to make some assumptions regarding the subjective beliefs μ that the agents hold on the probability of high income in a given period, how they are formed

[7]The optimal solution and the subsequent econometric analysis were conducted using the R programming language (R Core Team (2013)). The programs are available upon request.

[8]As Ballinger et al. (2003) and Feltovich and Ejebu (2014) notice, the no-borrowing constraint along with a positive third derivative of the utility function, imply motives for precautionary saving. Both conditions are satisfied in our experimental design.

and how are they updated during the experiment. The exogenous income follows a simple i.i.d. Bernoulli process. Nevertheless, the subjects have no information of the value of the parameters of this distribution during the ambiguity treatments. As the income generation process involves draws with replacement, the participants have the chance to obtain information that will allow them to update their beliefs regarding the parameters that characterise the distribution. We adopt a closed-form model of Bayesian learning with additive beliefs[9]. We assume Subjective Expected Utility preferences for the subjects. This is done for simplification reasons as the SEU model by definition assumes neutral attitudes towards ambiguity.[10] The decision maker holds some prior beliefs that are updated based on the relative frequencies that are observed from the sampling. As Zimper and Ludwig (2009) note, in this model of Bayesian learning with additive beliefs, additive posteriors converge to the same limit belief (to the true value of the distribution parameter). This model has initially appeared in the economics literature in Viscusi and O'Connor (1984) and Viscusi (1985). Briefly, the model assumes that the decision maker holds uniform priors regarding the composition of the urn, that is $\mu = Pr(High) = Pr(Low) = 0.5$ before being able to observe any draws. Then for every draw that is being observed, the prior beliefs are updated according to the Bayes rule and the posterior belief is given by:

$$\mu(\text{High}|I) = \frac{1+k}{2+n}$$

where I is the available information, k is the number of successes of High income and n is the total number of draws that has been observed so far.

One could assume that the distribution of the balls in the urn is uniform and calculate the optimal consumption path as if the probability of high income is always equal to 0.5 and ignoring any information provided by the draws. We prefer to adopt a learning model of

[9]The theoretical foundations of the model are presented in Appendix A.

[10]Controlling for averse or loving attitudes towards ambiguity would add two additional layers of complexity to the function mapping from consumption to monetary payoffs (Carbone and Duffy (2014)). If one wants to control for attitudes towards risk and ambiguity, she needs to appropriately extend the experimental design with tasks that will perplex an already complicated decision task (see for example Hey and Dardanoni (1988)). As our main objective is to understand the effects of ambiguity to saving decisions, we leave this for future work.

how beliefs are updated for two reasons. First, there is no psychological justification of why one should assume a uniform distribution and ignore all the available information. Then, as is shown in Figures 3 and 4, the learning model and the equal probabilities model, generate different optimal savings paths for a given stream of income. Figure 3 shows the two paths for an income stream where the income for at least half of the periods is High. Similarly, Figure 4 shows the path for the case of an income stream with at least half of the periods having Low income.

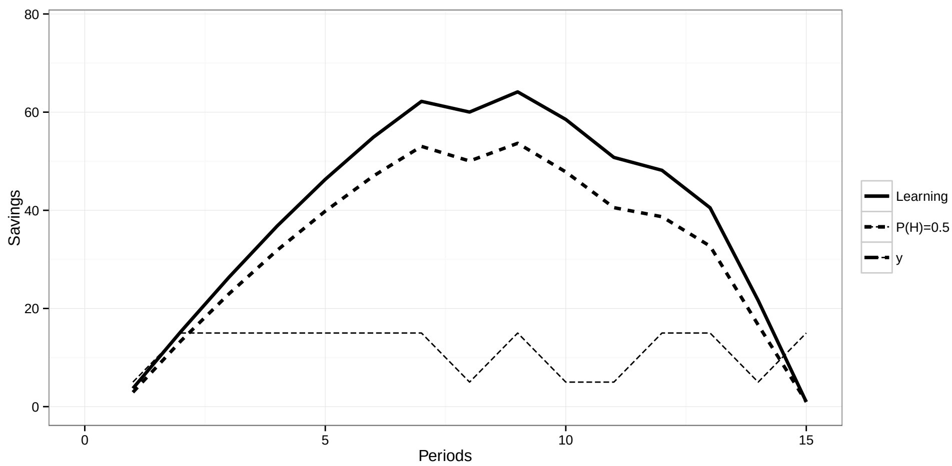


Figure 3: Learning vs. 50-50 cash-in-hand holdings (High Income Stream)

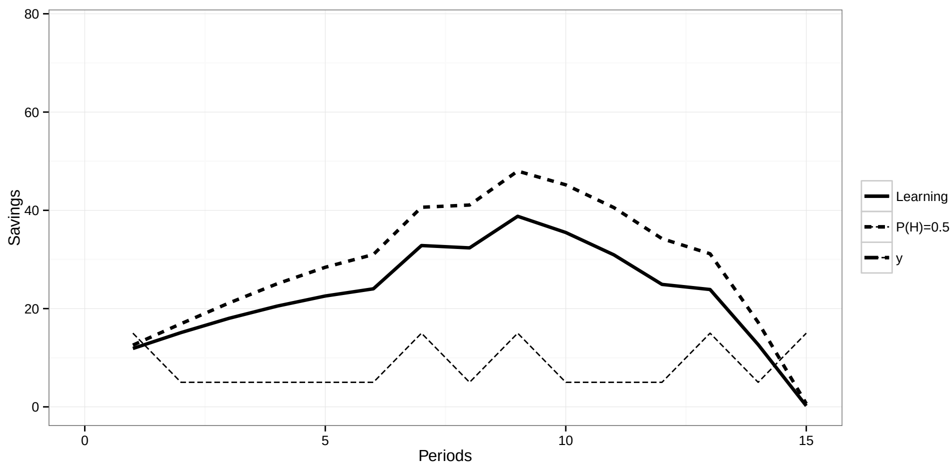


Figure 4: Learning vs. 50-50 cash-in-hand holdings (Low Income Stream)

4 Experimental Design and Procedures

In order to investigate the differences between individual and group planning within the intertemporal consumption framework, we design and conduct an economic experiment using a 2×2 factorial design, with two treatment variables: decision unit (individuals vs. groups) and decision environment (risk vs. ambiguity). Therefore, the experiment features four treatments in total: individual choice under risk (I-R), individual choice under ambiguity (I-A), group choice under risk (G-R) and group choice under ambiguity (G-A).

During an experimental life-cycle (henceforth sequence) there are 15 years (periods). At each period t, an individual (or a group) is endowed with some income expressed in experimental currency units (tokens). This income is determined based on the process described in section 3 and can be either *High* ($\overline{y}_t = 15$) or *Low* ($\underline{y}_t = 5$). The subjects, for each period, they can choose the proportion of income that they would like to consume (they were asked to decide how many of their available tokens they would like to convert into "points"), given that the residual will be saved and earn interest at a fixed rate of return equal to 0.20. As was mentioned before, there were no bequest motives (subjects were expected to consume the total amount of cash-on-hand at the last period of each sequence) and in addition, they could not borrow during a sequence. This task was performed twice, so each subject (or group) participated in two independent, 15-period sequences that we indicate as sequence 1 and sequence 2. Participants received written instructions that provided definitions for the meaning of *sequences* and *periods* which also clarified what was meant by "independence" of sequences[11]. The final payoff was determined by applying the random incentive mechanism, where one of the two sequences was randomly chosen and the accumulated consumption, transformed in monetary units at a fixed rate (two Euros per 100 points), was paid to the participants. Instructions also explained how to use the utility function (called "conversion function"), briefly

[11] During the experiment expressions like "income", "wealth", "consumption" or "utility" were carefully avoided.

pointing out some important features, such as the property of decreasing marginal utility[12]. As was described in section 3, the income at each period was determined by *i.i.d.* draws from an urn. In the risk treatments, the subjects were told in advance that inside the urn there were 10 black and white balls in equal proportions, representing high and low income respectively. During the ambiguity treatments, the same urn was used but without providing any information to the participants regarding its actual composition. At the beginning of the experiment, one participant was asked to publicly open the urn and count the balls. When drawing a ball, participants were asked to shuffle the contents of the urn and then pick one ball to show to everyone. The ball was then placed back into the urn so as not to alter the probability of future draws. When making a decision, subjects were made aware that tokens saved would produce interest (at a fixed rate of 0.2) which, in the next period, would be summed to savings and income to give the total of tokens available for conversion. Instructions also explained that all variables were integers. Participants were advised that interest would be rounded to the nearest integer, and examples were given to clarify this procedure. Finally, participants were told at different points of instructions that any savings left over at the end of the last period would be worthless.

4.1 Individual Decision Making

In the case of individual planning (I-R and I-A), subjects were randomly assigned to computer terminals. Any contact with others, apart from the experimenters, was forbidden. For each decision participants had *one* minute where they could try different conversions (using a calculator), however they were not permitted to confirm their decision before the end of the time span. This procedure was implemented to induce participants to think about their strategy and reduce noise in the data. The software included a calculator to allow participants to view the consequences of their decisions (in terms of future interest, savings and utility) and to compare

[12] Again, there was no explicit reference to decreasing marginal utility but to "increments at a decreasing rate".

alternative strategies.

4.2 Group Decision Making

During the group treatment (G-R and G-A) participants had to make the life-cycle decisions in pairs. Participants were randomly matched to groups at the beginning of the experiment. The identity of the members of the group was not revealed during the experiment. In the second sequence, a random matching rule was enforced, so that groups were formed at the beginning of each sequence and the same participants could not be counterparts more than once. This was implemented in an attempt to isolate the performance of groups to the greatest extent possible. As in the treatment with individuals, a strict no talking rule was imposed (with the exception of members within the group). Groups had a total of three minutes to discuss and confirm a decision; however, a choice could only be confirmed after the first minute. In order to limit the length of sessions, after the three minutes time, if no decision was confirmed by members, the computer would randomly choose between the last two proposals[13]. To facilitate interactions between members and increase information about group strategies, an instant messaging system was made available to chat within the group. Participants were informed about the fact that the software was recording all of their messages and that the chat system was available from the beginning to the end of each period. Participants could freely exchange messages with their counterpart but they were not allowed to reveal their identity, encourage their counterpart to share identifying information or use inappropriate language[14]. Instructions provided a detailed explanation of how to interact with one's counterpart and how to confirm a decision. Group members had to take turns in making proposals as well as take

[13]The software recorded all proposals. When members did not confirm a decision within three minutes, the computer would pick the last proposal of each member and then randomly choose one of those as representative of the group. This did not happen very frequently. We recorded 54 cases of "disagreement" out of 900 decisions (6%). Preliminary regressions suggested that disagreement was not a significant regressor.

[14]The messages from the chat were restricted only to discussions regarding the levels of wealth that the participants wished to consume. Therefore, no interesting data were recovered from the chat that could help us to infer anything regarding subjects' preferences.

turns as "first proposers", that is, who initiated the exchanges of proposals in a period[15]. The person whose turn it was to make a proposal, selected the available button labeled "Propose" which submitted it to their counterpart. After sending a proposal the turn then passed to the other group member, who had to make a counter-proposal. During this process, both members of the group had a calculator available to try different conversions and check the consequences of each of them. As mentioned above, counterparts could not confirm a group decision before one minute. For that reason, they could only use the "Propose" button; a "Confirm" button was only available after the one minute time limit. To confirm a proposal, a group member had to press the "Confirm" button; otherwise she could still make a counter-proposal and pass the turn the other member. After instructions were provided in both individual and group planning sessions, a questionnaire was distributed to test participants' understanding of the experiment. Participants were then given some time to practice with the software, in particular with the calculator and the system for group interaction. All sets of instructions included a graph of the utility function and two tables with examples of conversions and of the interest mechanism.

4.3 Payment

The final payoff was the conversion into money of the total of points accumulated in one sequence. The computer randomly determined which sequence would be used for payment. Instructions explained that points would be converted into money at a fixed rate of two Euros per 100 points. In the group treatments, both members of the group would receive the payoff calculated as described above. This design choice was made so as to not alter the framing of incentives between treatments. Also, the choice of not imposing a sharing rule or allowing participants to enter into bargaining on how to share the payoff, was motivated by considerations on how this might have altered the behaviour of participants during the experiment.

[15]In the first period of a sequence, the computer would randomly determine the "first proposer"; after that, counterparts would take turns exchanging proposals.

Experimental sessions were run at two experimental economics labs in Europe, known to prohibit deception, with participants being undergraduate students of various disciplines. The experiment was programmed and conducted using the z-Tree software Fischbacher (2007).

5 Findings

5.1 Deviations from Optimal Consumption

In the literature of life-cycle experiments there have been adopted two different definitions of optimality, the *unconditional* and the *conditional* level of consumption and savings (see Ballinger et al. (2003) and Carbone and Hey (2004)). The unconditional optimal path is given by the optimal consumption vector c^* which is calculated based on the assumptions regarding the agent's preferences, the values of the respective parameters, the income stream and the optimal level of wealth (given that all past consumption decisions were optimal). This definition of optimality is quite rigid and if an agent deviates from the optimal path at a given period, there is no way to converge to the optimal path in the future. The conditional optimal solution provides a more behaviourally plausible definition of optimality, as the optimal consumption path is calculated based on the actual available cash-on-hand (gross returns from savings of previous periods plus the endowment income y of that period) that a given subject has at the beginning of every period. In addition, this approach incorporates a measure of learning effects and improvement of choices along the life-cycle. For a given period t of the life-cycle, the decision maker is solving a reduced horizon problem of length $T - t + 1$ based on the available cash-on-hand that she has at the beginning of period t. At each period t we calculate the conditional optimal level of consumption given the actual level of the cash-on-hand holdings. Following this approach, the conditionally optimal consumption vector $\tilde{c}_i{}^*$ is calculated which is unique for every subject i. We therefore adopt the definition of conditional optimality upon which we base all the results presented below.

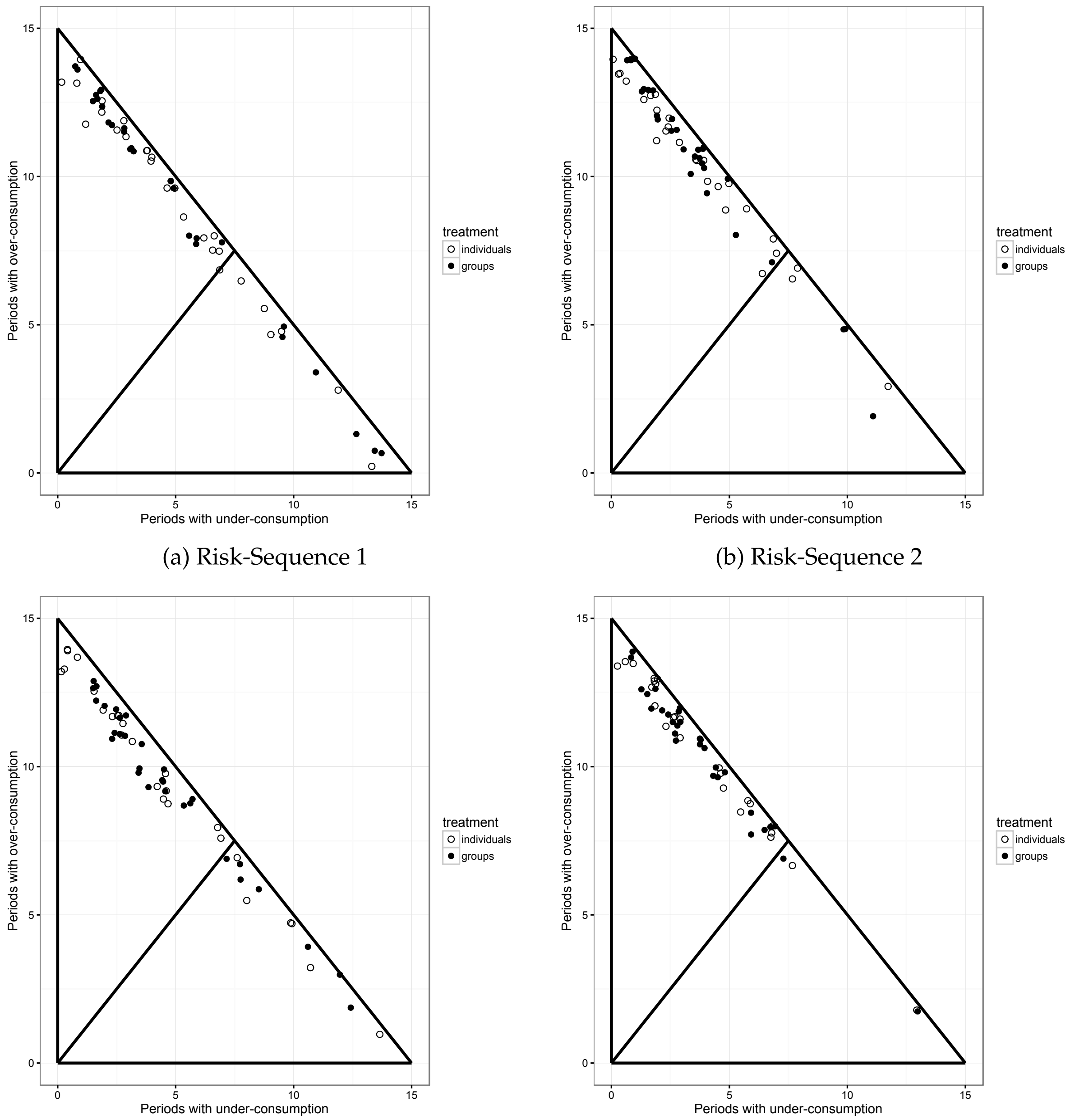
We have data from 170 participants (28 subjects in the I-R treatment, 26 subjects in the I-A, 28 groups in the G-R and 30 groups in the G-A) and each subject could only participate in one session. As a basic test of understanding, we expected all subjects/groups to consume all of their wealth during the last period of each sequence. Indeed, the vast majority of the participants passed this rationality test and we consequently excluded from our sample some *outlier* subjects that left in their saving accounts more than 9 units.[16]

Finding 1 *Both individuals and groups tend to systematically over-consume both under risk and under ambiguity compared to the predicted conditional optimal level of consumption.*

Evidence for this finding can be found in Figure 5. Figures 5a and 5b show the periods of overconsumption/underconsumption concerning the treatments under risk, for sequences 1 and 2 respectively, while Figures 5c and 5d present the same information for the treatments under ambiguity. In each Figure, the horizontal (vertical) axis represents the total number of periods during which a subject (or group) under-consumes (over-consumes). Points close to where the 45° line intersects with the hypotenuse correspond to agents that over-consume for roughly 50% of the rounds and under-consume for the rest, while subjects that behave according to the predicted optimal solution would be represented by points on the origin. Points above (below) the line represent individuals or groups who over-consume (under-consume) for at least half of the periods. There is extensive heterogeneity regarding behaviour and as it can be seen in both Figures, the majority of subjects tends to systematically over-consume for at least 10 out of the 15 periods. The average number of periods displaying an over-consuming behaviour under risk is 9.55 (individuals) and 9.35 (groups) in the first sequence, and 10.02

[16] From the sample we excluded the observations of 1 subject in sequence 1 and 2 subjects in sequence 2 in the I-R treatment, 3 subjects in sequence 1 and 1 subject in both sequences in I-A, 2 subjects in sequence 2 and 1 subject in both sequences in G-R and 1 subject in sequence 1 in G-A. We verified that including in our sample the observations of the participants who although failed the rationality test, they left in their saving accounts less than 9 units, does not change quantitatively the results that we report below.

Figure 5: Periods of over-consumption and under-consumption



(a) Risk-Sequence 1

(b) Risk-Sequence 2

(c) Ambiguity-Sequence 1

(d) Ambiguity-Sequence 2

(individuals) and 10.40 (group) in the second one. The respective number of periods for the ambiguity treatments are 9.20 and 9.25 for individuals and groups during the first sequence and 10.46 and 10.51 during the second. Figures 5b and 5d visually confirm this amplified over-consumption pattern during the second sequence.

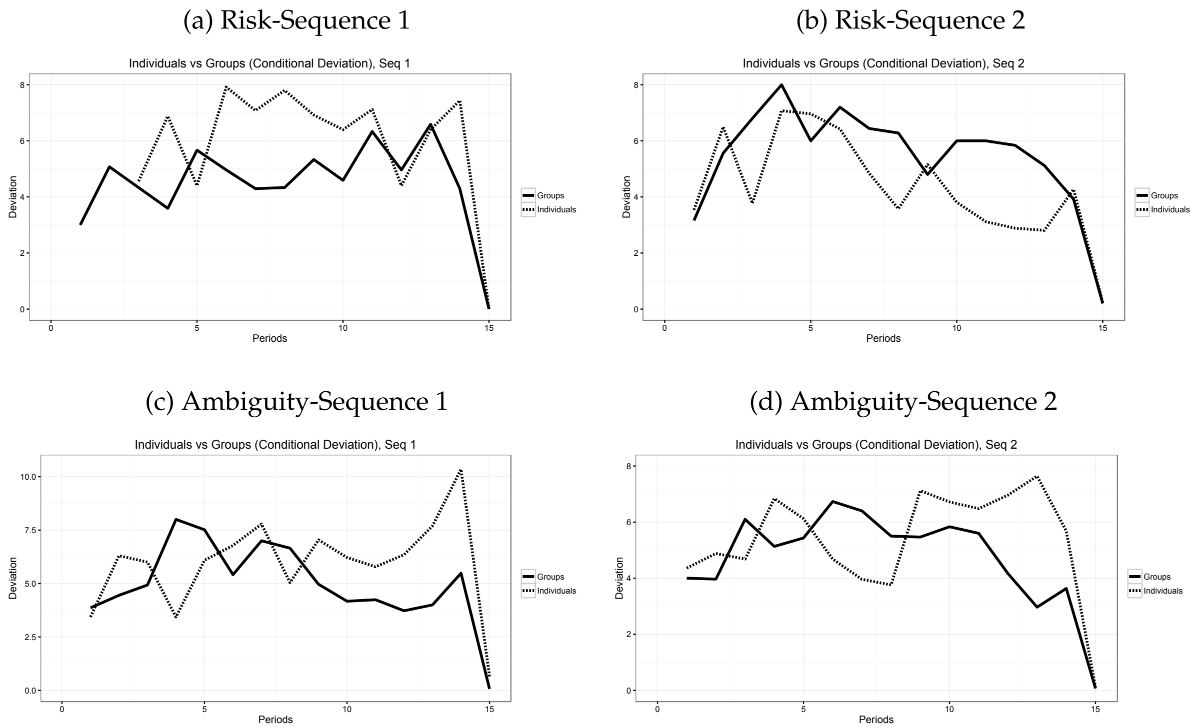
Finding 2 *Both groups and individuals substantially deviate from the predicted conditional optimal* *level of consumption both under risk and under ambiguity.*

Figure 6 depicts the mean absolute deviation of the actual consumption choices c observed in the experiment from the conditional optimal ($|c_t^*(w_t) - c_t|$), in every period, of the 15-period sequences, for both individuals and groups. Figures 6a and 6b illustrate the deviations of individuals and groups for the risk treatments, for sequences 1 and 2 respectively, while Figures 6c and 6d communicate the same information concerning the ambiguity treatments. The horizontal axis represents the periods of each sequence while the vertical axis the absolute deviation. The observations of a subject who never deviates from the optimal would coincide with the horizontal axis. From the Figures, it is apparent that there are clearly systematic differences between decision units within each decision environment regarding how much they deviate from the conditional optimal. First, both individuals and groups, in both decision environments and for the two sequences, begin by significantly deviating from the optimal level. On top of that, the average deviation has a positive sign, confirming the pattern of finding 1, highlighting subjects' difficulties to adopt a saving strategy that builds up for the first half of the sequence[17]. Focusing on the risk treatments, individuals seem to significantly deviate more compared to groups during the first sequence, a pattern which is later reserved during the second sequence. In the ambiguity treatments the pattern of behaviour is less clear for the first half of each sequence. Nevertheless, the gap between individuals and groups dramatically widens since groups substantially reduce their deviation from the optimal on the one hand, while individuals steadily increase theirs, during the last half of each life-cycle.

These different patterns call for a more formal comparison between treatments. To this end, we conduct a series of generalised least squares (GLS) random effects regressions with robust standard errors clustered at the individual level (similarly at group level for groups). We run regressions both within treatments in an effort to understand how different factors affect deviations from optimum, as well as between treatments in order to identify potential treatment

[17]Individuals (groups) exhibit positive deviation of 4.32 (2.62) consumption units under risk and deviation of 2.56 (3.17) units under ambiguity.

Figure 6: Deviations of groups and individuals from conditional optimal consumption



effects (both between individuals and groups and between risk and ambiguity). We first focus on the risk treatments. As dependent variable we use the conditional absolute deviation from the optimum[18]. The advantage of using absolute deviations is that the sign of the estimated coefficients can be interpreted as an indicator of the "direction" of the effect (i.e. a positive (negative) sign indicates increasing (decreasing) deviation from optimum). The first two columns of Table 1 report the results of regressions within the risk treatments, for individuals and groups respectively. In addition to a constant term, we include as explanatory variables the following: "period" which refers to the period number and captures the time trend, "seq2" which is a dummy variable indicating whether consumption decision was made during the second sequence, "income" which is the income the subjects received in each period, "wealth" which refers to the level of cash-on-hand at the beginning of period t, "gender", a dummy variable indicating of whether the subject is male, "gndrmx", a dummy variable indicating whether the

[18]We also conducted the regressions using the mean squared deviation from the conditional optimal as dependent variable. Although the results are magnified compared to those where absolute deviation has been used as the dependent variable, the qualitative results regarding the treatment effects remain the same. We report the results of these regressions in the supplementary material.

group was formed by a heterogeneous pair[19].

The constant term is positive and significantly different from zero. This term captures the deviation from the unconditional optimal at the beginning of the life-cycle $t = 1$ and the statistical significance confirms the hypothesis that both individuals and groups have difficulties in calculating the optimal consumption path. The coefficient of the sequence is not significantly different from zero for individuals implying that there is no effect from the experience of the first sequence in improving behaviour. On the contrary, this coefficient is significant and positive for groups indicating a further deviation from the conditional optimal in the second sequence. The rest of the explanatory factors seem to explain behaviour in a symmetric way for both individuals and groups. The coefficient of the time trend is significant and negative, showing that there is reduction to the deviations as subjects make choices towards the end of the sequences. Income plays a positive role as does wealth, indicating that an increase to either of these two measures , leads to further deviations from the optimum. Finally, there seem to be significant gender effects, where male subjects deviate less in the individual treatment while the same is true when heterogeneous groups are asked to make choices.

We then pool together the data from the I-R and G-R treatments to test whether there is a difference between individuals and groups. We estimate the model using the same explanatory variables with the only difference that we drop the "seq2 and we introduce the dummy variable "treatg" which indicates whether a decision was made by a group. In addition, we use the following control variables: "treatg $\times$ wealth", "treatg $\times$ period" and "wealth $\times$ period" which capture the interactions between treatment, wealth and period as well as their joint interaction. The results are reported in the third column of Table 1. Not surprisingly, the signs of the explanatory variables remain the same compared to the I-R and G-R treatments alone. Furthermore, the coefficient that captures the treatment effect is positive and statistically significant. This confirms that there is significant difference between individuals and groups and

[19] In the case the group consisted of one male and one female member, this dummy variable takes the value 1.

moreover, individuals seem to make choices that are closer to the benchmark. Also, all the interaction terms are significantly different from zero.

Table 2 reports the estimation of a similar set of regressions using the data from the ambiguity treatments. The first column includes the estimates for the I-A treatment, the second for the G-A and the third the coefficients of the the pooled I-A and G-A model. A similar pattern is observed in the within estimations as before. The main difference is that now the coefficient of the sequence is positive and significant for the individuals (compared to groups in the risk treatments) indicating a further deviation from the optimal during the second sequence. The same coefficient for the groups is statistically insignificant, implying no changes to the way the choices were made. The rest of the variables explain behaviour similarly to the risk treatments. In contrast to the risk treatments, focusing now on the pooled model (third column of Table 2), there is a significant and negative treatment effect coefficient, confirming that groups deviate less under ambiguity compared to individuals. It also worth's noting the magnitude of this treatment effect which in absolute terms, it is roughly four times bigger compared to the treatment effect in the risk treatments (-3.786 vs. 0.872). We summarise the results of the regressions in the following findings:

Finding 3 *Subjects significantly deviate from the conditional optimal path in both risk and ambiguity treatments. This deviation depends positively on the wealth and the income and negatively on the stage of the life-cycle. Groups improve their performance under risk while individuals worsen theirs under ambiguity during the second sequence. There are also significant gender effects with male and mixed groups deviating less from the conditional optimal.*

Finding 4 *There are significant treatment effects between treatments within a decision environment. Individuals perform better compared to groups under risk while groups perform better under ambiguity.*

We proceed by asking the question of whether there are any differences when the same decision unit makes choices in different decision environments. That is to say, we are interested to find out whether the introduction of an ambiguous decision environment has significant effects to the way individuals and groups choose. To this end, we pool the together the data from I-R and I-A treatments for individual choice and from G-R and G-A for groups. Table 3 reports the estimated coefficients for the two models where the same explanatory variables as before have been used. Note that the "treatg" variable has now been substituted by "treata", a dummy variable that indicates whether a choice was made in an ambiguous environment. The first column compares individuals under risk and ambiguity with the main variable of interest being "treata". This variable in the pooled I-R and I-A model, has a significant and positive value, indicating that individuals perform much worse in the ambiguity treatment compared to the risky one, implying that ambiguity has indeed significant effects on choices. On the contrary, as can be seen in the second column of Table 3, when we compare groups under risk and ambiguity, the treatment coefficient is significant and negative, implying that groups are much better planners under ambiguity rather than under risk. The effect of all the remaining explanatory variables remains the same as above.

Finding 5 *Individuals tend to deviate less from the conditional level of consumption when they plan under risk compared to ambiguity. On the contrary, groups deviate less in an ambiguous environment rather than in a risky one.*

Table 1: pooling effects regression estimates between actual and conditional optimal consumption (absolute deviation)

	Treatment I-R	Treatment G-R	Treatments I-R & G-R
(Intercept)	3.085***	2.422***	2.035***
	(0.123)	(0.119)	(0.237)
seq2	−0.005	0.991***	
	(0.068)	(0.100)	
period	−0.296***	−0.313***	−0.158***
	(0.005)	(0.008)	(0.023)
income	0.191***	0.088***	0.131***
	(0.003)	(0.004)	(0.009)
wealth	0.152***	0.174***	0.238***
	(0.001)	(0.001)	(0.013)
gndrm	−1.829***	−1.191***	−1.255***
	(0.042)	(0.057)	(0.181)
gndrmx		−1.675***	−1.902***
		(0.056)	(0.225)
treatg			0.872**
			(0.328)
treatg×wealth			−0.073***
			(0.019)
treatg×period			−0.203***
			(0.032)
period×wealth			−0.009***
			(0.001)
treatg×period×wealth			0.010***
			(0.002)
R^2	0.380	0.594	0.494

Table 2: pooling effects regression estimates between actual and conditional optimal consumption (absolute deviation)

	Treatment I-A	Treatment G-A	Treatments I-A & G-A
(Intercept)	2.568^{***}	3.328^{***}	5.015^{***}
	(0.090)	(0.093)	(0.279)
seq2	1.224^{***}	-0.100	
	(0.083)	(0.137)	
period	-0.368^{***}	-0.298^{***}	-0.603^{***}
	(0.007)	(0.006)	(0.029)
income	0.081^{***}	0.175^{***}	0.145^{***}
	(0.005)	(0.002)	(0.011)
wealth	0.226^{***}	0.109^{***}	0.039^{*}
	(0.001)	(0.000)	(0.015)
gndrm	-2.008^{***}	-1.067^{***}	-1.224^{***}
	(0.093)	(0.050)	(0.231)
gndrmx		-0.578^{***}	-0.402
		(0.057)	(0.264)
treatg			-3.786^{***}
			(0.360)
period$\times$wealth			0.016^{***}
			(0.001)
treatg$\times$wealth			0.210^{***}
			(0.019)
treatg$\times$period			0.491^{***}
			(0.041)
treatg$\times$period$\times$wealth			-0.027^{***}
			(0.002)
R^2	0.574	0.363	0.536

Table 3: pooling effects regressions between actual and conditional optimal consumption (comparison between decision units)

	Treatments I-R & I-A	Treatments G-R & G-A
(Intercept)	2.291^{***}	2.776^{***}
	(0.307)	(0.308)
tra	2.806^{***}	-1.455^{***}
	(0.391)	(0.353)
period	-0.214^{***}	-0.382^{***}
	(0.028)	(0.018)
income	0.147^{***}	0.124^{***}
	(0.011)	(0.009)
wealth	0.227^{***}	0.160^{***}
	(0.013)	(0.013)
gndrm	-1.275^{***}	-1.187^{***}
	(0.271)	(0.263)
period×wealth	-0.006^{***}	0.001
	(0.001)	(0.001)
tra×wealth	-0.168^{***}	0.100^{***}
	(0.019)	(0.018)
tra×period	-0.412^{***}	0.289^{***}
	(0.039)	(0.027)
tra×period×wealth	0.021^{***}	-0.014^{***}
	(0.002)	(0.002)
gndrmx		-1.042^{***}
		(0.254)
R^2	0.526	0.498
Num. obs.	1485	1665

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

The results above clearly indicate that individuals and groups behave in a substantial different way both within and between decision environments. One could argue that a potential explanation for this kind of differences is the discrepancy regarding the income streams in the various sessions.[20] For instance, when we compare the I-R and G-R treatments, the further divergence from the conditional optimal could have been the consequence of a larger number of "high" income periods in the G-R treatment which induced groups to consume more. Table 4 reports the average levels of income, consumption and wealth for all treatments. Although there seem to be differences regarding the distribution of income across treatments (first column), both Mann-Whitney-Wilcoxon (MWW henceforth) and χ^2 tests show that these differences are not statistically significant.[21] Therefore, differences in the distribution of income across treatments are not sufficient to explain the observed differences.

Table 4: Average levels of income, consumption and wealth levels (standard deviations in brackets)

Treatment	Income	Consumption	Wealth
I-R	9.88	12.36	23.59
s.d.	(5.00)	(7.63)	(10.67)
G-R	10.16	13.00	27.25
s.d.	(5.01)	(8.07)	(11.73)
I-A	9.44	12.03	25.20
s.d.	(5.02)	(9.88)	(13.30)
G-A	10.08	12.81	26.03
s.d.	(5.03)	(7.55)	(11.05)

5.2 Estimating Planning Horizons

In this section, we use a bounded rationality approach and estimate the apparent planning horizons used by the subjects (see among others Ballinger et al. (2003), Carbone and Hey (2004) and Ballinger et al. (2011)), as different levels of potential myopia may be able to explain differences in behaviour. During the experiment, subjects are required to solve a complex inter-

[20]Note that, since during the experiment there were actual draws from the urn, there was no way to implement the same income streams to all treatments.

[21]I-R vs. G-R: $p = 0.780$; I-R vs. I-A: $p = 0.550$; I-A vs. G-A: $p = 0.300$; G-R vs. G-A: $p = 0.900$. All reported p-values were generated using pairwise χ^2 tests.

temporal decision task and are expected to do so by employing their optimal plans using a "T-period" planning horizon, where T is equal to the 15 periods in each sequence. However, due to the complexity of the problem, some subjects tend to use simplifying rules, such as "using a shorten horizon which is then rolled forward[22]" to cover the actual length of the life-cycle. As noted in Ballinger et al. (2003) and Carbone and Hey (2004), this leads to dynamic inconsistency and sub-optimal choices. In particular, a subject using this kind of strategy (having a subjective horizon of τ) will behave in period t as if period $t+\tau+1$ were the last one (except for the last period, T, that will be correctly recognised as the end of the life-cycle). For example, a person with two periods planning horizon, will behave as if each period is the last-but-one, except for the last-but-one and last periods which are correctly recognised as the last two of the life-cycle. Hence, this strategy implies that in period t the subject will not use the optimal consumption function (policy function) that corresponds to period t. Instead, she will use the consumption function of period $T+1-\tau$ if t is smaller or equal to $T+1-\tau$, otherwise she will use the correct one. Following this reasoning, for each possible length of the planning horizon ($1 \leq \tau \leq T$[23]), the optimal solution has been computed, using the optimal consumption functions. The "apparent" planning horizon has been determined as the one in which the mean squared deviation from the optimal consumption is minimised. In other words, the "apparent" horizon is given by:

$$\hat{\tau} \equiv \arg\min_{\tau \in \{1,2,\ldots,T\}} \left(\sum_{t=1}^{T} ([c_t - {c_p}^*]^2) \right) \tag{6}$$

where c_t is the actual consumption of the subject for period t and ${c_p}^*$ with $p = \max\{1, t+\tau-15\}$ is the optimal level of consumption based on the optimal policy function for the respective horizon that the subject is optimising. As before, we estimate the horizons using the definition of conditional optimal consumption (the consumption that would be optimal given the cash-on-hand that the subject actually has in that period). Tables 5 and 6 report the average length

[22] Carbone and Hey (2004).

[23] In our experimental design, this τ may range from 1 (extreme myopic behaviour) to 15 (optimal behaviour).

of the planning horizons, the standard deviation and the maximum length of the horizons for both individuals and groups, for risk and ambiguity respectively.

Table 5: Planning Horizons - Risk (Conditional Optimal)

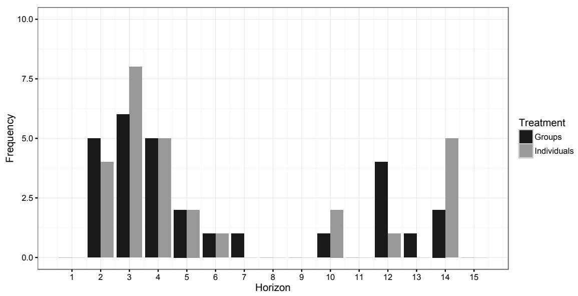
	Individuals		Groups	
	Sequence 1	Sequence 2	Sequence 1	Sequence 2
Average	6.07	4.43	6.07	3.86
s.d.	4.20	2.41	4.42	2.37
Max	14.00	12.00	14.00	12.00

Table 6: Planning Horizons - Ambiguity (Conditional Optimal)

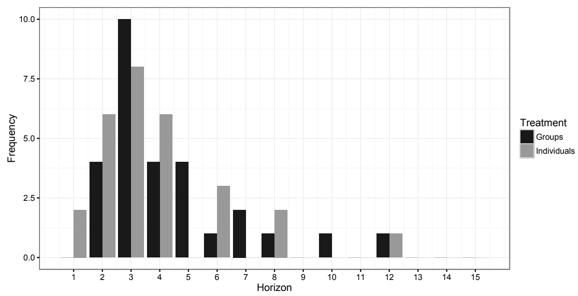
	Individuals		Groups	
	Sequence 1	Sequence 2	Sequence 1	Sequence 2
Average	4.81	4.23	4.23	3.87
s.d.	3.97	3.67	2.65	2.43
Max	14.00	14.00	12.00	15.00

Figure 7: Individual and Group Planning Horizons (Conditional Optimal)

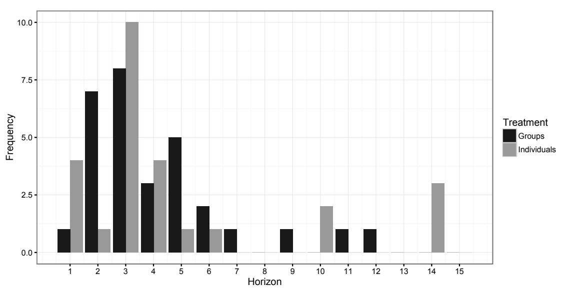
(a) Estimated Horizons - Risk (Seq 1)



(b) Estimated Horizons - Risk (Seq 2)



(c) Estimated Horizons - Ambiguity (Seq 1)



(d) Estimated Horizons - Ambiguity (Seq 2)

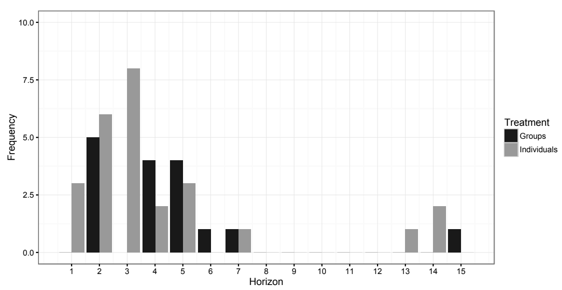


Figure 7 depicts the distribution of the estimated horizons. It is obvious from this figure that there is substantive heterogeneity between subjects. The distribution of the planning horizons is left skewed for all treatments, indicating that the majority of the subjects fail to apply a full-horizon plan for the whole life-cycle. The average planning horizon for both sequences is 5.25 (s.d.=3.31) periods for individuals and 4.97 (s.d. 3.40) periods for groups in the risk treatments and 4.52 (s.d.=3.82) for individuals and 4.05 (s.d.=2.54) periods for groups in the ambiguity treatments. At first sight, there seem to prevail two distinct patterns, that that estimated horizons in the risk treatments are longer and that both individuals and groups do not improve their planning capacity during the second sequence. Nonetheless, according to MWW tests, none of the between treatments comparisons seems to be statistically significant[24], nor any of the within treatments comparisons (compare first and second sequence) appears to be different, with the exception of the I-R treatment where subjects perform significantly worst during the second period concerning their planning capacity[25].

Finding 6 *There is extensive heterogeneity regarding the planning horizons in all treatments. The majority of the subjects is characterised by considerably myopic (short) horizons. In addition, there are no significant differences on the length of estimated horizons across treatments.*

5.3 Precautionary Saving and Aversion towards Risk and Ambiguity

Since neither differences in the distribution of income across treatments nor in differences regarding the planning horizons of the participants are able to explain the discrepancy in behaviour between individuals and groups, in this section we ask whether relaxing the assumption of uniform levels of risk and ambiguity neutrality as well as the uniformity of these attitudes across decision units, could explain the observed differences. We thus focus our analysis on the main variable of interest (i.e. consumption) and based on the notion of *precautionary* saving under risk and ambiguity (see Kimball (1990), Gollier (2001), Baillon (2016)), we aim

[24]I-R vs. G-R: $p = 0.452$; I-R vs. I-A: $p = 0.269$; I-A vs. G-A: $p = 0.620$; G-R vs. G-A: $p = 0.432$. All p-values reported were generated using rank-sum MWW tests for independent samples.

[25]I-R: $p = 0.008$; G-R: $p = 0.330$; I-A: $p = 0.178$; G-A: $p = 0.743$. All reported p-values were generated using rank-sum MWW tests for independent samples for the group treatments and signed-rank MWW tests for the individual ones.

to understand whether different levels in consumption decisions (and consequently distinct levels of divergence) may be explained by different levels of risk and ambiguity aversion that characterise individuals and groups.

The principle of precautionary saving states that an agent who is risk averse (and therefore has a convex marginal utility, $u''' > 0$)[26] will raise the level of optimal savings in the presence of uncertainty regarding the future income. Furthermore, based on a "more prudence" argument, higher levels of risk aversion should lead to higher levels of precautionary saving and therefore, to lower levels of consumption. Given that there are no significant differences regarding the distribution of income streams across treatments, one should not expect any significant differences in the average levels of consumption as well (see Table 4). We are however able to test whether there is a difference in the level of risk aversion between individuals and groups by simply comparing their consumption decisions. Nevertheless, MWW tests reveal that these differences in consumption, are indeed significant for all but the G-R vs. G-A pairwise comparisons, indicating that the levels of risk and ambiguity aversion may be different between individuals and groups[27].

To explain differences in the deviation from the optimal, we need to introduce two additional measures, namely *over-consumption* and *under-consumption*. Up to now, this deviation was measured in absolute terms. Nevertheless, as Feltovich and Ejebu (2014, footnote 15) highlight, using the absolute deviation is not a quite informative measure concerning the welfare implications of changes in the level of consumption. That is to say, an increase (decrease) in consumption when an agent over-consumes (under-consumes) is welfare-reducing (welfare-enhancing) and *vice-versa*. The two measures that will be used as dependent variables in the

[26] A decision maker having a positive third derivative is said to be *prudent*.

[27] I-R vs. G-R: $p = 0.025$;I-R vs. I-A: $p < 0.000$; I-A vs. G-A: $p < 0.001$; G-R vs. G-A: $p = 0.981$. All reported p-values were generated using rank-sum MWW tests.

subsequent regressions, are defined as follows:

$$Over - consumption = \begin{cases} c_t - c_t^*(w_t), & \text{if } c_t \geq c_t^*(w_t) \\ 0 & \text{otherwise} \end{cases}$$

and similarly

$$Under - consumption = \begin{cases} c_t^*(w_t) - c_t, & \text{if } c_t^*(w_t) > c_t \\ 0 & \text{otherwise} \end{cases}$$

Using these measures, we repeat similar GLS regressions, with both over-consumption and under-consumption as dependent variables and keeping the explanatory variables similar as above. Table 7 reports the estimates of these regressions, for all the pairwise comparisons between treatments, using the over-consumption measure (similarly Table 8 for under-consumption). The main coefficient of interest is "trt" which is a dummy variable that captures treatment effects[28]. A positive sign of "trt" indicates a higher level of over or under-consumption.

We first focus on the differences between I-R and G-R. Groups consume on average more (13.00 (s.d.=8.07) units compared to 12.36 (s.d.= 7.63) for individuals), an indicator of higher risk aversion for individuals due to potential precautionary saving. Furthermore, in section 5.1 it was shown that groups further deviate from the conditional optimal, so one would expect that the coefficient of "trt" for over-consumption will be positive, as an increase of consumption would reduce under-consumption, as well as deviation from the optimal. Indeed, evidence from the first column in both Tables 7 and 8, confirm this hypothesis, since groups have a coefficient of over-consumption equal to 2.519. The negative sign of under-consumption (-1.569) shows that groups under-consume less. Nevertheless, the magnitude of over-consumption outweighs this effect leading to higher deviation. The comparison I-R vs. I-A is particularly interesting as it focuses within the same decision unit and hence, allows

[28]In each pairwise comparison, the "trt" coefficient captures the treatment effect of the second element of the pair. For instance, in the I-R vs. G-R comparison, "trt" captures the effect of the G-R treatment, in I-R vs. I-A the effect of I-A and so on.

to identify the effects of introducing ambiguity to the future income. Since consumption in the I-A treatment is significantly lower compared to I-R, and given that the decision unit remains the same, one could argue that ambiguity aversion drives this result, validating the precautionary saving principle. Nevertheless, the lower consumption under ambiguity did not translate to lower deviation as well. Despite the negative coefficient of under-consumption for I-A (-0.723), a significant and positive effect of over-consumption (3.876) prevails, causing higher deviations under ambiguity. A reason for this result may be the inability of individuals to make correct judgments regarding the probability of high income, especially in the early stages of the experiment, and despite the higher savings levels, to miscalculate the optimal level of consumption, and consume at higher levels compared to those that optimality calls for. On the contrary, when we compare G-R to G-A we obtain a different result, where groups deviate less under ambiguity. The latter may confirm the finding usually observed in static choice experiments, where groups are better in probability judgment tasks compared to individuals. The lower levels of deviation can be explained by the negative coefficients of both over and under-consumption (-1.206 and -0.523 respectively) of the G-A treatment. In addition, the comparison of groups within environments, provides insights on the effects of ambiguity in consumption choices. A p-value of 0.981 indicates that there are no significant differences regarding the consumption levels of the two treatments and consequently, ambiguity aversion does not seem to have any particular effect in the levels of consumption, implying that groups have a neutral attitude towards ambiguity.

Finally, we turn to the last comparison between decision units, within the ambiguity treatments (I-A vs. G-A). Individuals now consume on average less compared to groups (12.25 (s.d.=9.88) vs. 12.81 (s.d.=7.55)). However, the reason for this potential precautionary saving from individuals is not straightforward and it may be attributed to either higher risk aversion, or higher ambiguity aversion or a combination of the two, from individuals. The lower divergence of groups, despite their higher level of consumption, is driven by both the lower levels

Table 7: Pooling effects regression estimates (over-consumption)

	Treatments I-R vs. G-R	Treatments I-A vs. G-A	Treatments I-R vs I-A	Treatments G-R vs. G-A
(Intercept)	1.878***	5.720***	3.121***	4.425***
	(0.253)	(0.285)	(0.155)	(0.295)
trt	2.519***	−2.173***	3.876***	−1.206**
	(0.386)	(0.407)	(0.190)	(0.384)
period	−0.145***	−0.598***	−0.167***	−0.405***
	(0.016)	(0.026)	(0.005)	(0.019)
income	0.245***	0.235***	0.223***	0.239***
	(0.009)	(0.010)	(0.002)	(0.008)
wealth	0.154***	−0.101***	0.144***	−0.025
	(0.013)	(0.017)	(0.002)	(0.014)
gndrm	−1.063***	−0.185	−2.159***	−0.485**
	(0.138)	(0.185)	(0.150)	(0.157)
gndrmx	−1.285***	0.085		−0.479**
	(0.171)	(0.211)		(0.162)
trt×wealth	−0.162***	0.114***	−0.266***	0.063***
	(0.019)	(0.022)	(0.004)	(0.018)
period×wealth	−0.011***	0.015***	−0.010***	0.005***
	(0.001)	(0.001)	(0.000)	(0.001)
trt×period	−0.230***	0.295***	−0.473***	0.155***
	(0.026)	(0.036)	(0.007)	(0.026)
trt×period×wealth	0.014***	−0.015***	0.026***	−0.007***
	(0.002)	(0.002)	(0.000)	(0.001)
R^2	0.228	0.205	0.213	0.185
Num. obs.	1545	1605	1485	1665

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

of over-consumption (-2.173) and that of under-consumption (-0.701) implying that groups are significantly better planners under ambiguity. As a concluding remark, we need to highlight the fact that if the subjects are actually violating the risk and ambiguity neutrality assumption, then the over and under-consumption measures are under-stated and therefore, the deviations are bound to be higher. The following findings summarise these results:

Finding 7 *Between decision environments and within decision units, individuals save more under ambiguity than under risk, implying that individuals exhibit ambiguity averse attitudes. On the contrary, ambiguity has no effect to the levels of savings for groups, implying ambiguity neutral attitudes for groups.*

Finding 8 *Within decision environments, individuals save more under risk compared to groups, implying that individuals are more risk averse. Similarly, individuals save more under ambiguity compared to groups, implying that individuals are either more risk averse, or more ambiguity averse, or both.*

Table 8: Pooling effects regression estimates (under-consumption)

	Treatments I-R vs. G-R	Treatments I-A vs. G-A	Treatments I-R vs I-A	Treatments G-R vs. G-A
(Intercept)	0.325^{***}	-0.924^{***}	0.330^{**}	-1.107^{***}
	(0.093)	(0.069)	(0.112)	(0.143)
trt	-1.569^{***}	-0.701^{***}	-0.723^{***}	-0.523^{**}
	(0.138)	(0.108)	(0.161)	(0.185)
period	-0.045^{***}	0.031^{***}	-0.000	-0.009
	(0.011)	(0.007)	(0.011)	(0.011)
income	-0.117^{***}	-0.108^{***}	-0.103^{***}	-0.104^{***}
	(0.003)	(0.002)	(0.003)	(0.004)
wealth	0.073^{***}	0.160^{***}	0.069^{***}	0.136^{***}
	(0.003)	(0.004)	(0.004)	(0.007)
gndrm	-0.408^{***}	-0.773^{***}	-0.869^{***}	-0.295^{***}
	(0.049)	(0.049)	(0.062)	(0.060)
gndrmx	0.076	-0.401^{***}		-0.251^{***}
	(0.054)	(0.044)		(0.054)
trt×wealth	0.075^{***}	0.031^{***}	0.063^{***}	0.042^{***}
	(0.005)	(0.005)	(0.006)	(0.008)
period×wealth	0.004^{***}	-0.002^{***}	0.003^{***}	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
trt×period	0.051^{***}	0.126^{***}	-0.027	0.158^{***}
	(0.015)	(0.011)	(0.015)	(0.015)
trt×period×wealth	-0.005^{***}	-0.007^{***}	-0.001^{*}	-0.008^{***}
	(0.001)	(0.000)	(0.001)	(0.001)
R^2	0.470	0.418	0.389	0.506
Num. obs.	1545	1605	1485	1665

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

6 Discussion and Concluding Remarks

We present results from an intertemporal choice experiment under risk and ambiguity where we compare individual and group decision making. By introducing a stochastic income generation process, keeping the level of interest rate constant, as well as controlling the level of utility derived from consumption, we are able to calculate the optimal path of savings/consumption choices for each lifecycle and for each income history and therefore, to study deviations from optimality. We study differences both within a decision unit (i.e. individuals (groups) under risk vs. individuals (groups) under ambiguity) and within a decision environment (i.e. individuals vs. groups under risk (ambiguity)). In our analysis, we take into consideration the fact that subjects may face difficulties in successfully solving complex, stochastic problems in a dynamic environment and therefore, we adopt the definition of *conditional* optimal as a benchmark, which allows for mistakes at the earlier periods of the lifecycle. Our data also allows us to estimate the apparent planning horizons of the subjects, assuming a *bounded* rationality approach.

Our main findings show that (1) both individuals and groups face difficulties in detecting the optimal decision path that this stochastic, dynamic problem implies, in either environments (risk and ambiguity); (2) groups tend to deviate less from the optimal choice compared to individuals under ambiguity, while on the contrary, they deviate more in a risky environments; (3) both individuals and groups are characterised by myopic planning horizons in both environments; and (4) the introduction of ambiguity, regarding the future level of income, provokes precautionary saving effects to individuals while there is zero effect for groups.

Our results seem to be in line with the main experimental findings in the literature of savings experiments that is, people tend to overconsume in the early stages of their lives, failing to build up the required wealth for smooth consumption across the lifecycle and that subjects are characterised by rather myopic planning horizons. We contribute to this literature by showing that these results hold also in the case of ambiguity. The novelty of our experimental design, allows to directly study the effects that ambiguity, regarding the future level of income, have to dynamic decision making. Furthermore, we investigate whether there are significant differences in the way that individuals and groups decide in this particular framework. Our findings confirm the usually observed in other studies phenomenon, where groups behave closer to the predictions that some kind of rationality defines, but in our case, only for the ambiguity treatment. We argue that this may be the effect of different attitudes towards risk and ambiguity and that our results extend the pattern that is frequently observed in static choice experiments, where groups tend to be less risk and ambiguity averse compared to individuals (or stating it differently tend to behave closer to risk and ambiguity neutrality) to its dynamic dimension.

These results are of interest both from a theoretical point of view and from a public policy perspective. Despite the fact that the theoretical literature on dynamic decision making under ambiguity is well advanced, only recently these theoretical developments have been applied to model behaviour in relevant applications like lifecycle savings decisions (Campanale

(2011), Peijnenburg (2015)). Furthermore, there is lack of empirical evidence regarding economic agents' behaviour during inter-temporal tasks, particularly in an ambiguous environment, both at individual and at group level. Although some recent studies have investigated collective choice and discounting behaviour (Jackson and Yariv (2014), Denant-Boemont et al. (2016)), the scope of this literature was to investigate preferences over time rather than to explore sequential group decision making upon the reception of new information as in our case. In this paper we have taken a first step towards understanding the effects of uncertainty regarding the future levels of income on optimal planning. From a public policy aspect, it is well established in the literature (Shapiro (2010), Carlsson et al. (2012)) that despite the various theoretical violations that groups (pairs or larger) exhibit, they tend to be more patient when making joint decisions rather than individual ones. In our study, we find that groups tend to behave closer to rationality when they plan under ambiguity, achieving in that way higher levels of welfare. This could have potential implications during the design of public policy, given that most of real-life economic decisions are taken in an ambiguous environment.

Our findings should be interpreted with some caution. As we are interested in the qualitative characteristics of inter-temporal choice under risk and ambiguity, we assume risk and ambiguity neutrality of the decision makers and we use their subsequent behaviour as our reference. The experimental design does not allow us to control for attitudes towards risk and ambiguity. A similar task would require a different design that would involve elicitation tasks (both for the attitudes and the beliefs) that would provide sufficient data in order to parametrically estimate the respective coefficients, as well as the subjective beliefs of the agents. Such a design would add additional levels of complexity to an already difficult decision task and probably would not allow us to focus on the pure effect that ambiguity has to planning, as well as to the potential differences between individuals and groups, as we are aiming to do at the current study. Despite this simplification assumption, it is a first step towards understanding the effects of ambiguity to inter-temporal consumption/savings problems. In our analysis, we make a speculation that the results might be driven by the existence of ambiguity non-neutral attitudes, phenomenon that is frequently observed among the standard experimental subject population. One can expect that the existence of ambiguity aversion would lead to different optimal saving and consumption decisions (e.g. precautionary saving) or that it wound intensify the observed deviations. As mentioned before, suitable adaptations are needed regarding the experimental design along with the assumed theoretical model that describes subjects' behaviour. We leave the above extensions for future work.

References

Ahn, D., Choi, S., Gale, D., and Kariv, S. (2014). Estimating Ambiguity Aversion in a Portfolio Choice Experiment. *Quantitative Economics*, 5(2):195–223.

Baillon, A. (2016). Prudence with Respect to Ambiguity. *The Economic Journal*, forthcoming.

Baillon, A., Bleichrodt, H., Keskin, U., L'Haridon, O., and Li, C. (2015). Learning Under Ambiguity: An Experiment Using Initial Public Offerings on a Stock Market. Working paper, University of Rennes 1 & University of Caen.

Baillon, A., Bleichrodt, H., Liu, N., and Wakker, P. (2016). Group Decision Rules and Group Rationality under Risk. *Journal of Risk and Uncertainty*, 52(2):99–116.

Baker, R., Laury, S., and Williams, A. (2008). Comparing Small-Group and Individual Behavior in Lottery-Choice Experiments. *Southern Economic Journal*, 75(2):367–382.

Ballinger, T., Hudson, E., Karkoviata, L., and Wilcox, N. (2011). Saving Behavior and Cognitive Abilities. *Experimental Economics*, 14(3):349–374.

Ballinger, T. P., Palumbo, M. G., and Wilcox, N. T. (2003). Precautionary Saving and Social Learning Across Generations: an Experiment. *The Economic Journal*, 113(490):920–947.

Bougheas, S., Nieboer, J., and Sefton, M. (2013). Risk-taking in social settings: Group and peer effects. *Journal of Economic Behavior & Organization*, 92(0):273 – 283.

Brown, A. L., Chua, Z. E., and Camerer, C. F. (2009). Learning and Visceral Temptation in Dynamic Saving Experiments. *Quarterly Journal of Economics*, 124(1):197–231.

Brunette, M., Cabantous, L., and Couture, S. (2015). Are Individuals More Risk and Ambiguity Averse in a Group Environment or Alone? Results from an Experimental Study. *Theory and Decision*, 78(3):357–376.

Campanale, C. (2011). Learning, Ambiguity and Life-cycle Portfolio Allocation. *Review of Economic Dynamics*, 14(2):339 – 367.

Carbone, E. (2005). Demographics and Behaviour. *Experimental Economics*, 8:217–232.

Carbone, E. and Duffy, J. (2014). Lifecycle Consumption Plans, Social Learning and External Habits: Experimental Evidence. *Journal of Economic Behavior & Organization*, 106:413 – 427.

Carbone, E. and Hey, J. D. (2004). The Effect of Unemployment on Consumption: an Experimental Analysis. *The Economic Journal*, 114(497):660–683.

Carbone, E. and Infante, G. (2014). Comparing Behavior Under Risk and Under Ambiguity in a Lifecycle Experiment. *Theory and Decision*, 57:313–322.

Carbone, E. and Infante, G. (2015). Are Groups Better Planners than Individuals? An Experimental Analysis. *Journal of Behavioral and Experimental Economics*, 57:112 – 119.

Carlsson, F., He, H., Martinsson, P., Qin, P., and Sutter, M. (2012). Household Decision Making in Rural China: Using Experiments to Estimate the Influences of Spouses. *Journal of Economic Behavior & Organization*, 84(2):525 – 536.

Charness, G., Karni, E., and Levin, D. (2007). Individual and Group Decision Making under Risk: An Experimental Study of Bayesian Updating and Violations of First-Order Stochastic Dominance. *Journal of Risk and Uncertainty*, 35(2):129–148.

Charness, G., Karni, E., and Levin, D. (2010). On the Conjunction Fallacy in Probability Judgment: New Experimental Evidence Regarding Linda. *Games and Economic Behavior*, 68(2):551 – 556.

Charness, G., Karni, E., and Levin, D. (2013). Ambiguity Attitudes and Social Interactions: An Experimental Investigation. *Journal of Risk and Uncertainty*, 46(1):1–25.

Charness, G. and Sutter, M. (2012). Groups Make Better Self-Interested Decisions. *Journal of Economic Perspectives*, 26(3):157–76.

Cohen, M., Gilboa, I., and Schmeidler, D. (2000). An Experimental Study of Updating Ambiguous Beliefs. *Risk, Decision and Policy*, 5 (2):123–133.

Curley, S., Yates, F., and Abrams, R. (1986). Psychological Sources of Ambiguity Avoidance. *Organizational Behavior and Human Decision Processes*, 38(2):230 – 256.

Denant-Boemont, L., Diecidue, E., and l'Haridon, O. (2016). Patience and Time Consistency in

Collective Decisions. *Experimental Economics*, page forthcoming.

Dominiak, A., Dürsch, P., and Lefort, J. (2012). A Dynamic Ellsberg Urn Experiment. *Games and Economic Behavior*, 75:625–638.

Duffy, J. (2014). Macroeconomics: A Survey of Laboratory Research. Technical report, University of California.

Ellsberg, D. (1961). Risk, Ambiguity and the Savage Axioms. *Quarterly Journal of Economics*, 75:643–669.

Epstein, L., Noor, J., and Sandroni, A. (2010). Non-Bayesian Learning. *The B.E. Journal of Theoretical Economics*, 10(1):1–20.

Epstein, L. and Schneider, M. (2007). Learning Under Ambiguity. *Review of Economic Studies*, 74(4):1275–1303.

Etner, J., Jeleva, M., and Tallon, J. (2012). Decision Theory Under Ambiguity. *Journal of Economic Surveys*, 26(2):234–270.

Feltovich, N. and Ejebu, O. (2014). Do Positional Goods Inhibit Saving? Evidence from a Life-cycle Experiment. *Journal of Economic Behavior & Organization*, 107:440 – 454. Empirical Behavioral Finance.

Fischbacher, U. (2007). z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics*, 10(2):171–178.

Gillet, J., Schram, A., and Sonnemans, J. (2009). The Tragedy of the Commons Revisited: The Importance of Group Decision-making. *Journal of Public Economics*, 93(56):785–797.

Gollier, C. (2001). *The Economics of Risk and Time*. MIT Press, Cambridge, Mass., London.

Halevy, Y. (2007). Ellsberg Revisited: An Experimental Study. *Econometrica*, 75(2):503–536.

Hey, J. and Pace, N. (2014). The Explanatory and Predictive Power of Non Two-Stage-Probability Models of Decision Making Under Ambiguity. *Journal of Risk and Uncertainty*, Forthcoming.

Hey, J. D. and Dardanoni, V. (1988). Optimal consumption under uncertainty: An experimental investigation. *The Economic Journal*, 98(390):105–116.

Jackson, M. O. and Yariv, L. (2014). Present Bias and Collective Dynamic Choice in the Lab. *American Economic Review*, 104(12):4184–4204.

Keck, S., Diecidue, E., and Budescu, D. (2014). Group Decisions under Ambiguity: Convergence to Neutrality . *Journal of Economic Behavior & Organization*, 103:60 – 71.

Keller, R., Sarin, R., and Sounderpandian, J. (2007). An Examination of Ambiguity Aversion: are two Heads Better than one? *Judgment and Decision Making*, 2(6):390–397.

Kimball, M. (1990). Precautionary Saving in the Small and in the Large. *Econometrica*, 58(1):53–73.

Kugler, T., Kausel, E. E., and Kocher, M. G. (2012). Are Group More Rational than Individuals? A Review of Interactive Decision Making in Groups. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(4):471–482.

Lahno, A. M. (2014). Social anchor effects in decision-making under ambiguity. Discussion Papers in Economics 20960, University of Munich, Department of Economics.

Levati, V., Napel, S., and Soraperra, I. (2016). Collective Choices Under Ambiguity. *Group Decision and Negotiation*, pages 1–17.

Marinacci, M. (2002). Learning from Ambiguous Urns. *Statistical Papers*, 43:145–151.

Masclet, D., Colombier, N., Denant-Boemont, L., and Lohéac, Y. (2009). Group and Individual Risk Preferences: A Lottery-choice Experiment with Self-employed and Salaried Workers. *Journal of Economic Behavior & Organization*, 70(3):470 – 484. Field Experiments in Economics.

Meissner, T. (2015). Intertemporal Consumption and Debt Aversion: an Experimental Study. *Experimental Economics*, pages 1–18.

Muthukrishnan, A., Wathieu, L., and Xu, J. (2009). Ambiguity aversion and the preference for established brands. *Management Science*, 55(12):1933–1941.

Nicholls, N., Romm, A. T., and Zimper, A. (2015). The Impact of Statistical Learning on Violations of the Sure-thing Principle. *Journal of Risk and Uncertainty*, 50(2):97–115.

Peijnenburg (2015). Life-cycle Asset Allocation with Ambiguity Aversion and Learning. Working paper.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Savage, L. (1954). *The Foundations of Statistics*. Wiley, New York.

Shapiro, J. (2010). Discounting for You, Me and We: Time Preference in Groups and Pairs. *mimeo*.

Shupp, R. and Williams, A. (2008). Risk Preference Differentials of Small Groups and Individuals. *The Economic Journal*, 118(525):258–283.

Stahl, D. (2014). Heterogeneity of Ambiguity Preferences. *The Review of Economics and Statistics*, 96(5):609–617.

Trautmann, S. and van de Kuilen, G. (2015). *Ambiguity Attitudes*, pages 89–116. John Wiley & Sons, Ltd.

Trautmann, S., Vieider, F., and Wakker, P. (2008). Causes of Ambiguity Aversion: Known versus Unknown Preferences. *Journal of Risk and Uncertainty*, 36(3):225–243.

Viscusi, W. (1985). A Bayesian Perspective on Biases in Risk Perception . *Economics Letters*, 17(1–2):59 – 62.

Viscusi, W. and O'Connor, J. (1984). Adaptive Responses to Chemical Labeling: Are Workers Bayesian Decision Makers? *The American Economic Review*, 74(5):942–956.

Zhang, J. and Casari, M. (2012). How Groups Reach Agreement in Risky Choices. *Economic Inquiry*, 50(2):502–515.

Zimper, A. and Ludwig, A. (2009). On Attitude Polarization under Bayesian Learning with non-additive Beliefs. *Journal of Risk and Uncertainty*, 39(2):181–212.

A Bayesian Learning with Additive Beliefs

In this Appendix we provide the formal Bayesian learning model we adopt which is the benchmark model of Zimper and Ludwig (2009). We consider the income generation process applied to our experimental design where an agent is uncertain about the probability of high income $P(H)$[29]. Nevertheless, she can observe n i.i.d. draws with replacement. We define a probability space $(\mu, \Omega, \mathcal{F})$ where μ stands for the subjective additive probability measure defined on the events of the event space $\mathcal{F}$. The state space is defined as $\Omega = \Pi \times S^{\infty}$ with generic element $\omega = (\pi, s^{\infty})$. The parameter space Π collects all the possible values of the true probability of (H) in any given trial. Similarly, the sample space $S^{\infty} \times_{i=1}^{\infty} \{H, L\}$ collects all the possible sequences of outcomes. It is assumed that after any given number of n trials, the agent knows the result of each of the trials. In addition, it is assumed that the agent cannot somehow observe the true parameter value of the distribution. Define $\tilde{\pi} : \Omega \to [0,1]$ such that $\tilde{\pi}(\pi, s^{\infty}) = \pi$ the random variable that defines at every state the true probability of the outcome (H). The decision maker holds priors over $\tilde{\pi}$ that are assumed to follow the Beta distribution with shape parameters $\alpha, \beta > 0$. The priors are given by:

$$\mu(\pi) = K_{\alpha,\beta}\pi^{\alpha-1}(1-\pi)^{\beta-1}$$

with $K_{\alpha,\beta} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$ and Γ the Gamma function.

Let $X_n : \Omega \to \{0,1\}, n = 1, 2, \cdots$ denote the random variable that takes the value 1 if the income is high (H) and zero otherwise in n trials. We define as I_n^k the event in $\mathcal{F}$ such that the outcome H has been realised k times out of n trials:

$$I_n^k = \{\omega \in \Omega | I_n(\omega) = k\}$$

Since it is assumed that the the random variable X_n is i.i.d. Bernoulli distributed, each I_n

[29]The probability of low income $P(L)$ is defined as the residual $P(L) = 1 - P(H)$.

is, conditional on the parameter-value π, binomially distributed with probabilities

$$\mu(I_n^k) = \binom{n}{k}\pi^k(1-\pi)^{n-k}\text{for } k \in \{0, \cdots, n\}$$

Each time that the decision maker observes a draw from the urn she receives information that allows her to update her prior beliefs. This happens with the application of the Bayes rule. The posterior that π is the true value conditional on the information I_n^k acquired till that point is given by:

$$\begin{aligned}
\mu(\pi|I_n^k) &= \frac{\mu(\pi \cap I_n^k)}{\mu(I_n^k)} \\
&= \frac{\mu(I_n^k|\pi)\mu(\pi)}{\int_{[0,1]} \mu(I_n^k|\pi)\mu(\pi)d\pi} \\
&= K_{\alpha+k,\beta+n-k}\pi^{\alpha+k-1}(1-\pi)^{\beta+n-k-1}
\end{aligned}$$

The agent's prior of the true value of the probability of H is given by the expected value of $\tilde{\pi}$ with respect to the prior distribution. In the case of a Beta prior, it is possible to show that this prior is equal to:

$$\mathbb{E}[\tilde{\pi}, \mu] = \frac{\alpha}{\alpha+\beta}$$

where α, β are the shape parameters of the Beta distribution. During the experiment, we assume that since there is no prior information on the proportion of H balls in the urn, the only reasonable prior that one can attach is 0.5 probability. Then, each draw from the urn provides the decision maker with additional information regarding the real values of the parameters of the distribution.

B Instructions - For Online Publication

B.1 Individual Decision Making

Welcome!

This is an experiment on decision making. The experiment will last about 1 hour and a half. Please read these instructions carefully as you have the chance to earn money depending on your decisions. If you have any questions please raise your hand. The experimenter will answer in private. You are not allowed to talk to other participants in the experiment.

The experiment consists of 2 independent "sequences", each one composed of 15 periods. Sequences are independent because there is no relation between them. This means that your choices in one sequence will not influence future sequences. However, please note that, within one sequence, your decision in each period will influence subsequent periods (for example, your decision in period 1 will have consequences for period 2 and so on).

At the beginning of each period you will receive an amount of tokens that will be available to you. You have to decide how many tokens you want to convert into points. You can convert a number of tokens between 0 and the amount available to you. The conversion function of tokens to points is reported in Figure 1 (Appendix). This figure shows graphically the conversion of tokens to points in a continuous interval. You may also look at Table 1 (Appendix) where some examples of conversions are provided. Please note that that the number of points obtained from the conversion increases as the number of tokens converted increases; however, increments are realized at a decreasing rate, that is, the difference in points obtained by converting 6 tokens rather than 5 is bigger than the difference between converting 16 tokens rather than 15. Finally, please note that the more tokens are converted in each period, the less tokens are saved for conversion in future periods. Please note that, before period 15 (the last period) is reached, tokens not converted will be saved for the next period. Savings will earn interest, thus increasing the amount of tokens available in the following period. When period 15 (the last period) is reached, any tokens left (that is, not converted) will be worthless.

Your payoff, at the end of the experiment, will be calculated on the decisions you have made in ONE of the above mentioned "sequences". This sequence will be randomly selected among the 2 played. This means that your payment will be calculated based on the decisions you made during the 15 periods composing the randomly selected sequence. In particular, your payment will be the conversion in Euros of the total amount of points earned in the selected sequence, using a conversion rate of 2 Euros each 100 points.

Periods and Decision Making

At the beginning of each period, you will be randomly assigned a number of tokens. This number may be "high" (15 tokens) or "low" (5 tokens). The probability of getting either of the two is unknown. It is important to note that the amount of tokens received in one period does not affect the chances of getting the same or the other amount in any following period. The number of tokens will be determined by a draw from a non-see-through bag containing coloured balls. There are only two colours, however the number of balls of either colour is unknown. A number of tokens (high or low) will be attributed to each of the two colours. The draw will determine the number of tokens for all participants in that period.

From period 1 to period 14, if you have any tokens saved, they will earn interest, at the rate of 20% ($r = 0.2$). Savings plus interest accumulated will increase the number of tokens available to you in the following period. Please remember that tokens not converted at the end of period 15 will be worthless. Table 2 (Appendix) is available to you, reporting some examples of calculation of interest.

At the beginning of each period you will be showed on the computer screen the total of tokens available, consisting in:

1. Tokens earned in the period: 15 or 5
2. Tokens saved in the previous period (S)
3. Interest earned on savings: S x 0.2 (not rounded)
4. Tokens available for conversion rounded to the nearest integer (for example, 3.4=3; 3.5=4 or 3.6=4): Tokens earned in the period (1.) + Tokens saved in the previous period (2) + Interest earned on savings (3.)
5. Total of points earned: sum of the points earned starting from period 1

Of course, in period 1 there will be no savings and no interest received, so the number of tokens available to you will be equal to 15 or 5 tokens.

Within this screen you will be asked to enter the number of tokens you wish to convert into points. You may change your decision in any moment before pressing the "confirm" button. When this button is pressed your decision will become irrevocable. You cannot move to the next decision before one minute from the beginning of the current period. To make your decision you may use a calculator to observe the consequences of your choice. Depending on the number entered, it is possible to see the related savings, interest earned on savings in the next period and the number of points earned from conversion. The use of the calculator will not make your choice final.

Once the first 15-period sequence has been completed, the following sequence will start. As explained above, the experiment involves making decisions through 2 sequences.

At the end of each sequence a summary of the choices made during the 15 periods will be provided.

Earnings

When the 2 sequences have been completed, your payment will be determined. One sequence will be randomly selected and you will receive the conversion in Euros of the total amount to points earned in the selected sequence.

If you have any questions, please raise your hand and an experimenter will be happy to assist you.

Right after these instructions a short quiz testing your comprehension of the experiment will take place followed by 3 minutes practice with the conversion function.

Appendix

Figure 1 - Conversion function

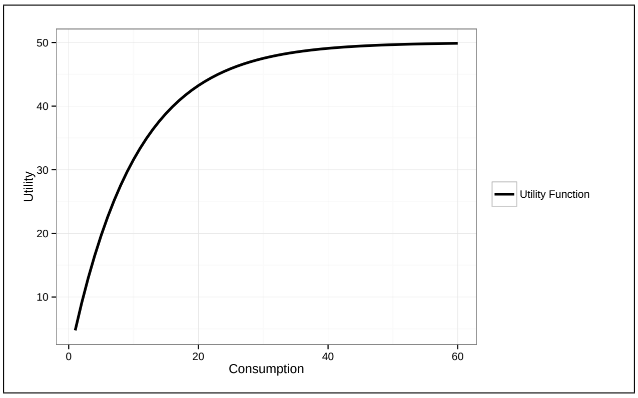


TABLE 1

Tokens Converted (G)	Points Earned
0	0
1	4.758129098
2	9.063462346
3	12.95908897
4	16.4839977
5	19.67346701
6	22.5594182
7	25.17073481
8	27.53355179
9	29.67151701
10	31.60602794
11	33.35644582
12	34.9402894
13	36.37341035
14	37.6701518
15	38.84349199
16	39.9051741
17	40.8658238
18	41.73505559
19	42.52156904
20	43.23323584
⋮	⋮
50	49.66310265
⋮	⋮
100	49.99773
⋮	⋮
150	49.9999847
⋮	⋮
200	49.9999999

$Punti = 50 - 50 * e^{-0.1*G}$

G = Tokens Converted

TABLE 2		
Tokens Saved	**Interest on saved Tokens**	**Tokens Saved + Interest**
0	0	0
1	0.2	1.2
2	0.4	2.4
3	0.6	3.6
4	0.8	4.8
5	1	6
6	1.2	7.2
7	1.4	8.4
8	1.6	9.6
9	1.8	10.8
10	2	12
11	2.2	13.2
12	2.4	14.4
13	2.6	15.6
14	2.8	16.8
15	3	18
16	3.2	19.2
17	3.4	20.4
18	3.6	21.6
19	3.8	22.8
20	4	24
⋮	⋮	⋮
50	10	60
⋮	⋮	⋮
100	20	120
⋮	⋮	⋮
150	30	180
⋮	⋮	⋮
200	40	240

Interest = 0,2 * S
S = Tokens Saved

B.2 Group Decision Making[30]

Welcome!

This is an experiment on decision making. You will be making decisions in cooperation with another participant whose identity will be unknown to you. The experiment will last about 1 hour and a half. Please read these instructions carefully as you have the chance to earn money depending on your decisions. If you have any questions please raise your hand. The experimenter will answer in private. You are not allowed to talk to other participants in the experiment.

The experiment consists of 2 independent "sequences", each one composed of 15 periods. Sequences are independent because there is no relation between them. This means that your choices in one sequence will not influence future sequences. However, please note that, within one sequence, your decision in each period will influence subsequent periods (for example, your decision in period 1 will have consequences for period 2 and so on).

During this experiment you will be part of a group composed of two individuals. The section "Groups and Decisions" explains how groups will be formed, how to interact within a group and reach a decision.

At the beginning of each period your group will receive an amount of tokens that will be available to you. You have to decide how many tokens you want to convert into points. You can convert a number of tokens between 0 and the amount available to you. The conversion function of tokens to points is reported in Figure 1 (Appendix). This figure shows graphically the conversion of tokens to points in a continuous interval. You may also look at Table 1 (Appendix) where some examples of conversions are provided. Please note that that the number of points obtained from the conversion increases as the number of tokens converted increases; however, increments are realized at a decreasing rate, that is, the difference in points obtained by converting 6 tokens rather than 5 is bigger than the difference between converting 16 tokens rather than 15. Finally, please note that the more tokens are converted in each period, the less

[30] The material referred to in the "Appendix" is the same for all sets of instructions and can be consulted in subsection 1 (Individual Decision Making).

tokens are saved for conversion in future periods. Please note that, before period 15 (the last period) is reached, tokens not converted will be saved for the next period. Savings will earn interest, thus increasing the amount of tokens available in the following period. When period 15 (the last period) is reached, any tokens left (that is, not converted) will be worthless.

Your payoff, at the end of the experiment, will be calculated on the decisions you have made in ONE of the above mentioned "sequences". This sequence will be randomly selected among the 2 played. This means that your payment will be calculated based on the decisions you made during the 15 periods composing the randomly selected sequence. In particular, your payment will be the conversion in Euros of the total amount of points earned in the selected sequence, using a conversion rate of 2 Euros each 100 points.

Each member of the group will receive this payoff.

Periods

At the beginning of each period, you will be randomly assigned a number of tokens. This number may be "high" (15 tokens) or "low" (5 tokens). The probability of getting either of the two is unknown. It is important to note that the amount of tokens received in one period does not affect the chances of getting the same or the other amount in any following period. The number of tokens will be determined by a draw from a non-see-through bag containing coloured balls. There are only two colours, however the number of balls of either colour is unknown. A number of tokens (high or low) will be attributed to each of the two colours. The draw will determine the number of tokens for all participants in that period.

From period 1 to period 14, if you have any tokens saved, they will earn interest, at the rate of 20% ($r = 0.2$). Savings plus interest accumulated will increase the number of tokens available to the group in the following period. Please remember that tokens not converted at the end of period 15 will be worthless. Table 2 (Appendix) is available to you, reporting some examples of calculation of interest.

Groups and Decisions

During each sequence you will be paired with another participant but you will not know his/her identity. This matching will be random. At the end of the first sequence, of 15 periods, new groups will be composed for the second sequence, using again random matching.

Participants matched with you in a group will never have the opportunity to know your identity. During the experiment is absolutely forbidden to reveal your identity to the other group member (or try to know the identity of other participants).

At the beginning of each period you will be showed on the computer screen the total of tokens available, consisting in:

1. Tokens earned in the period: 15 or 5
2. Tokens saved in the previous period (S)
3. Interest earned on savings: S x 0.2 (not rounded)
4. Tokens available for conversion rounded to the nearest integer (for example, 3.4=3; 3.5=4 or 3.6=4): Tokens earned in the period (1.) + Tokens saved in the previous period (2) + Interest earned on savings (3.)
5. Total of points earned: sum of the points earned starting from period 1

Of course, in period 1 there will be no savings and no interest received, so the number of tokens available to you will be equal to 15 or 5 tokens.

In the same screen described above you will be asked to interact with the other member of your group in order to make a decision. To do this the following procedure will be employed:

1. You will have to take turns interacting with the other member
2. In the first period, one of the members of the group will be randomly selected to start the interaction. In the periods following the first, members will take turns initiating the interaction.
3. By pressing the button "PROPOSE", the member of the group who begins the interaction will send his/her proposal to the other member and conclude his/her turn. After this,

he/she will have to wait the other member of the group to send his/her decision (accept the proposal or make a new one)

4. It will not be possible to make a group decision before 1 minute. However, during this time group members will be able to exchange proposals of conversion. At the end of the 1 minute time limit, each member of the group, during his/her turn, will also have the opportunity to confirm the proposal received, hence turning it into the group decision, which is irrevocable. The period is concluded when one of the group members confirms a proposal. Hence, the approval of the other member is not required.

5. Members will be able to keep interacting up to a time limit of 3 minutes. After this limit, if a group decision has not been made, the computer will randomly select one of the two members making his/her proposal the final decision of the group.

6. When the minimum time to make a group decision is over (1 minute), if the member whose turn it is to start interacting has not sent any proposal to his partner, the turn will automatically pass to the other member of the group.

Rules of Group Interaction

1. A group decision cannot be made before 1 minute since the start of the current period. This means that even if an agreement is reached, this decision cannot be confirmed before the minimum time limit of 1 minute is over.

2. On the screen used for group interaction, a calculator will be available to you to verify the consequences of your choice. Depending on the number of tokens entered, it is possible to see the related savings, interest earned on savings in the next period and the number of points earned from conversion.

3. A table, denominated "Group decision: current proposals" will be shown on screen. This table is composed of two rows containing the conversion proposals of each member of the group together with the related consequences. Your row is indicated by blue coloured characters.

4. Below this table a box will be available to enter your proposal of conversion, which may be confirmed by pressing the button "PROPOSE".

5. After 1 minute, that is, the minimum time allowed to make a group decision, at each turn a button labeled "CONFIRM" will be available. By pressing this button the group decision will be recorded (becoming irrevocable)

6. An instant messaging (IM) system will also be available and operative from the beginning to the end of the period. To use the chat simply write your message and press enter on the keyboard. This way, your message will be sent to your partner. Each message will be recorded. While using the chat system it is absolutely forbidden to:

   (a) Communicate one's identity in any way (name, student number, nicknames, etc.)

   (b) Ask other participants questions that could lead to the disclosure of identifying information

   (c) Use inappropriate language (insults, etc.)

   The experimenter will make sure that all the rules of chat usage are respected. A violation of one of these rules will cause the cancellation of the final payoff of the participant who committed the violation.

When the group decision has been made, the current period ends and a new period begins.

Once the first 15-period sequence has been completed, the following sequence will start. As explained above, the experiment involves making decisions through 2 sequences.

At the end of each sequence a summary of the choices made during the 15 periods will be provided.

Earnings

When the 2 sequences have been completed, your payment will be determined. One sequence will be randomly selected and you will receive the conversion in Euros of the total amount to points earned in the selected sequence.

If you have any questions, please raise your hand and an experimenter will be happy to assist you.

Right after these instructions a short quiz testing your comprehension of the experiment will take place followed by 3 minutes practice with the conversion function and 3 minutes practice with the group-interaction system.

Psychological Review
1996, Vol. 103, No. 4, 687–719

0033-295X/96/$3.00

Bias in Judgment: Comparing Individuals and Groups

Norbert L. Kerr
Michigan State University

Robert J. MacCoun
University of California, Berkeley

Geoffrey P. Kramer
Indiana University Kokomo

The relative susceptibility of individuals and groups to systematic judgmental biases is considered. An overview of the relevant empirical literature reveals no clear or general pattern. However, a theoretical analysis employing J. H. Davis's (1973) social decision scheme (SDS) model reveals that the relative magnitude of individual and group bias depends upon several factors, including group size, initial individual judgment, the magnitude of bias among individuals, the type of bias, and most of all, the group-judgment process. It is concluded that there can be no simple answer to the question, "Which are more biased, individuals or groups?," but the SDS model offers a framework for specifying some of the conditions under which individuals are both more and less biased than groups.

A great deal of research in social and cognitive psychology has been devoted to demonstrating what is probably an uncontroversial proposition: that human judgment is imperfect. What makes this work interesting and useful is that such imperfections often constitute more than random fluctuations around "rational," prescribed, or ideal judgments. Rather, humans consistently exhibit systematic biases in their judgments. Some of these biases seem to stem from self-enhancing or self-protective motives (e.g., Greenwald, 1980; Myers, 1980). Others may stem from general cognitive shortcuts or heuristics (e.g., Kahneman, Slovic, & Tversky, 1982). Still others seem to reflect an inappropriate sensitivity or insensitivity to certain types of information (e.g., underuse of base-rate information; Kahneman et al., 1982; Nisbett & Ross, 1980). Regardless of their sources, systematic judgmental biases can have serious consequences (cf. Dawes, 1988; Thaler, 1991), and identifying means of controlling such biases is an important challenge for psychology.

These questions have largely been the province of scholars of cognition, social cognition, and judgment and decision making, all of whom have understandably focused primarily upon the behavior of the individual judge. However, in many important instances, the judges who are potentially vulnerable to such systematic biases are groups rather than individuals. For example, typically juries (not individual jurors) must decide guilt or innocence; Congress (not individual lawmakers) must declare war; boards of directors (not individual directors) must decide corporate policy. In this article, we (as social psychologists) explore the following question: Are decision-making groups any less (or more) subject to judgmental biases than individual decision makers? For example, might we expect deliberating juries to be any less (or more) sensitive than individual jurors to proscribed extralegal information, such as the race of a victim? Our goal is to shed light on when groups are more biased than individuals, when individuals are more biased than groups, and most importantly, whether and why there are patterns in such comparisons.

We begin by discussing the concept of judgmental bias and advancing a simple taxonomy of bias effects. We then present an overview of the relevant empirical literature: namely, those studies that compare individual and group susceptibility to particular types of bias. This overview will demonstrate that there is no simple and general pattern in the literature. We then suggest that formal models that link individual and group judgment can usefully be applied to a theoretical analysis of this question. Davis' (1973) *social decision scheme* (SDS) model of group decision making is proposed as a promising basis for such an analysis. The parameters of the SDS model are then linked conceptually to information processing by individuals and by groups (Hinsz, Tindale, & Vollrath, in press). A small set of generic group-judgment processes are then identified from past theory and research using the SDS model. The effects of each of these processes for each type of bias are then explored within a series of "thought experiments" (Davis & Kerr, 1986), in which a number of variables of interest (e.g., group size; magnitude of individual bias) are systematically manipulated. When feasible, these thought experiments are augmented with relevant empirical illustrations of predicted patterns. Finally, an additional source of individual–group differences in biased judgment is

Norbert L. Kerr, Department of Psychology, Michigan State University; Robert J. MacCoun, Graduate School of Public Policy, University of California, Berkeley; Geoffrey P. Kramer, Department of Psychology, Indiana University Kokomo. Geoffrey P. Kramer is now writing and consulting in Ludington, Michigan.

Portions of this article were presented at the Conference on Formal Models of Psychological Processes (R. Hastie, Organizer), Nags Head, North Carolina, June 1990; at the International Conference on Theoretical Developments in Small Groups Research (E. Witte, Organizer), Hamburg, Germany, November 1992; and at the Annual Meeting of the Society for the Advancement of Socio-Economics, New York, March 1993.

Correspondence concerning this article should be addressed to Norbert L. Kerr, 433 Baker Hall, Department of Psychology, Michigan State University, East Lansing, Michigan 48824-1117. Electronic mail may be sent via Internet to kerr@pilot.msu.edu.

identified (and illustrated in prior research)—instances in which possession of certain information alters the group decision-making process itself.

Bias and Its Varieties

Defining (Systematic) Bias

We begin by defining what we mean by biased judgment. The concept of biased judgment assumes that one can specify a nonbiased standard of judgment against which actual human judgments can be compared (Funder, 1987; Hastie & Rasinski, 1988). The basis of that standard, the *normative model* of judgment, may be some formal logical system (e.g., syllogistic logic, probability theory, game theory, rational choice models). However, the normative model may also be based on convention. Good examples of the latter types of standards are the common law rules of evidence that proscribe jurors' use of certain information (e.g., a defendant's race or gender or physical appearance). It is not our purpose here to defend any particular normative models of judgment as ideal or unbiased nor is it to propose conditions for such normative models that are generally necessary or sufficient (see Hastie & Rasinski, 1988, for a discussion of related issues). Rather, we focus on a number of judgmental phenomena for which there are both reasonable and defensible normative models and convincing empirical demonstrations of bias, and we address the theoretical and empirical issue of whether individuals or groups are relatively more likely to exhibit those biases.

Our focus will be on systematic departures from a standard of judgment (i.e., patterned or orderly deviations from the normative standard; Einhorn, Hogarth, & Klempner, 1977; Funder, 1987). Thus, we will not be concerned here with individual versus group differences in the magnitude of unsystematic random error, in the consistency with which valid cues are applied or in ability to learn from diagnostic feedback (e.g., see Chalos & Pickard, 1985; Davis, Kerr, Sussmann, & Rissman, 1974; Einhorn et al., 1977; Laughlin & Shippy, 1983; Laughlin & Sweeney, 1977; Zajonc, 1962).

A Taxonomy of Systematic Biases

Hastie and Rasinski (1988) suggest that there are several distinctive logics for establishing a systematic bias in judgment. Their taxonomy of bias differs from others in the literature in that it distinguishes methods for demonstrating bias, rather than task domains (e.g., Pelham & Neter's, 1995, distinction between biases in persuasion vs. person perception vs. judgment under uncertainty) or the psychological origins of biases (e.g., Arkes', 1991, distinction among strategy-based, association-based, and psychophysically based errors). In this article, we will be concerned with three of Hastie and Rasinski's types of bias.

Judgmental Sins of Imprecision

The first and most straightforward type of bias is revealed by a direct comparison between judgment and criterion. A familiar example is research demonstrating that judges rarely alter their subjective probability judgments as much in response to new, diagnostic, and probabilistic information as Bayes' theorem prescribes (Edwards, 1968). Another example (of which we will make repeated use later) comes from research on prospect theory (Kahneman & Tversky, 1984). Participants are asked to choose between two courses of action with identical expected values (under an assumption of a linear utility function, such that the value of the n'th unit gained or lost is equal to the value of the first such unit). Unbiased judgment should result in indifference between the two choices. However, Kahneman & Tversky (e.g., 1984) have shown that this choice is biased by the way in which the choices are described or "framed"; for example, participants generally preferred Choice A with uncertain loss to Choice B with certain loss, but also prefer Choice C with certain gain over Choice D with uncertain gain (or, to use the customary terminology, participants seem to be risk seeking when the outcomes are framed as losses and to be risk averse when the outcomes are framed as gains). The magnitude of such a bias in judgment might be indexed by how much more popular Choice A was (e.g., percentage of participants choosing A) than the prescribed baseline (viz., participants preferred Choice A 50% of the time, indicative of indifference between A and B). Unbiased judgment in this example prescribes a specific and precise use of available information (viz., computation and comparison of expected utilities); biased judgment reflects systematic departure from this prescribed and precise use of information. For this reason, we will term this type of bias a judgmental *sin of imprecision* (SofI; Hastie & Rasinski, 1988, refer to such a contrast as a "direct assessment of criterion-judgment relationship").

At this point, we might introduce some useful notation. Let us suppose that the judgment task posed to individuals requires participants to choose among n possible responses. For example, in the prospect theory paradigm just discussed, researchers asked participants to choose between two outcomes (one certain and the other uncertain). We shall denote the distribution of individual judgments or decisions across these n alternatives with the vector $\boldsymbol{p} = (p_1, p_2, \ldots, p_n)$. In order for investigators to document that a judgmental sin of imprecision has occurred among individual judges, they must first specify how the judgments of perfectly unbiased individuals should be distributed. We shall denote this ideal criterion distribution as $\boldsymbol{I}$; for example, in our prospect theory example, $\boldsymbol{I} = (.5, .5)$. Unbiased judgment would require that (within the limits of sampling error) $\boldsymbol{p} = \boldsymbol{I}$; when $\boldsymbol{p} \neq \boldsymbol{I}$, bias would be indicated. In the latter case, the magnitude of the bias displayed by individual judges (b) could simply be indexed by $b = |\boldsymbol{p} - \boldsymbol{I}|$, where $|\boldsymbol{p} - \boldsymbol{I}|$ denotes the length of vector $(\boldsymbol{p} - \boldsymbol{I})$. (Of course, the direction in which judges depart from the criterion may also be important in understanding the causes of the biased judgment, but a simple scalar index of the magnitude of bias will be sufficient for the purposes of this article.) Thus, the assertion that individuals exhibit a sin of imprecision would require evidence justifying a rejection of the null hypothesis $H_0: b = 0$.

Now, in a like manner, we could ask r-person groups to perform the identical judgment task (e.g., we could ask 4-person groups to choose between certain and uncertain losses). Using upper-case letters to denote group variables, the magnitude of bias among group judges (B) would just be $B = |\boldsymbol{P} - \boldsymbol{I}|$, where $\boldsymbol{P} = (P_1, P_2, \ldots, P_n)$, the distribution of group judgments. Our primary interest in this paper is estimating and explaining the relative magnitude of individual and group bias. This may be indexed by RB = relative bias = $B - b$. When $RB = 0$, then groups and individuals exhibit an identical degree of bias. When $RB > 0$, groups are relatively more biased than individual judges; and when $RB < 0$, groups are relatively less biased than individuals.

Judgmental Sins of Commission

A key feature of a sin of imprecision is that the no-bias criterion is defined theoretically and the magnitude of bias is defined by the discrepancy between that criterion and human judgment.[1] The other two types of bias that we will consider use an empirical rather than a theoretical no-bias criterion. In one such type of bias, which we will term a judgmental *sin of commission* (SofC; and Hastie & Rasinski, 1988, call "use a bad cue"), the model of ideal, unbiased judgment holds that certain information is irrelevant or nondiagnostic for the required judgment. For example, the rules of evidence usually require that an unbiased juror pay no attention to a victim's race or a defendant's physical attractiveness in deciding whether or not the defendant is guilty. Bias is manifest when jurors use such information. This typically involves comparison of a condition in which the potentially biasing information is provided (e.g., jurors considering a stimulus trial with a physically attractive defendant) with a control condition in which either this information is missing (e.g., no information provided on defendant attractiveness) or different information is provided (e.g., the defendant is physically unattractive). We might call the former, experimental condition the high-bias condition and the latter, control condition the low-bias condition. A sin of commission has occurred when the judgments in these two conditions differ significantly.

Extending our earlier notation, a sin of commission by individual judges requires that the judgments of individuals in the high- (H) and low- (L) bias conditions differ, that is, $\boldsymbol{p}_H \neq \boldsymbol{p}_L$ (where $\boldsymbol{p}_H$ and $\boldsymbol{p}_L$ are the distributions of individual judgments in the high- and low-bias conditions, respectively) and, hence, that $b = |\boldsymbol{p}_H - \boldsymbol{p}_L|$ be greater than zero. A corresponding bias among groups would mean that $\boldsymbol{P}_H \neq \boldsymbol{P}_L$ (where $\boldsymbol{P}_H$ and $\boldsymbol{P}_L$ are the distributions of group judgments in the high- and low-bias conditions, respectively) and that $B = |\boldsymbol{P}_H - \boldsymbol{P}_L| > 0$. Again, in this article we are primarily interested in the relative magnitude of individual and group bias, $RB = B - b$.

It is worth noting that sins of imprecision can be documented using a sin of commission methodology. For example, rather than establishing the prospect theory's risk-seeking bias by comparing the popularity of the risky alternative with a chance, 50% baseline, one might instead compare the level of preference for the risky alternative in a condition using a loss frame with the same preference in a second condition utilizing a gain frame (e.g., see McGuire, Kiesler, & Siegel, 1987). This alters the goodness-of-fit statistical logic used when establishing a sin of imprecision to the more typical null hypothesis testing statistical logic used when establishing the other two types of judgmental sins (Hastie & Rasinski, 1988; Meehl, 1990). We note all this because, as we will show later, how one decides to demonstrate bias can affect the comparison of individual and group bias.

Judgmental Sins of Omission

We will term the third and final type of bias a *sin of omission* (SofO; what Hastie & Rasinski, 1988, call "miss a good cue"). This occurs when the judge fails to use information held to be diagnostic by the idealized model of judgment. For example, many studies have shown that judges frequently fail to use diagnostic base-rate information (see Nisbett & Ross, 1980). Likewise, people also tend to ignore situational constraints when explaining an actor's behavior (the correspondence bias, see Nisbett & Ross, 1980); for example, participants who read a written essay tend to ignore whether or not the writer chose voluntarily to take that position versus was randomly assigned that position when judging the writer's true feelings on the essay's topic (Jones & Harris, 1967).

A sin of omission has occurred when conditions differing on the availability of such useful information fail to produce reliably different judgments (e.g., no difference in judgment between participants given different levels of base-rate information). If we again refer to conditions differing in the availability of this prescribed information as the high and low conditions, then a sin of omission by individual judges requires that $\boldsymbol{p}_H = \boldsymbol{p}_L$ (where $\boldsymbol{p}_H$ and $\boldsymbol{p}_L$ are the distributions of individual judgments in the high and low conditions, respectively). Since it is the absence of an effect that signifies bias for this type of sin, the magnitude of differences in judgment between these conditions (i.e., $b = |\boldsymbol{p}_H - \boldsymbol{p}_L|$) serves to index not the magnitude of bias, but rather a lack of bias. So for a sin of omission to obtain, $b = 0$. Likewise, a sin of omission among groups would mean that $\boldsymbol{P}_H = \boldsymbol{P}_L$ and, thus, that $B = |\boldsymbol{P}_H - \boldsymbol{P}_L|$ should be zero. The reversal from the logic of detecting sins of commission requires that we reverse the terms in the definition of the relative magnitude of a SofO bias. That is, for sins of omission, $RB = b - B$ (instead of $B - b$, as in the SofI and SofC cases). When relative bias for the SofO case is defined in this way, positive values of relative bias still signify that groups are relatively more biased than individuals.[2]

[1] The criterion need not necessarily be a specific point value. For example, the conjunction error (Tversky & Kahneman, 1983) is indicated by a judgment falling anywhere above a maximum value specified by probability theory.

[2] In the text, we have made the simplifying assumption that the larger the difference between the high- and low-bias conditions, the more "accurate" judges are (i.e., the more appropriately they are using the available information). Of course, theoretically it is possible to specify not only that certain information should be used, but precisely how much of an impact such information should have. For example, not only should judges pay attention to base-rate information, but using certain normative models (e.g., Bayes' theorem), it is possible to specify exactly how much impact any particular piece of base-rate information should have. In such a case, it is also possible that bias could be revealed by judges paying too much attention to the information in question as well as too little. In such a case, one could not simply assume that the larger RB was, the more biased groups were relative to individuals. However, this theoretical complication does not negate the thrust of the present analysis (which makes the simplifying assumption that the more one uses prescribed information, the better).

Other Conceptions of Bias

Because theory and research on judgmental bias has grown so explosively over the past two decades, some further taxonomic distinctions will help bound our coverage of the topic.

First, we have excluded from our analysis a fourth logic for demonstrating bias described by Hastie and Rasinski (1988). This logic involves a comparison of the judgments reached by two or more sets of judges (e.g., men and women). In tasks where a single unbiased judgment can be assumed, reliable differences in judgment between such sets of judges implies that at least one of them is inaccurate (i.e., biased). In our present context, this logic can be readily generalized to comparisons of judgments by sets of groups instead of sets of individuals.

This logic may be used to examine what Kaplan and Miller (1978) call *trait biases,* that is, biases attributable to some stable characteristic or disposition of the judge. For example, mock jury studies have compared verdicts reached by sets of jurors classified as high versus low on traits like authoritarianism (Bray & Noble, 1978) or on prior beliefs about the probability that any given rape defendant is guilty (Davis, Spitzer, Nagao, & Stasser, 1978).

This fourth logic can be useful for probing the nature and consequences of traits, but we exclude it from our present analyses of bias because of its inherent inferential ambiguity. The mere fact that judgments by two or more sets of judges are reliably different is typically insufficient to unambiguously establish that nonnormative use of information has occurred. It is often possible, for example, that the judgment processes of the sets of judges differ in ways permitted by or irrelevant to the normative model of judgment. For example, jurors have broad discretion in evaluating the credibility of witnesses. Sex differences in trial verdict could simply reflect sex differences in the perceived credibility of a key witness, an effect that might not violate any normative model of juror judgment. Moreover, even if such problems could be avoided (e.g., by using tasks where the normative model prescribes that and how all available information is used), the locus of bias remains ambiguous under this logic; disagreement between two sets of judges could mean that one, the other, or both are making biased judgments. Thus, in the remainder of this article, we limit our analysis to cases involving the other three Hastie–Rasinski (1988) logics or what we have labeled judgmental sins of imprecision, commission, and omission.

An additional distinction involves group composition: specifically, the degree of homogeneity or heterogeneity in the group. Though group homogeneity often refers to the distribution of a personality or demographic trait across members, for our purposes it is defined with respect to exposure to information. Thus a homogeneous group is one in which each member of a group has been exposed to the same prescribed, proscribed, or neutral information set, although members may nevertheless differ with respect to their attention, encoding, and recall of that information. In a heterogeneous group, members differ quantitatively in their amount of exposure to the stimulus or qualitatively in the particular biasing stimuli to which they have been exposed (e.g., Kameda & Davis, 1990; Tindale, Sheffey, & Scott, 1993). Although we limit our analysis to cases involving homogeneous groups, at the conclusion of the article we briefly examine how heterogeneous grouping might influence the individual–group comparison (see Kerr & Huang, 1986; Tindale & Nagao, 1986).

Furthermore, our analysis does not address other senses in which individual and group decision processes may be more or less biased; for example, with respect to the representation of diverse viewpoints in the community, the perceived fairness of decision rules and the perceived legitimacy of the decision maker's mandate (see MacCoun & Tyler, 1988). We also limit our review and analysis to the potentially moderating effects of discussion in face-to-face small groups. Thus, we exclude the growing body of research in the experimental market paradigm, which examines the effects of simulated market transactions on the rationality of individual choice (see reviews by Camerer, 1992; Plott, 1986). Two characteristics generally distinguish that paradigm from the small-group paradigm examined here: In the former, judges generally make repeated individual choices without explicit group discussion, and they receive feedback on the effects of their choices, although that feedback is often lagged, noisy, and highly interdependent on the influences of other judges and exogenous factors (see the essays in Hogarth, 1990).

Finally, although this article's focus is upon biased use of information, it is important to recognize that there is nothing in our analysis of sins of omission or commission that requires this limitation. That is, our primary question could be rephrased from "are groups any less (or more) subject to SofC or SofO judgmental biases than individuals?" to "are groups any less (or more) likely to use a particular piece of information than individuals?"[3] Our presentation is couched within the framework of biased judgment (i.e., proscribed use of information), but it is worth remembering that the same analysis can be applied with profit to comparing unbiased judgment by individuals and groups. We return to this important point later.

Summary

Three qualitatively different forms of judgmental bias have been distinguished: differences between judgment and a particular judgment prescribed by a normative model (a sin of imprecision); differences in response to the availability of information that should be ignored by judges (a sin of commission); and failures to observe differences in judgment due to the availability of information that should be used by judges (a sin of omission).

In the following section we provide an overview of the relevant empirical literature, cross-categorized by the three types of bias defined above. This taxonomy of biases, per se, does not simply organize that literature. It is not the case, for example, that groups tend generally to be more biased than individuals for one type of bias, but less biased for another. However, distinguishing between these different types of bias is very useful for our subsequent theoretical analyses and for fitting particular empirical findings within those analyses.

[3] Thanks to Reid Hastie for making this important point.

Overview of Relevant Prior Research

Coverage

Several review essays have done an excellent job of comparing individual and group performance on various decision tasks (see Einhorn et al., 1977; Hastie, 1986; Hill, 1987; Laughlin & Ellis, 1986; McGrath, 1984; Vollrath, Sheppard, Hinsz, & Davis, 1989). The general consensus is that, on average, groups outperform individuals on such tasks, although groups typically fall short of the performance of their highest-ability members. These reviews have generally equated the quality of performance with accuracy, defined in terms of the distance between individual or group judgments and a value (e.g., judgments of weight or distance, arithmetic problems), what we have called sins of imprecision (see Hastie, 1986; Hastie & Rasinski, 1988). The accuracy criterion is most applicable to what Laughlin (e.g., Laughlin & Ellis, 1986) calls intellective tasks; that is, tasks where clear criteria exist for evaluating the quality of cognitive performance. But whether a task can be characterized as intellective depends on several factors: the existence of a normative theory of the task, the degree to which knowledge of the theory is shared by group members, and the degree to which the theory, once voiced, is accepted as valid by group members (Laughlin & Ellis, 1986; McGrath, 1984). Where previous reviews have focused primarily on tasks that are unambiguously intellective (e.g., arithmetic problems, deductive brain teasers, simple recognition and recall memory) most of the studies we review fall within a "grey area" marking the transition from pure intellective tasks to pure decision-making (McGrath, 1984) or judgmental (Laughlin & Ellis, 1986) tasks: tasks that have no demonstrably correct answer. Of course, when bias is a sin of commission or omission, it is not necessary that a task be clearly intellective. Bias can be demonstrated without recourse to a correct answer by comparing the performance of decision makers operating under alternative experimental conditions that should or should not, in a normative sense, influence outcomes. Besides their recurrence in the literature, what makes such studies interesting is that, in keeping with the quasi-intellective nature of their tasks, the normative standards for identifying bias may not be readily obvious to all decision makers.

Methodological Caveats

Comparisons of statistical significance levels across levels of analysis—individual versus group—can be hazardous. In many repeated-measures studies, the sample size at the group level of analysis is only $(1/r)$th as large as the individual-level sample size, where r is the group size. This implies that group-level effects will generally be tested at a much lower level of statistical power, and thus reliable individual effects might not be detected at the group level even when of equal or greater magnitude (Kenny, Kashy, & Bolger, in press). Thus, differential statistical power can artifactually make groups appear less biased than individuals with respect to sins of commission, where a null finding implies the absence of bias. It can make groups appear more biased than individuals with respect to sins of omission, where a null finding implies the presence of bias.

Ideally, one might compensate for studies with low statistical power by conducting a meta-analysis of the effects of group discussion on particular judgmental biases, but that goal seemed neither feasible nor appropriate given the paucity of available data. We have been able to locate fewer than 30 different empirical studies that directly examined both individual- and group-level biases. Across these studies, there is little consistency in decision tasks, procedures, group sizes, independent variables, dependent measures, and inferential statistical tests; in such cases, meta-analyses could not only be inappropriate, but quite misleading. Moreover, studies varied in their implementation of the individual–group comparison. In some studies, participants were randomly assigned to either an individual or a group condition in a between-subjects design, while other studies compared prediscussion, group-level, and postdiscussion judgments in a repeated-measures design. Many of these studies failed to report either explicit statistical tests of the individual–group comparison or the information needed to conduct such tests. Finally, our theoretical analysis, presented later, suggests that existing research provides quite spotty coverage of the relevant parameter space; as such, overgeneralization from observed empirical patterns provides a potentially misleading comparison of individual and group bias.

An Overview of Relevant Research

In Table 1, we use our trichotomous bias taxonomy (i.e., sins of commission, omission, or imprecision) to categorize the existing literature on individual versus group bias. We should emphasize that such a categorization is quite broad and almost certainly glosses over important psychological distinctions among judgmental phenomena. Indeed, the taxonomy's immediate purpose is to categorize experimental operations; whether it also categorizes distinct psychological process is an open question we explore throughout the article.

Across the three general types of bias we distinguish 15 judgmental phenomena that (a) seem to produce bias among individuals and (b) have been studied so as to permit some comparison of the relative susceptibility of individuals and groups to that bias. Eight of these can be classified as judgmental sins of commission: framing bias (e.g., Kahneman & Tversky, 1984; Tversky & Kahneman, 1981); preference reversals (i.e., inconsistencies in judgment across alternative ways of obtaining judgments; e.g., Lichtenstein & Slovic, 1971; Tversky, Sattath, & Slovic, 1988); theory-perseverance effects (i.e., overreliance on information that might once have been but is no longer diagnostic; e.g., Anderson, Lepper, & Ross, 1980; Nisbett & Ross, 1980); oversensitivity to irretrievable, "sunk" costs (e.g., Arkes & Blumer, 1985; Brockner & Rubin, 1985); jurors' use of legally irrelevant, extraevidentary information (see Dane & Wrightsman, 1982); nonindependence of judgments by jurors in trials with multiple, joined charges (e.g., Greene & Loftus, 1985); biasing effects on juror judgment of spurious attorney arguments (e.g., Wells, Miene, & Wrightsman, 1985); and the hindsight bias (e.g., Hawkins & Hastie, 1990). Three more phenomena can be classified as judgmental sins of omission: insensitivity to base-rate information (e.g., Kahneman, Slovic, & Tversky, 1982); underuse of situational information when making behavioral attributions, variously termed the dispositional bias, correspondence bias, or the fundamental attribution error (e.g., Jones & Harris, 1967; Nisbett & Ross, 1980); and

Table 1
Classification and Summary of Empirical Literature

Phenomenon	Studies	General effect of discussion
	Sin of commission	
Framing	Tindale et al. (1993)$\varnothing^1$ Kameda & Davis (1990)$\varnothing^1$ McGuire et al. (1987)↑ Paese et al. (1993)= Neale et al. (1986)↓	Mixed: Group discussion amplified bias in McGuire et al., attenuated bias in Neale et al., no effect in Paese et al.
Preference reversal	Mowen & Gentry (1980)↑ Irwin & Davis (1995)↓	Mixed: Groups more susceptible to choice/rank reversals but less susceptible to choice/match reversals than individuals.
Theory-perseverance effect	Wright & Christie (1990)↓$\varnothing^2$	Attenuation: Theory-perseverance effect eliminated in group-discussion and yoked-transcript conditions (but see Note $\varnothing^2$).
Weighing sunk costs	Whyte (1993)↑	Amplification: Groups were more influenced by the existence of past, sunk costs than individuals.
Extraevidentiary bias in juror judgments	Bray, Struckman-Johnson, Osborne, McFarlane, & Scott (1978)$\varnothing^2$ Carretta & Moreland (1983)$\varnothing^2 ?^1$ Hans & Doob (1976)↑ Izzett & Leginski (1974)$\varnothing^{2,3}$ Kaplan & Miller (1978)↓ Kerwin & Shaffer (1994)$\varnothing^4$ Kramer et al. (1990)↑ MacCoun (1990)↑ Thompson et al. (1981)= Zanzola (1977)↑	Mixed: Amplification is more common than attenuation.
Joinder bias in juror judgments	Tanford & Penrod (1984)= Davis et al. (1984)=	Mixed: No clear effect of group discussion.
Biasing effect of spurious attorney arguments	Schumman & Thompson (1989)↑	Amplification: Groups more susceptible than individuals.
Hindsight bias	Stahlberg, Eller, Maass, & Frey (1995)↓	Attenuation: Groups slightly less susceptible than individuals.
	Sin of omission	
Insensitivity to base rates	Argote, Seabright, & Dyer (1986)↑$?^2$ Argote, Devadas, & Melone (1990)↕$?^{2,3}$ Nagao, Tindale, Hinsz, & Davis (1985)↕$?^2$	Mixed: Good evidence that groups rely more heavily on individuating information, but no direct evidence that they rely less on base-rate information (and some to the contrary; see $?^3$).
Dispositional bias in attributions	Wright & Wells (1985)↓ Wittenbaum & Stasser (1995)↓	Attenuation: Appears that group discussion attenuates dispositional bias.
Underuse of consensus information in attributions	Wright et al. (1990)↓	Attenuation: Only group participants were affected by consensus information.
	Sin of imprecision	
Conjunction error	Tindale, Sheffey, & Filkins (1990)↕ Tindale, Filkins, Thomas, & Smith (1993)↕	Mixed: Groups made more conjunction errors than individuals when individual error rates were high, but fewer when individual error rates were low.
Use of representativeness heuristic	Stasson, Ono, Zimmerman, & Davis (1988)↕̸	Amplification? Individuals outperformed groups on one problem; no difference for second problem.
Use of availability heuristic	Stasson et al. (1987)↓	Attenuation?: Groups (especially when unanimous) marginally out-performed individuals.
Overconfidence (miscalibration)	Dunning & Ross (1992)↑ Sniezek & Henry (1989)↓ Plous (1995)=	Mixed: Groups are generally more confident than individuals, but whether this reflects overconfidence varies between studies.

Note. Amplification signifies a stronger bias among groups (or following group discussion) than among individuals (i.e., $RB > 0$). Attenuation signifies a weaker bias among groups (or following group discussion) than among individuals, $RB < 0$. Mixed signifies an inconsistent pattern of findings, such that for certain studies or analyses $RB > 0$, for others $RB < 0$.
↑ signifies that group discussion amplified individual bias.
↓ signifies that group discussion reduced or corrected individual bias.
↕ signifies that there were results indicating that group discussion both amplified and corrected individual bias.
= signifies that the magnitude of bias was comparable for individual and group judges.
$\varnothing$ signifies that although the study employed both individual and group judges and examined the bias phenomenon, the study's results are not informative for assessing the degree of relative bias for one of the following reasons:
$\varnothing^1$ Groups were not homogeneous with respect to exposure to potentially biasing information.
$\varnothing^2$ No clear bias effect for individuals for key dependent variables.
$\varnothing^3$ Bias observed only on dependent variable for which purported biasing information is not normatively proscribed.
$\varnothing^4$ The experimental design did not include a low-bias condition.
Overstruck (e.g., ↕̸) or paired (e.g. $\varnothing$ =) symbols signify combinations of the preceding conditions.
Symbols accompanied by question marks (?) reflect the following methodological or other ambiguities that cloud interpretation of the results:
$?^1$ Results might be attributed to differential power of statistical tests ($df_{error} = 255$ for individual bias tests but $df_{error} = 30$ for group tests).
$?^2$ Groups were more prone to use individuating information than individuals, a result that was interpreted as indicating that groups were also less sensitive to base-rate information. However, if the individuating information is diagnostic, one could alternatively conclude that groups make better use of this diagnostic information.
$?^3$ Access to base-rate information manipulated. When the individuating information was not diagnostic, groups were more likely to use base-rate information; when such information was diagnostic, no reliable effects on relative bias were observed.

underuse of consensus information in attribution (i.e., ignoring information about the proportion of people behaving similarly in a given situation; e.g., Nisbett & Borgida, 1975). Finally, 4 phenomena can be classified as judgmental sins of imprecision: the conjunction error (i.e., when the subjective probability of the conjunction of two events exceeds the minimum of the probabilities of the two isolated events; e.g., Tversky & Kahneman, 1983); use of the representativeness heuristic (i.e., overreliance on some representative or salient, but non-informative, feature of available information; e.g., Kahneman et al., 1982); use of the availability heuristic (i.e., overreliance on information that is readily available; e.g., Kahneman et al., 1982); and overconfidence in own accuracy at all but the most difficult problems (e.g., Lichtenstein, Fischhoff, & Phillips, 1982; Dunning, Griffin, Milojkovic, & Ross, 1990).

In Table 1 we list all those studies that we could identify that made an individual versus group comparison for each of these 15 biases. Unfortunately, certain features of a number of these studies made their results either uninformative or uninterpretable for assessing the degree of relative bias; the nature of these problems are described briefly in the table's footnotes. For each study, the relative degree of bias observed for group versus individual judges is summarized. The table reveals another unfortunate reality: frequently (viz., for 7 of the 15 bias phenomena), there is only one study making the key comparison.

Close inspection of Table 1 does not suggest a simple or coherent picture of the effects of group discussion on biases of judgment. There are several demonstrations that group discussion can attenuate, amplify, or simply reproduce the judgmental biases of individuals. And, although group amplification of bias seems to be the modal result, none of these three patterns appears predominant. Thus, research conducted to date indicates that there is unlikely to be any simple, global answer to the question, "Is group judgment more or less biased than individual judgment?"

In the face of such inconsistent findings, an obvious explanation might be that the effect of group discussion on relative bias is moderated by the nature of the judgmental bias under study. We are convinced that differences among general varieties of bias as well as specific bias phenomena must play an important role in an analysis of relative bias. The remainder of this article is largely devoted to justifying this conviction. Table 1 suggests that our trichotomous bias categorization alone cannot resolve the empirical discrepancies documented in the literature. It is not the case, for example, that groups generally attenuate sins of commission but amplify sins of omission, or vice versa. Rather, for each of these two categories, we find examples of group attenuation and examples of group amplification.

Likewise, comparison and contrast of the studies summarized in Table 1 do not suggest (to us, anyway) any simple task moderators that can organize this diverse literature. For example, two studies of decision framing suggest that groups are even more susceptible to framing than individuals (McGuire et al., 1987; Paese, Bieser, & Tubbs, 1993), yet another finds just the opposite pattern (Neale, Bazerman, Northcraft, & Alperson, 1986). Most studies find that jury deliberation accentuates the effects of extra-evidentiary attributes of trial participants, although one study (Kaplan & Miller, 1978) finds that it attenuates extralegal bias.

No doubt, ad hoc explanations for these discrepancies could be developed, invoking more subtle differences in tasks, procedures, or experimental design. But we believe that a more productive approach would be to start from first principles, building from established and verified theoretical principles regarding the processes by which individual responses are integrated into group judgments. In the following sections, we pursue such a strategy.

A Theoretical Analysis of the Relative Bias of Individuals Versus Groups

Why should groups be any more (or less) susceptible to judgmental biases than individuals? There have been a handful of attempts to provide a theoretical basis for an answer to this question, most in the context of juror versus jury decision making (e.g., Kalven & Zeisel, 1966; Myers & Kaplan, 1976). Most of these imply that bias should be stronger in groups than among individuals (although, see Kaplan & Miller, 1978, for a striking exception).

An Introductory Overview of the Social Decision Scheme (SDS) Model

Our approach to this theoretical problem is to use a formal model that links the product of individual judgment to the product of group judgment. Several such models have been developed specifically for jury decision making (e.g., Gelfand & Solomon, 1974; Klevorick & Rothschild, 1979; Penrod & Hastie, 1980). There are also some more general models that have been applied not just to juries but to other group decision tasks as well (e.g., Davis, 1973, 1980; Hoffman, 1979; Vinokur & Burnstein, 1973). Here we use one particularly influential model of the latter type—Davis' (1973, 1980) social decision scheme (SDS) model. (See Stasser, Kerr, & Davis, 1989, for a general introduction to the SDS model and its progeny; Davis, 1996; Kerr, 1981, 1982; Stasser & Davis, 1981.)

The SDS model suggests that the preferences of group members can be related to group decisions through simple functions, termed social decision schemes. A familiar example is a majority-rules decision scheme, which predicts that the group ultimately settles on the alternative initially favored by a majority of group members. Of course, some groups may not have an absolute majority favoring an alternative at the beginning of deliberation. To deal with such cases, not handled by the primary scheme, one must often posit some subscheme or subschemes (e.g., plurality wins; averaging) along with the primary decision scheme so that all possible distributions of initial preferences are accounted for.

Decision schemes need not be deterministic, predicting one particular group decision with certainty. Rather, they can (and usually are) probabilistic rules. For example, groups occasionally seem to operate under an equiprobability decision scheme for which all alternatives with at least one advocate have an equal chance of being selected as the group decision (e.g., Johnson & Davis, 1972).

Formally, a social decision scheme is a $m \times n$ stochastic matrix D, where n = the number of decision or judgment alternatives and m = the number of possible distributions of r group members across the n decision alternatives. It can be shown (see Davis, 1973) that

$$m = \binom{n+r-1}{r} = \frac{(n+r-1)!}{r!(n-1)!}. \tag{1}$$

For example, a 12-person jury making a guilty–not guilty choice can be distributed in 13 [=(12 + 2 − 1)!/(12! 1!)] possible ways, namely, (12G, 0NG), (11G, 1NG), ... (0G, 12NG). The d_{ij} element of the D matrix specifies the probability that a group beginning deliberation with the ith possible distribution of member preference will ultimately choose the jth decision alternative. Table 2 presents some possible social decision schemes for 11-person groups choosing between 2 alternatives. (We will soon have more to say about these particular decision schemes.)

All that is required to formally link the distribution of individual judgments or decisions, $(p_1, p_2, \ldots, p_n)$, to the distribution of group decisions, $(P_1, P_2, \ldots, P_n)$, is to link individual preference to the possible initial distributions of opinion in groups. If groups are composed randomly, it follows from the multinomial distribution that the probability, π_i, that the group will begin deliberation with the ith possible distribution, $(r_{i1}, r_{i2}, \ldots, r_{in})$, where $(r_{i1} + r_{i2} + \ldots + r_{in} = r)$, is just

$$\pi_i = \binom{r}{r_{i1} r_{i2} \ldots r_{in}} p_1^{r_{i1}} p_2^{r_{i2}} \ldots p_n^{r_{in}}. \tag{2}$$

If these probabilities and the distribution of group decisions are expressed as row vectors, $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_m)$ and $\boldsymbol{P} = (P_1, P_2, \ldots, P_n)$, we may relate the distribution of starting points of group decision making, $\boldsymbol{\pi}$, which Equation 2 shows to be a simple function of the distribution of individual preferences, $\boldsymbol{p}$, to group judgments, $\boldsymbol{P}$, with the following matrix–algebra equation:

$$\boldsymbol{P} = \pi D. \tag{3}$$

As this equation indicates, groups' final decisions depend on two things: (a) where group members begin deliberation, summarized by $\boldsymbol{\pi}$, which depends entirely on individual judgments (see Equation 2) and (b) the processes whereby group members combine their preferences to define a group decision, formally summarized by the social decision scheme matrix, D. The effect of any variable or process that affects the magnitude of relative bias could, in principle, be understood by tracing its effect on where groups begin deliberation, its effect on the process whereby groups reach their decisions, or both. As we shall see shortly, it is useful to distinguish between the simple case in which access to biasing information does not affect the group decision-making process versus where it does. Equation 2 also suggests that if we know how biased individuals are and can make intelligent guesses about the operative social decision scheme, it should be possible to use the SDS model to compare the magnitude of individual and group bias under various conditions of interest (e.g., different types of bias, different-sized groups, different social decision schemes). That is precisely the strategy followed in this article.

Before continuing with our application of the SDS model, we pause to characterize the possible nature of the group processes that are summarized by D.

Individuals and Groups as Information Processors

In this section we explore the question, how does group judgment differ from individual judgment? We attempt to show that every aspect of individual information processing may be al-

Table 2
Alternative Social Decision Schemes

Predeliberation splits		Simple majority		Proportionality		Equiprobability		Truth wins[a]		Strong asymmetry[b]	
G	NG	G	NG	G	NG	G	NG	G	NG	G	NG
11	0	1.00	.00	1.00	.00	1.00	.00	1.00	.00	1.00	.00
10	1	1.00	.00	.91	.09	.50	.50	.00	1.00	.99	.01
9	2	1.00	.00	.82	.18	.50	.50	.00	1.00	.23	.77
8	3	1.00	.00	.73	.27	.50	.50	.00	1.00	.14	.86
7	4	1.00	.00	.64	.36	.50	.50	.00	1.00	.10	.90
6	5	1.00	.00	.55	.45	.50	.50	.00	1.00	.08	.92
5	6	.00	1.00	.45	.55	.50	.50	.00	1.00	.05	.95
4	7	.00	1.00	.36	.64	.50	.50	.00	1.00	.04	.96
3	8	.00	1.00	.27	.73	.50	.50	.00	1.00	.03	.97
2	9	.00	1.00	.18	.82	.50	.50	.00	1.00	.02	.98
1	10	.00	1.00	.09	.91	.50	.50	.00	1.00	.01	.99
0	11	.00	1.00	.00	1.00	.00	1.00	.00	1.00	.00	1.00

Social decision schemes: Simple majority, Proportionality, Equiprobability, Truth wins, Strong asymmetry.

Note. G = guilty; NG = not guilty.
[a] Here, alternative NG is assumed to be "true." [b] Clearly, the asymmetry favors the NG alternative. See Appendix B for information on the function used to generate this D.

tered when groups are making judgments. We hope, thereby, to rectify a common misunderstanding about the nature of social decision schemes and to illustrate the difficulty of precisely specifying D a priori. We then introduce our present analytic approach: to explore the effect of several generic social decision schemes on relative bias.

A common metaphor in cognitive psychology is the human judge as an information processor who is provided with information, processes it in various ways, and outputs a response. A crude schematic model of the individual as information processor (adapted from Hinsz et al., in press) is sketched on the left-hand panel of Figure 1 (Intrapersonal Information Processing). The demands of the judgment task itself provide a context for all stages: They define what is and is not task-relevant information, which intrapersonal cognitive activities can reasonably be seen as task related, and the objective of the judgment task (i.e., task-relevant responses). Between stimulus (information) and response, many intermeshed cognitive activities occur, characterized here (crudely and nonexhaustively) by the processes of attention, encoding, storage, retrieval, and processing (e.g., counterargumentation; see Petty & Cacioppo, 1986) of information. Biased individual information processing is demonstrated by comparing the final response with some idealized criterion (when documenting a sin of imprecision of judgment) or with the responses of other individuals given somewhat different initial information (when documenting sins of commission or omission). In either case however, we can specify how randomly composed groups of size r would begin the process of group judgment (i.e., $\boldsymbol{\pi}$), knowing only the outcome of intraindividual information processing (i.e., $\boldsymbol{p}$). The link from individual judgment, $\boldsymbol{p}$, to the distribution of prediscussion group member judgments, $\boldsymbol{\pi}$, entails only sampling processes, not psychological processes.

When the judge is not an isolated individual but a group of r people, how is the information-processing task altered?[4] Some such changes are represented schematically in the right panel of Figure 1 (Interpersonal/Group Information Processing; see Hinsz et al., in press, and Levine & Resnick, 1993, for a more extensive analysis). First, the demands of the task are broadened. Illustrations of such possible new or altered demands are presented in Appendix A. For example, in the group context, members are often concerned with the task of maintaining or improving interpersonal relationships as well as the task of making a collective judgment (Thibaut & Strickland, 1956). As Janis' (1982) classic work on groupthink indicates, such group task demands can interfere with thorough, accurate judgment in groups. Similarly, group members are likely to be concerned about the impression they create as they contribute (or fail to contribute) to the collective task. In this vein, group discussion or deliberation may vary its style (Hastie, Penrod, & Pennington, 1983), with varying emphasis on thorough exchange and analysis of information versus consistently maintaining and defending one's initial preferences. And, unlike the individual judge, group members must work towards some level of consensus before producing a (collective) response (Miller, 1989; Stasser, Kerr, et al., 1989).

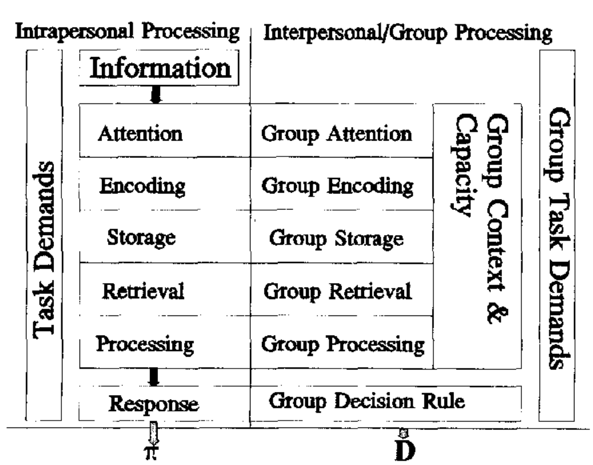


Figure 1. A schematic model of groups as information processors.

Likewise, shifting the judgment task from the individual to the group context potentially may modify every aspect of intraindividual information processing (signified by the regions labeled *Group Attention, Group Encoding,* etc., on the group side of the information-processing box in Figure 1). Some of these modifications stem from certain features of the group performance context; others from the differences between individuals and groups in information processing capacity. Again, Appendix A presents a (nonexhaustive) sampling of some of these potential modifications. For example, attentional processes may be altered because of the distraction created by the group context (Baron, 1986), because there is lowered motivation to attend due to the nonidentifiability of individual contributions to the collective product (Harkins & Szymanski, 1987; Latané, Williams, & Harkins, 1979) or to the possibility that other group members might pick up information that one has missed (Kerr, 1983; Harkins & Petty, 1982), or because the mere presence of other people facilitates simple (and inhibits complex) attentional performance (Zajonc, 1965). The reactions of other group members may also affect how one encodes available information; for example, by priming a schema (Higgins, Rholes, & Jones, 1977) or by providing a socially defined consensus on the meaning of new information (Festinger, 1954). The comments of fellow group members may serve as a cue to assist one's recall of task-relevant information. Obviously, the potential capacity for storage and retrieval of information is greater in the group context (cf. Hartwick, Sheppard, & Davis, 1982), particularly when responsibility for such storage and retrieval is distributed through some type of division of labor (Wegner, 1986). The process of articulating and defending one's position during group discussion may also give group members better access to and awareness of their cognitive processing strategies (much as Ericsson and Simon, 1993, have suggested that individual "talk aloud" or "think aloud" protocols provide more veridical data on cognitive processing than retrospective self reports). On the other hand, the group context can also impede the retrieval of relevant information. For example, Stasser and his colleagues (Stasser & Titus, 1985, 1987; Stasser, Taylor, et al., 1989; Stasser & Stewart, 1992) have shown that information that is unshared

[4] In this discussion, we have drawn liberally from Hinsz et al., (in press). See that article for a focused discussion of related issues.

between group members is (relative to shared information) unlikely to be elicited during group discussion. The group context may also alter the nature of the ultimate processing of available information. Clearly, there are facilitative possibilities, including (a) independent parallel processing of information (cf. Lorge & Solomon, 1955; Taylor, 1954); (b) the elusive "assembly bonus effect" (Tindale, 1992), the combination of different pieces of information that are separately inadequate to produce an accurate judgment but which together make a new, emergent solution possible; (c) fellow members catching and correcting one's errors (Shaw, 1932); (d) random error reduction simply through increasing the number of unbiased judgments being integrated (Zajonc, 1962); (e) for certain tasks, recognition by group members that an argument or position advocated by another group member is self-evidently correct (Laughlin & Ellis, 1986); or (f) the voicing of alternative positions during group deliberation might produce expectancy disconfirmation, which has been shown to undermine judgmental confidence and promote more systematic processing of information (e.g., Maheswaran & Chaiken, 1991). Conversely, some aspects of the group context might impair or further bias processing. For example, recent work on brainstorming (Diehl & Stroebe, 1987) suggests that when other group members are talking, production of one's own ideas is blocked. Or, if the initial, emerging consensus is for an inaccurate judgment, social-comparison processes could derail effective processing.

Finally, even in the rather unlikely event that the r members of the group were to independently process available information in parallel, they would still typically have to resolve differences in judgment to produce a consensual group response. This implicates the full range of social influence processes, from simple conformity to genuine persuasion to accepting a compromise judgment advocated by no group member to acceding to the judgment legitimized by an implicit or explicit decision rule (e.g., majority rules).

The social decision scheme matrix, D, does not (as some have suggested; Myers & Lamm, 1976) simply embody this final, effective social decision rule (i.e., final consensus requirement). Rather, D summarizes the totality of the modifications to information processing resulting from moving from the individual as judge to the group as judge (i.e., to all the processes symbolized on the right panel of Figure 1). Given the current state of knowledge, we cannot even anticipate all possible such modifications, much less specify a priori which such ones may arise and be important for any particular judgmental task. (We will return to a discussion of some such modifications later, though.) What we can do, though, is to identify certain generic social decision schemes that are of theoretical interest, have been shown to accurately summarize the group decision-making process for a sizeable range of interesting tasks, or both. We can then explore theoretically the implications of these Ds for the contrast of bias between individual and group judges.

To simplify our subsequent presentation, we restrict our attention to the simplest possible judgment task: one with only two choice alternatives (i.e., $n = 2$). Because we frequently use jury decision making to illustrate our ideas, we label those two alternatives G and NG (for guilty and not guilty). Although certain interesting processes can arise in cases where the response scale is multichotomous or continuous (Davis, 1996; Kerr, 1992), nearly all of the judgmental biases of interest here can be reduced to the simple dichotomous case by collapsing response categories. We occasionally note when our conclusions need to be qualified by this simplifying assumption.

Alternative Generic Social Decision Schemes

Proportionality D

In this article we focus on four generic social decision schemes. The first, the Proportionality D, is primarily of theoretical interest. This decision scheme assumes that the probability of a particular faction prevailing in the group is equal to the relative frequency of that faction (i.e., $d_{ij} = r_{ij}/r$). The proportionality decision scheme for an 11-person group choosing between two decision alternatives (i.e., $n = 2, r = 11$) is included in Table 2. To our knowledge, no research has ever found that a strict proportionality decision scheme actually provided an accurate summary of group decision making at any task (although one can imagine hypothetical social processes that would result in such a decision scheme.[5]). Nevertheless, two things make this decision scheme interesting and worth considering here: (a) its net effect is to reproduce exactly at the group level those judgments observed at the individual level (i.e., $P = p$ under the proportionality scheme); and (b) it serves as a theoretical boundary between two other classes of decision schemes that do have demonstrated empirical utility.

Majority D

The class of decision scheme for which there is the widest empirical support (see Stasser, Kerr, et al., 1989) is the *majority-wins* social decision scheme, of which the simple majority-wins D is a prototype; the ($n = 2, r = 11$) case is illustrated in Table 2. It has been shown that such a primary decision scheme (or a close relative like a two-thirds majority wins; cf. Davis et al., 1974) accurately summarizes the decision-making process of groups at many different tasks, including attitudinal judgments (Kerr et al., 1976), duplex bets (Davis et al., 1974), and jury decisions (see Davis, 1980, for a review). Laughlin (e.g., Laughlin & Ellis, 1986) has suggested that such a decision scheme generally applies to group decision making at judgmental tasks, which possess no clear criterion for the correctness of decision alternatives. Many aesthetic, political, ethical, and attitudinal judgments are, in this sense, judgmental tasks. The unifying feature of the generic majority social decision schemes is that they all exhibit "strength in numbers." That is, relatively large factions carry disproportionate influence; formally, if MC = a majority criterion (e.g., $MC = 0.5$ for a simple majority-rules scheme; $MC = 0.66$ for a two-thirds majority-rules scheme), then $d_{ij} > (r_{ij}/r)$ for $r_{ij}/r >$ MC. This reflects the underlying logic of Laughlin's hypothesis: when there is no

[5] For example, the proportionality scheme would summarize a group decision-making process in which groups simply endorsed the initial preference of a single, randomly selected member. Slightly less fanciful would be a process wherein each group member participated equally and the group was equally likely to endorse the position advocated in every argument expressed.

objective basis for evaluating the "correctness" or "accuracy" of a judgment (i.e., no widely shared and easily applied evaluative conceptual system), we must often rely on social consensus to define a valid response (cf. Festinger, 1954).

Equiprobability D

If majority-wins decision schemes exhibit disproportionate strength in numbers and if in a proportionality decision scheme faction strength is exactly equal to its proportional size, then we may also envision decision schemes in which there is little or no strength in numbers (e.g., where $d_{ij} < r_{ij}/r$ for relatively large r_{ij}). One such decision scheme is an equiprobability scheme, in which every alternative with at least one advocate is equally likely to become the group's final choice (see Table 2 for an example). Johnson and Davis (1972) and Davis, Hornik, and Hornseth (1970) found evidence that such a decision scheme accurately accounted for group probability matching judgments. Davis (1982) has speculated that this decision scheme might characterize group decision making under high task uncertainty. Kerr (1983) and Laughlin and Ellis (1986) speculate that such a decision process might arise when group members have very little commitment to or investment in their preferences, when maintaining group harmony is vital, or both. In support of the latter conjectures, Kerr (1992) found that as the importance of the issue being discussed by group members declined, so did factions' apparent strength in numbers.

Asymmetric D

Laughlin (1980; Laughlin & Ellis, 1986) has also suggested that for many tasks there is a widely shared consensus on the criteria for the evaluating group decisions. Simple mathematics problems nicely illustrate such intellective tasks; basic mathematical rules provide an objective basis for arguing that one solution is better than another. When certain conditions are met, Laughlin and Ellis (1986) suggest that particular alternatives are demonstrably correct. These conditions are (a) a conceptual evaluative system is shared among group members; (b) there is sufficient information available to the group to discover the "correct" response; (c) group members are able to recognize such a correct solution when it is presented in the group; and (d) any group member or members who favor the "correct" response have the ability, motivation, and time to demonstrate its correctness. The first of these criteria underscores an important point to which we will return. It is certainly possible to judge correctness within an abstract, formal logic with a few axioms, such as judging that a particular proof of a mathematical theorem is correct. However, when it becomes a matter of asserting and defending the correctness of one's judgment to others, "correctness" is largely a social construction. There must be some kind of reasonably clear and widely shared social consensus about what is and is not "correct" (and why) in order for one to convince others that one's preferred judgment is indeed the "correct" one. Moreover, the shared conceptual evaluative system underlying such judgments need not correspond to any particular normative (i.e., logically or empirically correct) system. For example, among Galileo's inquisitors, the assumption that everything in the universe revolved around the earth was clearly "correct," despite the clear empirical evidence he could provide that Jupiter had moons that revolved around it. To avoid confusion between the latter notion of normative correctness and the former, socially defined notion, we will continue to add quotes whenever we mean that an alternative is demonstrably "correct" in Laughlin's sense.

For highly demonstrable tasks, Laughlin has shown (e.g., Laughlin et al., 1976; Laughlin & Ellis, 1986) that all that is required for the group to choose the "correct" alternative is for there to be a single individual who advocates this alternative (a *truth-wins* social decision scheme; see Table 2). When the demonstrability conditions are not as fully met, advocates of the "correct" alternative may require some social support to prevail (a *truth-supported wins* decision scheme; e.g., Laughlin et al., 1975; Laughlin & Earley, 1982).

The distinctive feature of the decision schemes that summarize group judgment at intellective or quasi-intellective tasks is their asymmetry: Factions favoring the "correct" alternative are more likely to prevail than comparable (i.e., equally large) factions favoring an "incorrect" alternative. In order to explore the effects of such asymmetries for the relative bias of groups versus individuals, we constructed a strongly asymmetric decision scheme. The 11-person group version of this D is presented in the far right panel of Table 2 (see Appendix B). As one can see in the table, this decision scheme strongly favors alternative NG, with alternative G prevailing only when the initial support for alternative G is very large. This D characterizes likely group decision-making processes when alternative NG is a highly (although not wholly, as in "truth wins") demonstrably "correct" answer.

Summary

We have identified four generic social decision schemes: proportionality (in which a faction's strength is precisely equal to its relative size), majority-wins (in which large factions' strength is larger than their relative size; i.e., there is strength in numbers), equiprobability (in which large factions' strength is less than their relative size), and asymmetric decisions schemes in which an alternative is demonstrably "correct" within some social context.

The implications of these four generic social decision schemes for the contrast of individual versus group bias can be revealed via *thought experiments* (Davis & Kerr, 1986). Using the SDS model, we can not only compare the effects of different global processes of group judgment or decision making (by comparing and contrasting the generic Ds), but by varying other variables and model parameters (e.g., magnitude of individual bias; group size), we can also explore their effects on the central contrast of interest. Wherever enough information is available for an empirical study cited in Table 1 (e.g., the social decision scheme is estimated directly or can be safely assumed to be similar to one of the generic Ds we consider), we will compare the results of that study to those obtained in our thought experiments.

In carrying out our analysis, two general cases can be distinguished. In the first and simpler case (Case 1), a single, unitary process of group judgment, summarized by a single D matrix, describes all groups. That is, all groups transform initial group member preferences to group judgments utilizing the same basic group decision-making process. Under this case, exposure to potentially biasing information may alter individual preferences (i.e., $\boldsymbol{p}$), but it does not alter the process by which groups forge a collective judgment out of those member preferences (i.e., D itself). In the second case (Case 2), exposure to potentially biasing information again may (or may not) affect the process of individual judgment (and, hence, $\boldsymbol{p}$), but does affect the process of group judgment (i.e., D). We separately consider Case 1 and Case 2 below.

Case 1: Unitary Group-Judgment Process

We have now defined three qualitatively different types of bias and identified several (viz., four) different Ds (summary processes for group information processing or decision making) that are empirically or theoretically interesting. Below, using computer-assisted analyses, we examine the consequences for RB, the relative bias between individuals and groups, of each of the possible combinations of D and bias type (see Grofman, 1978, for a similar, but more limited analysis). Our objective is to map RB across the full domain of possible levels of individual bias and to see how RB may depend upon a variety of factors (such as the type of bias, the idealized standard of unbiased judgment, the magnitude of individual bias, group size, and the process of group decision making).[6]

Sins of Imprecision (SofI)

Proportionality D. As noted above, a proportionality social decision scheme simply reproduces the entire distribution of individual judgments in groups, such that $\boldsymbol{P} = \boldsymbol{p}$. Because groups reach precisely the same decisions as individuals under this decision scheme, there should never be any difference in the magnitude of bias (i.e., $RB = 0$, under all possible conditions, all possible group sizes, all possible ideal Is, etc.). However, this degenerate case is still useful as a theoretical baseline (Davis, 1969) and as means of introducing our method of conducting and plotting the results of our thought experiments.

Because we have restricted our attention to dichotomous judgments, the distribution of judgments by individual judges is fully expressed by p_G, the proportion of individuals favoring the first alternative (since $p_{NG} = 1.0 - p_G$). Our base analyses assumed that group size, r, was 11 (an odd number was chosen to obviate a subscheme for the majority D). Then, for every possible individual behavior (i.e., for any p_G value between 0.0 and 1.0), we calculated the expected distribution of group choices, $\hat{\boldsymbol{P}}_G$, under the assumption of a proportionality group decision-making process by using Equations 1–3 and the proportionality D in Table 2. Of course, for the proportionality decision scheme, because $\hat{P}_G$ always is equal to p_G, the plot of $\hat{P}_G$ as a function of p_G is the simple straight line depicted in Figure 2 and RB would always equal 0 (since $RB = B - b = |\hat{\boldsymbol{P}} - \boldsymbol{I}| - |\boldsymbol{p} - \boldsymbol{I}|$). The same would also be true for any other group size.

Simple-majority D. A rather different functional relationship resulted, though, when the simple-majority D in Table 2 was used in a similar computer-assisted analysis (see Figure 2). As Davis (1973) originally showed, the effect of a (symmetric)

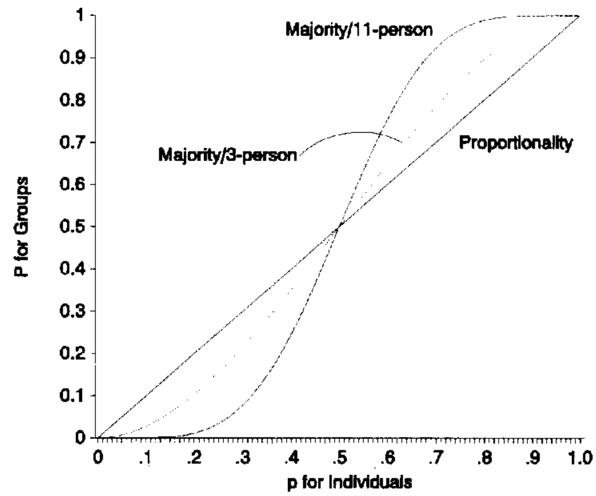


Figure 2. Predicted group judgment, $\hat{P}_G$, as a function of individual judgment, p_G, under proportionality and simple-majority social decision schemes.

group decision-making process that has strength in numbers, which gives large factions disproportional influence, is to make the more popular individual choice even more popular among groups. As Figure 2 shows (also see Davis, 1973; Davis & Kerr, 1986; Kerr & Huang, 1986), if alternative G is preferred by a minority of individuals and if groups follow a simple-majority decision scheme, alternative G is favored by an even smaller proportion of groups. Likewise, if G is preferred by most individuals, an even larger fraction of groups operating under a simple-majority D will endorse alternative G. In short, the popular responses become more popular and the unpopular responses less popular under a majority-wins decision scheme. One implication of this process under most conditions is the group polarization of mean individual preferences (Myers & Lamm, 1976; Kerr, Davis, Atkin, Holt, and Meek, 1975).

This pattern, although perhaps not intuitively obvious, is reflected in familiar experience. A vivid example is the way in which national difference of 6% in individual voters' preference for presidential candidates Bush and Clinton resulted in a 37% difference in popularity in the electoral college; of course, state electors are chosen using a majority-rules (or, more precisely, plurality-rules) decision rule (see Rosenwein & Campbell, 1992). This pattern is easily understood as a direct consequence of a familiar sampling principle: sample statistics are more stable in larger (e.g., group) than smaller (e.g., individual) samples. For example, consider a biased coin that produces "heads" with probability .60 on individual flips. However, suppose that instead of considering individual flips we considered "groups" of 1,000 flips. Now, the probability of getting more heads than tails (i.e., a majority of heads) in such a "group" would be considerably higher than .60 (approximately 1.0, in fact). As sam-

[6] In this article we present the full results of these computer-assisted analyses. Elsewhere (Kerr et al., 1996) we have presented a physical metaphor that can assist one's intuition about the consequences of our theoretical assumptions.

ple size decreases, the sampling error increases and "the proportion of samples getting more heads than tails" statistic declines until it reaches .60 when sample size equals one.

The foregoing clearly suggests that the larger the group is, the stronger the polarizing effect of group decision making. And indeed, when one compares the results of repeating our computer analysis with a smaller-sized group ($r = 3$), one confirms this suggestion (see Figure 2). This turns out to be a very general phenomenon: The effects of the social decision-making process summarized by a particular D matrix are generally more pronounced as the size of the group increases. The implication of this rule for our present discussion is that the magnitude of (any nonzero) RB tends to increase as group size increases (all else being equal), but whether groups or individuals are more susceptible to bias (i.e., the sign of RB) tends not to be affected by variations in group size.

Clearly, because a majority-wins process can produce differences between distributions of individual and group judgment, there is the potential for differences in the relative magnitude of bias exhibited by individuals and such groups (i.e., in RB). However, these effects depend crucially upon the nature of the original individual bias, b, which depends in turn upon what constitutes unbiased judgment (since $b = |\boldsymbol{p} - \boldsymbol{I}|$). To illustrate, we return to our prospect theory example, for which $\boldsymbol{I} = (p_{I,G}, p_{I,NG}) = (.50, .50)$. To simplify, we restrict our attention to the G alternative. That is, rather than calculating and plotting $RB = |B - b| = |\boldsymbol{P} - \boldsymbol{I}| - |\boldsymbol{p} - \boldsymbol{I}|$, we examine the related quantity $RB' = |\hat{P}_G - .5| - |p_G - .5|$. (In the dichotomous-alternative case, it is easy to show that two quantities RB' and RB are strictly proportional to one another (viz., $RB' = RB/\sqrt{2}$); thus, patterns in RB are fully captured by examining RB'.) In Figure 3A we plot the resulting RB' as a function of p_G (here group size = 11, and $\hat{P}_G$ is as predicted by the simple-majority D). As the figure shows, if there is any bias among individuals (i.e., $p_G \neq .5$), under these particular conditions groups would always show a larger bias than individuals (i.e., $RB > 0$; unless $p_G = 0.0$ or 1.0, where floor and ceiling effects prevent groups from being any more biased than individuals).

However, what if the model of unbiased judgment prescribed that alternative NG was the correct, unbiased choice, that is, $\boldsymbol{I} = (p_{I,G}, p_{I,NG}) = (0.0, 1.0)$? For example, in Kahneman and Tversky's (1972) well-known research demonstrating improper use of sample size information, we might force participants to choose between Hospital G (with 45 births per day) and Hospital NG (with 15 births per day) in response to the question posed by Kahneman and Tversky (viz., "Over the course of a year, which of these two hospitals will have more days on which more than 60% of the births were boys?"); because there is greater sampling error with a smaller sample, the normatively correct answer is Hospital NG. As Figure 3B shows, if most individuals get this question correct ($p_G < .50$) and 11-person groups operate under a simple-majority decision scheme, even more groups will make the correct judgment (i.e., $RB' > 0$), but if most individuals are wrong ($p_G > .50$), an even larger fraction of groups will be mistaken (i.e., $RB' < 0$). Given the symmetry in the simple-majority decision rule, precisely the opposite function would result had alternative G been the "correct" response (see Figure 3C).

If the idealized, nonbiased standard of judgment were not one of these standard possibilities (i.e., all for G, all for NG, or indifference between G and NG), the picture can become more complicated. For example, we also plotted the RB' versus p_G function when unbiased behavior requires individuals to favor alternative G 75% of the time, that is $\boldsymbol{I} = (.75, .25)$.[7] As Figure 3D shows, when individuals are relatively highly biased ($p_G < .5$), groups operating under the simple-majority decision scheme tend to exacerbate this bias, as in all the previous cases we have considered. However, if individual performance were less biased (i.e., $p_G > .5$), we see that the result of a majority rule is a complex function, with groups reducing bias in one region ($.5 \leq p_G \leq .67$) and increasing it in another ($p_G > .67$).

The moral of these stories should be clear: As far as sins of imprecision are concerned, a group decision-making process that gives disproportionate weight to numerically large factions (like the simple majority-rules D) does not have a single, simple effect on the relative magnitude of group versus individual bias. Generally, good (i.e., unbiased) individual performance results in even better group performance, whereas poor (biased) individual performance tends to be reflected in even poorer group performance. However, RB will also depend (systematically) upon how ideal–unbiased responding is defined, how biased individuals are, and how large the group is. This conclusion is not simply a theoretical curiosity. Strength-in-numbers group decision-making processes seem to characterize a wide domain of judgmental tasks. Theoretically, this should occur any time the normatively unbiased–ideal (or any other) response is not demonstrably correct (Laughlin & Ellis, 1986); this is a situation that seems likely to characterize many of the complex information processing tasks for which individual judgmental biases have been demonstrated. And, as noted above, various majority-wins Ds have been empirically validated more often and for more tasks than any other general type of social decision scheme.

Equiprobability. The net effect of an equiprobability social decision process (in which there is no strength in (nonunanimous) numbers) is precisely opposite to that produced by the majority process we have just considered (in which there is disproportionate strength in numbers; Davis, 1973). As we have seen, the latter tends to enhance among groups individual preference differences among individuals; the former tends to smoothe out among groups any individual preference differences among groups. This is evident in the plots of $\hat{P}_G$ versus p_G

[7] One has to be a bit creative (and perhaps dogmatic about what would constitute an unbiased judgment) to illustrate this case for a dichotomous judgment. For example, in a probability matching task (Zajonc, Wolosin, Wolosin, & Sherman, 1968) in which Light G is lit 75% of the time and Light NG is lit 25% of the time, one (arguable) definition of unbiased behavior would prescribe that the judge choose G 75% of the time and NG 25% of the time. However, with a multialternative response scale, one can easily identify judgmental tasks where the ideal–unbiased judgment takes on some particular intermediate value (e.g., a posterior-odds judgment prescribed by Bayes's theorem). The important point is that the same complex variation in RB modeled here for the theoretically precise and tractable dichotomous case would also obtain in the multialternative case under a majority-rules D and reasonable distributional assumptions (see Footnote 11; Kerr et al., 1975).

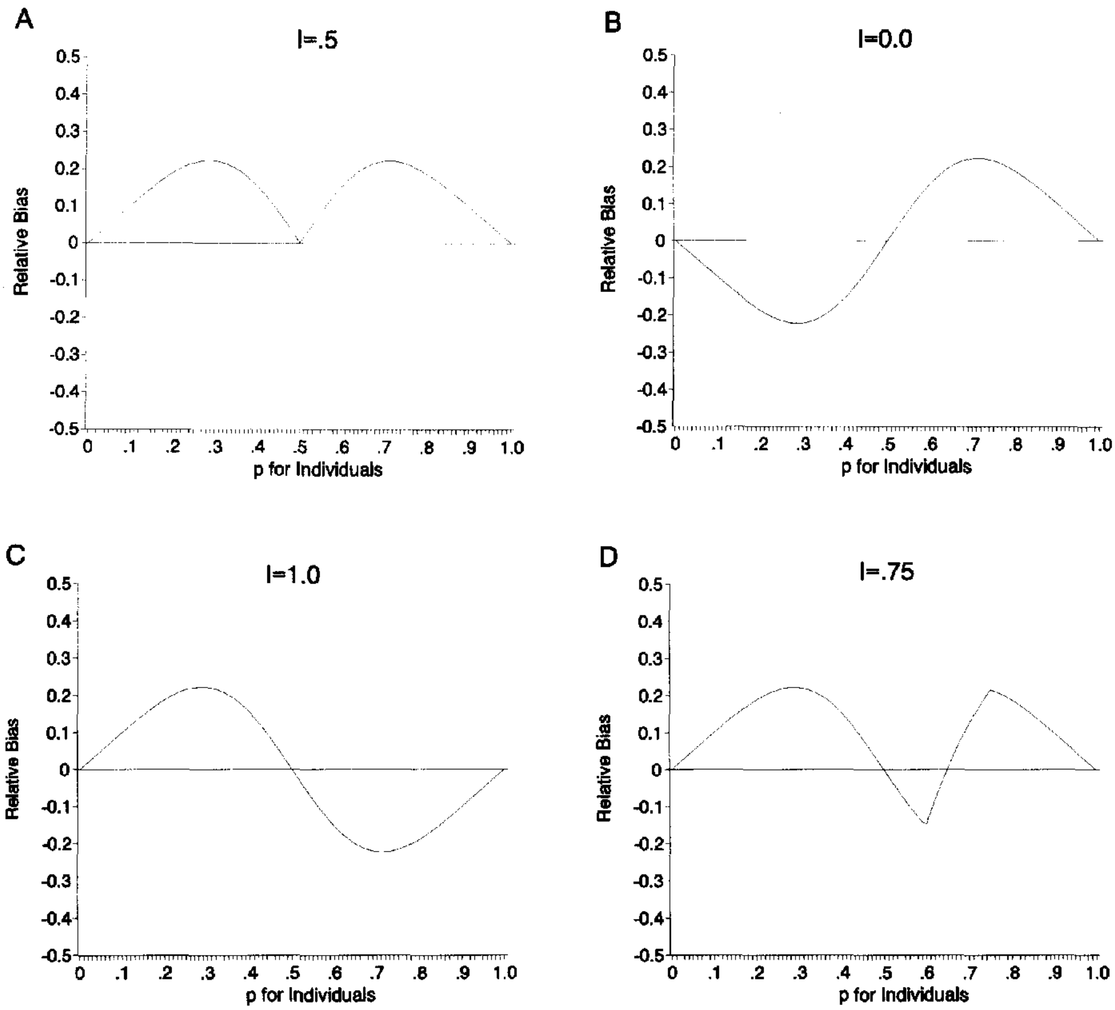


Figure 3. Plots of RB' versus p_G under a simple-majority D when $I = (p_{I,G}, p_{I,NG})$ is (a) (.50, .50), (b) (0, 1.0), (c) (1.0, 0), and (d) (.75, .25). RB' = relative bias; p_G = assumed proportion of individuals preferring G; D = the social decision scheme matrix; I = ideal criterion distribution; $p_{I,G}$ = ideal proportion of individual favoring G; $p_{I,NG}$ = ideal proportion of individuals favoring NG.

for 11- and 3-person groups operating under a strict equiprobability D (see Figure 4A). Relative to the proportionality baseline, these curves are qualitatively mirror images of the corresponding curves for majority rules (i.e., they reverse the qualitative effect of grouping; compare Figure 4A with Figure 2). No matter what individuals tend to prefer, the equiprobability decision process tends to pull groups towards a position of uniform preference across alternatives; here, towards (.5, .5). Once again, the larger the groups, the stronger this tendency.

Given the essential symmetry of the effects of majority and equiprobability decision schemes (around the impactless proportionality baseline), it should not be surprising that the results of the computer analyses deriving RB' under the equiprobability decision scheme produce precisely the opposite patterns to those we just observed for majority rules. So, the general tendency for majority rules to enhance bias in groups is reflected by a corresponding tendency for equiprobability to suppress bias in groups; exceptions to the former rule are (inverted) exceptions to the latter (compare Figure 3A with 4B and 3D with 4C).

It is not the widespread applicability of the equiprobability process that makes these results interesting. Indeed, as noted previously, this particular group decision-making process has only been empirically confirmed for a few rather unusual judgment tasks. What these analyses do demonstrate, though, is that even if many other important features of the judgment setting were held constant (such as judgment task, ideal–unbiased criterion, group size), one can reach exactly the opposite conclusions about the relative degree of bias of groups and individuals under different group decision-making processes. So, for example, if group members tenaciously defend their initial preferences and finally wear down opposing minority views (one type of social process consistent with majority wins), the prospect theory sin of imprecision we have used as a running example

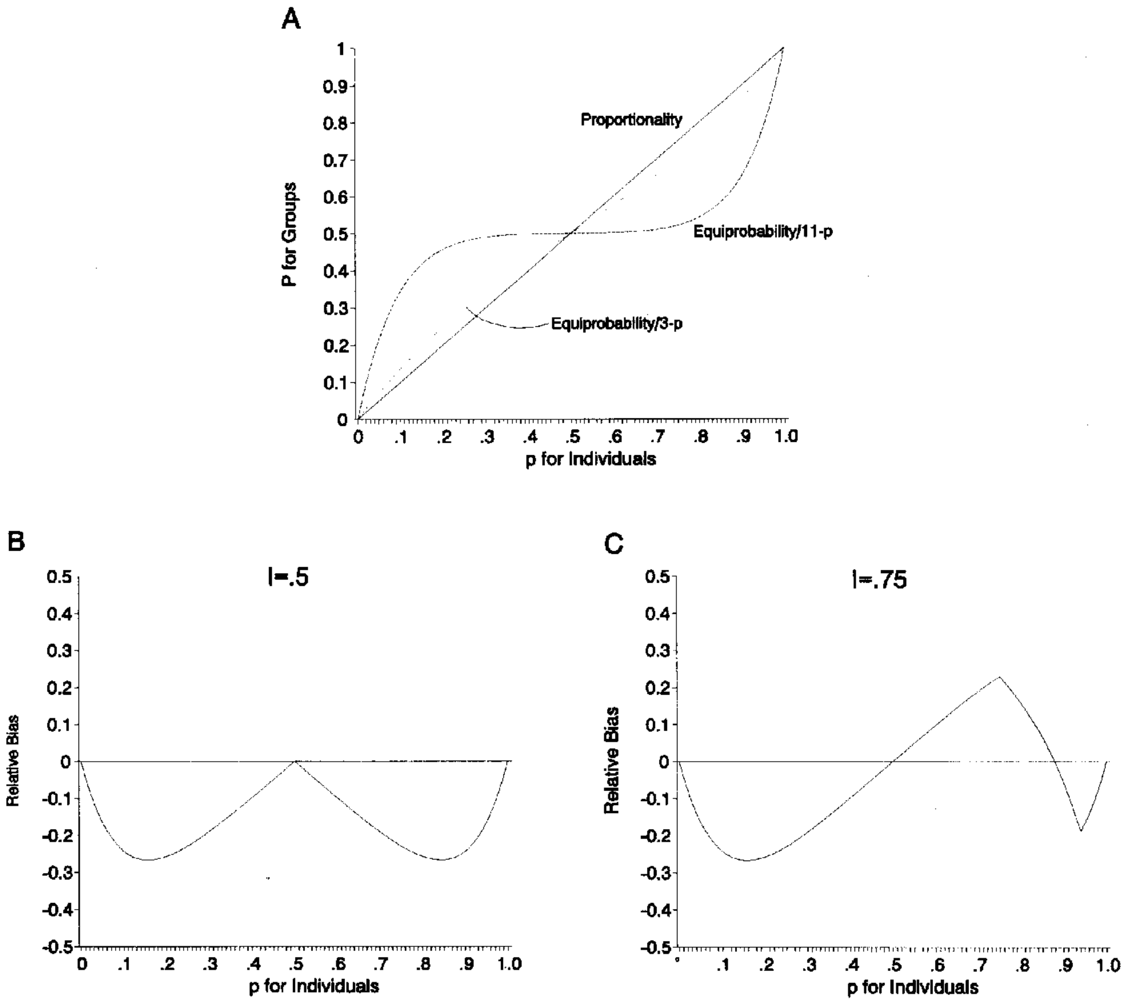


Figure 4. Plots of (a) $\hat{P}_G$ versus p_G, and RB' versus p_G when $I = (p_{I,G}, p_{I,NG})$ is (b) (.50, .50) and (c) (.75, .25) under an equiprobability D. $\hat{P}_G$ = predicted group judgment; p_G = assumed proportion of individual favoring G; RB' = relative bias; I = ideal criterion distribution; $p_{I,G}$ = ideal proportion of individuals favoring G; $p_{I,NG}$ = ideal proportion of individuals favoring NG; D = the social decision scheme matrix; 11-p = 11-person; 3-p = 3-person.

would generally be stronger among groups than individuals (Figure 3A). But if group members were (for whatever reason) uncommonly eager to accommodate opposing viewpoints, even to the point of entirely disregarding which viewpoints had many and which had few advocates (consistent with the equiprobability scheme), then groups should show a weaker bias than individuals (Figure 4B). The clear and crucial point is that *no conclusion about the relative bias of individuals and groups can be reached without careful specification of the operative group decision-making process.*

Strongly asymmetric D. Suppose in our running prospect-theory example that alternative G is the certain-loss alternative and alternative NG is the uncertain-loss alternative (with equal expected value). An expected-utility model would suggest that the ideal, unbiased, and correct choice would be indifference between these two alternatives. Further suppose that the uncertain NG alternative was highly (although not completely) "demonstrably correct" in the sense that Laughlin (e.g., Laughlin & Ellis, 1986) has used the term. This means that it would take very few risk-seeking group members to insure that the group as a whole would opt for the risk-seeking response (i.e., NG); this would be the effect of the group decision-making process embodied in the strongly asymmetric D matrix presented in Table 2.

It is very important to stress again that Laughlin's (e.g., Laughlin & Ellis, 1986) shared conceptual system for evaluation of alternatives that is the basis for the latter type of "correctness" can be but need not be the same as the normative model of judgment that underlies the psychologist's or experimenter's identification of the ideal–unbiased–correct choice (expected-utility theory, in the present case). These can be two wholly separate conceptual systems. The important features of the conceptual system of concern to Laughlin, which we might term the *functional model of judgment*, is that it is widely shared and accepted in the population of judges, and that it is appealed to by and will be persuasive to members of that population.[8] This may or may not be the same conceptual system shared, accepted, and revered by logicians, statisticians, game theorists, etc. (i.e., the normative model). To illustrate:

1. Although it would clearly be incorrect in the normative model that is English common law for a juror or a judge to treat a defendant's religion as evidence of guilt, such a bias might have been demonstrably "correct" in a courtroom in Nazi Germany.

2. It is hard to think of a logical reason for characterizing one of the following football options as normatively correct: Should we kick for the extra point and settle for a tie (choice G) or go for the risky two-point conversion to try and win the game (choice NG)? But, Laughlin and Earley (1982) provide evidence that within a widely shared, conceptual belief system (in which the essential—only?—point of competitive sports is to win), the go-for-the-win alternative is "demonstrably correct."

3. Shafir, Simonson, and Tversky (1993) recently argued that "the axioms of rational choice act as compelling arguments, or reasons, for making a particular decision *when their applicability has been detected,* not as universal laws that constrain people's choices" (p. 34; emphasis added). This suggests that in those studies where groups fail to show closer adherence to the rational-choice normative model (e.g., McGuire et al., 1987; Mowen & Gentry, 1980), these axioms either fail to be voiced, do not appear applicable, or do not appear compelling (i.e., are not the axioms of people's functional model).

Our first computer-assisted analysis using the strongly asymmetric decision scheme revealed that this D's net effect is qualitatively much like a majority-rules decision scheme, but with a quantitative exception. If you feel comfortable with the graphical results of the analyses, examine Figure 5A (and compare with Figure 2) and note that the inflection point of the function relating group to individual behavior (i.e., the function of $\hat{P}_G$ as a function of p_G) is displaced to the right (see Figure 5A), toward the pole that is incorrect under the functional model. If you prefer to think in metaphorical terms, the strong asymmetry basically acts like a strong pull toward the favored (here, NG) option, a pull that cannot be resisted in the group unless nearly all group members initially prefer the "incorrect" alternative (in the present instance, unless p_G begins to approach 1.0; see Kerr et al., 1996, for a further development of this metaphor).

The net results on RB' of introducing such an asymmetry likewise produces a similar distortion of the corresponding majority-wins function (e.g., compare Figure 5B with Figure 3D). The most interesting patterns emerge for the cases where the normative and functional conceptual systems are mutually reinforcing (or identical) versus mutually opposed to one another. Illustrating the former case, suppose the ideal–unbiased response is to choose the NG alternative, that is, $\boldsymbol{I} = (p_{I,G}, p_{I,NG}) = (0.0, 1.0)$. The extremely strong pull toward the NG alternative exerted by the strongly asymmetric D insures that groups are practically always closer to the ideal than individuals (i.e., $RB' < 0$; see Figure 5C); the rare exception occurs when practically no individuals favor the demonstrably correct alternative (see the tiny region where $RB > 0$ when $p_G \approx 1.0$, Figure 5C). This strongly suggests that when the judges subscribe to the same (or a functionally identical) logic as the experimenter, that is, when the participants' and the experimenter's conceptual systems coincide, groups should usually be much less biased (relative to the experimenter's criterion) than individuals.

On the other hand, a functional model's strong pull would insure that groups would practically always be farther from the ideal if it were the normatively defined *in*correct response that was, under that functional model, demonstrably correct, that is, if $\boldsymbol{I} = (p_{I,G}, p_{I,NG}) = (1.0, 0.0)$. (See Figure 5D.) Again, the only time the group could manage to overcome this pull would be when there is practically no one who favors the "demonstrably correct" alternative. When the judges' personal logic is functionally opposite to the experimenter's, groups should usually be much more biased (relative to the experimenter's criterion) than individuals.

One study cited in Table 1 provides a nice empirical illustration of the latter possibility. Tindale, Sheffey, and Filkins (1990) identified the number of persons in each of their 4-person groups who did not and did exhibit a conjunction error in an individual pretest and then determined whether the group itself committed such an error. In essence, this permitted Tindale et al. to estimate the operative D matrix to link individual performance, $\boldsymbol{p} = (p_{correct}, p_{error})$, to group performance, $\boldsymbol{P} = (P_{correct}, P_{error})$. This D matrix estimate is reproduced in Table 3. It is clear that there is a strong asymmetry in the matrix, which indicates that the normatively incorrect alternative (committing a conjunction error) exercises a strong functional pull in the groups. As we have just shown, when a functional model operates in opposition to the normative model, group discussion should typically exacerbate individual bias. And this was precisely the result Tindale et al. obtained.

Summary. When bias is defined as a sin of imprecision, RB has been shown to depend on a number of factors: what behavior is identified as ideal, on the degree of bias among individuals, (quantitatively) on the size of the group, and on the process whereby group judgments are reached. Examining the four generic social decision schemes, few generalizations about relative bias have been shown that hold across all (or even most) levels of the remaining factors. In fact, about the only such generalizations to emerge from our analyses arose when there was a strong asymmetry in the group decision process. When the asymmetry favors the choice prescribed by the normative model (i.e., the normative and functional models coincide), bias is nearly always lower among groups than individuals. However, when the asymmetry favors the other, normatively proscribed choice, bias is nearly always higher among groups than individuals. It has also been shown that (a) all other things being equal, processes that exhibit more-than-proportionate strength in numbers (e.g., majority wins) and those exhibiting less-than-proportionate strength in numbers (e.g., equiprobability) tend to lead to precisely opposite conclusions about relative bias; and (b) the direction of the difference in bias between individuals and groups is generally unaffected by group size, but this difference tends to get larger as group size increases.

[8] Actually, there need not be just a single, unitary functional model. It is quite possible that there could be more than one such functional model. Their combined effects would be summarized by a D that would most likely be asymmetric. It is the nonexistence of any such functional model or models that is indicative of purely judgmental tasks, typically summarized by a symmetric D with high strength in numbers.

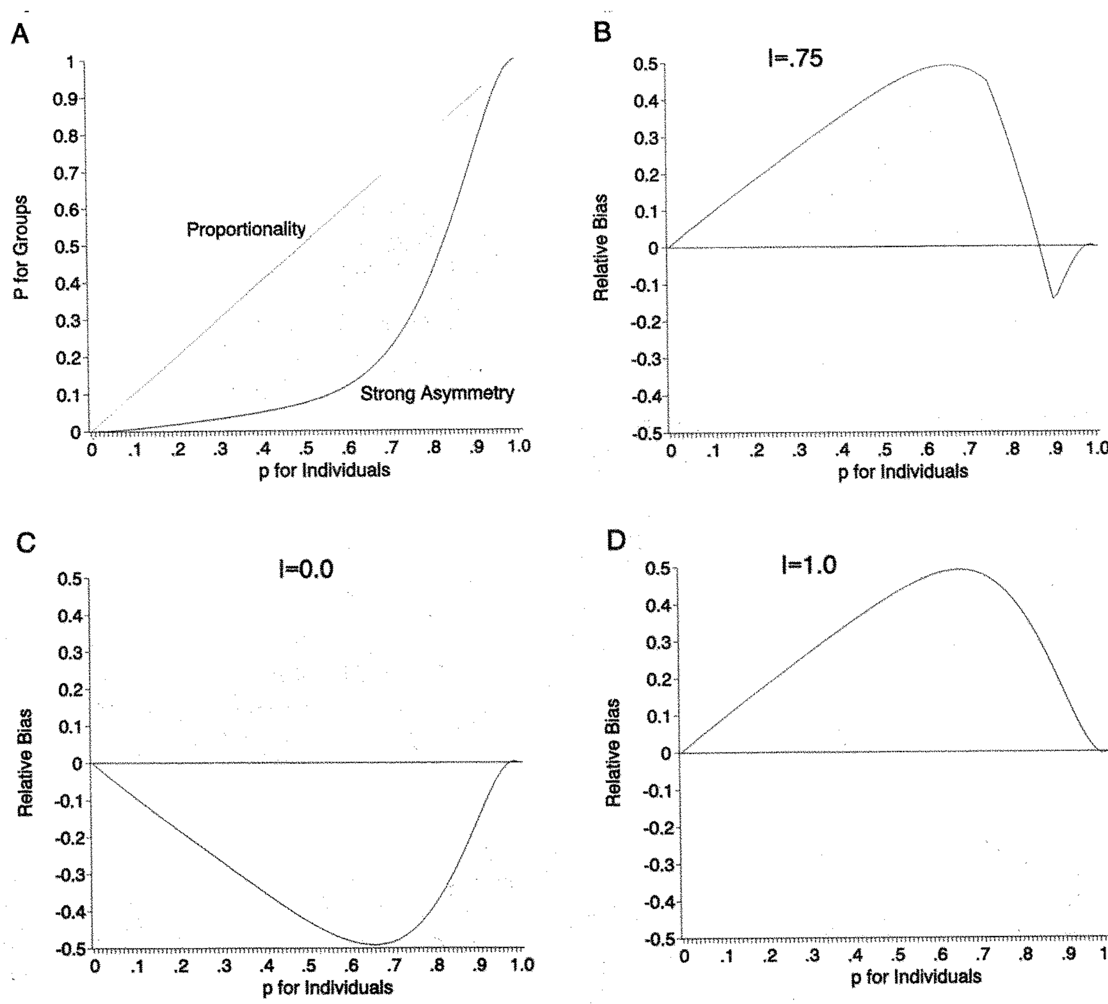


Figure 5. Plots of (a) $\hat{P}_G$ versus p_G, and RB' versus p_G when $I = (p_{I,G}, p_{I,NG})$ is (b) (.75, .25), (c) (0.0, 1.0), and (d) (1.0, 0.0) under the strongly asymmetric D. $\hat{P}_G$ = predicted group judgment; p_G = assumed proportion of individuals preferring G; RB' = relative bias; I = ideal criterion distribution; $p_{I,G}$ = ideal proportion of individuals favoring G; $p_{I,NG}$ = ideal proportion of individuals favoring NG; D = the social decision scheme matrix.

Sins of Commission (SofC)

As our review of the literature suggested, empirical comparisons of individuals and groups for judgmental sins of imprecision are not commonplace. This probably reflects the fact that unambiguously demonstrating such a bias is very difficult. To do so requires (a) a normative model sufficient to make a point prediction (e.g., Meehl, 1990); (b) an experimental paradigm that effectively controls all other sources of systematic judgmental error; and (c) a response scale that meets what may be severe psychometric requirements (e.g., interval or even-ratio level of measurement). Empirical demonstrations of sins of commission (and their converse, sins of omission) are far easier, and hence, more commonplace. The normative model need only make an ordinal prediction, other sources of systematic error can (with care) be equated in the high and low bias conditions, and valid conclusions may still be drawn without crawling out on shaky psychometric limbs. For these reasons, the exploration of the effect of individual bias, group size, etc., on RB in the SofC (and SofO) cases is more relevant to understanding existing (and potential) contrasts of individual and group judgment than the preceding analysis of the SofI case. However, the SofI analyses do provide a foundation for understanding these more typically studied judgmental biases.

Proportionality D. Once again, the proportionality decision scheme is most useful as a baseline (against which to compare the other Ds) and as a way of introducing our method of presentation for the computer-assisted analyses. In the judgmental sin of commission, judges use proscribed information. Because this type of bias involves comparing two groups of judges (which differ with respect to the availability or level of biasing information), there are now two "degrees of freedom" in the domain of possible individual bias effects: the popularity of the G alternative in the high-bias condition ($p_{H,G}$) and the corre-

Table 3
Estimated D Matrix From Tindale, Sheffey, and Filkins (1990; as Cited in Tindale, 1993)

Prediscussion splits			Proportion of groups avoiding and committing the conjunction error	
Correct	Error	N[a]	Correct	Error
4	0	16	.63	.37
3	1	65	.42	.58
2	2	166	.31	.69
1	3	189	.27	.73
0	4	131	.10	.90

Note. [a] Number of instances in which groups began discussion with this prediscussion split.

sponding value within the low-bias condition ($p_{L,G}$). Without any loss of generality (and in conformity to our labels), we assume that bias at the individual level means not just that $p_{H,G} \neq p_{L,G}$, but more specifically, that $p_{H,G} > p_{L,G}$. For example, suppose we are interested in whether exposure to incriminating pretrial publicity biases jurors (Carroll et al., 1986). The normative model of unbiased–ideal judgment would assert that after usual legal precautions have been taken (e.g., careful juror selection, judicial instructions to disregard information obtained outside of the courtroom), there should be no difference in conviction rate between those exposed to incriminating publicity—high- (potential) bias condition—and those exposed to no or nonincriminating publicity—low-bias condition, that is, $p_{H,G} = p_{L,G}$—whereas an elevated conviction rate in the former condition (i.e., $p_{H,G} > p_{L,G}$) would indicate that such information biased juror judgment. Again, we want to explore what happens to relative bias, *RB*, across the full domain of possible individual degrees of bias: in the SofC case, for all ($p_{H,G}$, $p_{L,G}$) such that $p_{H,G} > p_{L,G}$.

In Figure 6 we plot the individual bias surface, $b = p_{H,G} - p_{L,G}$ (>0 since, by assumption $p_{H,G} > p_{L,G}$). This surface is the baseline from which comparisons of bias in groups, B, is calculated (since $RB = B - b$). For any particular, possible behavior in the high- and low-bias individual conditions, that is, for any particular choice of ($p_{H,G}$, $p_{L,G}$), and any particular group decision-making process (assumed here in Case 1 to be a constant D), it is possible to predict, using the SDS model, what the corresponding behaviors would be in high- and low-bias groups, that is, $(P_{H,G}, P_{L,G}) = (\pi_H D, \pi_L D)$, where π_H and π_L are calculated from ($p_{H,G}$, $p_{L,G}$) using Equation 2. Thus, under the assumptions of our model, we can predict how biased groups should be for any particular pattern of individual bias.

The results of such calculations for the proportionality D are simple. Because this decision scheme simply reproduces precisely among groups what happens among individuals, $B = |P_{H,G} - P_{L,G}| = |p_{H,G} - p_{L,G}| = b$ and $RB = 0$; groups and individuals would necessarily be equally biased.

Simple-majority D. The earlier analysis of SofI suggested that a simple-majority decision process often exacerbated bias in groups, but that for certain ideal–criterion values, the opposite could also occur. A qualitatively similar conclusion emerges from our analyses of the SofC case. The RB' surface (with $r = 11$) is plotted in Figure 7A.[9] Of major interest, of course, is the departure of RB' values from 0; here, from the horizontal plane at $RB' = 0$ (see Figure 7B), groups are more biased than individuals where the surface is above this plane, and are less biased where the surface dips below this plane. This is a bit hard to see in the three-dimensional (3-D) plot, so in Figures 7C and 7D, we have plotted the intersection of the surface with the no-individual-group-difference plane (i.e., we plot the $RB' = 0$ contour function). In Figure 7C the contour function is displayed with the same orientation as the previous plots. Figure 7D is a two-dimensional (2-D) version of Figure 7C (imagine grasping the plane in Figure 7C by the edges and tilting it up; or imagine looking down on Figure 7C from above). This intersecting curve divides up the domain of possible individual behavior into regions where groups are more biased ($RB' > 0$ and $B > b$) and groups are less biased ($RB' < 0$, $b > B$) than individuals. As Figures 7C or 7D show, for the largest part of the domain of possible individual behaviors, the polarizing majority-rules process results in greater bias in groups.[10] The exceptions arise when all individuals, in both the high- and low-bias conditions, get too close to either pole. The polarizing effect of the majority-wins scheme may be likened to strong, symmetric pulls at both ends of the response scale (cf. Kerr, MacCoun, & Kramer, 1996). When individual preferences get somewhat extreme (i.e., p_G begins to approach either 0 or 1), group preferences get extremely extreme (i.e., P_G gets very close to 0 or 1; see Figure 2). Although hardly as catastrophic, this situation is a bit like wandering too close to a black hole: You get pulled in quickly and flattened at the event horizon (e.g., Hawking, 1988). If the high- and low-bias conditions, separated by a tangible distance, b, both get too near either pole, both conditions likewise get pulled (polarized) and their tangible b gets flattened to a less tangible B.[11]

[9] An identical analysis was run assuming a 3-person group. The resulting surface was precisely the same shape, but simply compressed on the vertical axis; that is, the magnitude of individual group differences were smaller, but qualitatively the same as in the analysis for 11-person groups.

[10] The more balanced the case being considered by the jury (i.e., the closer the overall conviction rate is to .5), the greater the exaggeration of bias in groups relative to individuals (MacCoun, 1990). In this sense, at least, Kalven and Zeisel's (1966) speculation that extralegal bias is "liberated" in groups considering very close cases is nicely confirmed.

[11] The simple, dichotomous-choice situation we have been considering has a hidden but crucial characteristic: The extremity of individual preference (e.g., proportion convicting) is negatively correlated with the skewness of the distribution. The net effect of applying a majority rule is to increase (among groups) the popularity of the modal position and to "pull in the tails" of the distribution (see Davis, 1973). If the distribution of individual preference is skewed, then this tends to move the mean closer to the mode of the distribution. It is very common, particularly for bipolar response dimensions, for individual opinion to "tail off" (i.e., be skewed) in the direction opposite to the generally preferred position. It follows that a majority-rules scheme predicts group polarization in such cases (Kerr et al., 1975), producing the patterns that we have shown in the text. However, under other possible (but less common) distributional assumptions, we would expect neither group polarization nor an exaggeration of bias in groups. For example, if the high- and low-bias distributions of individual opinion were both sym-

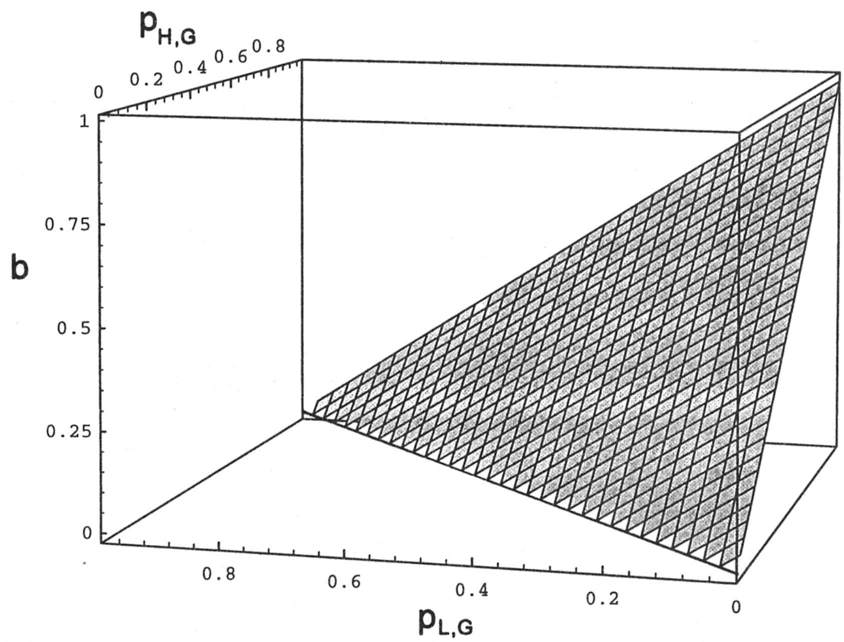


Figure 6. Plot of individual degree of bias, b, for a sin of commission as a function of the probability of individuals favoring the guilty (G) alternative in the high-bias and low-bias conditions (i.e., $p_{H,G}$ and $p_{L,G}$). $p_{H,G}$ = proportion of individuals favoring G in the high-bias condition; $p_{L,G}$ = proportion of individuals favoring G in the low-bias condition.

One thing this analysis teaches us is that whether groups are more or less biased than individuals does not only depend upon how biased individuals are (i.e., on the magnitude of b), but also upon the underlying base rate of behavior. For example, suppose that individual jurors exhibited a bias of 10% due to some extra-legal factor (e.g., jurors exposed to prejudicial pretrial publicity show a conviction rate 10% higher than those not so exposed). The current analysis shows that under the majority-wins decision scheme, this effect would be larger in juries if the overall trial evidence were fairly balanced (e.g., the condition means were 45% and 55%), but the same 10% effect among jurors would be attenuated within juries if the trial evidence were lopsided (e.g., the condition means were 5% vs. 15%). Thus, even if different investigators were examining precisely the same bias phenomenon using generally similar research paradigms, one could get completely opposite findings for the degree of relative bias with sufficiently different overall response base rates.

metric, then application of a simple-majority rule will reduce variability in each but will not produce any shift in the means. Similarly, even if both the high- and low-bias distributions were skewed, if they differed only in central tendency (viz., if they were identically skewed), then a majority-rules scheme would produce identical shifts and no net difference in the biasing effect for individuals and groups. And if extremity and skew were positively related (a possible but rather unusual pattern), a majority-wins process would produce depolarization and would generally attenuate bias in groups. So for more complex, multialternative response scales than we focus on in our current analyses, RB will depend on other features (e.g., skewness) besides the mean of the distribution of individual behavior.

One might attribute the "compression" of sins of commission that occur near the poles of the judgment dimension as a type of floor or ceiling effect: The polarizing effect of the majority-wins process should compress effects that begin near the poles of a bounded response scale. But it would be a mistake to dismiss this finding as simply a rare and easily recognized exception to the general rule. This is most strikingly illustrated by the clearest empirical demonstration that juries can attenuate juror bias, Kaplan and Miller (1978). Kaplan and Miller (1978, Study 3) examined the biasing effect of the nonevidentiary behavioral style of various people involved in presenting the case to jurors. In the part of the design of most direct interest to us, Kaplan and Miller contrasted a condition in which the defense attorney in a reenactment of an attempted manslaughter trial acted in a delaying, obnoxious manner versus a second condition in which it was the prosecutor who acted obnoxiously. The legally relevant and material evidence presented to the mock jurors was identical in both conditions. The normative model of judgment would prescribe that the jurors ignore the legally irrelevant factor of advocate's obnoxiousness when making the central judgment of guilt or innocence. However, Kaplan and Miller confirmed a judgmental sin of commission among mock jurors: Prior to jury deliberation the defendant was judged to be guiltier when his attorney acted obnoxiously than when the prosecutor acted obnoxiously. Kaplan and Miller also manipulated the overall strength of evidence against the defendant and, as one would expect, jurors were more likely to convict when the evidence against the defendant was stronger. Jurors then deliberated the case for 10 minutes in 12-person juries and then provided postdeliberation guilt judgments

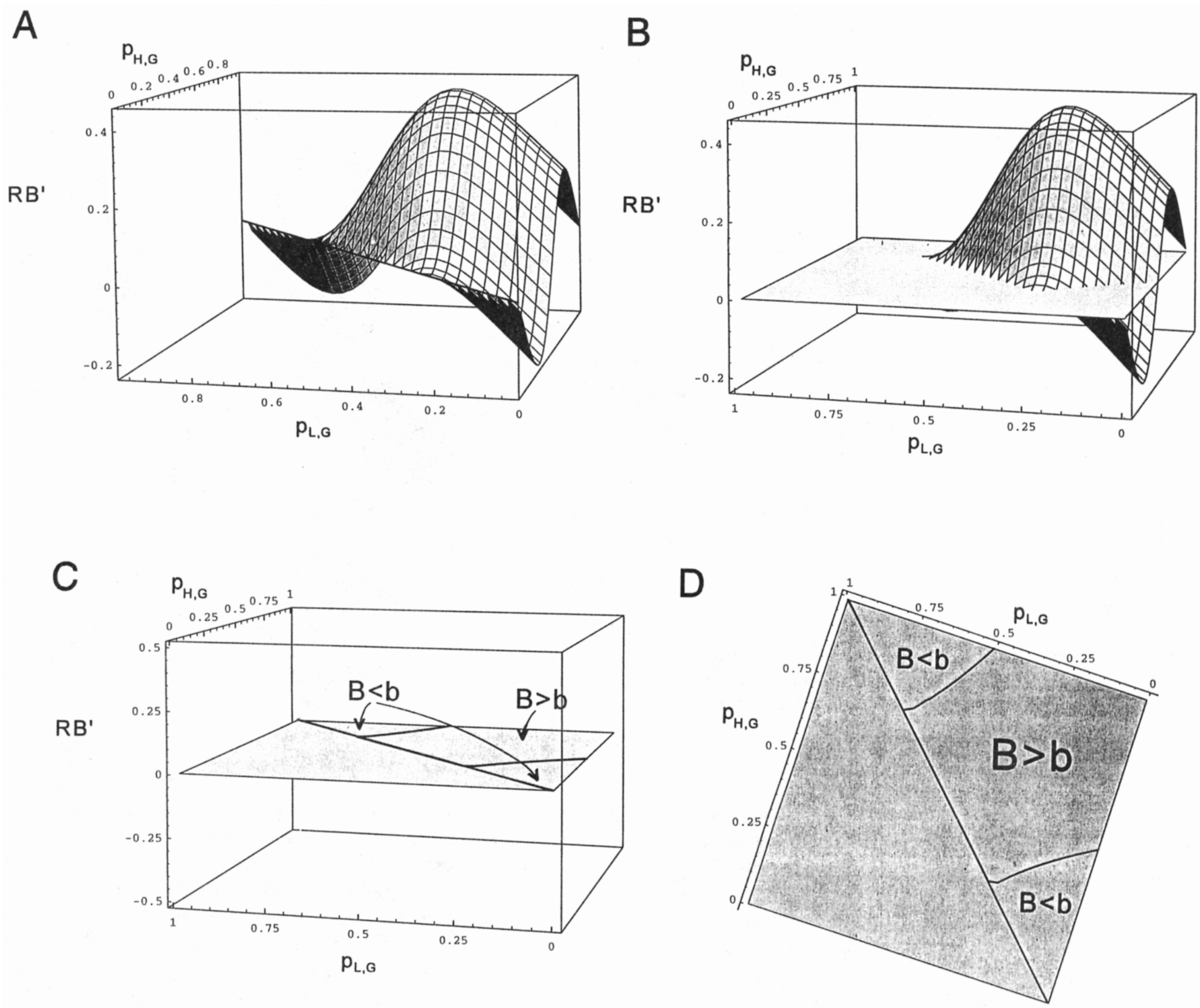


Figure 7. Plots of (a) RB' as a function of ($p_{H,G}$, $p_{L,G}$) and (b) the $RB' = 0$ contour function under a simple-majority D. RB' = relative bias; $p_{H,G}$ = proportion of individuals favoring G in the high-bias condition; $p_{L,G}$ = ideal proportion of individuals favoring G in the low-bias condition; D = the social decision scheme matrix; B = group bias; b = individual bias.

(which are typically very highly correlated with jury verdict; e.g., Kerr, 1981). The individual (i.e., predeliberation) and group (i.e., postdeliberation) judgments are contrasted in Figure 8. As the figure shows, following deliberation there was a significant polarization effect involving a shift toward greater guiltiness in the high guilt-appearance conditions and toward innocence in the low guilt-appearance conditions. However, there were no significant postdeliberation attorney obnoxiousness biases in postdeliberation judgments; juries were less biased than jurors (in our terminology, $RB < 0$).

Reasoning from information integration theory (e.g., Anderson, 1981), Kaplan has argued (1982; Kaplan and Miller, 1978; Kaplan & Schersching, 1980) that individual jurors' judgments are reached through the integration of many sources of information: personal predispositions (e.g., authoritarianism, general lack of sympathy for criminals, etc.), biasing extralegal factors (e.g., liking for defendant or victim, attitudes toward parties identified with them, like their counsel), and, most important, evidentiary factors. Furthermore, he has argued that jurors recognize, either without reminder or through judicial instructions, that biasing, extralegal material should not actually be considered. This recognition, he has argued, along with the greater amount of information available to the jury results in the content of jury deliberation being dominated by valid, acceptable information (viz., evidentiary material). Since most of the new information to which a juror is exposed during deliberation would not be biasing, deliberating jurors' verdicts should be influenced more by the evidence and less by personal biases during than before deliberation.

This line of argument suggests qualitative differences in the

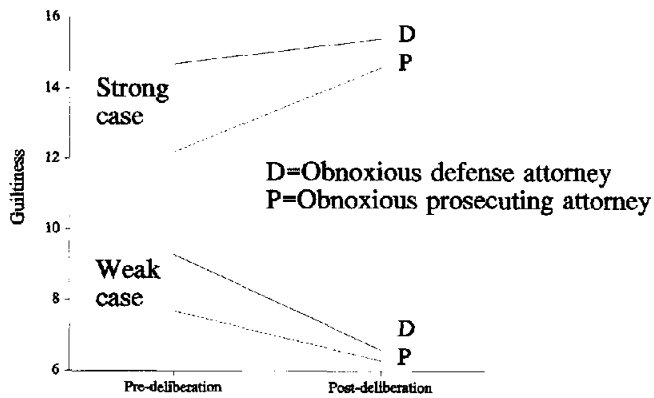


Figure 8. Predeliberation and postdeliberation guilt ratings from Kaplan and Miller (1978, Experiment 3).

process of individual and group judgment. However, our present analysis predicts exactly the pattern of results observed by Kaplan and Miller (1978). We know from much research that juries tend to follow one or another variation on a majority-rules decision scheme (Davis, 1980; Stasser, Kerr, et al., 1989). And our analysis has indicated (see Figure 7) that if all participants (in both the high and low bias conditions) are near a response pole (a result produced in Kaplan and Miller by the strength-of-evidence manipulation), then group judgment should be polarized and the error of commission bias in individuals should be attenuated in groups. Kaplan's theoretical assumption that jury deliberation successfully debiases juror thinking is not necessary.[12]

Another thing we learn from the present theoretical analysis is that whether groups are more or less biased than individuals can depend entirely on the way in which bias is defined (i.e., which type of judgmental sin is being shown). To see this, we return to our prospect-theory example (an instance in which Tversky & Kahneman [e.g., 1981] cleverly devised a paradigm that permitted a sin of imprecision demonstration). We assume further that besides the very general risk-seeking-for-loss bias (which Tversky and Kahneman attribute to nonlinear utility functions), our participants also bring another, more personal bias (e.g., our participant pool is the local chapter of Gamblers Anonymous, whose members are predisposed to take risks). Under these assumptions, individuals would clearly display a risk-seeking bias (i.e., $p_{risky\ alternative} > .5$), and, as we have seen earlier (see Figure 3A), groups operating under a majority-wins decision scheme would show an even larger bias. But suppose instead of comparing our participants' behavior to the unbiased criterion of indifference between alternatives, that is, $I = (.5, .5)$ in a sin of imprecision paradigm, we decide to contrast conditions in which the outcomes are framed as losses versus framed as gains (i.e., in a sin of commission paradigm; cf. McGuire et al., 1987). Under our assumptions, if we randomly assign our gambler participants to loss-frame (high-bias) and gain-frame (low-bias) conditions, we would expect both conditions to generally prefer risky alternatives (i.e., both $p_{H,risky\ alternative}$ and $p_{L,risky\ alternative} > .5$), because of the dispositional bias for risk seeking in this participant population, and, in addition, we would expect to observe the original sin of commission (i.e., $p_{H,risky\ alternative} > p_{L,risky\ alternative}$) that stems ultimately from the nonlinearity of the utility function. Now, assume that groups of our gamblers rather than individual gamblers were to serve as judges and the majority-rules scheme were to summarize those groups' decision making. If the dispositional bias were strong enough (i.e., $p_{H,risky\ alternative}$ and $p_{L,risky\ alternative}$ begin to approach the 1.0), our analysis shows that the original sin of commission effect would be attenuated within groups, exactly opposite to the conclusion that we reached in the SofI paradigm. The important conclusion is that the way in which a particular bias is defined (e.g., SofI vs. SofC) can result in diametrically opposite conclusions about whether groups or individuals are more apt to display that bias, even when we are talking about a single unitary bias phenomenon and the process of group decision making is identical within each type-of-bias paradigm.

Equiprobability D. For completeness sake, we confirmed in the SofC case what had been evident in the SofI case: The net effect of the equiprobability process is to invert the patterns for RB observed under the majority-wins process. Figure 9 presents the relevant results. Under equiprobability, only when all individuals are fairly extreme to begin with do groups magnify individual biases; the general rule (i.e., that $RB < 0$) elsewhere—that is, where at least one of the groups being compared is not extreme—is the opposite of the majority prediction in the same regions.

Strongly asymmetric D. The results of our computer analysis for the strongly asymmetric decision scheme are plotted in Figure 10. Recall that the strongly asymmetric D matrix assumes a functional model for which the second alternative, NG, is (nearly fully) demonstrably "correct." Those with exceptional spatial-reasoning abilities may be able to recognize the resulting RB' surface in Figure 10A as a distortion of the corresponding majority-wins surface (i.e., Figure 7A); those with less than exceptional (e.g., normal) spatial-reasoning abilities might want to forego the exercise of mentally manipulating 3-D surfaces and just skip ahead to the next paragraph, where the net result of a strong asymmetry in D is described. The surface for the strongly asymmetrical D is just the majority-wins surface stretched and distorted toward the back left-hand corner of the 3-D plot, that is, toward the (1.0, 1.0) pole. The net result of such a distortion is summarized in Figure 10D and may be meaningfully compared with the corresponding figure for the majority-wins plot (i.e., Figure 7D). Adding a strong asymmetry to the basic majority-wins model enlarges the $B < b$ area near the favored pole (here, where p_G approaches 0) and shifts and compresses the middle $B > b$ region toward the unfavored pole (i.e., toward the $p_G = 1.0$ pole; see Whyte, 1993, for a possible illustration of behavior in the latter region). In the present instance, the asymmetry is so strong that the other majority-rule $B < b$ region—at the top of Figure 7D—becomes vanish-

[12] Kaplan's RB prediction might well obtain if extralegal information were seen by nearly all jurors to be "demonstrably *in*correct." That is, if there were a functional model of juror decision making that both proscribed using such information and met the other requirements for such a functional model (e.g., widely shared, willingly advocated, etc.), and if advocates for conviction were largely to base their positions on such information, then we might expect the extralegal bias to be weaker among groups.

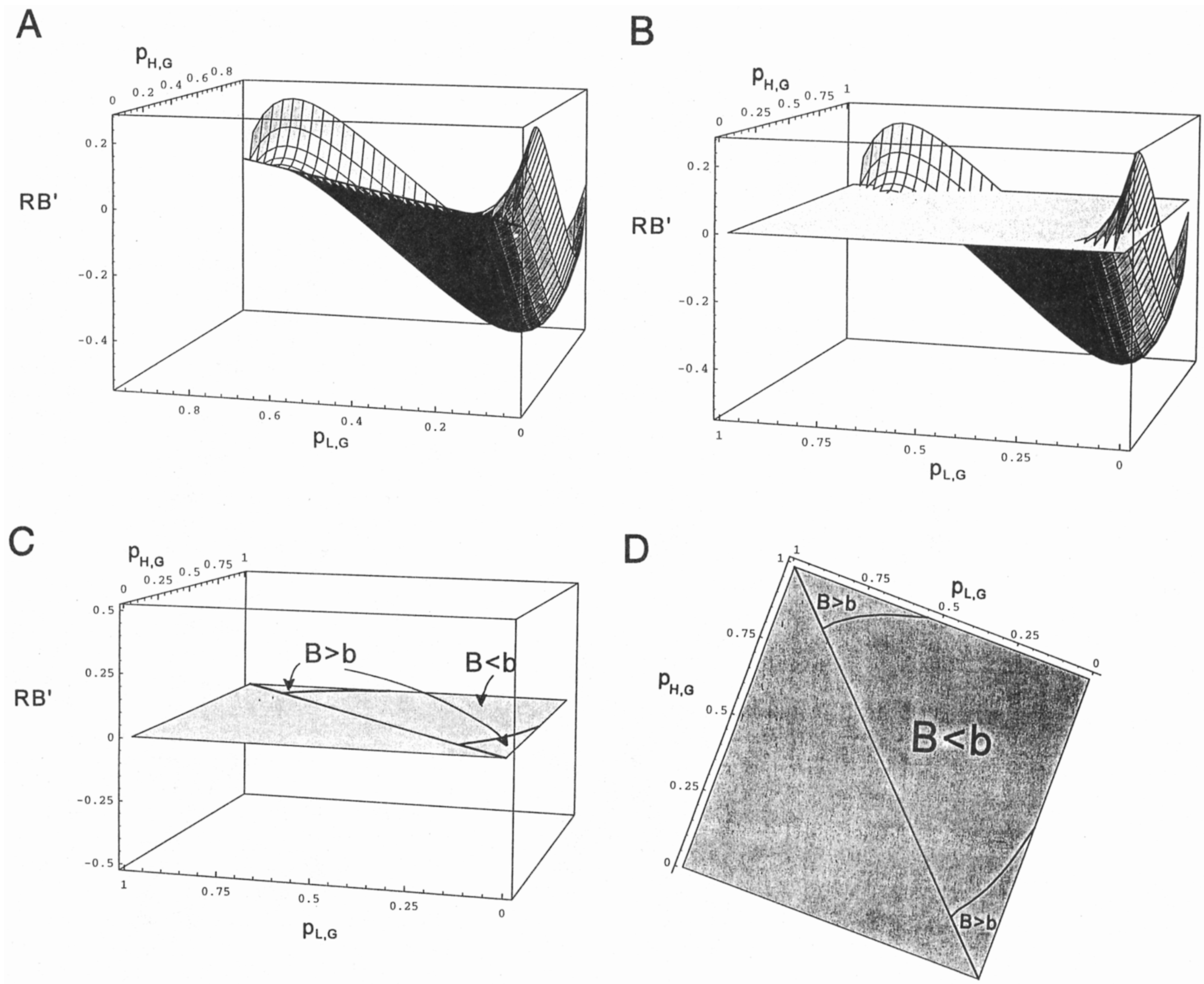


Figure 9. Plots of (a) RB' as a function of ($p_{H,G}$, $p_{L,G}$) and (b) the $RB' = 0$ contour function under an equiprobability D. RB' = relative bias; $p_{H,G}$ = proportion of individuals favoring G in the high-bias condition; $p_{L,G}$ = proportion of individuals favoring G in the low-bias condition; D = the social decision scheme matrix; B = group bias; b = individual bias.

ingly small. Had the asymmetry been less pronounced, this region could have survived as a small area.

What does this mean? That under Case 1 assumptions, when one direction is strongly favored by a functional model, judgmental sins of commission will more often than not be less pronounced among groups than among individuals. Only when there are very few individual advocates of that favored position to be found should we expect groups to be more biased than individuals (i.e., when mean individual judgment across the high- and low-bias conditions begins to approach the unfavored pole).[13]

The mock jury studies we have identified that examine both predeliberation juror and jury susceptibility to extralegal biases (see Table 1) each examine criminal, rather than civil, cases. Thus, it is important to note one "bias" that typically emerges during criminal jury deliberation—a so-called "leniency bias" (see MacCoun & Kerr, 1988, for a review). There is a reliable asymmetry in criminal jury deliberation that gives factions advocating acquittal better prospects of prevailing in the deliberation than equally sized factions favoring conviction. The net effect of this asymmetry is to make jury verdicts more lenient than juror verdicts (at least for reasonably close cases). MacCoun and Kerr (1988) have presented evidence that this effect is a product of common law norms for protecting the defendant

[13] This relatively simple pattern in the SofC case contrasts with the rather more complicated patterns that arose in the corresponding SofI case (see Figures 5A, 5B, and 5C). Once again, there is the potential for the same basic judgmental bias to produce very different RB values when it is framed as a sin of imprecision versus a sin of commission.

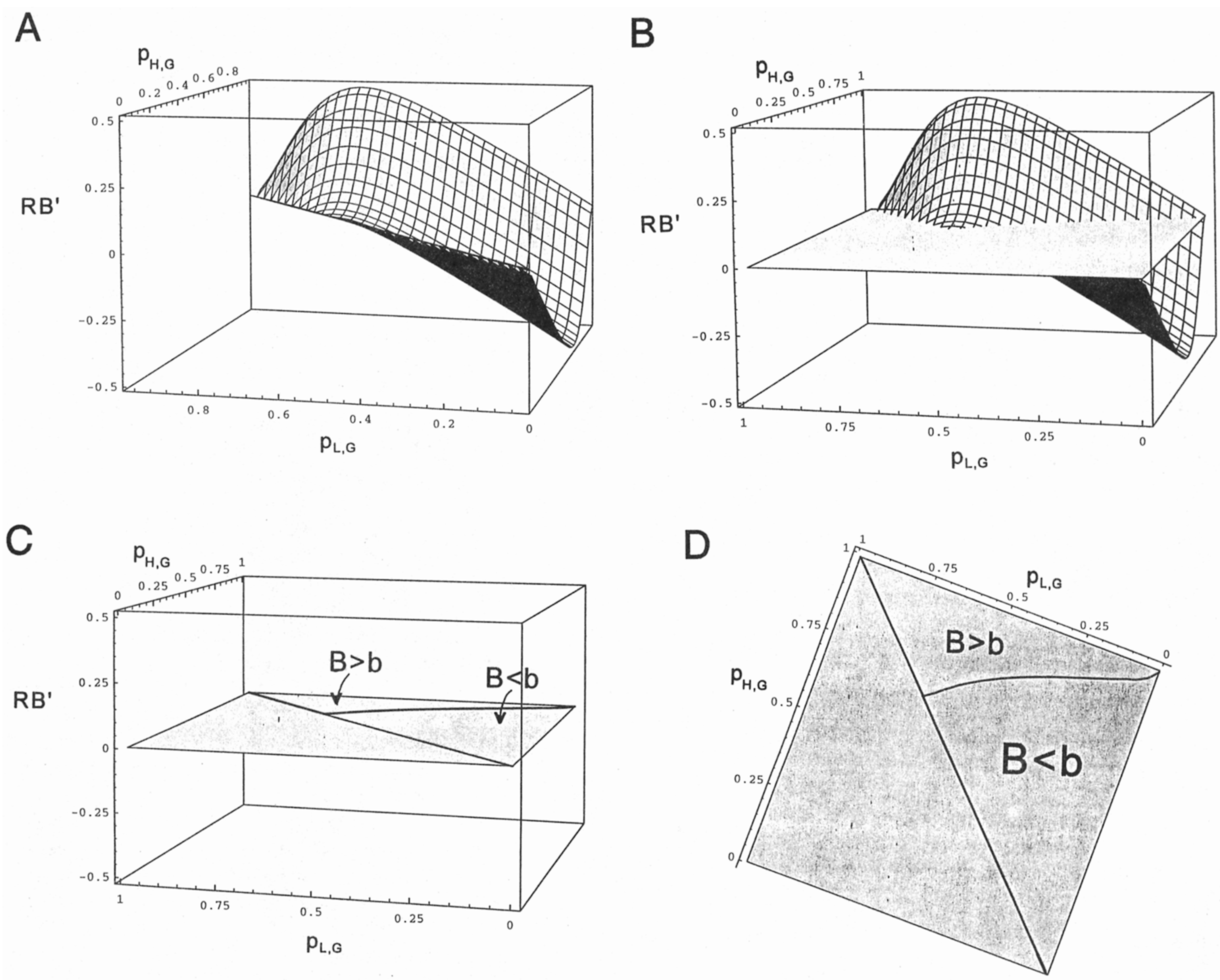


Figure 10. Plots of (a) RB' as a function of $(p_{H,G}, p_{L,G})$ and (b) the $RB' = 0$ contour function under the strongly asymmetric D. RB' = relative bias; $p_{H,G}$ = proportion of individuals favoring G in the high-bias condition; $p_{L,G}$ = proportion of individuals favoring G in the low-bias condition; D = the social decision scheme matrix; B = group bias; b = individual bias.

from false conviction, as reflected in the reasonable-doubt standard of proof and the presumption of innocence requirement. Because such defendant-protection norms (e.g., Davis, 1980) are prescribed by common law, the leniency "bias" is not, with our present definition of the term, really a bias at all. However, when examining the effects of deliberation on genuine extralegal biases, it is important to anticipate that shifts toward acquittal are likely to occur in juries, even in the absence of other biases.

Earlier we showed how juror bias should be attenuated in juries operating under a simple-majority decision scheme only for trials that produced extreme verdict distributions (i.e., p_G near 0.0 and 1.0). But according to the present analysis, a leniency "bias" would expand the former region upward from p_G near 0.0. Thus, if the leniency bias is strong (which appears to depend upon how juries are instructed; MacCoun & Kerr, 1988), we might expect jury deliberation to attenuate bias even if the overall conviction rate is fairly moderate. Interestingly, besides Kaplan and Miller's (1978) study (discussed earlier), the only other study not to observe a clear, unequivocal bias-enhancing effect of jury deliberation is Thompson, Fong, & Rosenhan, (1981). Casting their findings in our present terminology, Thompson et al. found $(p_{H,G}, p_{L,G}) = (p_{\text{pro-prosecution inadmissible evidence},G}, p_{\text{pro-defense inadmissible evidence},G}) = (.53, .38)$ and corresponding jury values of $(P_{H,G}, P_{L,G}) = (.39, .21)$. Statistically, one could not reject the hypothesis of equal degrees of bias by individuals and groups. This pattern of results is consistent with our model if Thompson et al.'s juries had a moderately strong leniency bias, that is, with such a degree of asymmetry, the observed $(p_{H,G}, p_{L,G})$ could lie close to the equal-bias contour curve.

Summary. When a majority-wins decision scheme (or some similar strength-in-numbers decision scheme) is likely to apply (e.g., for clearly judgmental tasks), moving from individual to group decision makers tends to exaggerate individual biases of commission in groups (unless floor–ceiling effects intrude).[14] Again, these patterns are reversed under group decision processes with disproportionately low strength in numbers (e.g., equiprobability). For tasks with demonstrably correct alternatives (e.g., clearly intellective tasks), sins of commission are usually less pronounced in groups (unless the "incorrect" positions are extremely popular, in which case the reverse can be true).

Sins of Omission (SofO)

As far as our central question about relative bias is concerned, this is a simply handled case. A sin of omission means that the high- and low-bias conditions do not differ. If individuals in these two conditions start at the same place (i.e., have the same π vector) and group decision making is summarized by the same process in each condition (Case 1's assumption of a single, constant D in all groups), then the SDS model predicts that groups in these two conditions must likewise end up at the same place. Hence, under Case 1 assumptions, a sin of omission in individuals must result in the same sin of omission in groups. (By identical logic, a sin of omission in groups must, under Case 1 assumptions, imply an identical sin of omission in individuals.) Thus, if there is any difference in individual and group susceptibility to a sin of omission (and, as our earlier overview indicated, there certainly are; e.g., Wright & Wells, 1985; Wright, Lüüs, & Christie, 1990), it can only mean that Case 1 assumptions have been violated. We consider Case 2 next.

Case 2: Varying Group-Judgment Processes

Our assumption in Case 1 of a single social decision-making process for all groups essentially means that for any given initial distribution of member preference (r_G, r_{NG}), the relative ability of each faction to prevail in the coming group discussion is not altered by exposure to potentially biasing information. So, for example, under Case 1 assumptions, receiving prejudicial pretrial publicity may increase a juror's chances of seeing the defendant as guilty, but it would not alter his or her ability, say, to resist a unanimous majority favoring acquittal in a jury beginning deliberation with a $(r_G, r_{NG}) = (11, 1)$ split (relative to a comparable juror in a low-bias condition who had seen no such publicity).

As we have seen, even under the simplifying Case 1 assumption, the relative degree of individual and group bias depends on a number of factors. But it is also quite possible that this assumption is false. Possession of certain information, per se, could actually alter the dynamics of the process of group judgment. In the preceding example, perhaps all other things (and especially the initial distribution of verdict preferences) being equal, a pro-conviction juror could be relatively more persuasive, more resistant to persuasion, or both in the group setting when the jury has been exposed to prejudicial pretrial publicity than when it has not been so exposed.

This is not a possibility in the sin of imprecision case, because there are no high- and low-bias conditions receiving different information for this type of bias. Rather, all individual and all group judges are given the same basic task information and their respective judgments are then compared to a standard defined by a normative model. It is, however, definitely a possibility for the other two types of bias; for both sin of commission and sin of omission it may be that $D_H \neq D_L$. Of course, there are literally an infinite number of ways that this could occur. Later, we speculate on the nature of such process differences, but first it is useful to examine a handful of studies that document that such differences do indeed occur.

Sins of Commission (SofC)

The most direct way to document such effects is to identify differences between the best-fitting or estimated D matrices for high- and low-bias groups. Unfortunately, there have been relatively few studies that have provided the data required to explore this possibility. Furthermore, the amount of data required to provide convincing statistical evidence can be prohibitive.[15] Nevertheless, we have identified three such studies that provide some fragmentary but suggestive evidence.

Two of these studies examine extralegal biases in jury decision making. MacCoun (1990) examined the effect of defendant physical attractiveness on the verdicts of 4-person mock juries. The biasing effect of defendant appearance information was somewhat greater among juries (an effect of 18% on conviction rate) than among individual jurors (an effect of 12%). This pattern is consistent with the predictions of the majority-wins primary scheme revealed in our Case 1 computer analyses.[16] The observed frequencies of initial split to final verdict transitions are presented in Panel A of Table 4; the relative frequencies (by row), provided in parentheses beneath the raw frequencies, represent the entries of the estimated D matrices for the high- and low-attractiveness defendant conditions.

As noted previously, besides the usual majority-rules primary scheme, a fair amount of research has documented an asymmetric subscheme for criminal jury deliberation (see Stasser, Kerr, & Bray, 1982; MacCoun & Kerr, 1988). When there is no strong initial majority, pro-acquittal factions are more likely to prevail than comparable pro-conviction factions. For example, one meta-analysis indicated that, on average, acquittal was

[14] It is noteworthy that the general amplification of bias produced by majority Ds occurs in precisely the same area (viz., nonextreme distributions of individual preference) where it would typically take a larger sample size to detect an effect among groups (see Nagao & Davis, 1980, Table 4). Again, power considerations can complicate the comparison of bias at the individual and group levels (Kenny et al., in press).

[15] For example, comparing D matrices for high- and low-bias 12-person juries requires estimation of at least 54 matrix entries (2 matrices, 13 rows and 2 columns per matrix). Even if one could assume relatively uniform π vectors (which is implausible, because for any given distribution of individual preference, there are many unlikely initial splits), in order to have a minimum of 10 groups per row per matrix, one would need 3,120 participants (2 matrices × 13 rows/matrix × 10 groups/row × 12 persons/group).

[16] Under random composition of juries, the relatively small, 12% effect for attractiveness among jurors resulted in a much larger proportion of juries with an initial majority for acquittal when the defendant was physically attractive (12/25 or 41%) than when he was unattractive (4/30 or 13%).

Table 4
Data on Effects of Extralegal Biasing Information Affecting Jury Leniency Bias

Predeliberation verdict split		Outcome of jury deliberation					
		High-bias condition			Low-bias condition		
G	NG	G	NG	Hung	G	NG	Hung
		MacCoun (1990)					
Panel A		Low-attractive defendant			High-attractive defendant		
4	0	1 (1.00)	0 (.00)	0 (.00)	1 (1.00)	0 (.00)	0 (.00)
3	1	7 (.47)	2 (.13)	6 (.40)	3 (.27)	2 (.18)	6 (.55)
2	2	3 (.30)	3 (.30)	6 (.40)	0 (.00)	4 (1.00)	0 (.00)
1	3	1 (.25)	3 (.75)	0 (0.0)	1 (.12)	5 (.62)	2 (.25)
0	4				0 (.00)	4 (1.00)	0 (.00)
		Kramer et al. (1990)					
Panel B		High-emotional publicity			Low-emotional publicity		
5	1	3 (.60)	0 (.00)	2 (.40)	1 (.00)	0 (.00)	0 (.00)
4	2	9 (.45)	0 (.00)	11 (.55)	3 (.33)	0 (.00)	6 (.67)
3	3	6 (.33)	4 (.22)	8 (.44)	0 (.00)	2 (.18)	9 (.82)
2	4	0 (.00)	5 (.42)	7 (.58)	0 (.00)	9 (.82)	2 (.18)
1	5	0 (.00)	1 (.33)	2 (.67)	0 (.00)	4 (1.00)	0 (.00)

Note. G = guilty; NG = not guilty. Numbers in parentheses are relative frequencies (by row).

about four times as likely as conviction for juries that were evenly split prior to deliberation (MacCoun & Kerr, 1988). Inspection of MacCoun's (1990) data in Table 4 suggests that this asymmetry bias was evident when the defendant was attractive but not when the defendant was unattractive. This is seen most directly in the evenly split juries (2G, 2NG). When the defendant was attractive, all such groups acquitted the defendant; when the defendant was unattractive, only 30% (3 of 10) of the juries acquitted. To a lesser extent, the comparison of pro- and anti-conviction majority juries with initial 3–1 splits reveals the same pattern. When the defendant was attractive, pro-acquittal majorities were more likely to prevail than comparably sized pro-conviction majorities (.62 − .27 = .35); the comparable figure in the unattractive defendant condition was somewhat lower (.75 − .47 = .28). When one drops hung juries, this trend is even clearer: (5/6 − 3/5 = .23) versus (3/4 − 7/9 = −.03). Although these differences are small and, given these small samples, not statistically significant, they are intriguing. They suggest that the usual willingness to give the defendant the benefit of the doubt (e.g., when there is no clear consensus in the group) is attenuated when that defendant is physically unattractive.

This suggestion is bolstered by a similar pattern of data in Kramer, Kerr, and Carroll's (1990) study of the biasing effects of pretrial publicity. The relevant estimated social decision schemes are presented in Panel B of Table 4. Kramer et al. examined (among other things) the biasing effects of emotional pretrial publicity.[17] And, just as in MacCoun's (1990) data, there are indications that the usual and legally prescribed pro-defendant leniency bias is attenuated when jurors were exposed to prejudicial emotional pretrial publicity. Consider the third row of the lower half of Table 4. In low-emotion publicity juries, when evenly split juries managed to reach verdicts, they were acquittals (2 acquittals, 0 convictions). However, an unusual trend in the opposite direction occurred in the high-emotion juries (6 convictions, 4 acquittals). Likewise in the fourth row, we see that acquittal was relatively less likely (particularly in comparison to hung juries) given an initial pro-acquittal majority in the high-emotion publicity juries than in the low-emotion juries. The association between initial split and verdict in these rows was marginally ($p < .07$) significant.

A third illustration uses the sin of commission approach to

[17] Again, the primary majority scheme may also contribute to the differences between jurors and juries in sensitivity to emotional publicity (which produced the stronger biasing effect among jurors). For example, pro-conviction initial majorities were quite a bit more likely to occur in high emotional pretrial publicity (PTP) groups (43% of the time) than in low-emotion juries (only 28%).

demonstrate the framing biases predicted by prospect theory. McGuire et al., (1987) found very weak framing effects among individuals but rather robust framing effects among groups. Qualitatively, this is fairly consistent with the predictions of a uniform-majority primary scheme. However, McGuire et al. also obtained direct estimates of the operative Ds in the high- and low-bias conditions (viz., the gain- and loss-frame conditions). These estimates, which are similar in structure to those described above for jury deliberations, are presented in the top panel of Table 5 and suggest that the stronger bias in groups could be attributed, at least in part, to different D matrices for the two framing conditions. The most striking differences can be seen in the middle two rows. Clearly, when the problem was framed as a gain problem, risk-averse factions (i.e., composed of those who initially favored taking the sure gain) usually prevailed (77% of the time), but this was true even when the faction was a minority of one (which prevailed 83% of the time). Exactly the opposite occurred for the loss framing; risk-seeking factions usually prevailed (80% of the time), even when it was a single advocate opposing a majority of two (75% of the time). Again, the sample sizes are small and the effects weak (distributional differences within the middle two rows resulted in $ps < .11$ by Fischer's exact tests).

McGuire et al. (1987) speculated that arguments advocating the attitudes toward risk predicted by prospect theory—namely, for risky choices under a loss frame and for more certain choices under a gain frame—are "demonstrably correct" in Laughlin's sense. That is, prospect theory may here be a potent functional model of collective judgment. This claim was bolstered by contrasting the above results, which were obtained in face-to-face discussion groups, with another condition in which all communication between group members took place via a computer network. The corresponding data for this latter condition are presented in the bottom panel of Table 5. It seems likely that it was more difficult for group members to communicate with one another in the computer-assisted-communication condition. Such difficulty seems likely to undermine group members' ability to advocate and recognize their shared conceptual system (which maintains that a sure gain is better than an uncertain one, and a chance to avoid a loss is preferable to a sure loss) and, thus, the "correctness" of the choices prescribed by that system. And indeed, there is little evidence of difference in the estimated D matrices between the two framing conditions.

Table 5
McGuire et al. (1987) Data on Effects of Framing Information Affecting Group Decision Making

Prediscussion preference distribution		Gain frame		Loss frame	
Risk averse	Risk seeking	Risk averse	Risk seeking	Risk averse	Risk seeking
		Group decision–face-to-face discussion			
3	0	1 (1.00)	0 (.00)	2 (1.00)	0 (.00)
2	1	5 (.71)	2 (.29)	2 (2.5)	6 (.75)
1	2	5 (.83)	1 (.17)	0 (.00)	2 (1.00)
0	3	1 (1.00)	0 (.00)	0 (.00)	3 (1.00)
		Group decision–computer-assisted discussion			
3	0	4 (.80)	1 (.20)	3 (1.00)	0 (.00)
2	1	0 (.00)	4 (1.00)	2 (.40)	3 (.60)
1	2	1 (.25)	3 (.75)	2 (.40)	3 (.60)
0	3	1 (.50)	1 (.50)	0 (.00)	2 (1.00)

Note. Numbers in parentheses indicate relative frequencies.

Sins of Omission (SofO)

To our knowledge, no studies have directly estimated D_H and D_L separately within an experimental demonstration of a sin of omission by groups. There is one study, however, that provides strong indirect evidence that the availability of prescribed information can alter the group decision-making process. Wright et al. (1990) found that individuals failed to use prescribed consensus information (i.e., when deciding whether a particular behavior was attributable to internal, dispositional factors or to external, situational factors, individuals paid no attention to whether many (high-consensus) or few (low-consensus) people acted as the target person did. An unbiased judge should attribute the target's behavior more readily to situational factors in the former, high-consensus case. Wright et al. (1990) found, however, that groups did use such consensus information properly: Group discussion appears to have corrected the sin of omission committed by individual judges. If the individuals tended to use consensus information properly but to do so very weakly and nonsignificantly, then this pattern could be understood to be the simple result of a majority-wins scheme. However, not only did individuals show no effect for consensus information, the nonsignificant trend was opposite to the prescribed direction. This pattern of results is hard to explain under Case 1 assumptions and strongly implies that Case 2 assumptions hold here (i.e., a different group decision-making process occurred in the high- and low-consensus information conditions). Other results by Wright et al. further supported this possibility. They found that not just face-to-face group discussion but reading a transcript of another group's discussion or a set of arguments generated by individuals also produced the prescribed use of consensus information. Working from Vinokur and Burnstein's (1973) persuasive arguments theory, they suggest that consensus information is highly potent ("demonstrably correct" in Laughlin's [Laughlin & Ellis (1986)] terminology) but that relatively few participants possess such information. Group discussion gives the opportunity for this potent information to be shared and, in Wright et al.'s, like many SofC and SofO paradigms, the information provided in the two experimental conditions prescribes opposite judgments. This suggests that groups beginning deliberation with ostensibly comparable initial splits are likely to move in opposite directions in the two experimental conditions (much as occurred in McGuire et al., 1987; see Table 5). In effect, this suggests that different D matrices described groups in these two conditions, and it was this difference in process that was (at least in part) responsible for reducing the SofO bias in groups.

General Discussion and Conclusions

The central question of this paper has been, "Which is more likely to make a biased judgment, individuals or groups?" Our overview of the relatively small and diverse empirical literature suggested that there was no simple empirical answer to this question. Even when we restrict our attention to particular bias phenomena (e.g., framing effects, preference reversals), there was frequently little consistency in the direction (i.e., sign) and magnitude of observed relative bias, RB.

Although there appeared to be no simple and general empirical answer to our question, the present theoretical analysis based on the social decision scheme model has revealed many partial answers, all of which begin with "Well, it depends . . . . " Even under the simplifying assumption that the same basic group process characterizes all groups (Case 1 assumption), we have shown that (and how) it depends jointly upon several factors. In particular, it depends on:

1. The size of the group: Generally, as group size increases, the sign of RB is unaffected, but its magnitude increases. (It can also be shown that the latter relationship between group size and RB is a monotonic, negatively accelerating one; cf. Latané, 1981).
2. The magnitude of individual bias: All other things being equal (and most particularly, under any one of several possible group processes), both the direction and magnitude of RB can vary as one varies only the magnitude of individual bias.
3. The location of the bias: All other things being equal, both the direction and magnitude of RB can change with the location in the response domain (e.g., the locations of $p_{H,G}$ and $p_{L,G}$ for sins of commission) of an individual bias of constant magnitude.
4. The definition of the bias: All other things being equal, one can come to diametrically opposite conclusions about RB depending on how bias has been defined (e.g., as a sin of imprecision vs. a sin of commission).
5. The normative ideal: As the ideal judgment shifts, RB can (for sins of imprecision) change both sign and magnitude, even if individual preference and group process remain constant.
6. The nature of the group process: Most important, all other things being equal, different group processes can produce dramatically different RBs. If the particular judgment task determined group process completely (and, as much research has shown, task features such as how judgmental–intellective the task is appear to have profound impact on the nature of the group decision-making process), then this factor at least would not contribute to variance in RB for any particular bias phenomenon. But since such situational, group, or personal factors as the importance of the task, the importance of intragroup harmony, or the judge's general level of uncertainty may also influence the nature of the group process, it is not safe to presume that group process is fixed by task demands.

Also note that all of these complex (but tractable) patterns assume that in any given group and task context, group process (as summarized by D) is constant. When we relax this assumption, many other patterns are possible. Given the extreme diversity of bias phenomena, group sizes, ways of operationalizing bias, experimental contexts, etc., in the empirical literature at a global level, it is, in retrospect, hardly surprising that this literature does not show a simple, consistent pattern of relative bias.

Of course, we ask more of a theory than that it correctly predicts that nature can be complex. A good theory ought to help reduce that complexity: by resolving apparent empirical anomalies, by posing informative new questions, by directing practical application, and by guiding where to look and what to look for. We conclude by discussing how well the present theoretical model satisfies these criteria.

Organizing Past Findings

The biggest stumbling block to applying our analysis retrospectively is in knowing exactly what kind of group processes to assume. In most of the bias phenomena that have been studied, there has been relatively little research using groups as judges, and almost none of this work has tested or estimated (Kerr, Stasser, & Davis, 1979) D-matrix summaries of group-decision process. The one clear exception is research examining jurors and juries committing judgmental sins of commission. Considerable research (see Davis, 1980; Kalven & Zeisel, 1966; MacCoun & Kerr, 1988; Stasser et al., 1982) has established that criminal juries' deliberations are summarized by a high-order majority primary decision scheme (e.g., initial two-thirds majorities nearly always prevail) with a subscheme (applying when there is no strong initial majority) that asymmetrically favors acquittal. From the preceding theoretical analysis, it follows that jury deliberation ought to amplify juror sins of commission unless the conviction rate for jurors is very extreme (i.e., $p_{H,G}$ and $p_{L,G}$ approach 0 or 1.0), although less extremity (i.e., $p_{H,G}$ and $p_{L,G}$ near or just below .50) could still result in attenuation of bias due to the leniency "bias." The relevant empirical literature (see Table 1) is basically consistent with this postdiction; generally, juries appear to be more sensitive to proscribed information than jurors, and the few exceptions to this rule appear to occur where the theory anticipates them (e.g., Kaplan and Miller, 1978, used cases with extreme conviction rates).

Our theoretical analysis can also be applied to organizing findings on topics other than the comparison of biased judgment in individuals and groups. For example, our analyses may have implications for the ongoing debate (e.g., Camerer, 1992; Hogarth & Reder, 1987; Thaler, 1991) about the descriptive validity of the rational-choice model, which plays such a central role in modern economics, political science, and public policy analysis. Many economists have disputed the significance of empirical violations of rational-choice assumptions, offering a number of reasons why laboratory demonstrations might underestimate human rationality in real-life settings. One such argument has been that collective decision making should cancel out judgmental errors. Though this may be correct for aggregate public opinion (Page & Shapiro, 1992), it is premised on a statistical analogy—the law of large numbers—that is clearly incompatible with actual interactive group decision making under some likely social decision schemes (e.g., simple majority wins, truth wins, truth-supported wins). More important, this argument does not apply to judgmental biases—the topic of this article—which are systematic rather than random. At best, our analyses offer an existence proof that collective rationality can sometimes be superior to individual rationality, but they also suggest that over a large and plausible region of relevant parameter space, group decision making actually exacerbates the biases observed in individual decisions.

Posing New Questions

Of course, successful prediction is generally more satisfying than apparently successful postdiction. The preceding theoretical analysis suggests many new and testable hypotheses that ought to be systematically tested (e.g., that the effect of jury deliberation on an extralegal sin-of-commission bias will depend on the overall strength of evidence against a defendant). Moreover, when a particular D can be estimated or confidently assumed, this approach does not make only ordinal predictions that one condition will be more biased than another, but makes specific point predictions over an entire domain of model parameters.

Our original question was essentially posed in terms of the outcome of individual versus group judgment. But our analysis (like so many other previous theoretical analyses of group phenomena; e.g., Zajonc, 1965; Myers & Lamm, 1976; Burnstein & Vinokur, 1977) refocuses our attention away from outcome and toward process. That is, away from the question, "which is more biased, individual or group judgment?" and toward the question, "what are the processes whereby individual preferences are translated into group preferences?" And a social decision scheme perspective on the latter broad question raises several other fundamental questions:

1. "What factors determine the operative social decision scheme, D?" A number of such factors have been identified (including the judgmental–intellective nature of the task, task uncertainty, task importance), but undoubtedly many more remain to be identified. One promising way of identifying such factors may be to examine variables (like those in Appendix A) that are known and can be shown to affect group information processing.
2. "When and why will the availability of certain information alter D; that is, when and why must we abandon the simplifying Case 1 assumptions of a single social decision scheme describing all groups?" Ultimately, this raises fundamental questions of how individuals pool and use information in groups. Once again, processes like those listed in Appendix A seem like good starting points for research. For example, groups seem to have difficulty accessing information that is not widely shared among members. This suggests that we need to be concerned not only with the mean impact of biasing information on individual preference, but on how that biasing information is distributed among group members. It is conceivable for a bit of biasing information to have a clear effect on mean individual judgment, yet, because it is not widely shared among group members, to have little effect in the group. Such a pattern could be manifest by the process of group decision making (i.e., as summarized by the d_{ij}) being different for those with versus without such information (i.e., by the need to make Case 2 assumptions).
3. "What do patterns in the operative decision scheme tell us about the existence and nature of a functional model of group judgment?" We suspect that clear asymmetries in D are especially interesting and informative in this regard. Such asymmetries usually suggest that there is some kind of functional model operating: what Laughlin (e.g., Laughlin & Ellis, 1986) termed a "shared conceptual system" and Tindale (e.g., 1993) termed a "shared representation."
4. "What does the process of individual judgment imply about the functional model of group judgment (and vice versa)?"
5. "When a particular normative model is defensible and groups' functional model departs substantially from that normative model, how can groups be induced to modify their functional model toward the normative model?"

At present we have only fragmentary answers to these fundamental questions.

As noted earlier, our present analysis has been focused on biased judgment, but there is nothing in that analysis to preclude applying it to exploring the relative degree to which individuals versus groups use any information, proscribed or not. If one can determine whether individuals use certain information and one can estimate the relevant D matrix or matrices, then one can show whether groups or individuals are more likely to make use of a piece of diagnostic information. So, for example, Davis et al. (1974) found that in choosing between bets, groups were even more sensitive to relevant bet parameters than were individuals, a result that followed directly from the additional finding that these groups operated under a majority social decision scheme.

Directing Practical Application

Although we have argued that groups will amplify bias under some conditions but attenuate it under others, readers will note that we predict enhanced bias within a region of the parameter space that is likely to characterize many real-world group decision tasks; specifically, sins of commission by groups operating under majority-rule decision schemes (as long as individual judgment is not too extreme). Again, these decision schemes tend to apply to judgment tasks with no clearly shared conceptual scheme for defining right or wrong answers (Laughlin & Ellis, 1986; McGrath, 1984; Stasser, Kerr, et al., 1989); prominent examples include jury decision making, hiring decisions, risky investment decisions, and foreign policy decisions of the type examined by Janis (1982).

Thus, our analyses might be taken to imply that group decision making is ill-advised for this large and important class of real-world decision tasks. But quite apart from many other reasons for preferring group to individual decision makers,[18] our analysis also suggests a strategy for mitigating the bias-amplifying tendencies of groups at such tasks. Ultimately, bias as we have conceived it reflects decisions about whether and how to

[18] For example, compared to individuals, groups tend to attenuate unsystematic, random errors (see Page & Shapiro, 1992; Zajonc, 1962), to better satisfy ideals of procedural fairness and legitimacy (e.g., MacCoun & Tyler, 1988), and to enhance the chances of collective mobilization (Rosenwein & Campbell, 1992). Furthermore, a number of explicit procedures have been recommended for encouraging thorough information gathering and consideration and for promoting unbiased judgment in groups (e.g., Janis, 1982; Stasser & Stewart, 1992).

use information. Group discussion can modify such decisions made by individual members. Presumably, in the absence of a compelling functional model of judgment identifying one particular alternative as "correct," these decisions too are made under a majority–strength-in-numbers decision scheme. This suggests that if the majority of individuals recognize and accept the normative use of particular information, that groups are more likely to choose to use that information properly. This reasoning underscores the value of teaching principles of rational, normative judgment through general education and special training (e.g., Arkes, 1991; Nisbett, 1993; Shafir et al., 1993)—of making defensible normative models into operative functional models for most (but not necessarily all) individuals.

Guiding Where to Look and What to Look for

The SDS analysis advanced here offers a conceptual framework within which to identify and analyze individual–group differences in the use of normatively significant information. But, in addition, it provides very useful methodological tools, the foremost of which is using one or more D matrices to summarize the process of group information processing. Several methods for estimating or competitively testing potential D matrices have been developed (see Kerr et al., 1979). Using such methods, it is possible empirically to determine whether one can or cannot safely make Case 1 assumptions, to detect asymmetries that are the signatures of interesting functional models, and to make not just qualitative but quantitative predictions about relative bias. Thus, estimation of the operative D matrix ought to be a routine feature of empirical studies comparing individual and group bias. Such methods can be very usefully augmented by direct manipulations of such factors as the composition of the group (e.g., Tindale et al., 1993) or the content (e.g., Wright et al., 1990) and communication modality (e.g., McGuire et al., 1987) of intragroup communication and by direct assessments of the demonstrability of correctness of judgment alternatives (e.g., Laughlin & Ellis, 1986).

Conclusions

As long as we continue routinely to rely upon groups to make important decisions, it is important to minimize demonstrable bias in group judgment. We have attempted to show several things in this article: (a) that there can be no simple answer to the question, "Which is more biased, individuals or groups?"; (b) that the social decision scheme model offers a framework for identifying and analyzing individual versus group differences in judgment; and (c) that using that framework, we can now specify some of the conditions under which groups are both more and less biased than individuals.

References

Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology, 39,* 1037–1049.

Anderson, N. H. (1981). *Foundations of information integration theory.* New York: Academic Press.

Argote, L., Devadas, R., & Melone, N. (1990). The base-rate fallacy: Contrasting processes and outcomes of group and individual judgment. *Organizational Behavior and Human Decision Processes, 46,* 296–310.

Argote, L., Seabright, M. A., & Dyer, L. (1986). Individual versus group use of base-rate and individuating information. *Organizational Behavior and Human Decision Processes, 38,* 65–75.

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin, 110,* 486–498.

Arkes, H., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes, 35,* 125–140.

Baron, R. S. (1986). Distraction-conflict theory: Progress and problems. In L. Berkowitz (Ed.), *Advances in experimental social psychology,* Vol. 19 (pp. 1–40). New York: Academic Press.

Bray, R. M., Struckman-Johnson, C., Osborne, M. D., McFarlane, J. B., & Scott, J. (1978). The effects of defendant status on decisions of student and community juries. *Social Psychology, 41,* 256–260.

Brockner, J., & Rubin, J. (1985). *Entrapment in escalating conflict: A social psychological analysis.* New York: Springer-Verlag.

Burnstein, E., & Vinokur, A. (1977). Persuasive argumentation and social comparison as determinants of attitude polarization. *Journal of Experimental Social Psychology, 13,* 315–332.

Camerer, C. (1992). The rationality of prices and volume in experimental markets. *Organizational Behavior and Human Decision Processes, 51,* 237–272.

Carretta, T. R., & Moreland, R. L. (1983). The direct and indirect effects of inadmissible evidence. *Journal of Applied Social Psychology, 13,* 291–309.

Carroll, J. S., Kerr, N. L., Alfini, J., Weaver, F., MacCoun, R., & Feldman, V. (1986). Free press and fair trial: The role of behavioral research. *Law and Human Behavior, 10,* 187–202.

Chalos, P., & Pickard, S. (1985). Information choice and cue use: An experiment in group information processing. *Journal of Applied Psychology, 70,* 634–641.

Dane, F., & Wrightsman, L. (1982). Effects of defendants' and victims' characteristics on jurors' verdicts. In N. Kerr & R. Bray (Eds.), *The psychology of the courtroom.* New York: Academic Press.

Davis, J. H. (1969). *Group performance.* Reading, MA: Addison-Wesley.

Davis, J. H. (1973). Group decision and social interaction: A theory of social decision schemes. *Psychological Review, 80,* 97–125.

Davis, J. H. (1980). Group decision and procedural justice. In M. Fishbein (Ed.), *Progress in social psychology* (pp. 157–229). Hillsdale, NJ: Erlbaum.

Davis, J. H. (1982). Social interaction as a combinatorial process in group decision. In H. Brandstätter, J. H. Davis, & G. Stocker-Kreichgauer (Eds.), *Group decision making.* London: Academic Press.

Davis, J. H. (1996). Group decision making and quantitative judgments: A consensus model. In E. Witte & J. H. Davis (Eds.), *Understanding group behavior: Consensual action by small groups* (Vol. 1, pp. 35–60). Hillsdale, NJ: Erlbaum.

Davis, J. H., Hornik, J., & Hornseth, J. P. (1970). Group decision schemes and strategy preferences in a sequential response task. *Journal of Personality and Social Psychology, 15,* 397–408.

Davis, J. H., & Kerr, N. L. (1986). Thought experiments and the problem of sparse data in small-group performance research. In P. Goodman (Ed.), *Designing effective work groups* (pp. 305–349). New York: Jossey-Bass.

Davis, J. H., Kerr, N. L., Sussmann, M., & Rissman, A. K. (1974). Social decision schemes under risk. *Journal of Personality and Social Psychology, 30,* 248–271.

Davis, J. H., Spitzer, C. E., Nagao, D. H., & Stasser, G. (1978). Bias in social decisions by individuals and groups: An example from mock juries. In H. Brandstätter, J. H. Davis, & H. Schuler (Eds.), *Dynamics of group decisions* (pp. 33–52). Beverly Hills, CA: Sage.

Davis, J. H., Tindale, R. S., Nagao, D. H., Hinsz, V. B., & Robertson,

B. (1984). Order effects in multiple decisions by groups: A demonstration with mock juries and trial procedures. *Journal of Personality and Social Psychology, 47,* 1003–1012.

Dawes, R. M. (1988). *Rational choice in an uncertain world.* San Diego, CA: Harcourt Brace Jovanovich.

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward solution of a riddle. *Journal of Personality and Social Psychology, 53,* 497–509.

Dunning, D., Griffin, D. W., Milojkovic, J., & Ross, L. (1990). The overconfidence effect in social prediction. *Journal of Personality and Social Psychology, 58,* 568–581.

Dunning, D., & Ross, L. (1992). *Overconfidence in individual and group prediction: Is the collective any wiser?* Unpublished manuscript, Cornell University.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment.* New York: Wiley.

Einhorn, H. J., Hogarth, R. M., & Klempner, E. (1977). Quality of group judgment. *Psychological Bulletin, 84,* 158–172.

Ericsson, K. A., & Simon, H. A. (1993). Protocol analysis: Verbal reports as data. Cambridge, MA: MIT Press.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7,* 117–140.

Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin, 101,* 75–90.

Gelfand, A. E., & Solomon, H. (1974). Analyzing the decision making processes of the American jury. *Journal of the American Statistical Association, 70,* 305–309.

Greene, E., & Loftus, E. F. (1985). When crimes are joined at trial. *Law and Human Behavior, 9,* 193–207.

Greenwald, A. G. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist, 35,* 603–618.

Grofman, B. (1978). Judgmental competence of individuals and groups in a dichotomous choice situation: Is a majority of heads better than one? *Journal of Mathematical Sociology, 6,* 47–60.

Hans, V. P., & Doob, A. N. (1976). Section 12 of the Canada Evidence Act and the deliberations of simulated juries. *Criminal Law Quarterly, 18,* 235–254.

Harkins, S., & Petty, R. E. (1982). Effects of task difficulty and task uniqueness on social loafing. *Journal of Personality and Social Psychology, 43,* 1214–1230.

Harkins, S., & Szymanski, K. (1987). Social loafing and social facilitation: New wine in old bottles. In C. Hendrick (Ed.), *Group processes and intergroup relations. Review of personality and social psychology* (Vol. 9, pp. 167–188). Newbury Park, CA: Sage.

Hartwick, J., Sheppard, B., & Davis, J. H. (1982). Group remembering: Research and implications. In R. A. Guzzo (Ed.), *Improving group decision making in organizations: Approaches from theory and research* (pp. 41–72). New York: Academic Press.

Hastie, R. (1986). Review essay: Experimental evidence on group accuracy. In B. Grofman & G. Guillermo (Eds.), *Information pooling and group decision making* (Vol. 2, pp. 129–157). Greenwich, CT: JAI Press.

Hastie, R., Penrod, S., & Pennington, N. (1983). *Inside the jury.* Cambridge, MA: Harvard University Press.

Hastie, R., & Rasinski, K. A. (1988). The concept of accuracy in social judgment. In D. Bar-Tal & A. W. Kruglanski (Eds.), *The social psychology of knowledge* (pp. 193–208). Cambridge, England: Cambridge University Press.

Hawking, S. W. (1988). *A brief history of time.* New York: Bantam Books.

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events when outcomes are known. *Psychological Bulletin, 107,* 311–327.

Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology, 13,* 141–154.

Hill, G. W. (1987). Group versus individual performance: Are N + 1 heads better than one? *Psychological Bulletin, 91,* 517–539.

Hinsz, V. B., Tindale, R. S., & Vollrath, D. A. (in press). The emerging conceptualization of groups as information processors. *Psychological Bulletin.*

Hoffman, L. R. (1979). *The group problem solving process: Studies of a valence model.* New York: Praeger.

Hogarth, R. M. (1990). *Insights in decision making: A tribute to Hillel J. Einhorn.* Chicago: University of Chicago Press.

Hogarth, R. M., & Reder, M. W. (1987). *Rational choice: The contrast between economics and psychology.* Chicago: University of Chicago Press.

Irwin, J. R., & Davis, J. H. (1995). Choice/matching preference reversals in groups: Consensus processes and justification-based reasoning. *Organizational Behavior and Human Decision Processes, 64,* 325–339.

Izzett, R. R., & Leginski, W. (1974). Group discussion and the influence of defendant characteristics in a simulated jury setting. *Journal of Social Psychology, 93,* 271–279.

Janis, I. L. (1982). *Groupthink* (2nd ed.). Boston: Houghton-Mifflin.

Johnson, C. D., & Davis, J. H. (1972). An equiprobability model of risk-taking. *Organizational Behavior and Human Performance, 8,* 159–175.

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology, 3,* 1–24.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases.* New York: Cambridge University Press.

Kahneman, D., & Tversky, A. (1972). Subjective probability: An analysis of decision under risk. *Cognitive Psychology, 3,* 430–454.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39,* 341–350.

Kalven, H., & Zeisel, H. (1966). *The American jury.* Boston: Little-Brown.

Kameda, T., & Davis, J. H. (1990). The function of the reference point in individual and group risk decision making. *Organizational Behavior and Human Decision Processes, 46,* 55–76.

Kaplan, M. F. (1982). Cognitive processes in the individual juror. In N. Kerr & R. Bray (Eds.), *The psychology of the courtroom* (pp. 197–220). New York: Academic Press.

Kaplan, M. F., & Miller, L. E. (1978). Reducing the effects of juror bias. *Journal of Personality and Social Psychology, 36,* 1443–1455.

Kaplan, M. F., & Schersching, C. (1980). Reducing juror bias: An experimental approach. In P. Lipsitt & B. Sales (Eds.), *New directions in psycholegal research.* New York: Van Nostrand-Reinhold.

Kenny, D. A., Kashy, D. A., & Bolger, N. (in press). Data analysis in social psychology. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed.). New York: McGraw-Hill.

Kerr, N. L. (1981). Social transition schemes: Charting the group's road to agreement. *Journal of Personality and Social Psychology, 41,* 684–702.

Kerr, N. L. (1982). Social transition schemes: Model, method, and applications. In H. Brandstätter, J. Davis, & G. Stocker-Kreichgauer (Eds.), *Group decision making.* London: Academic Press.

Kerr, N. L. (1983). Motivation losses in task-performing groups: A social dilemma analysis. *Journal of Personality and Social Psychology, 45,* 819–828.

Kerr, N. L. (1992). Group decision making at a multialternative task: Extremity, interfaction distance, pluralities, and issue importance. *Organizational Behavior and Human Decision Processes, 52,* 64–95.

Kerr, N. L., Atkin, R., Stasser, G., Meek, D., Holt, R., & Davis, J. H.

(1976). Guilt beyond a reasonable doubt: Effects of concept definition and assigned decision rule on the judgments of mock jurors. *Journal of Personality and Social Psychology, 34,* 282–294.

Kerr, N. L., Davis, J. H., Atkin, R., Holt, R., & Meek, D. (1975). Group position as a function of member attitudes: Choice shift effects from the perspective of social decision scheme theory. *Journal of Personality and Social Psychology, 35,* 574–593.

Kerr, N. L., & Huang, J. Y. (1986). Jury verdicts: How much difference does one juror make? *Personality and Social Psychology Bulletin, 12,* 325–343.

Kerr, N. L., MacCoun, R. J., & Kramer, G. P. (1996). When are N heads better than one?: Bias in individuals vs. groups. In E. Witte & J. H. Davis (Eds.) *Understanding group behavior: Consensual action by small groups* (Vol. 1, pp. 105–136). Hillsdale, NJ: Erlbaum.

Kerr, N. L., Stasser, G., & Davis, J. H. (1979). Model-testing, model-fitting, and social decision schemes. *Organizational Behavior and Human Performance, 23,* 339–410.

Kerwin, J., & Shaffer, D. R. (1994). Mock jurors versus mock juries: The role of deliberations in reactions to inadmissible testimony. *Personality and Social Psychology Bulletin, 20,* 153–162.

Klevorick, A. K., & Rothschild, M. A. (1979). A model of the jury decision process. *Journal of Legal Studies, 8,* 141–161.

Kramer, G. P., Kerr, N. L., & Carroll, J. S. (1990). Pretrial publicity, judicial remedies, and jury bias. *Law and Human Behavior, 14,* 409–438.

Latané, B. (1981). The psychology of social impact. *American Psychologist, 36.* 343–356.

Latané, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology, 37,* 822–832.

Laughlin, P. R. (1980). Social combination process of cooperative, problem-solving groups at verbal intellective tasks. In M. Fishbein (Ed.), *Progress in Social Psychology* (Vol. 1, pp. 127–155). Hillsdale, NJ: Erlbaum.

Laughlin, P. R., & Earley, P. C. (1982). Social combination models, persuasive arguments theory, social comparison theory, and choice shift. *Journal of Personality and Social Psychology, 42,* 273–280.

Laughlin, P. R., & Ellis, A. L. (1986). Demonstrability and social combination processes on mathematical intellective tasks. *Journal of Experimental Social Psychology, 22,* 17–189.

Laughlin, P. R., Kerr, N. L., Davis, J. H., Halff, H. M., & Marciniak, K. A. (1975). Group size, member ability, and social decision schemes on an intellective task. *Journal of Personality and Social Psychology, 31,* 522–535.

Laughlin, P. R., Kerr, N. L., Munch, M., & Haggerty, C. A. (1976). Social decision schemes of the same four-person groups on two different intellective tasks. *Journal of Personality and Social Psychology, 33,* 80–88.

Laughlin, P. R., & Shippy, T. A. (1983). Collective induction. *Journal of Personality and Social Psychology, 45,* 94–100.

Laughlin, P. R., & Sweeney, J. D. (1977). Individual-to-group and group-to-individual transfer in problem solving. *Journal of Experimental Psychology: Human Learning & Memory, 3,* 246–254.

Levine, J. M., & Resnick, L. B. (1993). Social foundations of cognition. *Annual Review of Psychology, 44,* 585–612.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). New York: Cambridge University Press.

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology, 89,* 46–55.

Lorge, I., & Solomon, H. (1955). Two models of group behavior in the solution of eureka-type problems. *Psychometrika, 20,* 139–148.

MacCoun, R. J. (1990). The emergence of extralegal bias during jury deliberation. *Criminal Justice and Behavior, 17,* 303–314.

MacCoun, R. J., & Kerr, N. L. (1988). Asymmetric influence in mock jury deliberation: Jurors' bias for leniency. *Journal of Personality and Social Psychology, 54,* 21–33.

MacCoun, R. J., & Tyler, T. R. (1988). The basis of citizens' perceptions of the criminal jury: Procedural fairness, accuracy and efficiency. *Law and Human Behavior, 12,* 333–352.

Maheswaran, D., & Chaiken, S. (1991). Promoting systematic processing in low-motivation settings: Effect of incongruent information on processing and judgment. *Journal of Personality & Social Psychology, 61,* 13–25.

McGrath, J. E. (1984). *Groups: Interaction and performance.* Englewood Cliffs, NJ: Prentice-Hall.

McGuire, T. W., Kiesler, S., & Siegel, J. (1987). Group and computer-mediated discussion effects in risk decision making. *Journal of Personality and Social Psychology, 52,* 917–930.

Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry, 1,* 108–141.

Miller, C. E. (1989). The social psychological effects of group decision rules. In P. B. Paulus (Ed.), *Psychology of group influence* (2nd ed., pp. 327–355). Hillsdale, NJ: Erlbaum.

Mowen, J. C., & Gentry, J. W. (1980). Investigation of the preference-reversal phenomenon in a new product introduction task. *Journal of Applied Psychology, 65,* 715–722.

Myers, D. G. (1980). *The inflated self.* New York: Seabury Press.

Myers, D. G., & Kaplan, M. F. (1976). Group-induced polarization in simulated juries. *Personality and Social Psychology Bulletin, 2,* 63–66.

Myers, D. G., & Lamm, H. (1976). The group polarization phenomenon. *Psychological Bulletin, 83,* 602–627.

Nagao, D. H., & Davis, J. H. (1980). Some implications of temporal drift in social parameters. *Journal of Experimental Social Psychology, 16,* 479–496.

Nagao, D. H., Tindale, R. S., Hinsz, V. B., & Davis, J. H. (1985, August). *Individual and group biases in information processing.* Paper presented at the meeting of the American Psychological Association, Los Angeles.

Neale, M. A., Bazerman, M. H., Northcraft, G. B., & Alperson, C. (1986). 'Choice shift' effects in group decisions: A decision bias perspective. *International Journal of Small Group Research, 23,* 33–42.

Nisbett, R. E. (1993). *Rules for reasoning.* Hillsdale, NJ: Erlbaum.

Nisbett, R. E., & Borgida, E. (1975). Attribution and the psychology of prediction. *Journal of Personality and Social Psychology, 32,* 932–943.

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice Hall.

Paese, P. W., Bieser, M., & Tubbs, M. E. (1993). Framing effects and choice shifts in group decision making. *Organizational Behavior and Human Decision Processes, 56,* 149–165.

Page, B. I., & Shapiro, R. Y. (1992). *The rational public: Fifty years of trends in Americans' policy preferences.* Chicago: University of Chicago Press.

Pelham, B. W., & Neter, E. (1995). The effect of motivation of judgment depends on the difficulty of the judgment. *Journal of Personality and Social Psychology, 68,* 581–594.

Penrod, S., & Hastie, R. (1980). A computer simulation of jury decision making. *Psychological Review, 87,* 133–159.

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change.* New York: Springer Verlag.

Plott, C. R. (1986). Rational choice in experimental markets. In R. M. Hogarth & M. W. Reder (Eds.), *Rational choice: The contrast be-*

tween economics and psychology (pp. 117–144). Chicago: University of Chicago Press.

Plous, S. (1995). A comparison of strategies for reducing interval overconfidence in group judgments. *Journal of Applied Psychology, 80,* 443–454.

Rosenwein, R. E., & Campbell, D. T. (1992). Mobilization to achieve collective action and democratic majority/plurality amplification. *Journal of Social Issues, 48,* 125–138.

Schumann, E. L., & Thompson, W. C. (1989). *Effects of attorneys' arguments on jurors' use of statistical evidence.* Unpublished manuscript, University of California at Irvine.

Shafir, E., Simonson, I., & Tversky, A. (1993). Reason-based choice. *Cognition, 49,* 11–36.

Shaw, M. E. (1932). Comparison of individuals and small groups in the rational solution of complex problems. *American Journal of Psychology, 44,* 491–504.

Sniezek, J. A., & Henry, R. A. (1989). Accuracy and confidence in group judgment. *Organizational Behavior and Human Decision Processes, 43,* 1–28.

Stahlberg, D., Eller, F., Maass, A., & Frey, D. (1995). We knew it all along: Hindsight bias in groups. *Organizational Behavior and Human Decision Processes, 63,* 46–58.

Stasser, G., & Davis, J. H. (1981). Group decision making and social influence: A social interaction sequence model. *Psychological Review, 88,* 523–551.

Stasser, G., Kerr, N., & Bray, R. (1982). The social psychology of jury deliberations: Structure, process, and product. In N. Kerr & R. Bray (Eds.), *The psychology of the courtroom* (pp. 221–256). New York: Academic Press.

Stasser, G., Kerr, N. L., & Davis, J. H. (1989). Influence processes and consensus models in decision-making groups. In P. Paulus (Ed.), *Psychology of group influence* (2nd ed.). Hillsdale, NJ: Erlbaum.

Stasser, G., & Stewart, D. (1992). Discovery of hidden profiles by decision-making groups: Solving a problem versus making a judgment. *Journal of Personality and Social Psychology, 63,* 426–434.

Stasser, G., Taylor, L. A., & Hanna, C. (1989). Information sampling in structured and unstructured discussion of three- and six-person groups. *Journal of Personality and Social Psychology, 57,* 67–78.

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during group discussion. *Journal of Personality and Social Psychology, 48,* 1467–1478.

Stasser, G., & Titus, W. (1987). Effects of information load and percentage shared information on the dissemination of unshared information during discussion. *Journal of Personality and Social Psychology, 53,* 81–93.

Stasson, M. F., Ono, K., Zimmerman, S. K., & Davis, J. H. (1988). Group consensus processes on cognitive bias tasks: A social decision scheme approach. *Japanese Psychological Research, 30,* 68–77.

Tanford, S., & Penrod, S. (1984). Social inference processes in juror judgments of multiple-offense trials. *Journal of Personality and Social Psychology, 47,* 749–765.

Taylor, D. W. (1954). Problem solving by groups. *Proceedings of the XIV International Congress of Psychology* (pp. 218–219). Amsterdam: North Holland Publishing.

Thaler, R. H. (1991). *Quasi rational economics.* New York: Russell Sage.

Thibaut, J., & Strickland, L. (1956). Psychological set and social conformity. *Journal of Personality, 25,* 115–129.

Thompson, W. C., Fong, G. T., & Rosenhan, D. L. (1981). Inadmissible evidence and juror verdicts. *Journal of Personality and Social Psychology, 40,* 453–463.

Tindale, R. S. (1992). Assembly bonus effect or typical group performance? A comment on Michaelsen, Watson, and Black (1989). *Journal of Applied Psychology, 77,* 102–105.

Tindale, R. S. (1993). Decision errors made by individuals and groups. In N. J. Castellan (Ed.), *Individual and group decision making: Current issues* (pp. 109–124). Hillsdale, NJ: Erlbaum.

Tindale, R. S., Filkins, J., Thomas, L. S., & Smith, C. M. (1993, November). *An attempt to reduce conjunction errors in decision-making groups.* Paper presented at the annual meeting of the Society for Judgment and Decision Making, Washington, DC.

Tindale, R. S., & Nagao, D. H. (1986). An assessment of the potential utility of "scientific jury selection": A "thought experiment" approach. *Organizational Behavior and Human Decision Processes, 37,* 409–425.

Tindale, R. S., Sheffey, S., & Filkins, J. (1990). *Conjunction errors by individuals and groups.* Paper presented at the meetings of the Society for Judgment and Decision Making, New Orleans, LA.

Tindale, R. S., Sheffey, S., & Scott, L. A. (1993). Framing and group decision-making: Do cognitive changes parallel preference changes? *Organizational Behavior and Human Decision Processes, 55,* 470–485.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211,* 453–458.

Tversky, A., & Kahneman, D. (1983). Extensional vs. intuitive reasoning: The conjunction fallacy in probability judgments. *Psychological Review, 90,* 293–315.

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review, 95,* 371–384.

Vinokur, A., & Burnstein, E. (1973). Effects of partially shared persuasive arguments on group-induced shifts: A group-problem-solving approach. *Journal of Personality and Social Psychology, 29,* 305–315.

Vollrath, D. A., Sheppard, B. H., Hinsz, V., & Davis, J. H. (1989). Memory performance by decision-making groups and individuals. *Organizational Behavior and Human Decision Processes, 43,* 289–300.

Wegner, D. (1986). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior* (pp. 185–208). New York: Springer-Verlag.

Wells, G. L., Miene, P. K., & Wrightsman, L. S. (1985). The timing of the defense opening statement: Don't wait until the evidence is in. *Journal of Applied Social Psychology, 15,* 758–772.

Whyte, G. (1993). Escalating commitment in individual and group decision making: A prospect theory approach. *Organizational Behavior and Human Decision Processes, 54,* 430–455.

Wittenbaum, G. M., & Stasser, G. (1995). The role of prior expectancy and group discussion in the attribution of attitudes. *Journal of Experimental Social Psychology, 31,* 82–105.

Wright, E. F., & Christie, S. D. (1990). *The impact of group discussion on the theory-perseverance bias.* Unpublished manuscript, St. Francis Xavier University.

Wright, E. F., Lüüs, C. E., & Christie, S. D. (1990). Does group discussion facilitate the use of consensus information in making causal attributions? *Journal of Personality and Social Psychology, 59,* 261–269.

Wright, E. F., & Wells, G. L. (1985). Does group discussion attenuate the dispositional bias? *Journal of Applied Social Psychology, 15,* 531–546.

Zajonc, R. B. (1962). A note on group judgments and group size. *Human Relations, 15,* 177–180.

Zajonc, R. B. (1965). Social facilitation. *Science, 149,* 269–274.

Zajonc, R. B., Wolosin, R. J., Wolosin, M. A., & Sherman, S. J. (1968). Individual and group risk taking in a two-choice situation. *Journal of Experimental Social Psychology, 4,* 84–106.

Zanzola, L. (1977). *Effects of pretrial publicity on the verdicts of jurors and juries.* Unpublished master's thesis, Northern Illinois University.

Appendix A

Illustrations of Processes Arising in Interpersonal–Group Context

Group task demands
- Concern with maintaining group harmony
- Impression-management concerns
- Need to achieve consensus (satisfy decision rule)
- Style of deliberation
- Etc. . . .

Effects of group context and capacity on information processing

Group attention
- Distraction
- Social loafing
- Social facilitation
- Etc. . . .

Group encoding
- Social priming
- Social consensus on meaning
- Etc. . . .

Group storage
- Multiple, parallel storage
- Transactive, distributed memory
- Etc. . . .

Group retrieval
- Socially cued recall
- Multiple parallel recall
- Transactive, distributed recall
- Retrieval bias against unshared information
- Etc. . . .

Group processing
- Parallel processing
- Assembly effect bonuses
- Error checking
- Integration of multiple judgments
- Demonstrability of solutions
- Production blocking
- Social comparison of judgments
- Etc. . . .

Group consensus requirements–Processes
- Compliance/conformity
- Compromise alternatives
- Accede to implicit or explicit decision rule
- Etc. . . .

Appendix B

Generating the Asymmetric Social Decision Scheme Matrix

The function used to generate the elements of a strongly asymmetric social decision scheme was

$$d_{iA} = \begin{cases} \alpha - \dfrac{\alpha(\chi - \alpha)^{\beta}}{\alpha^{\beta}} & \text{when } \chi \le \alpha, \\[2ex] \dfrac{(1-\alpha)(\chi - \alpha)^{\beta}}{(1-\alpha)^{\beta}} + \alpha & \text{when } \chi \ge \alpha, \end{cases} \tag{B1}$$

where

$$(r_{iA}, r_{iB}) = (r + 1 - i, i - 1),$$

$$\chi = \frac{(i-1)}{r}, \text{ and}$$

$$\beta = \frac{1}{2K+1}.$$

In these computations, α and K were free parameters, where $0 \le \alpha \le 1.0$ and K could take on a nonnegative integer value. When $\alpha = .50$, the D matrix was symmetric. For $\alpha < .5$, the resulting asymmetry favored alternative G, while $\alpha > .5$ resulted in an asymmetry favoring alternative NG. The value of α also represented the inflection point in the function in Equation 4. In the present study, the strongly asymmetric D matrix used $\alpha = .85$. The smaller the K value, the smoother, less inflected the function in Equation 4. When $K = 0$, this function simply becomes the proportionality decision scheme (i.e., $\beta = 1$ and $d_{iA} = \chi$, the proportion of group members advocating alternative A); as $K \rightarrow \infty$, this function approaches a step-function breaking at α; for example, simple majority is produced when $\alpha = .5$ and $K = \infty$. In the strongly asymmetric D used in this article, $K = 5$.

Received June 9, 1995
Revision received February 8, 1996
Accepted March 8, 1996 ■

Advanced Review

Are groups more rational than individuals? A review of interactive decision making in groups

Tamar Kugler,[1*] Edgar E. Kausel[2] and Martin G. Kocher[3,4,5]

Many decisions are interactive; the outcome of one party depends not only on its decisions or on acts of nature but also on the decisions of others. Standard game theory assumes that individuals are rational, self-interested decision makers—that is, decision makers are selfish, perfect calculators, and flawless executors of their strategies. A myriad of studies shows that these assumptions are problematic, at least when examining decisions made by individuals. In this article, we review the literature of the last 25 years on decision making by groups. Researchers have compared the strategic behavior of groups and individuals in many games: prisoner's dilemma, dictator, ultimatum, trust, centipede and principal–agent games, among others. Our review suggests that results are quite consistent in revealing that group decisions are closer to the game-theoretic assumption of rationality than individual decisions. Given that many real-world decisions are made by groups, it is possible to argue that standard game theory is a better descriptive model than previously believed by experimental researchers. We conclude by discussing future research avenues in this area.

How to cite this article:
WIREs Cogn Sci 2012, 3:471–482. doi: 10.1002/wcs.1184

INTRODUCTION

People make decisions all the time. Furthermore, they make decisions in which outcomes depend not only on what they do or on acts of nature, but also on the decisions of others. Such decisions are called interactive decisions, and economists refer to them as games. If a student sells her old bike to her friend, she plays a price bargaining game. When drivers maneuver their car in heavy traffic, they play a route selection game. If one lends money to a coworker upon request, one plays a trust game. Traditional game theory, the science of rational behavior in interactive settings,[1] makes a few assumptions—mostly based on the concept of the *homo economicus*. First, it assumes that people have complete, exact knowledge of their interests and preferences.[2] Second, rational human beings are assumed to possess the ability to flawlessly calculate what actions would best serve these interests.[3] The third assumption is that people are self-interested, in the sense that they care only about their own material payoff.[4–6] A final assumption in game theory is that of common knowledge; each player knows the rules of the game, that others are also rational and selfish, and that everybody knows that everybody knows the rules, and so on and so forth.[7,8]

If one accepts these assumptions, comparing the behavior of individual decision makers and the behavior of unitary groups[a] seems almost dull. When there is a unique game-theoretic equilibrium or optimal choice, both individuals and unitary groups should follow the normative prediction, and their choices should not differ at all.[b] It is therefore not surprising

*Correspondence to: tkugler@eller.arizona.edu

[1]Department of Management and Organizations, University of Arizona, Tucson, AZ, USA

[2]Faculty of Economics and Business, University of Chile, Santiago, Chile

[3]Department of Economics, University of Munich, Munich, Germany

[4]Department of Economics, University of Gothenburg, Gothenburg, Sweden

[5]CESifo Munich, Germany

that researchers in economics have traditionally overlooked the study of group decision making. For example, the *Handbook of Experimental Economics Results,*[9] devotes no attention to how groups make decisions, *despite* the fact that interactive decisions in the real world are often made by groups. Boards of directors (not individual managers) decide on corporate strategy; congresses (not individual legislators) declare war on other countries; families (not individual family members) decide about budget allocation; and work design in organizations is evolving from an individual task to a group task.[10,11]

Considering the enormous recent body of literature on individual behavior in interactive contexts, it becomes clear that, while traditional game theory is still very useful as a normative theory, it fares less well as a descriptive tool. If game theory is expected to provide a realistic account of human behavior, its assumptions have to be adjusted. One should take into account heterogeneity in levels of rationality,[12–14] different extents of other regarding preferences,[15–18] and different forms of uncertainty attitudes among decision makers.[19] Once these assumptions are integrated into classical game theory, the analysis of unitary group decision making becomes interesting and important. Therefore, investigating group decisions in games has slowly picked up in the late 1990s and after the turn of the century, leading Camerer[20] in his widely used textbook on *Behavioral Game Theory* to conclude that the study of group decisions making is among the top 10 research programs in behavioral and experimental economics.

The main purpose of this review article is to survey the existing results on differences between individuals and groups in interactive tasks. The review reveals that groups tend to behave in these environments in a way that is more rational (as defined by the game-theoretic assumptions described above) than individuals do. Often related to this, groups seem to be more strongly motivated by payoff maximization—although we also refer to the cases where this is not so. Finally, groups seem to be more competitive than individuals—a behavioral tendency that can backfire in certain classes of decision settings.

The remainder of the article is structured as follows. We first discuss briefly some classical comparisons between group and individual decisions in non-strategic situations (games against nature). Next, we center our attention on the focus of this article: a review comparing individual and group decisions in interactive settings. Finally, we discuss the findings and their implications for behavioral game theory, and provide some avenues for future research.

RESEARCH ON GROUP DECISION MAKING IN NON-INTERACTIVE SETTINGS

A significant amount of research has been conducted on decisions against nature (decisions in non-interactive situations). Understanding differences of decisions made by individuals and by groups in non-interactive situations is a prerequisite for discussing differences in interactive settings. However, the literature is so vast that we can only discuss certain important aspects. Many of the studies in non-interactive contexts show that individuals make choices that differ from normative models.[21] This departure from normative models is systematic: decisions made by individuals are routinely biased, as demonstrated convincingly by the heuristics and biases research program.[22] On the basis of this framework, a number of researchers examined whether groups exhibit stronger or weaker biases than individuals do, with mixed findings.[23,24]

Groups tend to do better than individuals in many domains, such as the *hindsight bias* (also known as the 'I-knew-it-all-along' effect[25]), where the bias is significantly smaller for groups compared with individuals,[26] or the *overconfidence* effect,[27] the tendency to be more confident in judgments than what accuracy warrants, where group interaction has been found to reduce the standardized overconfidence effect by 24%.[28] Groups also make fewer errors than individuals in certain *risky choices,*[29] faster and better decisions in uncertain environments,[30] earn more in *risky portfolio decisions,*[31] exhibit less *myopic loss aversion,*[32] and are less prone to *information cascades.*[33] On the other hand, groups are more prone to biases such as the *decoy effect*[34] and *escalation of commitment* (sunk cost fallacy[35]). In some other domains no differences are found between individual and group decisions (*expected utility maximization*[36]; *fund management performance*[37]; *stock market performance*[38]).

The general question of whether groups make riskier choices than individuals is unresolved. Early literature provided evidence that groups tend to polarize individual attitudinal judgment—an effect that is known as the *risky shift.*[39] It contrasts the intuitive conjecture that groups tend to moderate extreme positions and was initially demonstrated in many different settings.[24] More recent research, however, shows inconclusive results. While some find evidence supporting the risky shift,[40,41] other studies found that group decisions are more risk averse than individual decisions,[42–44] or found no differences.[45,46]

A distinction that is particularly useful in the context of non-interactive decisions is the one between *intellective tasks* and *judgmental tasks*.[c] Intellective tasks have a clear ex post evaluation criterion for the quality of performance, whereas judgmental tasks do not. Intellective tasks can be further differentiated with respect to their *demonstrability*,[47] the degree to which the knowledge of the solution to the task is recognized by group members once it is voiced in the group discussion (also referred to as *truth wins*). In intellective tasks, groups typically perform better than individuals.[48,49] This is particularly the case for decision tasks that are easily demonstrable (*eureka* tasks[50]).

There have been several attempts in the literature to identify differences in process that lead to differences in decisions. For example, the concept of *groupthink*[51–53] proposed that, in order to minimize conflict and maintain cohesiveness, group members are less critical in analyzing or assessing ideas. This, in turn, leads to defective decision making. Despite its popularity and intuitive appeal, findings supporting groupthink are sparse.[54–57] Similarly, it has been argued that *group polarization*[58] results during group discussion from self-categorization, group membership, social comparison, or persuasive argumentation.[59–63]

Having briefly reviewed some results in group decision making in non-interactive tasks—with a focus on decisions making under uncertainty—as well as studies on how groups tend to come to their decisions, we now turn our attention to the main part of the paper—a review of empirical studies on group decision making in interactive (strategic) settings.

REVIEW OF STUDIES ON INTERACTIVE DECISION MAKING IN GROUPS

We structure this part of the review around four subsections. In the first, we examine findings in prisoner's dilemma (PD) games, which have been extensively researched. We then present results from ultimatum and dictator games. Next, we examine trust games and other lesser known sequential games. We end this section by reviewing simultaneous games other than the PD game.

PD Games

Insko, Schopler, and their colleagues first examined group behavior in PD games, perhaps because of this game's enormous popularity in previous decades. Nearly all of their studies show that groups defect in PD games more often than individuals.[64–70] The authors identify two primary motives for groups to compete more in a PD game.[71] The *social support of shared self-interest* (or, *greed*) hypothesis argues that groups are greedier than individuals because group members provide each other with support for acting in a selfish, ingroup-oriented way. The *schema-based distrust* (or, *fear*) hypothesis postulates that in contrast to individuals, groups *expect* their opponents to act greedily, and therefore want to protect themselves against the possibility of being exploited. If indeed groups have more negative expectations regarding the behavior of the group that they are interacting with than the individuals' expectations regarding other individuals, then groups are less likely than individuals to cooperate in hope that the opponent will cooperate as well (behavior that results in higher payoff for both players[d]).

An additional motive for groups to compete more in a PD game is the *identifiability* hypothesis, which proposes that in interindividual interactions players assume that they are identifiable and thus can be held 'accountable' if they make a competitive or selfish choice.[72] In intergroup interactions responsibility for a choice is by its very nature obscured. Therefore, group membership provides a chance to evade accountability, and it thus makes it easier for group members to propose and make a competitive choice.

Insko, Schopler, and their colleagues also studied different variations of the PD game, identifying some factors that increase the magnitude of the discontinuity effect.[69] One such factor is (un)constrained communication—the possibility given to different parties to communicate with one another before making their decisions (i.e., allowing intergroup discussion in addition to intragroup discussion). Unconstrained communication is very effective in reducing competition between individuals.[73] However, the schema of distrust of out-groups makes communication less effective and credible, thus reducing its benefits in intergroup interactions.[66] A second factor is procedural interdependence, which refers to the interrelationship between own-group member choices and outcomes. For example, because in a majority rule group members' individual decisions are combined into a collective group decision, they cannot be traced back to the individual members. This creates procedural interdependence among group members. Wildschut and colleagues[74] found that groups that are procedurally interdependent are more competitive, because this feature creates a 'shield of anonymity', facilitating self-interested behavior. This finding is clearly related to the above-mentioned identifiability hypothesis.

The discontinuity effect in the PD game has been replicated and extended by others. Charness et al.[75] found that when group membership is made salient, group members become more competitive. Morgan and Tindale[76] reported that groups behave more competitively than individuals, and that a single group member wanting to defect caused the whole group to defect in over 50% of the cases. Both groups and individuals did not seem sensitive to whether the opponent in the game is a group or an individual. Takemura and Yuki[77] took a cross-cultural perspective and replicated the result in Japan, a society that is believed to be lower in trust than western societies.

Kugler et al.[78] reported a study designed to differentiate between fear and greed as motives for competitive behavior in the PD game. In addition to this game, they compared individual and group behavior in the chicken game and the stag hunt game. The two other games acted as controls, given that in the chicken game greed is a reason to compete while fear leads to cooperating; in contrast, in the stag hunt game only fear is a reason to compete, but greed should lead to cooperative behavior. They found that the discontinuity effect was present to similar extent in the two games that include only one motive for competitive behavior, but that it was significantly stronger when both motives are present (in the PD game). They also reported that the size of the discontinuity effect increases dramatically when individuals and groups are given the possibility to engage in between player (i.e., intergroup) free discussion before the decisions are made.

Ultimatum and Dictator Games

Bornstein and Yaniv[79] compare the behavior of individuals and three-person groups in the ultimatum game.[80] In this game, two players bargain over the allocation of a pie. It is meant to capture a simplified and stylized form of 'take it or leave it' bargaining. The first player (the proposer) proposes an allocation to the second player (the responder), who then gets to either accept the proposal, in which case the allocation takes place as proposed, or reject the allocation, in which case both players get nothing. The game-theoretic prediction (subgame-perfect equilibrium[81]) based on standard assumptions (payoff maximization) states that, given that the responder prefers any positive payoff over zero, and the proposer knows this, she will propose to keep almost everything for herself—offering only the minimal unit to the responder—and the responder will agree to this proposal. It is clear that behavioral findings from individual play do not support this prediction. Proposers offer on average 40% of the pie (with a median of 50%), and responders often reject offers lower than 30%.[82,83] Bornstein and Yaniv demonstrated that groups made and possibly accepted smaller proposals in this game—so group behavior was closer to the rational and selfish (game-theoretic) prediction than individuals. In a similar study, Robert and Carnevale[84] showed that groups made lower offers in the ultimatum game. Further, if group members had the opportunity to participate in the game again, this time as individuals, their offers remained lower. This suggests that the group process changed individual preferences or individual beliefs about the acceptance threshold of the opponents. They also found that the most competitive members of each group had the largest effect on the group's decision (i.e., the group offers were best predicted by the offer made by these individuals). The article focused only on the proposer and therefore the authors did not test whether group responders behaved differently than individual responders. Similar results in a structured voting environment without direct group interaction are provided by Elbittar et al.[85] These results are consistent with the *social support for shared self-interest* explanation and with the *groups are more rational* explanation.

A similar 'game', the dictator game[86] helps distinguishing further between fear and greed as motives for group decisions. This allocation task is similar to the ultimatum game, except for the fact that the responder does not get to accept or reject the allocation offer (so strictly speaking, it is not a game). It is important to note that in this game fear of the opponent should not guide behavior of groups (or individuals), because responders cannot reject the allocation. Therefore, if groups allocate less to others, only selfishness (or greed) can explain the results.

Experiments using the dictator game yield mixed results. Cason and Mui[87] reported a tendency of groups to be more generous in giving than individuals, whereas Luhan et al.[88] found significantly smaller transfers by groups than by individuals playing the role of the dictator. Luhan and colleagues argued that the differences in these findings may be due to two reasons. First, these authors used groups of three members, whereas Cason and Mui used groups of two members. With fewer members per group, the 'shield of anonymity' explanation is reduced, and so is the 'social support of shared self-interest'. Second, Cason and Mui used a procedure where participants could be easily identified (i.e., groups were called to the front of a main room to receive feedback and payment and were then excused to the hallway). As this procedure was common knowledge, it reduced further the effects of the two motives above, and may have enhanced a need to publicly obey social norms of generosity. Finally, while in Cason and Mui the discussion engaged by group members was face-to-face, in the study by Luhan and colleagues it was computer-mediated, increasing anonymity even further.

Trust Games and Other Sequential Games

The trust game,[89] also known as the investment game, is another two-player game that shares some similarities with ultimatum and dictator games. In this game, the first player (the trustor) receives an initial endowment and gets to choose how much of this endowment, if any, to send to a trustee. The amount sent to the trustee is multiplied by a commonly known factor (often tripled) before being given to the trustee. The trustee then gets to return any part of the money back to the trustor. Following a backward-induction logic, the trustee has no reason to return any of the money she receives. Knowing that, the trustor has no reason to send anything to begin with. This game captures a wide-spread definition of trust as 'a psychological state comprising the intention to accept vulnerability based on positive expectations of the intentions or behavior of another' (Ref 90, p. 395). Results from individual behavior indicated that despite the game-theoretic prediction, experimental trustors send on average half of the initial endowment, and trustees return slightly less than what was sent before tripling.[20]

Kugler and colleagues[91] showed that groups of three people sent on average lower amounts than individuals did. They also analyzed asymmetric interaction of individual trustors with group trustees and group trustors with individual trustees. However, the amounts that are sent to individual trustees do not differ significantly from those sent to group trustees. Cox[92] showed that trustees return smaller amounts in this game. Song[93] found that group trustors (using a consensus rule) exhibited lower *psychological trust*—expectations of reciprocity—than individuals, but higher *behavioral trust*—amount sent—when controlling for psychological trust. Song also found that group trustees sent back less money than individual trustees, thus replicating Cox's main result.

The centipede game[94] is a repeated trust game: two players repeatedly bargain over the allocation of an increasing pie. They alternate in deciding whether to stop the game or transfer the decision to the other player. In the standard version of this game, every time the decision is transferred, the size of the pie increases. However, if a player decides to transfer the decision, which could result in the other player stopping the game, then the first player will end up with a lower payoff than she would have gotten, had she stopped the game one step earlier. Given that the game has a finite number of steps, backward induction predicts that the game will stop on the first step, giving both players small payoffs and foregoing a much higher level of overall efficiency. McKelvey and Palfrey[95] reported that for individuals this is rarely the case: only 37 of the 662 games ended with the first player taking the money at the first decision node, while 23 games ended with both players transferring at every node. Bornstein et al.[96] showed that groups stopped the game significantly earlier than individuals do. Once again, this means that group behavior is closer to the game-theoretic prediction. Using a constant-sum variant of the centipede game, they also demonstrated that groups were less altruistic in the game and also less prone to reasoning errors.

The principal–agent game, sometimes called the gift-exchange game[97,98] is modeled to capture the problem of incomplete contracts in the labor market. In the game there are two players: a principal and an agent. The principal determines a wage. In return, the agent decides on an effort level. Effort is costly to agents, but results in increased efficiency and therefore a higher profitability for the principal. As the game is designed in a way that makes it impossible for principals to enforce effort levels, agents are expected to choose the lowest level of effort, once wage is determined. Therefore, principals have no reason to pay agents more than minimal wages.

Contrary to this prediction, Fehr and colleagues[98] find that principals award agents with 42% of the surplus (and payments significantly above the minimal wage), and the average effort chosen by the agents is significantly higher than the effort predicted by standard theory. Kocher and Sutter[99] reported that groups chose lower wages than individuals in the role of principals, but only when communication was computer-mediated—they failed to find differences between individuals and groups who discussed their decisions face-to-face. In terms of the agents' effort, there were no differences between groups whose discussion was computer-mediated and individuals, whereas groups who communicated face-to-face decided on higher effort levels than individuals.

Cooper and Kagel[100] examined group behavior in a signaling game. They showed that groups play more strategically than individuals do. The increased strategic play is a result of the ability of groups to put themselves in the position of another player, and therefore adjust their behavior to the other's strategies. This leads to positive learning transfers, an ability of groups to generalize their learning regarding the game to similar situations with other parameters (i.e., the groups learn more than just the correct behavior, they learn the principles leading to this behavior, and can implement them in related situations). Individuals, on the other hand, exhibited less strategic play and no learning transfer.

Bosman et al.[101] investigated group behavior in the power-to-take game (a variant of the ultimatum game) and report no differences between individuals and groups. In this game, a taxing agency (take authority) decides how much of the endowment of another player (the responder) to take. The responder then gets to choose to agree, or burn his endowment or parts of it. This results in reduced or zero income for the responder and a smaller income for the take authority. Just like in the ultimatum game, game theory predicts that the taxing agency will take all the endowment except for a minimal unit, but experimental results show that takes are lower (on average, 58.5% of the whole endowment), and responders are willing to burn the endowment for large takes: when the taxing agency takes 80% or more of the endowment, the responder typically destroys most of her endowment (62.4% on average[102]). The finding by Bosman and colleagues[101] of no differences between individuals and groups is in contrast to most of the results surveyed above.

Müller and Tan[103] compared the behavior of individuals and groups in a sequential Stackelberg market game. In this game two players sequentially set quantities for production. Both have the same costs of production, and their total output is restricted by the market demand. Interestingly, this study reports behavior of groups to be farther away from the subgame-perfect equilibrium of the stage game than that of individuals. First-mover groups set quantities that are lower than first-mover individuals and lower than predicted by standard theory. There is also research on group versus individual behavior in common pool resource problems. These problems are characterized by the tragedy of the commons (i.e., by a tendency to be overused). Gillet et al.[104] showed that groups are less myopic than individuals in an isolated resource extraction problem, but are more competitive than individuals in a strategic setting, where other users can extract the same resource.

Simultaneous Games

With the exception of the PD game, all the games surveyed above are two-player sequential games (where one player chooses an action first, and the second player observes this action before making her choice). In contrast, Kocher et al.[105,106] and Sutter[107] investigated individual and group decisions in (simultaneous) beauty-contest games. In the beauty-contest game (named after a note by economist JM Keynes who likened the stock market to a beauty contest in one of his famous treatises; also referred to as the guessing game), decision makers simultaneously select a number from 0 to 100. The winner, who receives a fixed prize, is the player who chooses a number closest to p times the mean of the numbers chosen by all participants (p is known to all players beforehand and can range between 0 and 1). The game is then repeated a number of trials, which varies across studies. For $0 < p < 1$, the unique nash equilibrium of the game is zero, which can be obtained by a process of iterated elimination of weakly dominated strategies.[e] The beauty-contest game is commonly used to measure the depth of reasoning of a player and learning dynamics. Usually p is set at 2/3, and the game is repeated four times. The main finding of the papers on the beauty-contest game is that, although individuals and groups do not differ in their choices in the first round, groups choose lower numbers (i.e., closer to equilibrium) than individuals do in rounds 2, 3, and 4—so groups appear to converge to the equilibrium faster than individuals. Furthermore, they found that groups adapt much faster to the feedback regarding the choices of other players. When interacting with individuals, the authors found that groups outperform individuals in terms of payoffs (being able to guess correctly the choices of individuals). Furthermore, larger groups converge quicker to the equilibrium than smaller groups, and less able participants are more likely to select themselves into a decision making group. Nevertheless, despite adverse selection, groups learn faster than individuals also in the self-selection experiments.

Van Vugt et al.[108] investigated an n-person step-level public goods game, which is another example of a simultaneous game. Just like in the PD game, they find that groups cooperate less than individuals do. Cox and Hayne[109] tested the interesting case where rationality, as defined by the game-theoretic prediction, is teased apart from competitiveness. They look at common value auctions, where competitive behavior leads to overbidding (the *winner's curse*) and therefore lower payoffs, and find that groups are more prone to experiencing the winners curse. However, their result emerged only after participants had a chance to gain experience with the task, and only when group members shared the same information. Sutter et al.[110] analyzed laboratory license auctions (a combination of a private value and a common value auction format) with individuals and groups, and their conclusion is similar. Groups are more likely to overbid than individuals. In contrast, Casari et al.[111] reported that in a company takeover experiment groups placed better bids than individuals and substantially reduced the winner's curse. Likewise, Sheremeta and Zhang[112] found less overbidding of groups in Tullock contests than individuals, and Cheung and Palan[113] provided evidence that groups are less prone to create bubbles than individuals on a stock market based on a double auction mechanism.

Another setting that is related to cooperation is coordination. Feri et al.[114] studied six different coordination games, where either individuals or teams interact with each other. They found that teams coordinate much more efficiently than individuals and, thus, are able to achieve higher levels of payoff. In a related coordination game, the stag hunt game, Charness and Jackson[115] showed that the voting rule in the group plays an important role in shaping group choices between the risk-dominant and the payoff-dominant equilibrium.

DISCUSSION

The main purpose of the article was to review a large set of studies on interactive decision making by unitary groups. On the basis of the literature surveyed here, it is fair to conclude that the majority of experimental findings reveal that group behavior in games is more in line with rational and selfish predictions than individual behavior is. Trying to understand the implications of this statement, let us have a closer look at the decision process in an interactive game.

The games used in the reviewed literature are laboratory decision tasks, and participants are not likely to have experience with them. Some of the games are complex, or require processing of substantial amounts of new information. Therefore, the first objective a decision agent faces is to get a full and coherent picture of the decision problem. It is not surprising that groups are superior to individuals in this aspect. Understanding the rules and structure of the game is often an intellective task, and groups are provided with better information processing capabilities, as well as opportunities to catch and correct errors of other group members through discussion—something not available to individuals. It is likely, therefore, that groups understand the structure and rules of the decision tasks better than individuals do.

Once the rules of the games are clear, players have to decide on a strategy. To do so, they first need to construct beliefs regarding the behavior of the other player (or players) in most games and in most roles. Note that this point is unique to interactive decisions, and constructing realistic beliefs is a crucial step in selecting the right strategy. Results supporting the schema-based distrust hypothesis point out that groups may have different beliefs regarding the behavior of other players and expect other players to be greedier. Therefore, fear of the opponent's behavior may cause groups to believe that other players will choose certain strategies. Individuals who are less afraid of the behavior of the other players may have a different probability distribution over the possible acts of others. Overall, the literature is vague regarding the construction of beliefs. Only few of the studies measure beliefs regarding the behavior of the other player explicitly. Kugler and colleagues[91] measured expectations of others' behavior in the trust game and showed that individuals expect higher returns than groups do. Song[93] found a similar result: individuals have higher expectations of others' trustworthiness (i.e., expectations of reciprocity) than groups. Wildschut et al.[116] found that groups are as affected by manipulations of opponents' expectations as individuals. Sutter et al.[117] studied strategic thinking and behavior of individuals and groups in a set of one-shot normal-form games, as well as explicitly elicit beliefs. They found that groups are more likely to play strategically than individuals.

A promising direction for future research is to conduct group experiments with mixed designs—-groups and individuals playing against each other. To the extent that players are sensitive to the nature of their opponents, and expect groups and individuals to behave differently, they should choose different behavioral strategies when facing groups or individuals. Thus one can infer beliefs from actions. For example, using a PD game, Wildschut and colleagures[116] found that of all possible combinations, actions are most competitive in the group-on-group condition; actions are least competitive in the indivdiual-on-individual condition; and group-on-indivdiual conditions are in between. They conclude that the discontinuity effect is a joint function of acting *as a* group and interacting *with a* group. However, both Kugler and colleagues,[91] and Morgan and Tindale[76] failed to find such an effect.

Once players finished analyzing the game structure and considering the opponents' expected behavior and its consequences, they need to decide on their own strategy. At this point, groups differ from individuals not only in the information they accumulated and processed, but also in their aggregated preferences (social or otherwise). The social support of shared self-interest hypothesis supplies one explanation to why groups may have different aggregated preferences than individuals. Specifically, it seems that the dynamics that lead to aggregation of individual preferences into group preferences allow group members to express more greed and less altruism toward the other players, thus making groups more similar to the normative player modeled by standard game theory—a player who cares only about her payoffs, and has no preferences regarding the payoffs of other players involved in the game. Kugler and colleagues[91] sketched a model based on individual social preferences. Specifically, they extended Fehr and Schmidt's[17] inequity aversion model to groups, and argued that based on this model groups are likely to be more selfish.

It is important to qualify the general conclusion that groups are more rational and selfish than individuals. Two exceptions make groups appear sometimes even less rational or selfish than individuals. First, if less selfish behavior can create large profits, and the worst-case payoff is not particularly low, the temptation to secure the larger payoff might take over, even at the risk of not succeeding, and groups might become less selfish. This can occur in games with high potential efficiency gains like the gift-exchange game.[99] However, Bornstein and colleagues[118] presented contrasting evidence in the centipede game, where higher efficiency gains are foregone by groups. Second, groups may become less rational than individuals in highly competitive settings. Auction fever and the proneness to the winner's curse are examples, and groups have indeed been shown to perform worse in auctions than individuals.

It is clear that there is still much that is not understood regarding the process that leads to groups (usually) behaving more rationally and selfishly than individuals in interactive tasks. Future research will have to systematically address many variables before we have a better understanding of the processes underlying this phenomenon. Specifically, we will need to address the impact of variables such as group size, testing whether two-person groups differ from groups of three or more, and what happens when groups are enlarged. They should also be aware of the higher costs of group decision making compared to individual decisions making. Further attention should be paid to within-group interaction and communication, examining the apparent differences between face-to-face communication and computer-mediated communication. Similar attention should be drawn to the official decision rules within groups, investigating whether groups vote, use unanimity rules or have no explicit rules. In addition, there is a need for a better classification of the decision tasks—differentiating between sequential or simultaneous games, two-player and *n*-player games, and other factors such as the complexity of the rules, whether the game requires substantial analysis and strategizing, whether it is played once or repeatedly, and whether there is a possibility to learn over time. Researchers will have to face the task of analyzing group discussion content in order to learn more about group dynamics, and find support for the theoretic claims presented in this section. Content analysis of group discussion is not a trivial task, and therefore not done in most of the studies. Finally, it will be important to develop theoretic models of the group interaction. Economic theory is surprisingly silent about decision making of unitary groups, but ultimately it will be crucial to rigorously model the decision making process of unitary groups.

CONCLUSION

Important decisions are often made by groups that have more previous experience, increased processing capabilities, the ability to monitor each other for mistakes, and share information regarding the task and the expected behavior of others. Therefore, groups (mostly) act as more rational and selfish players, which means that their behavior is more in line with the theoretic predictions. Game theory based on standard assumptions may be, after all, a much better descriptive theory than currently believed.

NOTES

[a] A unitary group is a group that has to come up with a joint decision and does not face any internal conflicts of interests in terms of payoffs.

[b] Naturally, there are situations with multiple equilibria where the type of the decision maker could, in principle, matter also according to traditional game-theoretic analysis, but that would confine the object of study to a very small subset of research questions.

[c] This distinction is more difficult to make for interactive settings, because they typically involve elements of judgmental as well as intellective tasks. For example, forming beliefs regarding the expected behavior of others is close to a judgmental task, while optimizing with respect to those beliefs is more intellective in nature. Therefore, a precise distinction is sometimes hard for games.

[d]We refer to 'players' as decision makers throughout the review in order to avoid confusion. Note that 'players' can either be individuals or groups.
[e]To illustrate this, imagine that a player believes all other players choose randomly over the range of options. She should then choose p times 50, the expected mean, to win the prize. If all players choose this strategy, then a sophisticated player should choose p^2 times 50, and so on, until the only strategy left is to choose zero.

REFERENCES

1. Osborne MJ. *An Introduction to Game Theory, International Edition*. New York, NY: Oxford University Press; 2009.
2. Kreps DM. *A Course in Microeconomic Theory*. Princeton, NJ: Princeton University Press; 1990.
3. Gerrard B. *The Economics of Rationality*. London: Routledge; 1993.
4. Binmore KG. *Does Game Theory Work? The Bargaining Challenge*. Cambridge, MA: MIT Press; 2007.
5. Binmore KG. *Natural Justice*. New York, NY: Oxford University Press; 2005.
6. Gintis H. Behavioral ethics meets natural justice. *Polit Philos Econ* 2006, 5:5–32.
7. Aumann RJ. Backward induction and common knowledge of rationality. *Games Econ Behav* 1995, 8:6–19.
8. Binmore KG. A note on backward induction. *Games Econ Behav* 1996, 17:135–7.
9. Plott CR, Smith VL. *Handbook of Experimental Economics Results*. 1st ed. Amsterdam, The Netherlands: North Holland; 2008.
10. Katzenbach JR, Smith DK. *The Wisdom of Teams: Creating the High-Performance Organization*. Boston, MA: HarperBusiness; 1994.
11. Morgeson FP, Campion MA. Work design. In: Borman WC, Ilgen DR, Klimoski R, eds. *Handbook of Psychology: Industrial and Organizational Psychology*. Hoboken, NJ: Wiley; 2003, 423–452.
12. Costa-Gomes MA, Crawford VP. Cognition and behavior in two-person guessing games: an experimental study. *Am Econ Rev* 2006, 96:1737–1768.
13. Crawford VP, Iriberri N. Level-K auctions: can a nonequilibrium model of strategic thinking explain the winner's curse and overbidding in private-value auctions? *Econometrica* 2007, 75:1721–1770.
14. Nagel R. Unraveling in guessing games: an experimental study. *Am Econ Rev* 1995, 85:1313–1326.
15. Bolton GE, Ockenfels A. ERC: a theory of equity, reciprocity, and competition. *Am Econ Rev* 2000, 90:166–193.
16. Charness G, Rabin M. Understanding social preferences with simple tests. *Q J Econ* 2002, 117:817–869.
17. Fehr E, Schmidt KM. A theory of fairness, competition, and cooperation. *Q J Econ* 1999, 114:817–868.
18. Fehr E, Klein A, Schmidt KM. Fairness and contract design. *Econometrica* 2007, 75:121–154.
19. Abdellaoui M, Baillon A, Placido L, Wakker PP. The rich domain of uncertainty: source functions and their experimental implementation. *Am Econ Rev* 2011, 101:695–723.
20. Camerer CF. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press; 2003.
21. Hastie R, Rasinski KA. The concept of accuracy in social judgment. In: Bar-Ta D, Kruglanski AW, eds. *The Social Psychology of Knowledge*. Cambridge, UK: Cambridge University Press; 1988, 193–208.
22. Gilovich T, Griffin DW, Kahneman D. *Heuristics and Biases: The Psychology of Intuitive Judgement*. Cambridge, UK: Cambridge University Press; 2002.
23. Kerr NL, Tindale RS. Group performance and decision making. *Annu Rev Psychol* 2004, 55:623–655.
24. Kerr NL, MacCoun RJ, Kramer GP. Bias in judgment: comparing individuals and groups. *Psychol Rev* 1996, 103:687–719.
25. Fischhoff B. Hindsight is not equal to foresight: the effect of outcome knowledge on judgment under uncertainty. *J Exp Psychol Hum Percept Perform* 1975, 1:288–299.
26. Stahlberg D, Eller F, Maass A, Frey D. We knew it all along: hindsight bias in groups. *Organ Behav Hum Decis Process* 1995, 63:46–58.
27. Camerer CF, Lovallo D. Overconfidence and excess entry: an experimental approach. *Am Econ Rev* 1999, 89:306–318.
28. Sniezek JA, Henry RA. Accuracy and confidence in group judgment. *Organ Behav Hum Decis Process* 1989, 43:1–28.
29. Charness G, Karni E, Levin D. Individual and group decision making under risk: an experimental study of Bayesian updating and violations of first-order stochastic dominance. *J Risk Uncertain* 2007, 35:129–148.
30. Blinder AS, Morgan J. Are two heads better than one? Monetary policy by committee. *J Money Credit Bank* 2005, 37:789–811.
31. Rockenbach B, Sadrieh A, Mathauschek B. Teams take the better risks. *J Econ Behav Organ* 2007, 63:412–422.

32. Sutter M. Are teams prone to myopic loss aversion? An experimental study on individual versus team investment behavior. *Econ Lett* 2007, 97:128–132.
33. Fahr R, Irlenbusch B. Who follows the crowd - groups or individuals? *J Econ Behav Organ* 2011, 80:200–209.
34. Slaughter JE, Bagger J, Li A. Context effects on group-based employee selection decisions. *Organ Behav Hum Decis Process* 2006, 100:47–59.
35. Teger AI. *Too Much Invested to Quit*. New York, NY: Pergamon Press; 1980.
36. Bone J, Hey J, Suckling J. Are groups more (or less) consistent than individuals? *J Risk Uncertain* 1999, 18:63–81.
37. Prather LJ, Middleton KL. Are N+1 heads better than one? The case of mutual fund managers. *J Econ Behav Organ* 2002, 47:103–120.
38. Barber BM, Heath C, Odean T. Good reasons sell: reason-based choice among group and individual investors in the stock market. *Manag Sci* 2003, 49:1636–1652.
39. Stoner JA. *A Comparison of Individual and Group Decisions Involving Risk*. Unpublished Master's thesis, Boston: Massachusetts Institute of Technology; 1961.
40. Keck S, Diecidue E, Budescu D. Group decision making under ambiguity. *INSEAD* 2011. Available at: http://www.insead.edu/facultyresearch/faculty/personal/ediecidue/research/documents/2011-08-12-GroupDecisions.pdf&ei=JR1NT_2AHcLjiALU2KHCDw&usg=AFQjCNHSUG1BfcFi2OnOp6lleWbOSfqAAg&sig2=RphBuNH-DEjQlF0fm9avog. (Accessed February 28, 2012).
41. Zhang J, Casari M. How groups reach agreement in risky choices: an experiment. *Econ Inq* 2012, 50: 502–515. doi:10.1111/j1465-7295.2010.00362.x.
42. Baker RJ, Laury S, Williams AW. Comparing small-group and individual behavior in lottery-choice experiments. *South Econ J* 2008, 75:367–382.
43. Masclet D, Colombier N, Denant-Boemont L, Lohéac Y. Group and individual risk preferences: a lottery-choice experiment with self-employed and salaried workers. *J Econ Behav Organ* 2009, 70:470–484.
44. Shupp RS, Williams AW. Risk preference differentials of small groups and individuals. *Econ J* 2008, 118:258–283.
45. Brunette M, Cabantous L, Couture S. *Comparing Group and Individual Choices Under Risk and Ambiguity: An Experimental Study*. Nottingham: International Centre for Behavioural Business Research; 2011. Available at: http://ideas.repec.org/p/bbr/workpa/15.html. (Accessed October 3, 2011).
46. Harrison GW, Lau MI, Rutström EE, Tarazona-Gómez MP. Preferences over social risk. *SSRN eLibrary* 2005. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=714922. (Accessed October 3, 2011).
47. Laughlin PR, Ellis AL. Demonstrability and social combination processes on mathematical intellective tasks. *J Exp Soc Psychol* 1986, 22:177–189.
48. Hastie R. Experimental evidence on group accuracy. In: Grofman B, Owen G, eds. *Information Pooling and Group Decision Making*. Greenwich, CT: JAI Press; 1986, 129–157.
49. Levine JM, Moreland RL. *Small Groups: Key Readings*. 1st ed. New York, NY: Psychology Press; 2006.
50. Laughlin PR. *Group Problem Solving*. Princeton, NJ: Princeton University Press; 2011.
51. Janis IL. *Victims of Groupthink*. Boston, MA: Houghton Mifflin; 1972.
52. Janis IL. *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*. 2nd ed. Boston, MA: Houghton Mifflin Company; 1982.
53. Janis IL. *Crucial Decisions*. New York, NY: Free Press; 1989.
54. Flowers ML. A laboratory test of some implications of Janis's groupthink hypothesis. *J Pers Soc Psychol* 1977, 35:888–896.
55. Fodor EM, Smith T. The power motive as an influence on group decision making. *J Pers Soc Psychol* 1982, 42:178–185.
56. Neck CP, Moorhead G. Jury deliberations in the trial of US v. John DeLorean: a case analysis of groupthink avoidance and an enhanced framework. *Hum Relat* 1992, 45:1077–1091.
57. Turner ME, Pratkanis AR. Twenty-five years of groupthink theory and research: lessons from the evaluation of a theory. *Organ Behav Hum Decis Process* 1998, 73:105–115.
58. Krizan Z, Baron RS. Group polarization and choice-dilemmas: how important is self-categorization? *Eur J Soc Psychol* 2007, 37:191–201.
59. Baron R, Kerr N. *Group Process, Group Decision, Group Action*. 2nd ed. Buckingham: Open University Press; 2003.
60. Burnstein E, Vinokur A. Persuasive argumentation and social comparison as determinants of attitude polarization. *J Exp Soc Psychol* 1977, 13:315–332.
61. Isenberg DJ. Group polarization: a critical review and meta-analysis. *J Pers Soc Psychol* 1986, 50:1141–1151.
62. Myers DG. Polarizing effects of social comparison. *J Exp Soc Psychol* 1978, 14:554–563.
63. Turner JC. *Social Influence*. Milton Keynes, UK: Open University Press; 1991.
64. Insko CA, Hoyle RH, Pinkley RL, Hong GY, Slim RM, Dalton B, Lin Y-HW, Ruffin PP, Dardis GJ, Bernthal PR, et al. Individual-group discontinuity: the role of a consensus rule. *J Exp Soc Psychol* 1988, 24:505–519.

65. Insko CA, Schopler J, Drigotas SM, Graetz KA, Kennedy JF, Cox C, Bornstein G. The role of communication in interindividual-intergroup discontinuity. *J Confl Resol* 1993, 37:108.
66. Insko CA, Schopler J, Kennedy JF, Dahl KR, Graetz KA, Drigotas SM. Individual-group discontinuity from the differing perspectives of Campbell's realistic group conflict theory and Tajfel and turner's social identity theory. *Soc Psychol Q* 1992, 55:272–291.
67. Schopler J, Insko CA. The discontinuity effect in interpersonal and intergroup relations: generality and mediation. *Eur Rev Soc Psychol* 1992, 3:121–151.
68. Schopler J, Insko CA, Graetz KA, Drigotas SM, Smith VA. The generality of the individual-group discontinuity effect: variations in positivity-negativity of outcomes, players' relative power, and magnitude of outcomes. *Pers Soc Psychol Bull* 1991, 17:612–624.
69. Wildschut T, Pinter B, Vevea JL, Schopler J, Insko CA. Beyond the group mind: a quantitative review of the interindividual-intergroup discontinuity effect. *Psychol Bull* 2003, 129:698–722.
70. Wolf ST, Insko CA, Kirchner JL, Wildschut T. Interindividual-intergroup discontinuity in the domain of correspondent outcomes: the roles of relativistic concern, perceived categorization, and the doctrine of mutual assured destruction. *J Pers Soc Psychol* 2008, 94:479–494.
71. Schopler J, Insko CA, Drigotas SM, Wieselquist J, Pemberton MB, Cox C. The role of identifiability in the reduction of interindividual-intergroup discontinuity. *J Exp Soc Psychol* 1995, 31:553–574.
72. Rabbie J, Lodewijkx H. Conflict and aggression: an individual-group continuum. *Adv Group Proc* 1994, 11:139–174.
73. Wichman H. Effects of isolation and communication on cooperation in a two-person game. *J Pers Soc Psychol* 1970, 16:114–120.
74. Wildschut T, Lodewijkx HF, Insko CA. Toward a reconciliation of diverging perspectives on interindividual-intergroup discontinuity: the role of procedural interdependence. *J Exp Soc Psychol* 2001, 37:273–285.
75. Charness G, Rigotti L, Rustichini A. Individual behavior and group membership. *Am Econ Rev* 2007, 97:1340–1352.
76. Morgan PM, Tindale RS. Group vs individual performance in mixed-motive situations: exploring an inconsistency. *Organ Behav Hum Decis Process* 2002, 87:44–65.
77. Takemura K, Yuki M. Are Japanese groups more competitive than Japanese individuals? A cross-cultural validation of the interindividual–intergroup discontinuity effect. *Int J Psychol* 2007, 42:27–35.
78. Kugler T, Becker W, Mai M. *Beyond the Discontinuity Effect: Fear, Greed and Competition Between Groups*. Tucson, AZ: University of Arizona; 2012.
79. Bornstein G, Yaniv I. Individual and group behavior in the ultimatum game: are groups more 'rational' players? *Exp Econ* 1998, 1:101–108.
80. Güth W, Schmittberger R, Schwarze B. An experimental analysis of ultimatum bargaining. *J Econ Behav Organ* 1982, 3:367–388.
81. Selten R. Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit. *Zeitschrift für die gesamte Staatswissenschaft* 1965, 121:301–324.
82. Camerer C, Thaler RH. Anomalies: ultimatums, dictators and manners. *J Econ Perspect* 1995, 9:209–219.
83. Thaler RH. Anomalies: the ultimatum game. *J Econ Perspect* 1988, 2:195–206.
84. Robert C, Carnevale PJ. Group choice in ultimatum bargaining. *Organ Behav Hum Decis Process* 1997, 72:256–279.
85. Elbittar A, Gomberg A, Sour L. Group decision-making and voting in ultimatum bargaining: an experimental study. *B E J Econ Anal Policy* 2011, 11:11–53. doi:10.2202/1935-1682.2631.
86. Forsythe R, Horowitz JL, Savin NE, Sefton M. Fairness in simple bargaining experiments. *Games Econ Behav* 1994, 6:347–369.
87. Cason TN, Mui VL. A laboratory study in group polarisation in the team dictator game. *Econ J* 1997, 107:1465–1483.
88. Luhan WJ, Kocher MG, Sutter M. Group polarization in the team dictator game reconsidered. *Exp Econ* 2009, 12:26–41.
89. Berg J, Dickhaut J, McCabe K. Trust, reciprocity, and social history. *Games Econ Behav* 1995, 10:122–142.
90. Rousseau DM, Sitkin SB, Burt RS, Camerer C. Not so different aftger all: a cross-discipline view of trust. *Acad Manag Rev* 1998, 23:393–404.
91. Kugler T, Bornstein G, Kocher MG, Sutter M. Trust between individuals and groups: groups are less trusting than individuals but just as trustworthy. *J Econ Psychol* 2007, 28:646–657.
92. Cox JC. Trust, reciprocity, and other-regarding preferences: groups vs. individuals and males vs. females. In: Zwick R, Rapoport A, eds. *Experimental Business Research*. Boston: Kluwer; 2002, 331–350.
93. Song F. Intergroup trust and reciprocity in strategic interactions: effects of group decision-making mechanisms. *Organ Behav Hum Decis Process* 2009, 108:164–173.
94. Rosenthal RW. Games of perfect information, predatory pricing and the chain-store paradox. *J Econ Theory* 1981, 25:92–100.
95. McKelvey RD, Palfrey TR. An experimental study of the centipede game. *Econometrica* 1992, 60:803–836.
96. Bornstein G, Kugler T, Zamir S. One team must win, the other need only not lose: an experimental study of an asymmetric participation game. *J Behav Decis Mak* 2005, 18:111–123.

97. Charness G. Responsibility and effort in an experimental labor market. *J Econ Behav Organ* 2000, 42:375–384.

98. Fehr E, Kirchsteiger G, Riedl A. Does fairness prevent market clearing? An experimental investigation. *Q J Econ* 1993, 108:437–459.

99. Kocher MG, Sutter M. Individual versus group behavior and the role of the decision making procedure in gift-exchange experiments. *Empirica* 2007, 34:63–88.

100. Cooper DJ, Kagel JH. Are two heads better than one? Team versus individual play in signaling games. *Am Econ Rev* 2005, 95:477–509.

101. Bosman R, Hennig-Schmidt H, van Winden F. Exploring group decision making in a power-to-take experiment. *Exp Econ* 2006, 9:35–51.

102. Bosman R, Van Winden F. Emotional hazard in a power-to-take experiment. *Econ J* 2002, 112: 147–169.

103. Müller W, Tan F. *Team Versus Individual Play in a Sequential Market Game*. Tilburg, The Netherlands: Tilburg University; 2011. Available at: www.tilburguniversity.edu/research/institutes-and-research-groups/center/phd_stud/tan/Mueller_Tan.pdf &ei=KflMT6DoJMXkggf9wPzRAg&usg=AFQjCNF 6Iv5UNX6IMrRvtB8uCJov5qc0sA&sig2=3m3gJyIe 7_eIgAvDwLSMsQ. (Accessed October 3, 2011).

104. Gillet J, Schram A, Sonnemans J. The tragedy of the commons revisited: the importance of group decision-making. *J Public Econ* 2009, 93:785–797.

105. Kocher MG, Sutter M. The decision maker matters: individual versus group behavior in experimental beauty-contest games. *Econ J (London)* 2005, 115:200–223.

106. Kocher MG, Strauß S, Sutter M. Individual or team decision-making–causes and consequences of self-selection. *Games Econ Behav* 2006, 56:259–270.

107. Sutter M. Are four heads better than two? An experimental beauty-contest game with teams of different size. *Econ Lett* 2005, 88:41–46.

108. Van Vugt M, De Cremer D, Janssen DP. Gender differences in cooperation and competition: the male-warrior hypothesis. *Psychol Sci* 2007, 18:19–23.

109. Cox JC, Hayne SC. Barking up the right tree: are small groups rational agents? *Exp Econ* 2006, 9:209–222.

110. Sutter M, Kocher MG, Strauß S. Individuals and teams in UMTS-license auctions. *Oxford Econ Papers* 2009, 61:380–394.

111. Casari M, Zhang J, Jackson C. *When Do Groups Perform Better than Individuals? A Company Takeover Experiment*. Bologna, Italy: University of Bologna; 2011. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1873267. (Accessed October 3, 2011).

112. Sheremeta RM, Zhang J. Can groups solve the problem of over-bidding in contests? *Soc Choice Welfare* 2010, 35:175–197.

113. Cheung SL, Palan S. Two heads are less bubbly than one: team decision-making in an experimental asset market. *Exp Econ*. Forthcoming. doi:10.1007/s10683-011-9304-6.

114. Feri F, Irlenbusch B, Sutter M. Efficiency gains from team-based coordination—large-scale experimental evidence. *Am Econ Rev* 2010, 100:1892–1912.

115. Charness G, Jackson MO. Group play in games and the role of consent in network formation. *J Econ Theory* 2007, 136:417–445.

116. Wildschut T, Insko CA, Pinter B. Interindividual-intergroup discontinuity as a joint function of acting as a group and interacting with a group. *Eur J Soc Psychol* 2007, 37:390–399.

117. Sutter M, Czermak S, Feri F. *Strategic Sophistication of Individuals and Teams in Experimental Normal-Form Games*. University of Innsbruck; 2010. Available at: http://econpapers.repec.org/paper/innwpaper/2010-02.htm. (Accessed February 28, 2012).

118. Bornstein G, Kugler T, Ziegelmeyer A. Individual and group decisions in the centipede game: are groups more 'rational' players? *J Exp Soc Psychol* 2004, 40:599–605.

FURTHER READING

Baron R, Kerr N. *Group Process, Group Decision, Group Action*. Buckingham: Open University Press; 2003.

Camerer CF, Loewenstein G, Rabin M. *Advances in Behavioral Economics*. Princeton, NJ: Princeton University Press; 2003.

Journal of Economic Behavior & Organization 70 (2009) 470–484

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/econbase

Group and individual risk preferences: A lottery-choice experiment with self-employed and salaried workers

David Masclet[a,b,*], Nathalie Colombier[a], Laurent Denant-Boemont[a], Youenn Lohéac[a,c]

[a] *Department of Economics, CREM, University of Rennes, 7 place Hoche, 35065 Rennes, France*
[b] *CIRANO, 2020 Rue University, Montreal, Quebec, H3A 2A5 Canada*
[c] *ESC Bretagne Brest, 2 avenue de Provence, CS 23812, 29238 Brest Cedex 3, France*

ARTICLE INFO

Article history:
Received 14 April 2007
Received in revised form 30 October 2007
Accepted 28 November 2007
Available online 22 January 2009

JEL classification:
C91
C92
C93
D81
D70

Keywords:
Decision under risk
Individual decision
Group decision
Self-employment

ABSTRACT

This paper focuses on decision making under risk, comparing group and individual risk preferences in a lottery-choice experiment. In the *individual* treatment, subjects make choices individually; in the *group* treatment, each subject placed in a group made lottery choice via voting. In the *choice* treatment, subjects choose whether to be on their own or in a group. The originality of this research lies in the fact that we introduced variability in socio-demographic characteristics by recruiting salaried and self-employed workers. Our main findings indicate that groups are more likely than individuals to choose safe lotteries. Our results also show that individuals risk attitude is correlated with both the type and the sector of employment.

1. Introduction

In many real life situations, important decisions are made by (small) groups such as production units, boards of directors, committees rather than by a single individual. This then raises the question of how the preferences of different group members are combined to produce the group decision. In spite of the fact that many important decisions are made collectively, economics has devoted little empirical attention to group decision-making. In this paper, we contribute to this literature by comparing group and individual decision-making. More precisely, we focus on decision-making under risk and compare group and individual risk preferences in a lottery-choice experiment inspired by Holt and Laury (2002). In this seminal paper, Holt and Laury used the results of a simple lottery choice experiment to determine the degree of risk aversion. Subjects were successively confronted with the following treatments: a real lottery with low payments (less than four Euros in both outcomes), a hypothetical lottery with high payments (the low payment outcomes multiplied by 20, 50 or 90), a real lottery with the same high payments, followed by the same real lottery with low payments as at the start of the sequence.

* Corresponding author at: Faculté de Sciences Economiques-CREM, Université de Rennes, 7 place Hoche CS 86514 35065 Rennes, France. Tel.: +33 2 23 23 33 18; fax: +33 2 23 23 35 09.
E-mail address: David.Masclet@univ-rennes1.fr (D. Masclet).

0167-2681/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jebo.2007.11.002

Holt and Laury's most important results are that subjects exhibit risk aversion even for low payments and that risk aversion increases sharply as the scale of payoff increases (for real payoffs only).

In our experiment, the risk preferences of groups and individuals are compared by implementing three treatments over eight independent sessions. In the individual treatment (Ind), subjects were asked to choose between playing two lotteries, one "safe" and one "risky", with varying probabilities of obtaining the higher monetary payoff. In the group treatment (Group), each individual was placed in an anonymous group of three and voted over which lottery was chosen. If no unanimous decision was reached in the vote, players were informed of other group members' choices in the current vote, and then voted again. The voting rounds continued until agreement was reached or until five rounds were completed. If five rounds were completed without agreement, then the lottery option was randomly chosen by the computer. Finally, in a third treatment, called the choice treatment (Choice), subjects were asked to state a maximum willingness to pay for making their decisions alone instead of choosing in the group of three people (and thus express their preference over the first two treatments).

The originality of our research lies in the fact that we introduced variability in socio-demographic characteristics by recruiting "real people", including not only students who are typically viewed as the standard subject pool used by experimenters, but also self-employed workers and salaried workers. Indeed, student samples exhibit limited variability in some key characteristics such as age or occupation that may be highly correlated with risk attitude. However, as Harrison and List (2004, p. 1009) noted, these last years, "more and more experimentalists are recruiting subjects in the field rather than in the classroom."[1] Introducing variability in sociodemographic characteristics among subjects allows us to investigate whether contextual effects are robust to the introduction of sociodemographic variables. In addition, it allows us to compare the relative influence of contextual (individual versus group decisions, prior experience, simultaneous versus sequential context) and non-contextual variables (sociodemographic variables) on risk decision. Are individuals more likely to be influenced in their decision by the context or by their intrinsic individual characteristics?

Moreover, do sociodemographic characteristics interact with these contextual variables? Our experiment seeks to provide a first experimental evidence of the link between risk attitude and employment status. In fact, several theoretical research studies emphasize the importance of unobservable factors such as attitudes toward risk and preferences for autonomy in the decision between self-employment and working for others. Partly drawing on Knight's (1921) classic work, Kihlstrom and Laffont (1979) and Rees and Shah (1986) posit that less risk adverse individuals are more likely to choose self-employment. In addition, models by Rees and Shah (1986) and Blanchflower and Oswald (1998) examine other aspects of self-employment such as "the flexibility associated with hours worked and the independence entailed," and "the non pecuniary utility from being independent and one's own boss" (Blanchflower and Oswald, 1998, p. 31). However, there exists very little empirical evidence on the importance of these characteristics in the self-employment decision. In particular, we do not know whether attitudes toward risk or preferences for autonomy play a major role or only a minor role relative to those of human and social environment. In a recent study, using data on Finns born in 1966, Ekelund et al. (2005) found that risk-seekers are significantly more likely to choose self-employment. However in contrast to this paper, in our study, the direction of causality is from self-employment to risk attitude.

The main findings of our study are, consistent with previous work, that groups exhibit more risk aversion than individuals for high-risk lotteries. In addition, our results indicate a further explanation for group decision-making by showing that relative risk-loving subjects (those who are less risk-averse than the other two group members) are more willing to change their vote to conform to the group average risk decision than were relatively risk-averse players. Finally, apart from the context, our results show that a large part of risk attitude is explained by socio-demographic characteristics. In particular, individuals' risk attitude seems to be strongly correlated with both the type and the sector (private or public) of employment. Those who are self-employed tend to be significantly less risk averse than others. In addition salaried workers employed in the private sector tend to take significantly more risk than salaried workers from the public sector.

The remainder of this paper is organized as follows. Section 2 summarizes the relevant previous research comparing groups and individuals. Our experimental design is presented in more detail in Section 3, and Section 4 presents and interprets the results of the experiment. Finally Section 5 summarizes and concludes.

2. Previous literature

A number of empirical results based on natural data concerning team versus individual decisions can be found in the existing literature (financial decisions in Prather and Middleton, 2002, productivity in Hamilton et al., 2003, and betting in Adams and Ferreira, 2007), but the majority of results have come from experimental economics. A recent, growing experimental literature has explored differences between individuals and teams (or between teams of different size) with respect to many different kinds of decisions: beauty-contest games (Kocher and Sutter, 2005; Kocher et al., 2006; Sutter, 2005),

[1] For example, Smith et al. (1988) conducted a large series of experiments not only with student subjects but also with professional and business people from the Tucson community as subjects. Another example are the experiments of Cummings et al. (1995), who used individuals recruited from churches in order to obtain a wider range of demographic characteristics than one would obtain in the standard college setting. Blondel et al. (2007) compared risk aversion and time preference of drug users and non-drug users in order to identify some differences.

centipede games (Bornstein et al., 2004), ultimatum games (Bornstein and Yaniv, 1998), dictator games (Cason and Mui, 1997), signaling games (Cooper and Kagel, 2005), policy decisions (Blinder and Morgan, 2005), location and pricing (Barreda et al., 2002), and auctions (Cox and Hayne, 2006; Sutter et al., 2008). Experimental results on the type of choice that interests us here, risky decisions, can be found in Bone (1998), Bone et al. (1999, 2004), Shupp and Williams (2008), Baker et al. (2008), Bateman and Munro (2005), Rockenbach et al. (2007), and Harrison et al. (2007). The main issues considered in this literature are whether teams make better decisions and whether they are more rational than individuals. No consensus has been reached regarding either question, with results depending on the kind of game under consideration. This conclusion is similar to that reached in social psychology regarding differences in group and individual decisions. In their meta-analysis of replies to Choice Dilemma Questionnaires, Kerr et al. (1996, p. 693) stress that group discussion can "attenuate, amplify, or reproduce the judgment biases of individuals depending on the group decision making process". Rockenbach et al. carried out an experiment where individuals and groups (not consisting of the same subjects) make lottery choices and evaluations. The common effects observed in the literature regarding expected utility theory (the common ratio and preference reversal effects) are found for both individuals and groups (as in Bone et al., 1999). However, teams accumulated significantly more expected value than did individuals, and at significantly lower total risk. In an experiment comparing the risk preferences of two real spouses, both separately and together, Bateman and Munro replicated the result of equal rationality in group and individual decisions mentioned above, but found that joint choices are more risk averse than those made by individuals. Shupp and Williams evaluate risk aversion via certainty equivalent ratios (certainty equivalent/expected value) elicited using a maximum willingness to pay mechanism for lotteries. They find that groups exhibit lower risk aversion than individuals for lotteries with high winning probabilities, but are increasingly risk-averse as winning probabilities fall. Comparing the decisions of the same subjects both alone and in groups, Shupp and Williams stress that group discussion led to greater risk-aversion for lotteries with low winning probabilities. Using Holt and Laury's method (with payoffs raised by a factor of 10), Baker et al. observe the same phenomena as Shupp and Williams. Last, in a paper on preferences over social risk, Harrison et al. also appeal to the same method (with payoffs raised by a factor of 25) and conclude that there are no differences in risk aversion between individuals and groups (consisting of the same subjects). However the general conclusion of this literature on risky decisions with few exceptions (Harrison et al.) is that groups tend to be more cautious than individuals.

Many of the above experiments consider the individual and group treatments independently (i.e. the same individual participates in only one of the two treatments). As such, we cannot examine the behavior of the same individual in two different decision environments. However, if the same subject participates in both treatments, a new problem arises: the order of the treatments. Shupp and Williams include the same subjects in two treatments but do not control for order effects. Baker et al. use an individual-group-individual sequence of decisions and find that subjects were more risk-averse (for high-risk lotteries) in groups than in the first individual treatment and that the group treatment significantly affected the decision in the second individual treatment: individuals exhibited greater risk-aversion than in first individual treatment.

A second point is that the great majority of group decisions in this experimental literature are based on informal discussion (cheap talk) so that the decision-making process within the group cannot be analyzed. It would be very interesting to open the "black-box" to discriminate between the many hypotheses regarding how groups make decisions. The data in Rockenbach et al. are consistent with an excess-risk vetoing rule. In a signaling game experiment, Cooper and Kagel introduced discussion between the two (anonymous) team members via an instant messaging system that recorded discussions. Their analysis of the dialogue between team members leads them to conclude that teams exhibit strong positive cross-game learning whereas individuals show negative cross-game learning, which is consistent with adaptive learning models with a growing number of sophisticated learners.

Compared to the existing literature, we analyze differences in risk aversion between individual and group (three-member) treatments, composed of the same subjects. We take the order effect between these two treatments into account and look into the "black-box" of the group to analyze the decision process leading to unanimous choices. Last, we propose a new approach to the analysis of the taste for autonomy.[2]

3. The experimental design

The experimental procedure is based on that of Holt and Laury. The experiment was computerized and the scripts were programmed using the z-tree platform (Fischbacher, 2007). We recruited 144 subjects among students, salaried workers and self-employed workers. Roughly 43% of our participants were salaried workers or self-employed. The remaining subjects were students who constituted our benchmark population in the experience. The students were recruited from undergraduate courses in business, literature and economics at the University of Rennes (France). None of the subjects had participated in an economics experiment previously. The salaried workers were recruited by phone or by email from public and private

[2] Kocher et al. (2006) appears to be the first contribution considering the taste for autonomy in decision-making. Before individual or group decisions, Kocher et al. (2006) asked individuals to choose between the two decision procedures and to explain their choice. In their experimental beauty-contest game, about 60% of subjects preferred to act in teams (and teams won significantly more often than did individuals). Their analysis of the causes and consequences of self-selection showed that both individuals and team members were satisfied with their chosen role, but for different reasons.

Table 1
Standard payoff matrix.

Decision	Option A				Option B			
	Prob. p	Payoff	Prob. $(1-p)$	Payoff	Prob. p	Payoff	Prob. $(1-p)$	Payoff
1	10%	40 euros	90%	32 euros	10%	77 euros	90%	2 euros
2	20%	40 euros	80%	32 euros	20%	77 euros	80%	2 euros
3	30%	40 euros	70%	32 euros	30%	77 euros	70%	2 euros
4	40%	40 euros	60%	32 euros	40%	77 euros	60%	2 euros
5	50%	40 euros	50%	32 euros	50%	77 euros	50%	2 euros
6	60%	40 euros	40%	32 euros	60%	77 euros	40%	2 euros
7	70%	40 euros	30%	32 euros	70%	77 euros	30%	2 euros
8	80%	40 euros	20%	32 euros	80%	77 euros	20%	2 euros
9	90%	40 euros	10%	32 euros	90%	77 euros	10%	2 euros
10	100%	40 euros	0%	32 euros	100%	77 euros	0%	2 euros

sectors. Finally, self-employed workers were recruited among self-employed farmer, artisan, shopkeeper and professional workers with the help of the Chamber of Commerce of Rennes.[3]

Our overall design consists of eight sessions (with 18 subjects each) of a lottery choice experiment with three treatments. Our first treatment, called the "individual treatment", is based on 10 sequential choices between two lotteries, one "risky" (with payoffs of €77 and €2) and one "safe" (with payoffs of €40 and €32), with probabilities ranging from 10% to 100% (see Table 1). As noted by Holt and Laury, the payoffs for the safe lottery (Option A) are less variable than those for the risky lottery (Option B). In both options the probabilities for the first of the 10 sequential decisions are 10% for the high payoff and 90% for the low payoff. The difference in the expected payoffs between the two lotteries is such that only an extreme risk-seeker would choose Option B. As the probability of the high payoff outcome increases B becomes more attractive relative to A, and at some point subjects will switch their preference. Towards the end of the decision sequence, even the most risk averse subjects should switch over to option B. Contrary to Holt and Laury, the ten decisions were not presented simultaneously, as in Table 1, but shown sequentially and randomly. The "individual" (sequential) treatment consists of 10 successive periods, with a different decision each period. This procedure allows us to measure the differences between group and individual decision-making for each of the ten individual decisions. To test whether introducing sequential framing may affect decisions, we also had a variant of the individual treatment with a simultaneous framing, labeled "simultaneous individual treatment". This treatment is identical to the simultaneous high payoff treatment presented in Holt and Laury.

In the "group" treatment, subjects were placed in anonymous groups of three and presented with the same 10 decisions as in the individual treatment. After each period, the groups were randomly reshuffled. Group members voted over lotteries to try to reach a unanimous decision. If a unanimous decision was not reached, players continued to another vote after being informed of the votes of the other group members. Voting continued until unanimous agreement was reached or until five voting rounds were completed. If no agreement was reached after five votes, the option was randomly chosen by the computer.

The "choice" treatment consists of two stages within each period. In the first stage, each individual is endowed with 10 units (with 2 units corresponding to 1 euro) and is asked how much she would be prepared to pay to make her lottery choice alone, with the proviso that only the three individuals with the highest bids (among the 18 players of the session) will be allowed to play the individual treatment while the others play the group treatment. The price paid by each winner corresponded to the fourth highest bid. In the second stage, subjects were asked to choose between options A and B, alone or by group, depending on the outcome of the previous stage.

In sessions 1–4, subjects initially undertook the individual treatment, followed by the Group treatment and the Choice Treatment. To account for potential order effects, as noted by Harrison et al., we ran two additional sessions (sessions 5 and 6) with a different sample of subjects that began with the group treatment followed by the individual treatment. Finally, we ran two additional sessions (sessions 7 and 8) to test whether presenting the 10 decisions sequentially instead of simultaneously induces a potential "framing effect". In sessions 7–8, subjects initially played the 10 decisions of the simultaneous individual treatment, followed by the sequential individual treatment.

At the end of the experiment, the outcome of each treatment was determined by the random selection of a single decision for each treatment. To control for wealth effects, subjects were informed that only one of the two treatment payoffs would be chosen for the payment at the end of the experiment. On average, a session lasted for about an hour and 20 min, including the initial instructions and payment of subjects. Each participant earned €45 on average plus a lump sum of €3.

[3] Salaried workers were recruited via posters, by phone, and among parents of students to take part in an experiment in economics. Recruitment of self-employed workers involved contacting the Chamber of Commerce of Rennes and the "Club des créateurs et repreneurs d'entreprise d'Ille et Vilaine", an economic club of entrepreneurs sponsored by the Chamber of Commerce, who helped us in contacting potential participants by emails. All participants were recruited "topic blind". Hence participants did not know that the focus of the experiment would be on risk attitude. Experiments were conducted in the same way with students and non-students. Finally there were no differences in payoffs and fees between participants.

Table 2
Risk aversion classification based on lottery choices.

Number of safe choices	Range of relative risk aversion $U(x) = (x^{1-r}/(1-r))$	Risk preference classification	Proportion of choice		
			Indiv treat (1)	Group treat (2)	Choice treat (3)
0–1	$r < -0.95$	Highly risk loving	0.93	–	–
2	$-0.95 < r < -0.49$	Very risk loving	0.93	–	–
3	$-0.49 < r < -0{,}15$	Risk loving	0.93	0.68	–
4	$-0.15 < r < 0.15$	Risk neutral	10.19	0.87	–
5	$0.15 < r < 0.41$	Slightly risk averse	10.19	4.36	4.17
6	$0.41 < r < 0.68$	Risk averse	24.07	31.30	13.89
7	$0.68 < r < 0.97$	Very risk averse	25.93	36.82	45.83
8	$0.97 < r < 1.37$	Highly risk averse	7.41	18.22	27.78
9–10	$1.37 < r$	Stay in bed	19.45	7.75	8.33

4. The experimental results

4.1. Individual and group decisions

Table 2 provides interesting information on the lottery choice frequencies for all treatments. Consistent with Holt and Laury's results, it indicates that in all treatments, most of players are risk averse and choose on average more than four safe options. Table 2 also indicates differences between treatments. The proportion of safe choice is higher in the group and choice treatments than in the individual treatment. For example 36.8% and 45.8% of subjects chose seven safe options, respectively in the group and choice treatments, while this proportion was only of 25.9% in the individual treatment.

Fig. 1a shows the proportion of A choices in sessions 1–4 for each of the 10 decisions listed in Table 1. Fig. 1b displays the corresponding data for sessions 5–6. The horizontal axis represents the decision number, which corresponds to the probability of the higher payoff. The dashed line shows predicted behavior under risk neutrality: A is chosen for the first four decisions, and subsequently B.

In Fig. 1a and b, the percentage choosing the safe option A falls as the probability of the higher payoff increases. The average numbers of "safe" choices (option A) for the individual treatments are 6.6 and 6.5, respectively, for sessions 1–4 and 5–6. Further, groups tend to report higher levels of risk aversion for most of the decisions, except for lotteries with high probability of the larger payoff where groups are actually less likely to choose the safe lottery A. This is consistent with the result of Baker et al. (2008). The average numbers of "safe" choices for the group treatment are 7.11 and 7.03, in sessions 1–4 and 5–6, respectively. A Mann–Whitney test on the total number of "safe" lottery choices over the first 10 periods rejects the null hypothesis of equal means between the individual and group treatments ($z = -1.892$; $p = 0.058$). A similar test over periods 11–20 produces similar results ($z = -1.864$; $p = 0.0623$). These results indicate that groups are more likely than individuals to choose safe lotteries for decisions with low winning percentages. The average number of safe choices is 7.2 in the Choice treatment.[4] A Wilcoxon matched-pairs signed-rank test for differences between the number of safe choices in the group and choice treatments finds no significant difference ($z = -1.628$). Finally, a Wilcoxon rank-sum test cannot reject the null hypothesis of equal distributions between the two individual treatments ($z = 0.120$) as well as between the two group treatments ($z = 0.363$), which shows that "prior experience" has no significant effect.

Fig. 1c shows the proportion of A choices in sessions 7–8 for the simultaneous and sequential individual treatments. It indicates that both the simultaneous and sequential individual treatment show the same patterns: the percentage choosing the option A falls as the probability of the higher payoff increases. The average numbers of "safe" choices are 6.5 and 6.6, respectively for the simultaneous and the sequential individual treatments. A Mann–Whitney test on the total number of "safe" lottery choices over the ten decisions accepts the null hypothesis of equal means between these two treatments ($p > 0.1$).[5]

Table 3 provides a more formal support for these results. It shows the results of a random effect probit model, using "safe choice" (lottery A) as the dependent variable.[6,7] The right-hand side variables include the probability of winning the larger amount (0.1–1.0) and dummy variables for group treatment, choice treatment, prior experience and framing effect. We also included an interaction variable "Group*prob" between the group variable and the winning probability and an interaction variable called "choice*win the auction" between the choice variable and the fact of winning the auction in the choice treatment.

4 The average number of safe choices under the choice treatment is 6.9 for individuals who decided alone.

5 One difference between the simultaneous and sequential individual treatments is the higher proportion of subjects switching back from B to A more than once. However in both treatments, this proportion is rather low.

6 The variable "safe choice" takes the value 1 if the lottery A is chosen for a given decision and zero otherwise. In all estimates, except estimate (6), this variable is not corrected for inconsistent choices (i.e. multiple switches). In contrast, in estimate (6), we corrected for inconsistent choices by considering the first switch only in the analysis. Finally, we also ran additional estimates excluding multiple switchers (available on request). Both estimates provide results similar to those obtained without correction.

7 In all estimates except estimate (5), the variable "safe choice" corresponds to the group's final response in the group treatment. In estimate (5) this variable indicates the group's initial responses instead of the group's final responses.

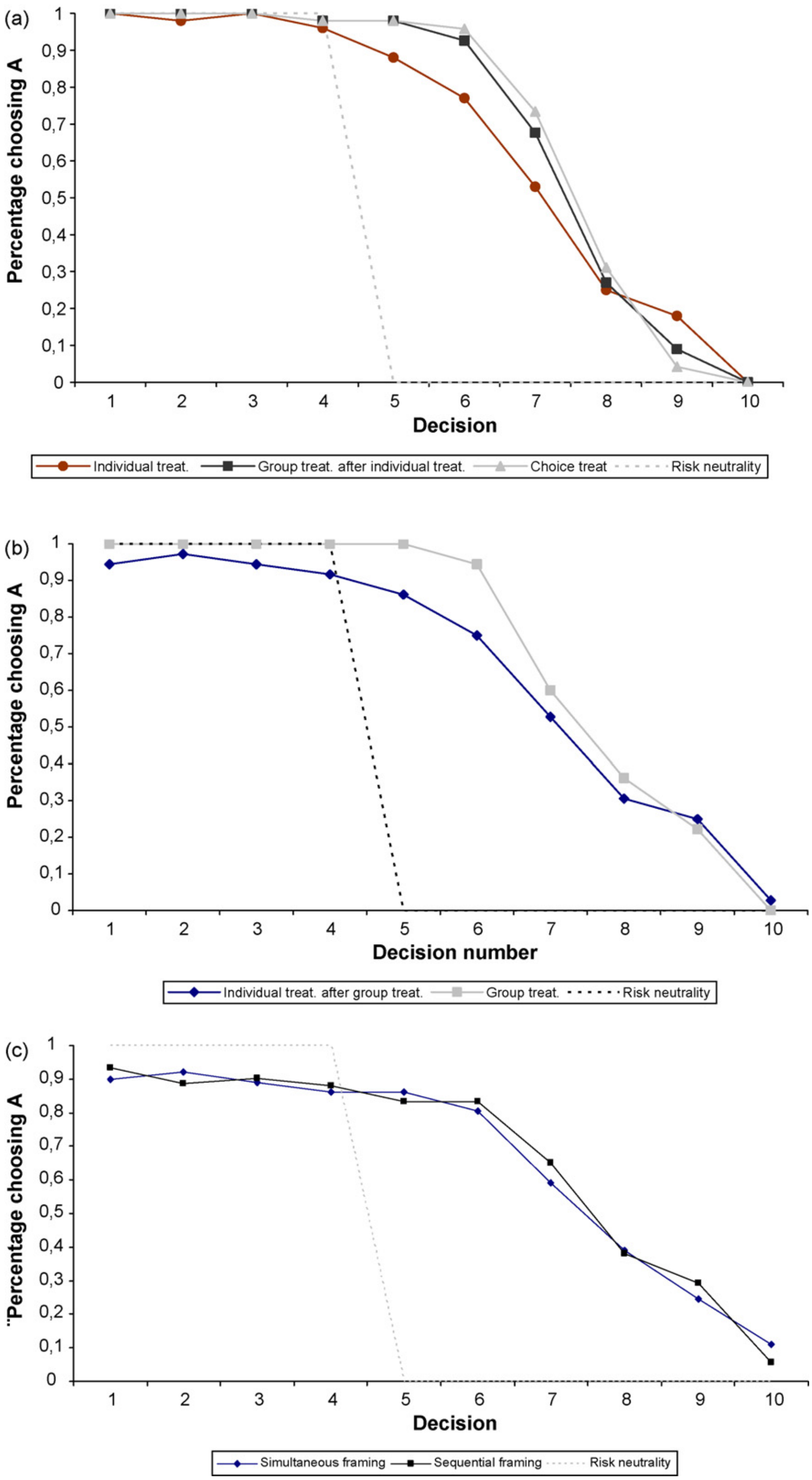


Fig. 1. (a) The proportion of safe choices in each decision (sessions 1–4). (b) The proportion of safe choices in each decision (sessions 5–6). (c) The proportion of safe choices in each decision (sessions 7–8).

Table 3
The probability of safe choice: random effects probit: contextual variables.

	Indiv. and group treatments (sessions 1–6)						Ind, group and choice treatments (sessions 1–6)		Ind treatments sessions 7–8
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
p (higherpayoff)	−0.069***	−.0904***	−0.074***	−0.062***	−0.057***	−0.090***	−0.079***	−0.079***	−0.045***
	(0.0036)	(0.008)	(0.003)	(0.003)	(0.0023)	(0.005)	(0.003)	(0.003)	(0.0033)
Group	0.363***	0.651***	0.432***	5.204***	−0.028	6.440***	0.476***	0.466***	
	(.0111)	(0.186)	(0.095)	(0.644)	(0.082)	(0.851)	(0.087)	(0.097)	
Prior exper.			−.0108	−0.002	−0.093	−0.034			
			(0.192)	(0.212)	(0.176)	(0.278)			
Interaction:				−0.064***		−0.080***			
Group × prob				(0.008)		(0.011)			
Choice treat.								0.166	
								(0.127)	
Choice × win auct								0.308	
								(0.255)	
Simul Fram.									−0.0022
									(0.133)
Constant	4.904***	6.169***	5.229***	4.377***	4.092***	6.431***	5.622***	5.605***	3.203***
	(0.270)	(0.578)	(0.259)	(0.269)	(0.194)	(0.445)	(0.238)	(0.239)	(0.292)
Numb of ind	72	36	108	108	108	108	108	108	36
Numb of obs	1440	720	2160	2160	2160	2160	2790	2790	720
Log-Like.	−362.58	−167.07	−539.05	−496.72	−669.95	−401.26	−648.62	−645.40	−265.23
Sigma_u	0.536	1.406	0.8159	0.894	0.786	1.338	0.822	0.825	1.014
	(0.088)	(0.208)	(0.082)	(0.091)	(0.076)	(0.128)	(0.081)	(0.08)	(0.108)
Rho	0.223	0.664	0.399	0.444	0.382	0.641	0.403	0.405	0.507
	(0.223)	(0.066)	(0.048)	(0.050)	(0.045)	(0.044)	(0.047)	(0.047)	(0.053)

Standard errors in parentheses.
*** Significant at 1% 0.405.

Columns (1) to (4) reveal that the probability of safe choice falls as the probability of the higher payoff increases, and increases when decisions are collective, suggesting the importance of decisions made in a group and underlining the importance of context.[8] The "prior experience" variable is not significant, which confirms our previous results. Finally, the estimated coefficient on the interaction variable "Group*prob" shows that groups become progressively more risk-averse as the probability of the higher payoff falls. Column (6) indicates that similar results are obtained when one considers only the first risky switch in the analysis. Similar results were also obtained excluding multiple switchers (available on request). Columns (7) and (8) show similar results when including the choice treatment in the analysis. Controlling for selection treatment in estimate (8) indicates that selection has no significant effect. Finally, estimate (9) shows that presenting the 10 decisions sequentially instead of simultaneously does not induce any significant framing effect. In the next sub-section, we investigate whether such context effects are robust to the introduction of demographic variables.

4.2. The role of demographic variables in risk decisions

Are previous results affected by the introduction of socio-demographic variables in the analysis? Moreover does risk strongly vary across individuals? Our results indicate a strong heterogeneity among individuals. We observed that the self-employed tend to report a lower level of risk aversion than other populations for most of the decisions. The average numbers of safe choices for the individual treatment are 5.5, 6.7, and 6.6, respectively for self-employed workers, salaried workers, and students. A Wilcoxon rank-sum indicates that the self-employed significantly choose less safe lottery than students ($p < 0.05$). A similar test also indicate that self-employed take more risk than salaried workers ($p < 0.1$). Finally this test indicates no differences between salaried workers and students ($p > 0.1$). This result confirms other empirical analysis (Ekelund et al.).

A detailed analysis of the data also reports differences among salaried workers depending on the choice of the sector (public or private) of employment. Private sector workers report on average 6 safe choices against 7.4 for the public sector employees. A Wilcoxon rank-sum test rejects the null hypothesis of equal distributions between public and private sector employees ($p < 0.1$).

In summary, these results indicate that both the type and the sector of employment are significant determinants of choosing the safer option. To check, we estimated a probit model in Table 4, which yields a measure of the effect of each socio-demographic variable on the probability of choosing a safe option. Indeed it might also be possible that differences among salaried workers, self-employed workers, and students reflect in fact other demographic differences such as gender, age, occupation or education differences.

The estimates include several standard demographics (age, gender, marital statute, education) and some dummy variables to control for the type and the sector (public private) of employment.

Table 4 shows that standard demographic variables have no significant effect on the probability of choosing the safe option. In contrast, variables concerning the choice of type (self/paid) and sector (public/private) of employment significantly affect the decision of choosing the safe option. Both estimate (1) and (2) indicate that self-employed workers tend to be less risk averse than others.[9] In estimate (3), the coefficient associated with the variable «self-employed» is also negative and significant (as opposed to being a salaried worker, which is the omitted category). Turning next to salaried workers, estimate (4) shows that public sector employees are more likely than private sector employees to choose less risk options. Last, estimate (5), (6), and (7) indicate that introducing demographic variables does not change the influence of contextual variable.

4.3. Voting decisions and the determinants of collective choice

We now focus on voting decisions in the group treatment. The vote procedure consists of five rounds of voting. If no unanimous decision was reached during a vote, players went on to the next round of voting after being informed of choices of the other group members in the previous vote. The rounds continued until agreement was reached or until five rounds had been completed. If five votes were completed without agreement, then an option was randomly chosen by the computer. We first consider the evolution of disagreements within groups. Disagreement occurs when the group makes decisions unanimously and when some group member deviates from the average decision. A number of configurations are possible. First, the subject who disagrees is more risk-loving than the other group members and chooses lottery B while the other two group members choose lottery A. Second, the subject is more risk-averse if he chooses lottery A and the other two subjects choose lottery B. There are also two intermediate situations: the subject is weakly more risk averse if he chooses lottery A (or weakly more risk-loving if he chooses lottery B) and one of the two other group members chooses lottery B (lottery A). We then analyze the determinants of collective choice. Fig. 2a and b display the evolution of disagreements within groups in each round of voting and for each of the ten decisions listed in Table 1, for sessions 1–4 and 5–6.

[8] Estimate (5) indicates no significant differences between the individual responses and initial responses in the group treatment. This additional result compared to previous results using final decisions indicates that collective negotiation needs several iterations to converge to a less risky decision.

[9] We acknowledge that such variables should be interpreted cautiously because of a potential endogeneity problem. Indeed, one alternative interpretation of this result is that risk-seeking types would tend to choose self-employment with the implication that self-employment would be endogenous in the choice model.

Table 4
The probability of safe choice: random effects probit: non-contextual variable.

	Individual treatment (sessions 1–8)				Indiv. and group treatments (sessions 1–6)	Indiv., group and choice treatments (sessions 1–6)	All sessions and all treatments
	All indiv	All indiv	Salaried and SE only	Salaried workers only	All indiv	All indiv	All indiv
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Socio-demographic variables							
Men	−0.134	−0.134	−0.262	−0.1073	−0.127	−0.184	−0.201
	(0.190)	(0.190)	(0.211)	(0.245)	(0.197)	(0.205)	(0.229)
Age in years	−0.002	−0.002	−.0042	−.0196*	0.006	0.001	0.002
	(0.008)	(0.011)	(0.009)	(0.011)	(0.015)	(0.014)	(0.009)
Married/couple	−0.208	−0.209	−0.1718	−0.1663	−0.287	−0.311	−0.217
	(0.381)	(0.388)	(0.407)	(0.828)	(0.414)	(0.355)	(0.353)
Graduate	0.169***	0.169***	0.138**	0.0749	0.158*	0.123	0.174
	(0.065)	(0.066)	(0.059)	(0.077)	(0.092)	(0.077)	(0.035)
Major is literature	−0.156	−0.156			−0.281	−0.351	−0.034
	(0.395)	(0.395)			(0.464)	(0.485)	(0.425)
Major is busi/eco.	0.153	0.153			0.427*	0.317	0.318
	(0.200)	(0.201)			(0.213)	(0.222)	(0.218)
Salaried worker		−0.0034			0.330	0.299	−0.065
		(0.275)			(0.450)	(0.394)	(0.249)
Public sector employ.				0.838***			
				(0.2610)			
Self-employed w.	−0.671**	−0.673*	−0.668**		−0.727*	−0.661**	−0.697*
Contextual variables	(0.338)	(0.383)	(0.299)		(0.382)	(0.310)	(0.369)
p (higher payoff)	−0.056***	−0.056***	−0.0405***	−.0396***	−0.0625***	−0.063***	−0.055***
	(0.002)	(0.002)	(0.0031)	(0.003)	(0.003)	(0.003)	(0.002)
Group					4.153***	4.620***	5.075***
					(0.533)	(0.473)	(0.457)
Prior experience					0.0094	−0.088	−0.0456
					(0.227)	(0.231)	(0.248)
Interaction:					−0.051***	−0.056***	−0.063***
Group × proba.					(0.007)	(0.006)	(0.005)
Constant	3.727***	3.71***	2.861***	3.869***	3.783***	4.207***	3.485***
	(0.394)	(0.399)	(0.474)	(0.702)	(0.492)	(0.425)	(0.430)
Observations	1440	1440	620	480	2160	2880	3600
Number of ind	144	144	62	48	108	108	144
Log-Likelihood	−489.62	−489.62	−255.40	−192.678	−521.05	−635.86	−911.96
Sigma_u	0.913	0.913	0.627	0.577	0.841	0.863	0.957
	(0.094)	(0.094)	(0.110)	(0.124)	(0.087)	(0.087)	(0.084)
Rho	0.455	0.455	0.282	0.249	0.414	0.428	0.478
	(0.051)	(0.051)	(0.0711)	(0.080)	(0.050)	(0.049)	(0.043)

Standard errors in parentheses.

* Significant at 10%.

** Significant at 5%.

*** Significant at 1%.

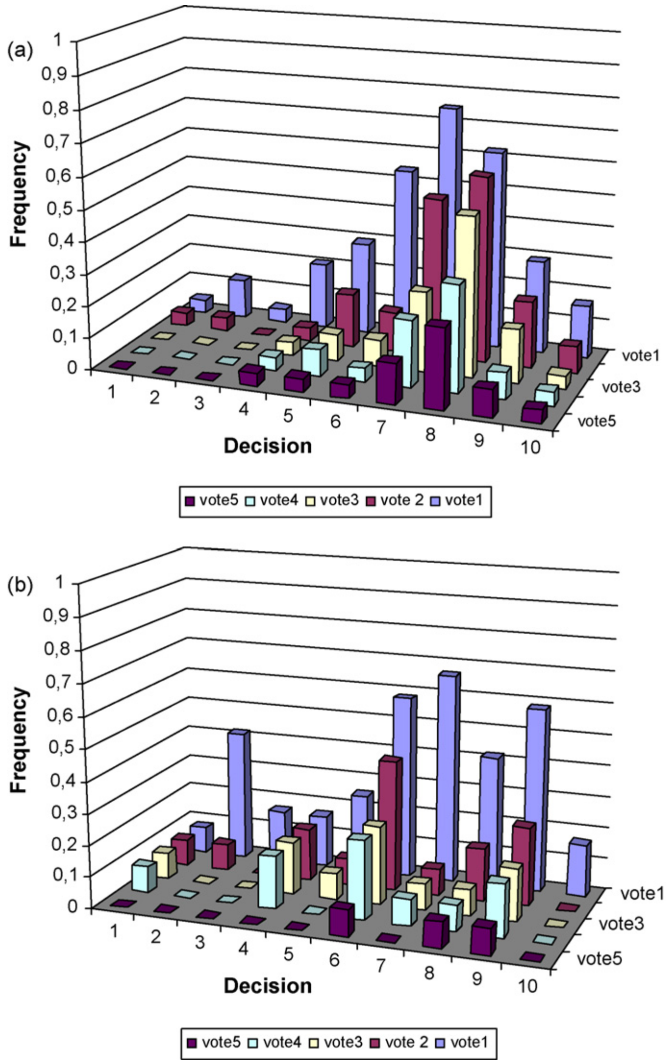


Fig. 2. (a) Frequency of disagreements for each decision in treatment 4 (sessions 1–4). (b) Frequency of disagreements for each decision in treatment 4 (sessions 5–6).

As expected, unanimous group decisions were more difficult for the intermediate probabilities (decisions 5–8). The figures also show that the probability of disagreement decreases with the number of voting rounds. For example, the probability of disagreement is 75% in vote 1 of decision 7 in sessions 1–4 and decreases to 12% in vote 5. Groups therefore required several rounds of voting to reach unanimous decisions. Most unanimous decisions involved the safe lottery (A), and the average number of A lotteries chosen increases with the number of votes. For example, the probability of choosing lottery A increases from 66% in vote 1 in sessions 1–4 (63% in sessions 5–6) to 70% (71%) in vote 5. This result is of interest because it suggests that decision-making under the unanimity rule produces safer choices.

One possible reason for this result might be that risk lover players would be less reluctant than others to converge to a safer choice. To investigate in more detail this possibility, we considered to what extent relative risk lovers were less reluctant than others to change their vote by considering the relationship between the probability of changing a vote between two rounds of voting and the individuals' risk attitude relative to the group's average risk attitude. The groups defined on the horizontal axis are determined as described above.

Fig. 3 shows that the probability of changing a decision between two rounds of voting depends on the relative risk attitude compared to the rest of the group. The probability of changing a decision is greater when the two other group members have voted for the same lottery. Fig. 3 also shows that relative risk lovers are more likely to move to a less risky choice than are the relatively risk-averse to move to a more risky choice.

Table 5 provides a formal support of this result via a random effects probit on the probability of changing a decision between two votes. The key independent variable is "relative risk lover", which equals one if the subject chose the risky option while at least one of the other group members chose the safer lottery. The coefficient on this variable is interpreted

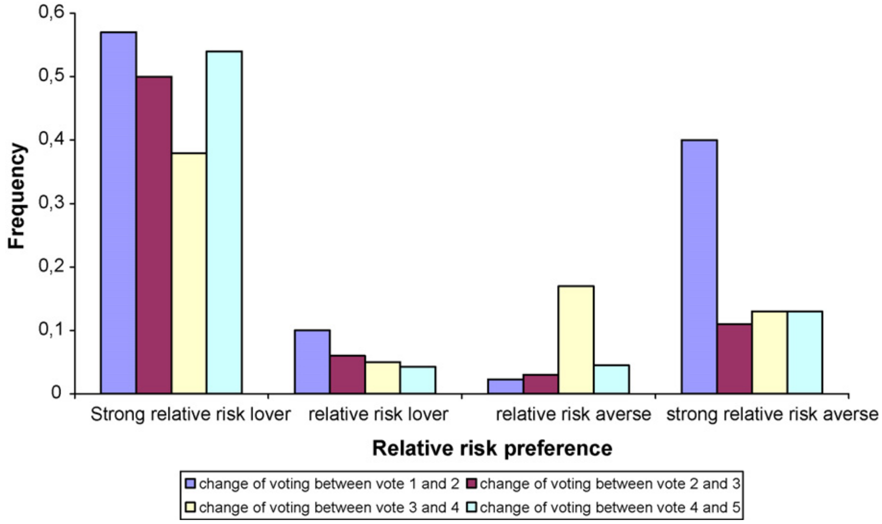


Fig. 3. Change of vote decision as a function of relative risk aversion.

Table 5
Changing lottery decision between two votes: random effects probit.

	(1)	(2)
Relative risk lover	1.0766***	1.031***
	(0.2925)	(0.295)
Men		−.0981
		(0.141)
Age in years		0.0112
		(0.007)
Married/couple		.0693
		(.241)
Graduate		−0.0724
		(0.052)
Self-employed		0.0402
		(0.225)
Constant	−0.661***	−0.7059***
	(0.089)	(0.218)
Observations	476	476
Log likelihood	−291.00	−289.07
Sigma_u	0.347	0.285
	(0.109)	(0.124)
Rho	0.107	0.075
	(0.060)	(0.06)

Standard errors in parentheses.
*** Significant at 1%.

in relation with the omitted variable "relative risk averse". The second specification includes additional variables to control for demographics.

The results in Table 5 confirm our previous results. The estimated coefficient on "relative risk lover" is positive and significant at the 5% level so that the probability of switching is higher if the subject is more risk loving than other group members. Relative risk lovers change their votes to safer options more often than the relatively risk averse change their vote to riskier options. This result indicates that groups converge toward less risky decisions because subjects who were relatively less risk averse were more likely to change their vote in order to conform to the group average decision.[10]

[10] This result can be related to the existing literature on conformity. Jones (1984) presented an economic theory of conformity where peoples' tendency to conform persists even after an initial "social influence" is removed. Persistence of social conformity is explained by tradition and internalization of social values. Bernheim (1994) also presented a model of conformity, assuming that individuals care about intrinsic utility but also about status. When status is sufficiently important relative to intrinsic utility, some individuals may be willing to conform to a standard behaviour (social norm) because a small departure from this standard may seriously impair their status. Some studies have also investigated the importance of conformity in the context of team production (Kandel and Lazear, 1992; Barron and Gjerde, 1997). These studies show how conformity influences cooperation within teams through peer pressure. Peer pressure refers to a psychological pressure felt by agents when they compare their action with the actions taken by theirs colleagues. Peer pressure leads individuals to conform to social norms. Finally, in a recent study, Levitt and List (2007) investigated the relationship between conformity and

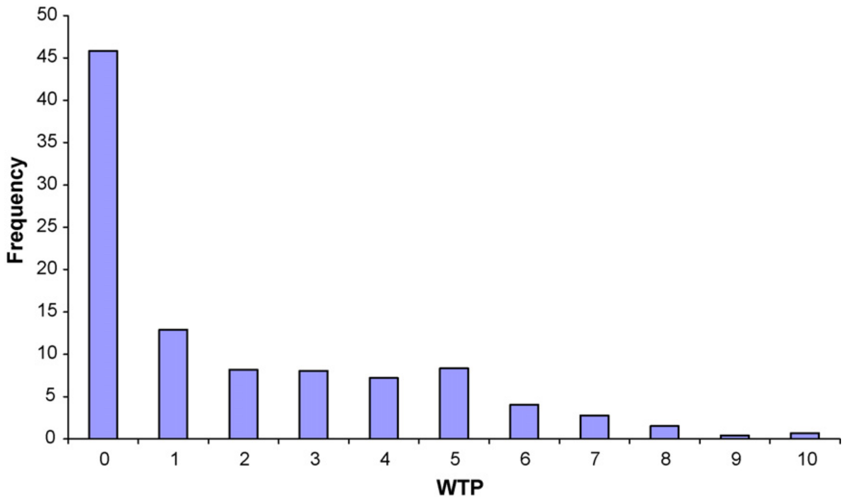


Fig. 4. Frequency of bids.

4.4. Willingness to pay (WTP) and risk aversion

In this section, we consider the determinants of bidding in the first stage of treatment "Choice". Remember that in this treatment, only the three players with the highest bids (among the 18 players of the session) are allowed to play the individual treatment while the others play the group treatment. Fig. 4 presents the frequency of bids in the first stage of the choice treatment.

In 45% of cases subjects chose the minimum bid of 0 units and fewer than 1% chose the maximum bid of 10 units. The average bid is 1.9 units over all participants, with figures of 5.71 units for those who decided individually and 1.14 units for those who decided in groups. Table 6 provides a formal analysis of the determinants of WTP, using a Tobit model to control for censoring. The right-hand side variables include several socio-demographic variables and a variable measuring risk aversion (the number of safe choices under the individual treatment). Finally we also include a variable controlling for previous conflicts within groups by considering whether the lottery was randomly chosen by the computer in the previous period in the case of disagreement.

Table 6 shows that several socio-demographic variables are significant. Both men and older individuals are more likely to propose a higher bid. Interestingly, individuals who are married/in couple would be less likely to propose a higher bid. A possible interpretation of this result is that married/in couple people are more likely than others to take group decisions. Finally, we find a positive and significant coefficient associated with the variable "self-employed worker", indicating that self-employed workers' bids are significantly higher. This result is consistent with the interpretation in term of willingness to decide alone as suggested by Rees and Shah (1986) and Blanchflower and Oswald (1998) who consider "the flexibility associated with hours worked and the nonpecuniary utility from being independent and one's own boss" as strong determinants of self-employment (Blanchflower and Oswald, 1998, p. 31). Last, Table 6 shows that risk aversion is negatively and significantly related to current bids, indicating that less risk averse individuals are more willing to escape from the tyranny of group decision-making, especially since they tend to compromise more frequently in the group treatment.

5. Conclusion and discussion

The main findings of our study are that both context and socio-demographic variables significantly influence the choice of risky options. Our results indicate that age, gender, or marital statute do not significantly influence the probability of choosing the safe option. In contrast, both the type and the sector (private or public) of employment seem to influence risk decisions significantly. Our results show that the self-employed report lower level of risk aversion than other individuals for most of the decisions. In addition, consistent with previous literature, we also observe that public sector employees are generally more likely than private sector employees to choose the safer option (Bellante and Link, 1981). As it is argued by Bozeman and Kingsley (1998), risk avoidance is not necessarily a determinant of job choice but may be a consequence of remuneration schemes. Indeed employees who have low expectation that good performance will be rewarded (in the public

---

moral concerns. Levitt and List presented a model that assumes that individual choices depend not only on financial implications but also on non-monetary moral costs (or benefits) that may vary across people and that may be also influenced by several factors such as context or scrutiny. In particular, moral concerns may depend on the process by which the decision is reached (negotiation, discussion, vote, etc.. . .).

Table 6
WTP for deciding alone: random effects Tobit regressions.

Socio-demographic variables	
Men	0.473*
	(.281)
Age in years	0.101***
	(0.0101)
Married/couple	−1.762***
	(0.406)
Graduate	−0.0285
	(0.0967)
Self-employed worker	2.587***
	(0.378)
Contextual variables	
Risk aversion	−0.135***
	(0.036)
Random vote in $t-1$	0.495
	(0.368)
Constant	2.587
	(0.3788)
Nb. Obs	720
Nb. Ind.	72
Log likelihood	−1039.03
Left-censored	330
Sigma_u	3.00
	(0.173)
Rho	2.05
	(0.078)

Standard errors in parentheses.
* Significant at 10%.
*** significant at 1%.

sector) may tend to perceive lesser risk taking than employees who have high expectation that good performance will be rewarded (in the private sector).

Turning to the influence of contextual variables, our results indicate that decisions are influenced neither by prior experience nor by the framing of the experience (sequential or simultaneous framing, i.e. the way of presenting the decisions). On the contrary, our data reveal that groups are more likely to choose safe lotteries than are individuals. Introducing socio-demographic variables does not change these effects.

We then provide new insights on group decision-making by showing that relative risk-lovers (subjects who are less risk averse than the other two group members) were more likely to change their position than were relatively risk averse players who were reluctant to change their position. Finally, our results indicate that less risk averse individuals (including self-employed workers) were more likely than others to propose a higher bid to escape from the tyranny of group decision-making.

Our results about the importance of context and socio demographic variables can be interpreted in relation to the recent paper of Levitt and List. In this paper, the authors show that individual choices depend not only on financial concerns but also on the context in which decisions are embedded, the way the participants are selected, and the nature and extent of scrutiny. These dimensions have particular implications for lab experiments that seek to investigate differences across groups.

There are number of possible explanations for the group decision making. We have presented our results in terms of willingness to conform to the group. Alternatively, group decision-making may derive from strategic or non-strategic motives (altruism, fairness). While it is difficult to distinguish cleanly between theories, we note that our experimental design rules out strategic reasons since all participants were rematched after each period. Also social preferences cannot explain why voting rounds differ from one another in probability of disagreement.[11]

Our findings are of interest in the context of previous researches. As discussed in the paper, previous results on the effect of collective decision-making on risky choice are mixed. Harrison et al. find no group effect, while Baker et al. indicate that groups are on average more likely than individuals to choose safe lotteries for low winning probabilities. A possible explanation for these differences is that the rule by which decisions are reached is different. In both our work and that of Baker et al., subjects were asked to make unanimous decisions, inducing less risk-averse subjects to converge toward less risky decisions. On the contrary, in Harrison et al., each group voted for the lottery he preferred under a majority voting rule.

[11] Another alternative explanation of our experimental results is that the background risk of choosing randomly an option if no agreement is obtained after vote 5 might interact with individual risk aversion. However, if this were the case, one should observe a higher level of disagreement in the last voting round compared to the previous rounds. On the contrary, our results indicate that the probability of disagreement decreases over time with the number of voting rounds, which is more consistent with our conjecture of a dynamic conformity to the group.

There are grounds to expect that choices will be different under the unanimity rule than under the majority rule. Indeed it seems reasonable to assume that the unanimity rule induces more pressure toward uniformity in groups than the majority rule. Moreover, the alternatives over which the group decides may also be influenced not only by the decision rule but also by the amount of information individuals have about one another's preferences. These are likely fruitful areas for additional research on group decision-making. Concerning socio-demographic variables, our results can be related to previous research on occupational choice. Indeed some theoretical research showed the importance of unobservable factors such as attitudes toward risk or preferences for autonomy in the occupational choice (Kihlstrom and Laffont, 1979; Rees and Shah, 1986). However, there exists very little empirical evidence on the importance of these characteristics in the employment decision. One reason is that such characteristics are generally unobservable. In this regard, additional experiments may provide new empirical evidence that the value placed on the stability of employment depends to some extent on the individual's degree of risk aversion.

Acknowledgments

We are grateful to Glen Harrison, Jayson Lusk, Claude Montmarquette, the participants at the ESA North American Meeting (Tucson) and the participants at the Risk Attitude Workshop. We are also grateful to two anonymous referees. Thanks to Elven Priour for programming the experiment presented in this paper, to the ANR Risk Attitude (Agence Nationale de la Recherche) for a grant to support this research and to Pierre-Jean Richard from the "Club des créateurs et repreneurs d'entreprise d'Ille et Vilaine" and the Chamber of Commerce of Rennes for help in recruiting self-employed participants.

References

Adams, R.B., Ferreira, D., 2007. Moderation in groups: evidence from betting on ice break-ups in Alaska. Available at SSRN: http://ssrn.com/abstract=594501.
Baker, R.J., Laury, S.K., Williams, A.W., 2008. Comparing group and individual behavior in lottery-choice experiments. Southern Economic Journal 75, 367–382.
Barreda, I., Gallego, A.G., Georgantzis, N., Andaluz, J., Gil, A., 2002. Individual vs. group behaviour in discrete location-and-pricing experiments. LINEEX Working Paper 37/02.
Barron, J.M., Gjerde, K.P., 1997. Peer pressure in an agency relationship. Journal of Labor Economics 15, 234–254.
Bateman, I., Munro, A., 2005. An experiment on risky choice amongst households. Economic Journal 115, C176–C189.
Bellante, D., Link, A., 1981. Are public sector workers more risk averse than private sector workers? Industrial and Labor Relations Review 34, 408–412.
Bernheim, D.K., 1994. A theory of conformity. Journal of Political Economy 102, 841–877.
Blanchflower, D.G., Oswald, A.J., 1998. What makes an entrepreneur? Journal of Labor Economics 16, 26–30.
Blinder, A., Morgan, J., 2005. Are two heads better than one? An experimental analysis of group versus individual decision making. Journal of Money, Credit and Banking 37, 789–812.
Blondel, S., Loheac, Y., Rinaudo, S., 2007. Rational decision of drug users: an experimental approach. Journal of Health Economics 26, 643–658.
Bone, J., 1998. Risk-sharing CARA individuals are collectively EU. Economic Letters 58, 311–317.
Bone, J., Hey, J., Suckling, J., 1999. Are groups more (or less) consistent than individuals? Journal of Risk and Uncertainty 8, 63–81.
Bone, J., Hey, J., Suckling, J., 2004. A simple risk-sharing experiment. Journal of Risk and Uncertainty 28, 23–38.
Bornstein, G., Yaniv, I., 1998. Individual and group behavior in the ultimatum game: are groups more "rational" players? Experimental Economics 1, 101–108.
Bornstein, G., Kugler, T., Ziegelmeyer, A., 2004. Individual and group behavior in the centipede game: are groups (again) more rational players? Journal of Experimental Social Psychology 40, 599–605.
Bozeman, B., Kingsley, G., 1998. Risk culture in public and private organizations. Public Administration Review 58, 109–118.
Cason, T.N., Mui, V.-L., 1997. A laboratory study of group polarisation in the team dictator game. Economic Journal 107, 1465–1483.
Cooper, D.J., Kagel, J.H., 2005. Are two heads better than one? Team versus individual play in signaling games. American Economic Review 95, 477–509.
Cox, J.C., Hayne, S.C., 2006. Barking up the right tree: are small groups rational agents? Experimental Economics 9, 209–222.
Cummings, R.G., Harrison, G.W., Rutström, E.E., 1995. Homegrown values and hypothetical surveys: is the dichotomous choice approach incentive compatible? American Economic Review 85, 260–266.
Ekelund, J., Johansson, E., Jarvelin, M.-R., Lichtermann, D., 2005. Self-employment and risk aversion—evidence from psychological test data. Labour Economics 12, 649–659.
Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Experimental Economics 10, 171–178.
Hamilton, B.H., Nickerson, J.A., Owan, H., 2003. Team incentives and worker heterogeneity: an empirical analysis of the impact of teams on productivity and participation. Journal of Political Economy 111, 465–497.
Harrison, G.W., List, J.A., 2004. Field experiments. Journal of Economic Literature 42, 1009–1055.
Harrison, G.W., Lau, M.I., Rutström, E.E., Tarazona-Gomez, M., 2007. Preferences over social risk. Working Paper 05-06. Department of Economics, College of Business Administration, University of Central Florida.
Holt, C.A., Laury, S.K., 2002. Risk aversion and incentive effects. American Economic Review 92, 1644–1655.
Jones, S., 1984. The Economics of Conformism. Basil Blackwell, Oxford/New York.
Kandel, E., Lazear, E.P., 1992. Peer pressure and partnerships. Journal of Political Economy 100, 801–817.
Kerr, L.N., MacCoun, R.J., Kramer, G.P., 1996. Bias in judgment: comparing individuals and groups. Psychological Review 103, 687–719.
Kihlstrom, R.E., Laffont, J.-J., 1979. A general equilibrium entrepreneurial theory of firm formation based on risk aversion. Journal of Political Economy 87, 719–748.
Knight, F., 1921. Risk, Uncertainty and Profit. Houghton Mifflin, Boston.
Kocher, M., Sutter, M., 2005. The decision maker matters: individuals versus group behaviour in experimental beauty-contest games. Economic Journal 115, 200–223.
Kocher, M., Strauβ, S., Sutter, M., 2006. Individual or team decision-making—causes and consequences of self-selection. Games and Economic Behavior 56, 259–270.
Levitt, S.D., List, J.A., 2007. What do laboratory experiments measuring social preferences tell us about the real world? Journal of Economic Perspectives 21 (2), 153–174.
Prather, L.J., Middleton, K.L., 2002. Are $N+1$ heads better than one? The case of mutual fund managers. Journal of Economic Behavior and Organization 47, 103–120.
Rees, H., Shah, A., 1986. An empirical analysis of self-employment in the UK. Journal of Applied Econometrics 1, 95–108.
Rockenbach, B., Sadrieh, A., Mathauschek, B., 2007. Teams take the better risks. Journal of Economic Behavior and Organization 63, 412–422.

Shupp, R.S., Williams, A.W., 2008. Risk preference differentials of small groups and individuals. Economic Journal 118, 258–283.
Smith, V.L., Suchanek, G.L., Williams, A.W., 1988. Bubbles, crashes, and endogenous expectations in experimental spot asset markets. Econometrica 56, 1119–1152.
Sutter, M., 2005. Are four heads better than two? An experimental beauty-contest game with teams of different size. Economic Letter 88, 41–46.
Sutter, M., Kocher, M., Strauβ, S., 2009. Individuals and teams auctions. Oxford Economic Papers 61, 380–394.

THE REVIEW OF ECONOMIC STUDIES

Review of Economic Studies (2010) **77**, 882–913

0034-6527/10/00411011$02.00
doi: 10.1111/j.1467-937X.2009.00594.x

Moderation in Groups: Evidence from Betting on Ice Break-ups in Alaska

RENÉE ADAMS

University of Queensland and ECGI

and

DANIEL FERREIRA

London School of Economics, CEPR and ECGI

First version received August 2005; *final version accepted August* 2009 *(Eds.)*

We use a large sample of guessed ice break-up dates for the Tanana River in Alaska to study differences between outcomes of decisions made by individuals versus groups. We estimate the distribution of guesses conditional on whether they were made by individual bettors or betting pools. We document two major distinctions between the two sets of guesses: (1) the distribution of guesses made by groups of bettors appears to conform more to the distribution of historical break-up dates than the distribution of guesses made by individual bettors, and (2) the distribution for groups has less mass in its tails and displays lower variability than the distribution for individuals. We argue that these two pieces of evidence are consistent with the hypothesis that group decisions are more moderate, either because groups have to reach a compromise when their members disagree or because individuals with extreme opinions are less likely to be part of a group.

1. INTRODUCTION

Many economic decisions are made by groups. In firms, corporate boards and management committees make important decisions. In households, final decisions depend on the relative bargaining power of each family member. However, despite their obvious relevance for understanding economic choices, little is known about the differences between the decisions made by groups and the ones made by individuals.

Since individuals differ in the information they have, deliberation within groups presumably leads to information sharing among group members. If there are no conflicts of interest within a group, group decision making should be based upon all relevant information. This suggests that groups should make more predictable decisions than individuals, in the sense of relying more on "hard evidence" collected by information pooling. Some also argue that groups are more prone to reason-based choice, that is, group members favour choices that are easy to justify (see, for example, Barber, Heath and Odean, 2003). The importance of finding "good reasons" naturally leads groups to rely more on tangible evidence.

To be able to reach a consensus, groups need to balance individual opinions. This implies that groups should make less extreme decisions than the ones made by individuals. We call this a *compromise effect*. Groups may also choose to restrict membership of individuals with extreme opinions, or individuals with extreme opinions may choose not to join groups. We call this a *membership effect*. Taken together, these arguments imply that groups should make more moderate decisions than individuals.[1]

We develop a simple model in which these two effects interact and lead to moderation in groups that are endogenously formed. In order to avoid compromise, some individuals choose not to be group members.[2] Thus, the key insight from the model is that endogenous membership amplifies the moderating effect of groups.

Some experimental studies document cognitive biases specific to group decision making that may result in more extreme outcomes. For example, there is evidence that groups sometimes make riskier choices than individuals, a phenomenon that has been labelled "risky shifts" (Wallach and Kogan, 1965; Stoner, 1968). A related empirical regularity is known as "group polarization", which happens when group judgment deviates significantly from the average of the pre-discussion individual judgments, in the direction of higher "extremity" (Moscovici and Zavalloni, 1969; Kerr, 1992). Finally, there is the phenomenon of "groupthink" (Janis, 1982), which can be described as a "...dysfunctional mode of group decision-making characterized by a relentless striving for unanimity, resulting in a reduction in independent critical thinking" (Forbes and Milliken, 1999).

Such biases could counteract the moderating effects we describe above. Groupthink and group polarization could lead groups to discard some evidence aggregated through information pooling, which may make group decisions less predictable, while group polarization and risky shifts could lead groups to make more extreme decisions than the ones made by individuals. Research on small groups sometimes suggests that "small-group judgments tend to be more volatile and extreme" (Surowiecki, 2005, p. 176). Thus, the question "Are decisions made by groups more or less predictable and extreme than the ones made by individuals?" is, in principle, an open one.

The main contribution of this paper is to complement the experimental literature on group behaviour by presenting field evidence consistent with the idea that the outcomes of decisions made by individuals and groups differ systematically. We use data from an unusual betting game played in Alaska by a large number of individuals in 2002. The Nenana Ice Classic, 2002, named after the city of Nenana, started in 1917 when a group of surveyors for the Alaska Railroad made bets with each other on the date the ice on the Tanana River would break up. Nowadays, bets can be made on the Tanana River ice break-up date by buying tickets. The ticket holders whose guess is closest to the exact minute of break-up win the Jackpot. In 2002, the amount that could be won was $304,000.

Our primary data consist of the Nenana Ice Classic's *List of Guesses* for 2002. This is the official compilation of guesses that serves to determine the winners of the Ice Classic. It lists the names of each bettor in the order of predicted break-up date, hour and minute. Using bettors' names, we classify guesses into those made by individual bettors or by groups of bettors (*betting pools*). To examine whether the betting behaviour of groups is more or less extreme than that of individuals, we use estimates of the distribution of guesses conditional on this information.[3]

1. These ideas are well-known in the social psychology literature on groups. According to Levine and Moreland (1998), there are two mechanisms for the resolution of disagreement among group members: (i) *change of position*: members either compromise or adopt the position of others, and (ii) *redefinition of group boundaries*: non-conforming members leave the group.

2. Kocher, Strauß and Sutter (2006) provide experimental evidence that unwillingness to compromise and the desire to preserve autonomy are the main reasons why individuals choose not to join groups.

3. We discuss potential problems with our classification in detail in Sections 4 and 6.

There are many advantages of using this game to study differences between group and individual decision making. First, it is a real game that participants most likely understand and choose to play voluntarily. The rules are very simple, the game has been around for a long time, and all participants have access to the same information necessary for an educated guess. The tickets are accompanied with brochures that provide a tabulation of the historical break-up dates and times since 1917, and weather and ice condition information can be obtained from the Nenana Ice Classic website, as well as from online and print services. Thus, concerns about whether agents understand the problem or have the right incentives to behave in the expected manner are minimal.

Second, the decisions we examine are easy to measure and the dataset is very large. We are able to use 290,051 (virtually all) guesses for the 2002 break-up. Since we analyse essentially the entire population of guesses, we are able to obtain very precise estimates.

Of course, our data also have their limitations. Because we compare group and individual decision making in the context of a wagering game, our results may not generalize to situations in which participants do not behave in risk-seeking ways. It is also difficult to determine whether groups perform better or worse than individuals, because there is no obvious measure of good performance in this game. As Thaler and Ziemba (1988), Farrell *et al.* (2000) and Forrest, Simmons and Chesters (2002) point out, this issue is generally difficult to address in the context of wagering games. In this case, bettors face a trade-off between risk and return: betting on the most obvious dates would most likely imply a smaller prize, if right, because the prize is shared by all winners. Without knowledge of risk preferences, it is difficult to say that one guess is "better" than another one. This issue is complicated by the fact that the objectives of betting pools may be different from those of individuals, and that factors other than the expected return on the bet may play a role in bettors' objectives, such as, for example, fun.[4]

Because we rely on non-experimental data, we also cannot control for all aspects one may think are important. In particular, we cannot perfectly disentangle the direct effects of groups on outcomes from the effects of individuals' characteristics on group membership. Thus, we primarily provide evidence for the *joint hypothesis* that groups make more moderate decisions, either because they require a compromise among their members, or because individual characteristics may affect group membership, or both. However, we also provide some evidence that selection on individual characteristics alone is not sufficient to explain our findings.

Our main findings can be summarized as follows. First, the distribution of group guesses appears to conform more to the distribution of historical break-up dates than the distribution of individual guesses. In this sense, group guesses are more predictable. For example, individual bettors place more weight on earlier spikes in the historical break-up distribution than on later spikes, which is consistent with the idea that individual bettors may be biased towards earlier break-up dates. Second, we find that the group guess distribution places less weight on the tails and displays lower variability than the distribution for individuals.

By means of individual fixed-effects regressions, we also show that individuals who bet both alone and in groups are more likely to make their extreme guesses alone, suggesting that the moderating effect of groups works not only through the sorting of similar people

4. The results may also not generalize to situations in which utility functions and payoffs are invariant across individuals and groups or to situations in which individuals cannot choose to act alone as well as with a group. In addition, unlike in many experimental studies, per capita payoffs are not equal across individual bettors and betting pools. While our results seem to hold even if we restrict ourselves to groups and examine variation in group size, it is possible that having the choice to bet alone changes the way individuals negotiate about the pool's betting behaviour.

into the same groups. Although the fixed-effects regressions cannot hold the characteristics of the other group members fixed, it goes a long way towards illuminating the mechanism that leads to moderation in groups. Because we focus on a subsample of individuals who bet both alone and in groups, these results rule out the possibility that only people with moderate bet preferences choose to join groups. In this subsample, moderation in groups occurs either because of compromise or because group members cannot "agree" upon an extreme guess, and consequently choose to place their extreme guesses alone or leave groups that have guess preferences that are too different from their own.

To confirm that our findings are not special to the year 2002, we replicate our tests using 2008 data. Although we cannot use the same approach to classify bettors into individuals versus pools as in 2002, we were able to obtain a list of "registered pools" from the Ice Classic office, which was not available for 2002. A major advantage of using the 2008 data is that this subsample of pools is free from measurement error, which enables us to assess the impact of measurement errors on our results. Overall, the data from 2008 confirm the main results found in 2002: the guess distribution for pools displays a lower variance than the guess distribution for individuals.

Our evidence suggests that groups behave more conservatively than individuals. This statement, by itself, does not say anything about the relative *rationality* of group versus individual decision making. However, because group decisions appear to differ systematically from individual decisions and many economic decisions are made by endogeneously formed groups, our findings have economic implications, at least for situations similar to those we study.

The paper is structured as follows. In Section 2 we discuss the related literature. We present a simple model that illustrates the main ideas in Section 3. Section 4 describes the data. We discuss factors that are likely to affect decisions and present our results in Section 5. We address selection and other data-related issues in Section 6. We present the results for the 2008 data and their implications for potential problems with measurement errors in Section 7. We conclude in Section 8.

2. RELATED LITERATURE

The literature on wagering games, such as sports betting and lotteries, is extensive. Its primary focus has been to analyse the efficiency of betting markets and to estimate bettors' utility and demand functions (for surveys of this literature, see, for example, Sauer, 1998; Williams, 1999). Empirical papers in this literature have generally relied on aggregate data on bets. In contrast, we are able to identify the entire betting portfolios of individual bettors.

Few empirical papers in economics explicitly compare group to individual decision making. The main questions in the ones that do are the relative rationality and performance of groups versus individuals. Bone, Hey and Suckling (1999) find no evidence that groups conform more to expected-utility maximizing behaviour than individuals. Using field-data on mutual fund performance, Prather and Middleton (2002) also find no evidence that differences in fund performance can be attributed to differences in group versus individual decision making. Work by Blinder and Morgan (2005) documents contrasting results. Using experiments that are specifically designed to simulate how decisions about monetary policy are conducted in central banks, they provide experimental evidence that groups make both faster and better decisions than individuals. In experimental beauty-contest games, Kocher and Sutter (2005) find that groups do not appear to be more rational than individuals, but they tend to learn faster. In a similar setting, Sutter (2005) finds that larger groups perform better. None of these papers addresses our hypothesis that groups are more moderate.[5]

Barber *et al.* (2003) provide field evidence that stock clubs, relative to individual investors, are more likely to engage in reason-based choice when picking stocks. Their findings are consistent with the hypothesis that group decisions are more predictable on the basis of hard evidence (i.e. "good reasons"). Adams, Almeida and Ferreira (2005) find that stock returns are significantly more variable for firms run by powerful CEOs. They stress that in firms in which the CEO has less discretion, decisions are more clearly the product of consensus among the top executives. Bär, Kempf and Ruenzi (2005) provide evidence that team management reduces portfolio risk in mutual funds. Rockenbach, Sadrieh and Mathauschek (2007) present experimental evidence that for a given expected return, groups, relative to individuals, are more likely to invest in portfolios with lower standard deviations. These findings are broadly consistent with the ones in this paper.

The social psychology literature on group decision making has also analysed the specific effect of group processes on different dimensions of group decisions, such as their extremity and riskiness. As discussed by Moscovici and Zavalloni (1969), a natural hypothesis is that the "group consensus" (the final choice made by a group) represents "an averaging, a compromise among individual positions". This idea is supported by a number of experimental research findings, such as those of Kogan and Wallach (1966), who find that group judgment represents the average of the prior individual judgments, even when consensus is achieved via group discussion of each prior judgment.

While this evidence is consistent with a moderating effect of groups, as we described in the introduction, there is also experimental evidence for cognitive biases, such as "risky shifts", "group polarization", and "groupthink", which may make group decisions less predictable and more extreme than individual ones. Risky shifts in groups occur when groups make riskier choices than those that would represent a compromise between the choices of the individuals comprising the group (Wallach and Kogan, 1965). Groups may also display "cautious shifts" (Stoner, 1968). Eliaz, Ray and Razin (2006) develop a model that links the phenomenon of choice shifts to well-known failures of expected-utility theory.

Group polarization arises when "members of a deliberating group move toward a more extreme point in whatever direction is indicated by the members' predeliberation tendency" (Sunstein, 2002). See also Cason and Mui (1997), who are among the first to incorporate the idea of group polarization in economics. For a recent model of group polarization, see Glaeser and Sunstein (2007).

Finally, the phenomenon of "groupthink" (Janis, 1982) may lead groups to ignore "hard evidence" at their disposal to irrationally pursue unanimity, which may lead to an increase in the riskiness and unpredictability of group decisions. Bénabou (2008) provides a formal model of this phenomenon.

Ultimately, whether group decisions are more or less risky than individual decisions is an empirical issue. It is possible that group polarization, risky shifts or groupthink may attenuate or even reverse the *compromise effect* of group decision-making. What is less clear from the experimental literature is how important these biases are when group formation is endogenous. Because the issue of group membership has received little attention in the previous literature,

5. From a theoretical perspective, the most closely related argument is in the work of Sah and Stiglitz (1986, 1991). In their models, because group members may disagree, group decision making entails a *diversification-of-opinions* effect. The final group decision will be a compromise, which reflects the different opinions of the group members.

in the next section we develop a simple model to clarify the relation between the *compromise* and the *membership effect*.

3. A SIMPLE MODEL OF MODERATION IN GROUPS

In this section, we provide a simple model to illustrate the two effects that may lead to moderation in naturally occurring groups: the *compromise effect* and the *membership effect*. Although these effects are quite intuitive, the model helps us clarify the relationship between them. In particular, we show that they are closely related: the membership effect decreases the variability of group decisions only when there is compromise in decision making.[6]

We model a situation in which group members must agree upon a single decision when all members have access to the same material information. We assume that there are no strategic communication issues.[7]

In order to make the analogy with our empirical application clear, we develop a model of a betting game in which the main decision is to make a guess x. A *bettor* (group or individual) must guess the realization of a single draw from the distribution of D. In a standard "guess-the-number" game, bettors win if they guess the exact number that is chosen. We call a possible draw of D a *break-up date*, and denote it by d. Break-up dates are integers drawn from a finite support $\mathcal{D} \equiv \left\{-\overline{D}, ..., 0, ..., \overline{D}\right\}$. For simplicity, we assume that the distribution of dates is symmetric with zero mean, and denote the probability of a break-up date d by $p_d > 0$, for all $d \in \mathcal{D}$.[8]

There is a large number of individuals, whose types $t \in \mathcal{D}$ represent the date they would like to guess, all else constant. That is, we assume that each individual has a preference for a specific guess. This is particularly realistic in our empirical application.[9] In general, group members could disagree on decisions due to differential information, heterogeneous priors, overconfidence, and the like, but assuming a preference for decisions is the simplest way of generating disagreement among group members.

For simplicity, we assume that the distribution of t in the population of individuals who may become group members is the same as the distribution of break-up dates:

Assumption 1. $\Pr(t) = p_t$, *all* $t \in \mathcal{D}$.

The main role of this assumption is to simplify notation, but incidentally it also implies that bettors do not have biased preferences, either for extreme or conservative guesses. Thus, in our model moderation can only arise due to the decision-making process and membership decisions.[10]

We now define the utility of an individual of type t who guesses as a member of a group of size n (an individual who bets alone is equivalent to a group of size $n = 1$). We assume that each individual only has enough funds to make a single guess, which costs c. If guesses

6. Although we model the membership effect as a consequence of individuals' decisions to be part of a group, it can also be modelled as the consequence of groups' decisions to restrict membership.

7. Alternatively, see Ottaviani and Sorensen (2006) for a model of strategic information transmission that also has implications for the variability of decisions.

8. Symmetry simplifies notation and facilitates the computation of equilibria.

9. For example, the sole winner of the 2008 Ice Classic was an Anchorage woman who placed guesses on the exact time that she was born.

10. Relaxing this assumption has no important implications for the results; it merely introduces a group "bias" towards some dates by assumption. Our results should be understood as being "net" of such biases.

are made by a group of size n, the group as a whole enters n guesses of a single date x.[11] We denote by $x_n(t_1, ..., t_n) : \mathcal{D}^n \to \mathcal{D}$ the rule that determines the guess of a group of size $n \geq 1$ with member types given by $(t_1, ..., t_n)$.

An individual t whose group makes a guess according to $x_n(t_1, ..., t_n)$ enjoys utility:

$$U_t = V[x_n(t_1, ..., t_n), t] - c + f - |t - E_t[x_n(t_1, ..., t_n)]|, \tag{1}$$

where $V[x_n(t_1, ..., t_n), t]$ is the expected payoff to a group guess made according to the rule $x_n(t_1, ..., t_n)$ conditional on knowing one's own preferred guess t, c is the cost of one guess, f is a positive number, and $E_t[x_n(t_1, ..., t_n)]$ is the expected group guess conditional on t. This utility function is valid both for individual bettors ($n = 1$) and for members of proper groups ($n > 1$).

This utility function can be broken down into two parts: the first one, $V[x_n(t_1, ..., t_n), t] - c$, measures the net pecuniary payoffs from betting, while the second one, $f - |t - E_t[x_n(t_1, ..., t_n)]|$, measures net non-pecuniary or psychological payoffs. The term f can be interpreted as a direct benefit from betting, or the value of "fun". Because such types of lotteries have negative expected monetary returns, it is necessary to assume either that bettors derive utility from betting directly or that they are risk-lovers. Therefore, f can also be interpreted as the (negative of the) risk premium if bettors are risk-lovers, in an alternative certainty-equivalent representation.

The term $|t - E_t[x_n(t_1, ..., t_n)]|$ represents the costs from not guessing one's preferred date t. Because the group guess $x_n(t_1, ..., t_n)$ when $n > 1$ will not necessarily be equal to the individual's preferred guess t, group bettors are usually worse off than individuals due to the need for compromise. We assume that compromise is more costly the farther the group guess is from the individual's preferred guess. Thus, we need to explain why individuals choose to join or remain as members of groups.

There are many ways of modelling group formation. Here we choose a simple framework that is sufficient to illustrate the main effects we wish to emphasize. We assume that each individual is randomly assigned either to a group of size $N > 1$ or to no group at all.[12] If assigned to a group, an individual must choose whether to stay with that group or to leave and make an individual guess, or perhaps not bet at all. We denote the (endogenous) group size after membership decisions are made by $n \leq N$. Each individual chooses whether to stay after learning his own type t but before the group decides the group guess $x_n(t_1, ..., t_n)$ (that is, we assume that individuals cannot leave their groups after group guesses become known). If an individual chooses to leave a group, he pays a "defection" cost of y. An individual who leaves a group pays this cost regardless of whether he bets alone or chooses not to bet. We normalize the utility (net of defection costs) of not making a guess to zero. Similarly, we could also have assumed that the "fun" value of betting in groups is larger for groups than individuals, with identical results.

If J is the jackpot and w_d is the number of winning tickets on date d, the payoff to one guess on date d is given by J/w_d. Let T be the total number of guesses. The jackpot is usually a fraction z of the revenue from bets cT, that is, $J = zcT$. In the Ice Classic, z is roughly 0.5.

11. This assumption is made for simplicity. Allowing groups to guess multiple dates creates no difficulties for the analysis, because the utility function (defined below) depends on the expected group guess. However, the analysis would change if groups could make fewer than n guesses. This would create an additional benefit from group membership, which is currently not modelled.

12. Strictly speaking, it is not necessary that groups are randomly formed, but only that group membership is not correlated with t.

We can now characterize the equilibrium distribution of guesses. A crucial assumption we make here is that there is a (potentially very) large number of unmatched individuals (i.e. individuals that are required to bet alone or not at all) relative to the number of groups. We focus only on equilibria with "large" numbers of guesses, so that the number of groups is relatively small compared to the total number of guesses.[13] Under these assumptions, the equilibrium distribution of guesses and the expected pecuniary payoffs to guessing a given day are determined by the behaviour of individual bettors alone. Thus it is possible to describe the equilibrium distribution of guesses without reference to $x_n(t_1, \ldots, t_n)$, unless $n = 1$, because the "marginal" players are individual bettors.

We consider a rational expectations equilibrium in which: (i) all players know the number of guesses of each date d in equilibrium, w_d^*, and the equilibrium jackpot, J^*, (ii) all players consider themselves too small to affect the equilibrium jackpot and the number of guesses of each date, (iii) all players form $E_t[x_n(t_1, \ldots, t_n)]$ in a Bayesian rational manner, and (iv) all individual bettors choose their guesses (i.e. choose their $x_1(t)$) in order to maximize their expected utilities as defined in (1).

The utility of an individual player of type t guessing date d in equilibrium is

$$U_t^* = p_d \frac{J^*}{w_d^*} - c + f - |t - d|. \tag{2}$$

It is straightforward to see that, in a rational expectations equilibrium, it must be that $x_1(t) = t$ for all $t \in \mathcal{D}$, all guesses from unmatched players should yield the same expected payoff, and (due to free entry of bettors) expected payoffs should be zero, implying that:[14]

$$p_d \frac{zcT^*}{w_d^*} = c - f, \tag{3}$$

which implies that the fraction of guesses of date d must be

$$\pi_d^* \equiv \frac{w_d^*}{T^*} = kp_d, \tag{4}$$

where $k \equiv zc/c - f$. Thus, the equilibrium distribution of guesses should be proportional to the true distribution of break-up dates. Because the condition above must hold for all d, it is clear that the only possible equilibrium requires $k = 1$. We assume that the game managers choose z and c so that this condition is satisfied.[15]

Result 1: *In a rational expectations equilibrium with a (sufficiently) large number of individual bettors: (1a) the proportion of guesses of a given date d should be equal to the true probability of break up on that date:* $\pi_d^* = p_d$, *for all* $d \in \mathcal{D}$; *and (1b) all guesses should yield the same net pecuniary payoff:* $p_d \frac{J^*}{w_d^*} - c = -f$, *all* $d \in \mathcal{D}$.

Now we turn to the question of individual versus group betting.

13. This assumption is realistic in our empirical application.

14. For simplicity, we ignore possible discontinuities here that may arise because of the discreteness of the setup. Formally, the problem can be made continuous by allowing for mixed strategies for the participation decision.

15. This does not eliminate multiple equilibria: the total number of guesses T is indeterminate. Thus, choosing $k = 1$ only guarantees that a rational expectations equilibrium is possible. Because any feasible profit for the game managers can be achieved with $k = 1$, there is no reason to assume that they will choose a different k.

Case 1: Exogenous membership.

We start by analysing the simplest case in which groups are formed exogenously and defection is not possible (i.e. the defection cost is $y = \infty$). When a group of n individuals has to decide on a single guess x, there must be a rule to determine the group's choice. Let t_i denote the preferred break-up date of member i of a given group. We assume a simple rule to model compromise:

Assumption 2. *A group's guess is the average of each individual's preferred guesses:*

$$x_n\,(t_1, ..., t_n) = \frac{1}{n}\sum_{i=1}^{n} t_i. \tag{5}$$

Clearly, given Assumptions 1 and 2, the unconditional expected group guess is $E[x_n(t_1, ..., t_n)] = 0$. If we denote the unconditional guess variance for an individual by σ^2, we have that the unconditional variance of a group guess is σ^2/n.[16] Although this is a straightforward consequence of the assumption that groups compromise, we state it as a result due to its importance for the empirical part:

Result 2 – The compromise effect: *If group size is exogenously determined, and if a group's guess is the average of its members' preferred guesses, then the variance of group guesses is lower than the variance of each of the preferred guesses of its members.*

Case 2: Endogenous membership.

The exogenous membership case neatly illustrates the compromise effect. When group membership is endogenous, however, the membership decisions of individuals can also affect the statistical distributions of guesses. Here we develop a very simple model of endogenous membership decisions and their effects on group moderation.

An individual of type t matched to a group of a still unknown size n needs to estimate the group bet $x_n\,(t_1, ..., t_n)$ before deciding whether to stay or leave the group. From the individual t's perspective, $x_n\,(t_1, ..., t_n)$ is a random variable because both n and $t_i \neq t$ are unknown. We assume that an individual cannot learn the types of the other group members before decisions are made. We economize on notation by denoting $x_n\,(t_1, ..., t_n)$ by x_n.

The equilibrium utility of a type t individual when deciding whether to stay with the group is

$$U_t^*\,(\text{stay}) = E_t\left[p_{x_n}\frac{J^*}{w_{x_n}^*}\right] - c + f - |t - E_t\,[x_n]|\,. \tag{6}$$

If he leaves, his expected utility is

$$U_t^*\,(\text{leave}) = \max\left\{p_t\frac{J^*}{w_t^*} - c + f, 0\right\} - y. \tag{7}$$

16. More generally, the compromise effect can be modelled as a weighted average of individual guesses, in which some members could be more influential than others. The qualitative Result 2 (below) would still go through.

The individual will compare these two utilities when deciding whether to stay or leave. Result 1 simplifies the problem: because the net expected pecuniary payoffs are always equal to $-f$ regardless of the date one guesses, the individual will choose to stay only if

$$|t - E_t [x_n]| \leq y. \tag{8}$$

We can now state the conditions for an equilibrium. Let $s(t) : \mathcal{D} \to [0,1]$ be a function that assigns a probability that a player will stay with his group for each possible type in $\mathcal{D}$ and let $F(x)$ be the distribution of x_n. The functions $s^*(t)$ and $F^*(x)$ represent strategies and beliefs in equilibrium if and only if $F^*(x)$ is equal to the true distribution of x_n implied by Assumptions 1 and 2 and the equilibrium strategy profile $s^*(t)$, and

$$s^*(t) = \begin{cases} 1 \text{ if } \left|t - E_t^*[x_n]\right| < y \\ 0 \text{ if } \left|t - E_t^*[x_n]\right| > y \\ \theta \in [0,1] \text{ if } \left|t - E_t^*[x_n]\right| = y \end{cases}, \tag{9}$$

where

$$E_t^*[x_n] = \int_{-\overline{D}}^{\overline{D}} x_n dF^*(x_n \mid t), \text{ all } t \in \mathcal{D}. \tag{10}$$

We restrict attention to symmetric equilibria, that is, to equilibria where type t's equilibrium strategy is the same as type $-t$'s equilibrium strategy. Once restricted to symmetric equilibria, it is possible to show that a unique equilibrium exists, and has a "threshold property":[17]

Result 3 – The membership effect: *There exists a unique symmetric equilibrium. The symmetric equilibrium is characterized by a unique integer t^* such that individuals with $t \in (-t^*, t^*)$ choose to stay as group members, while $t \notin \left[-t^*, t^*\right]$ leave the group.*

This result shows that endogenous membership generates a truncation of the distribution of guesses, such that individuals with extreme guess preferences (either too high or too low) choose not to participate. As a consequence, groups will never make extreme guesses. Intuitively, members leave the group in order to avoid the negative effects of compromise. Individuals with more extreme guess preferences are more likely to suffer from compromise and thus will choose to leave the group and bet alone (or not at all).[18]

Proof. Notice that

$$E_t^*[x_n] = E_t^*\left[\frac{t + (n-1)e}{n}\right], \tag{11}$$

where e is the average type of all other individuals who stay and n is the size of the group after membership decisions are made. Symmetry implies that $E^*[e] = 0$ for all groups and

17. Asymmetric equilibria may also exist and will all have a similar threshold property, although an asymmetric one. Without any ad hoc reasons for choosing an equilibrium, the unique symmetric equilibrium we describe below is appealing because it is the "average" or "median" equilibrium, if, for example, we consider each equilibrium to be equally likely.

18. Note that the exogenous membership case is a special case of the threshold equilibrium in which $t^* > \overline{D}$. This is the unique equilibrium for sufficiently high y.

also that e is mean-independent of group size n. Thus, condition (10) simplifies to

$$\begin{aligned} E_t^*\left[x_n\right] &= tE^*\left[\frac{1}{n}\right] + E^*\left[\frac{(n-1)}{n}\right]E^*\left[e\right] \\ &= tE^*\left[\frac{1}{n}\right] \equiv t\alpha. \end{aligned}$$

We drop the conditioning on t for the expectation because n and e are independent of t. Let q_s denote the probability that a randomly chosen original group member (other than the member making this calculation) chooses to stay in an equilibrium. Define

$$\alpha\left(q_s\right) \equiv E^*\left[\frac{1}{n}\right] = \lim_{q \to q_s} \sum_{j=1}^{N} \frac{N-1!}{(N-j)!\,(j-1)!} \frac{q^{j-1}\left(1-q\right)^{N-j}}{j}. \tag{12}$$

It can be shown that $\alpha\left(q_s\right)$ is continuous and strictly decreasing in q_s (since increases in q_s make larger groups more likely, so that $E\left[1/n\right]$ is smaller), with $\alpha\left(0\right) = 1$ and $\alpha\left(1\right) = 1/N$.[19]

Using condition (8), an individual of type t will choose to stay with probability 1 in equilibrium if

$$\left|\left[1-\alpha\left(q_s\right)\right]t\right| < y. \tag{13}$$

If a symmetric equilibrium exists, this condition implies that there must be a threshold t^* such that all $|t| < t^*$ choose to stay while all $|t| > t^*$ choose to leave. A symmetric equilibrium is then fully characterized by a triplet $\left(t^*, q_s^*, \theta^*\right)$ such that the conditions below are satisfied:

$$q_s^* = p_0 + 2p_1 + \dots + 2\theta^* p_{t^*}, \tag{14}$$

$$t^* = \left\lfloor \frac{y}{1-\alpha\left(q_s^*\right)} \right\rfloor, \tag{15}$$

$$\theta^* = \begin{cases} \theta \in [0,1] \text{ if } t^* = \frac{y}{1-\alpha\left(q_s^*\right)} \\ 1 \text{ if } t^* < \frac{y}{1-\alpha\left(q_s^*\right)} \end{cases}, \tag{16}$$

where the lower brackets denote the integer part function, that is, $\lfloor x \rfloor$ is the largest integer less than or equal to x.

Define the correspondence $f : [0,1] \to [0,1]$ as

$$f\left(q_s\right) = p_0 + 2p_1 + \dots + 2\lambda p_{\left\lfloor \frac{y}{1-\alpha(q_s)} \right\rfloor},$$

where

$$\lambda = \begin{cases} \tau \in [0,1] \text{ if } \left\lfloor \frac{y}{1-\alpha(q_s)} \right\rfloor = \frac{y}{1-\alpha(q_s)} \\ 1 \text{ if } \left\lfloor \frac{y}{1-\alpha(q_s)} \right\rfloor < \frac{y}{1-\alpha(q_s)} \end{cases}.$$

19. The minimum group size n is 1 (and not zero), because we are considering the perspective of a member of type t assuming that he will stay with the group.

For each $q_s \in [0, 1]$, the set $f(q_s)$ is non-empty and convex, and the graph of f is closed. Thus, Kakutani's theorem applies, so there exists at least one fixed point $q_s' \in f\left(q_s'\right)$. For each q_s', condition (15) implies a unique $t^*\left(q_s'\right)$, and consequently condition (14) implies a unique $\theta^*\left(q_s'\right)$.

To confirm that the equilibrium is indeed unique, notice that f is non-increasing in q_s (in the sense that if $x > y$, $\max_\lambda f(x) \le \min_\lambda f(y)$), which implies that the graph of f crosses the 45 degree line only once, implying a unique q_s^*. ‖

4. DATA

Bets can be placed on the Tanana River ice break-up date either by depositing tickets in red cans, which are located in more than 200 locations throughout Alaska, or by mailing in a guess. The tickets cost \$2.00 (in 2002) and are sold from 1 February to midnight on 5 April.

To determine the winner, a tripod is erected on the river ice. The tripod is connected to a clock so that a tripwire will signal the exact minute of break-up. To prevent bettors from tampering with the tripod, the tripod is guarded by watchmen 24 hours a day. The ticket holders whose guess is closest to the exact minute of break-up win the jackpot.

Our data consist of the Nenana Ice Classic's *List of Guesses* for 2002. This is the official compilation of guesses which serves to determine the winners of the Ice Classic. It lists, in order of date, hour and minute, the names of each bettor who guessed the ice would break at that time. Thus, we can track bettors' guesses using their names. In addition, bettors' names can be used to identify betting pools, which makes it possible for us to compare the betting strategies of individuals to those of groups. For example, we can classify bettors with names such as "7 Lucky Ladies", "Fat Freddie's 2 p.m. Coffee Club", "Chilly Dogs & Co." and "Gene Pool 2002" as pools.

Of course, we cannot tell if a given bettor is a pool if the pool's guesses are registered under the name of an individual. In 1997, for example, a winning pool containing two members was listed under an individual's name. We also cannot tell if a bettor is an individual if the bettor uses an imaginative pseudonym that appears to be a pool name. However, we believe that bettors will not "disguise" their identities very frequently for several reasons. First, pools have an incentive not to bet using an individual pool member's name to avoid potential conflicts over the distribution of the jackpot, especially since pool members may bet on their own as well as with the pool. Using a name distinct from an individual's name also makes it easier for a pool to make multiple guesses, because the pool need only enter its name on each ticket and register the pool by sending in a list of pool members to the Nenana Office. If multiple guesses are listed under an individual's name who acts as a representative of the pool and that individual also bets alone, the names of all other pool members must be listed on the back of *each* ticket belonging to the pool. Finally, it is clear from the names of some of the pools that part of the fun of being in a pool is to come up with a betting name. Because the Ice Classic tickets require bettors to fill in a name and mailing address (see the example at http://www.nenanaakiceclassic.com/TicketsBrochures.html), we also think that individuals betting alone are more likely to use their real names and addresses than to use pseudonyms. Thus, although we may sometimes misclassify pools as individuals if they bet under an individual's name and vice versa, we believe that such misclassification may add noise to our estimates but not systematically bias them. Nevertheless, we try to address potential measurement error in our pool classification in Sections 6 and 7.

We were only able to obtain a hard copy of the *List of Guesses*, which contained roughly 295,667 guesses.[20] We scanned the data using the scanning software OmniPage. Since the quality of the type in the hardcopy varied, and the scanning process introduced certain systematic errors (e.g. sometimes replacing the letter “M” with “14”), we subjected the data to an extensive cleaning process. It was straightforward to clean the guesses since they are listed in order by date and time in the *List of Guesses.* We end with a sample of 294,176 usable guess dates and 284,724 usable guess times.[21]

To clean the names, we developed a cleaning program which we describe in more detail in the Appendix. After applying the program, we also checked the bettors’ names by hand to ensure that they were spelled uniformly. Since each guess is entered on a separate ticket, it is possible that name changes could occur which would make it difficult to track bettors’ betting profiles. Since guesses made by betting pools may be entered by different people in the pool, we believed it to be more likely that spelling changes would occur for betting pools than for individuals.[22] Thus, we focussed primarily on making names with four or more words uniform. To make the bettors’ names consistent, we changed names that occurred in multiple variations, for example, “Dave & Linda Huffaker” and “Linda & Dave Huffaker”, to one of the forms in which the name occurred.

Because of the size of the dataset, it was impossible to examine each name to determine whether it corresponded to a pool or not. We proceeded as follows to identify individuals and pools. First, we classified names that contained obvious pool identifiers as pools. Such identifiers consisted of words such as “pool”, “group”, “company”, “team”, “family”, and the word “and”, as well as symbols such as “&” and “/”, which were often used to separate individual names in a betting pool. Remaining observations were classified as individuals if the first word in the name appeared in the list of first names in the 2000 census. The observations not classified were then checked by hand to see if they belonged to pools with unusual names, e.g. “Alaska’s Point of View”. Checking by hand also enabled us to classify as individuals bettors who listed their last names first or who abbreviated their first names. With this procedure, we identify a total of 3093 pools. The most common pool identifiers are the character “&” (1030 pools), the words “co” (591) and “pool” (395), and the character “/” (243 cases). After cleaning, we are able to classify 290,051 guess dates belonging to 34,656 different bettors and 281,018 guess times according to whether they were entered by individuals or pools.[23]

20. The *List of Guesses* has 1479 pages with approximately 200 bets per page, except for page 1479 which contains only 67 bets.

21. It is easier to clean guess dates than guess times for several reasons. First, there are fewer dates in a year than minutes in a day, which makes it easier to infer correct dates (for example, by referring to the original text). Second, scanning errors often occur at the end of strings; for example, the last character is often dropped. Dates are provided in the format dd/mm/yyyy, where yyyy is identical for all bets, so scanning problems in the last four characters do not pose a problem. However, scanning errors at the end of times make it impossible to infer the correct time. Finally, times are provided in the format hh:mm A or hh:mm P, where A refers to a.m. and P to p.m. In general, it is impossible to rectify scanning error in the letter A or P, which means that the correct time cannot be inferred for those observations. We attribute the fact that we lose fewer guess times in our 2008 data to improvements in scanning software since 2002.

22. The Nenana Ice Classic also allows bettors to rubber stamp their names and addresses on tickets. Individuals are probably more likely to have such rubber stamps which makes name changes across bets less likely for individuals.

23. We have information on 553 non-missing guess dates and 498 non-missing guess times for dates between 1 February and 5 April 2002. Of these, 48 date and 45 time guesses were made by pools. These guesses may potentially be considered irrational, since the guessed break-up date occurs during the time in which tickets are sold, which means that the submission date for the bet potentially succeeds the guessed date. However, we do not exclude these from our analysis because we have no way of verifying whether they are irrational or not. Furthermore, we view this as further evidence that individuals place more weight on tails than pools do.

Because our distinction between individuals and pools is not error free, we analyse similar data for 2008 in Section 8. The 2008 data has special features which we can use to validate our bettor classification method, as well as to perform robustness checks. As we will discuss in Section 8, we do not find evidence of substantial mismeasurement induced by our classification procedure.

As is to be expected, pools make more guesses than individuals.[24] The mean (standard deviation) number of guesses entered by 3093 pools was 24.03 (77.70), with a maximum of 2158. The mean (standard deviation) number of guesses entered by 31,563 individuals was 6.84 (15.15), with a maximum of 1161.

5. ANALYSIS AND RESULTS

5.1. *Factors influencing guesses*

There are several key factors that we expect should play a role in a bettor's guesses. Since the ice break-up date is determined by environmental factors, information about weather and ice conditions should be important, although the extent to which the break-up day can be predicted is, of course, limited. Many bettors do appear to actively search out this information in order to make their guesses. For example, many of them call the Nenana Ice Classic office, asking about weather and snow conditions, according to the manager of the Ice Classic, Cherrie Forness (*Arctic Science Journeys*, 1997). Newspaper articles describe how bettors gather environmental information for their guesses (e.g. *The Seattle Times*, 1986; Richards, 1995; *Arctic Science Journeys*, 1997). The Nenana Ice Classic also posts a record of historical (since 1989) and current ice measurements together with the historical break-up dates on its website (www.nenanaakiceclassic.com); the historical break-up dates together with their frequencies are printed on brochures accompanying Ice Classic tickets. In addition, Alaskan television stations include updates about the thickness of the ice on the Tanana River in their weather coverage as the deadline for betting approaches (Finkel, 1998). Since bettors can only rely on information on current environmental conditions up until the day that betting closes, we expect bettors to take both current and historical information into account.

5.2. *The distribution of guesses in 2002*

Because we will compare the distributions of pool and individual guesses to examine whether pools are more moderate in their betting behaviour, it is useful to discuss some features of the overall distribution of guesses. Although it is difficult to say that one guess is "better" than another without knowledge of individual risk preferences, we can characterize the expected return and the shape of the guess distribution by making some simplifying assumptions.

In the Ice Classic, the optimal betting strategy for a single bettor depends on the strategies of all other bettors because the jackpot is evenly divided if there are multiple winners. According to the model in Section 3, if we assume a rational expectations equilibrium with free entry of bettors, all guesses should yield exactly the same net expected payoff in equilibrium. This is a straightforward consequence of our model (see Result 1b). Thus, the net expected payoff

24. The fact that bettors we classify as pools make more guesses than bettors we classify as individuals suggests that our classification reflects bettors' identities reasonably well and should lessen concerns about measurement error due to misclassification.

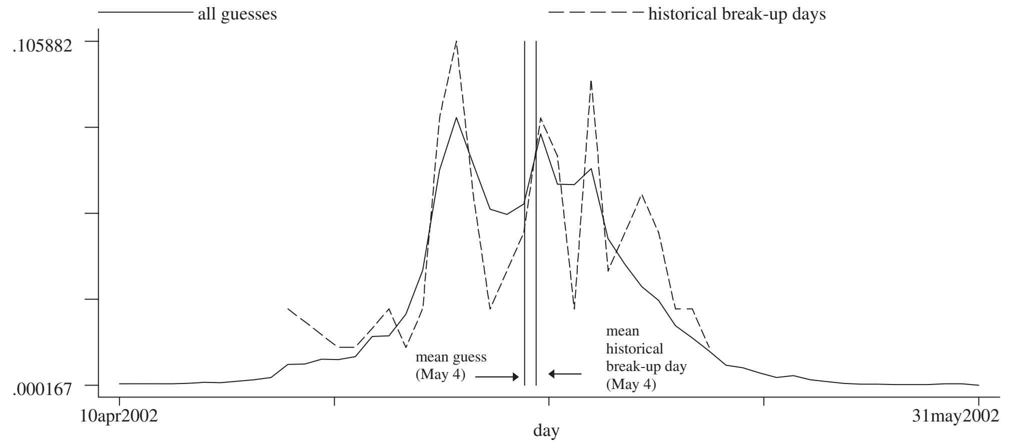


FIGURE 1

Comparison of frequency distribution of guessed dates in 2002 to historical break-up days (entire sample: 294,170 observations; historical dates from 1917 to 2001; only guesses from 10 April to 31 May displayed)

of any guess should be equal to −\$0.97, which is the total jackpot (\$304,000) divided by the total number of guesses (295,667) minus the cost of a ticket (\$2). As in many other wagering games, bettors are willing to participate in the Ice Classic even though the expected payoff is negative, either because they exhibit risk-seeking behaviour or they derive satisfaction from playing (i.e. "fun"). If bettors are risk-neutral, as in our model, our estimates suggest that the value of "fun" (f) per guess is −\$0.97.

These numbers are only meaningful if the actual betting behaviour is similar to the one implied by equilibrium play. In our model, this requires bettors to know the distribution of break-up dates. In reality, bettors do not know the true distribution of break-up dates. However, they do know the historical distribution of break-up dates, which we plot in Figure 1 together with the distribution of 2002 guesses.[25] If we are willing to assume that, as a first approximation, bettors view the historical distribution as the true one, Result 1a implies that the actual distribution of guesses should mimic the historical distribution of break-ups exactly ($\pi^*_d = p_d$, for all $d \in \mathcal{D}$). Because many assumptions of this simple model will not hold, we do not expect the two distributions to be exactly the same in our data. Still, the similarities are striking, as we can see from Figure 1, which suggests that this simple model provides a good approximation for the behaviour of bettors in this game. Figure 1 strongly suggests that bettors take the historical data into account when making their guesses, which gives rise to the unusual shape of the guess frequency distribution. In particular, the sample frequency appears to mimic all but the last spike in the historical data on 11 May.

To get a feeling for how well our model works in explaining the guess distribution in 2002, consider the historical mean break-up date (4 May). The historical probability of break-up on that day is roughly 7%, which is also the proportion of 2002 guesses for that day (see Figure 1). Thus, for the strategy of guessing the mean day, the model's prediction is fulfilled exactly. Looking at all dates, the theoretical prediction is off by at most 3.7 percentage points, and in most cases significantly less.

25. To make the graph easier to read, we report only the distribution of guesses between 10 April 2002 and 31 May 2002, a period which contains 99.53% (292,807) of the guesses.

TABLE 1
Summary statistics

Variable	Observations	Mean	Std. dev.	Min	Max
Historical break-up days	85	124.74 (4 May)	5.95	110 (20 Apr)	140 (20 May)
		Guess days			
Entire sample	294,170	124.04 (4 May)	6.64	33 (2 Feb)	363 (29 Dec)
Individual guesses	215,713	124.03 (4 May)	6.84	33 (2 Feb)	363 (29 Dec)
Pool guesses	74,332	124.25 (4 May)	5.94	33 (2 Feb)	320 (16 Nov)
Historical break-up time	85	858.34	281.38	4	1404
		Guess times			
Entire sample	284,718	839.87	260.67	0	1439
Individual guesses	208,765	838.44	263.89	0	1439
Pool guesses	72,247	844.21	251.22	0	1439

Notes: *Guesses* consist of guesses of the day, hour and minute that the ice on the Tenana River was going to break in 2002. We exclude six guesses of days prior to the date guesses could be made by buying tickets (1 February). *Historical break-up days* are the days the ice broke from 1917 to 2001. *Historical break-up time* is the minute within a day the ice broke from 1917 to 2001. *Individual guesses* are guesses made by individual bettors. *Pool guesses* are guesses made by betting pools. All dates are expressed in the number of days since 31 December. Standard deviations of days are measured in number of days. Time is measured in terms of minutes in a day, from 0 to 1439. Standard deviations of time are measured in minutes.

The evidence displayed in Figure 1 is quite striking. Despite its simplicity, the main prediction (Result 1a) of the equilibrium model in Section 3 seems to be supported by the evidence. In spite of the very low stakes involved and the rather lighthearted nature of the game, bettors appear to take the historical evidence seriously. Of course, it is impossible to determine exactly which factors affect bettors' choices, but Figure 1 suggests that history is an important one. In particular, the dates of 30 April, 5 May, and 8 May are clear attractors and guesses have a tendency to cluster around them.

5.3. *Pools versus individuals*

In Table 1, we present summary statistics for the historical ice break-up days from 1917 to 2001, the historical break-up times, the entire sample of guesses, and the guesses made by individuals and pools. All dates are expressed in the number of days since 31 December 2001, and all times are expressed in minutes of the day.[26]

The historical mean break-up day is 4 May, which is also the mean guess for the entire sample. Although prior literature has documented that groups may perform differently than individuals, *on average* pools did not behave much differently than individuals in our data. The mean guessed date for individuals (124.03) is earlier than for pools (124.25), but both fall within the same day. Thus, although the difference between the mean guess for individuals and for pools is statistically significant (the t-statistic is 7.8), it is arguably not economically significant. The historical mean break-up time is at 2:18 p.m. (858.34 minutes), while the mean guessed time in 2002 was 2:00 p.m. In 2002, the break-up time was at 9:27 p.m. Similarly to the pattern for the dates, the individuals' mean guessed break-up time is earlier (1:58 p.m.)

26. In total, six guesses were for dates prior to the opening of the ticket selling season, one of which was made by a pool. We exclude these guesses from our analysis, since it is irrational to guess a date that has already passed. Results do not change if these guesses are included.

than for pools (2:04 p.m.). These differences are again statistically significant, but whether a 6-minute difference is substantial is less clear.[27]

The break-up date in 2002 was 7 May. Individuals and pools display similar rates of ex post success: roughly 6% of each guessed 7 May. The ex post performance of both groups and individuals is quite good: the proportion that guessed the right day is significantly higher than the historical probability of break-up on that day (which is 2.35%). Interestingly, the largest difference between guess and historical probabilities for all days in our sample occurs on 7 May.

Result 2 from the model implies that the variance of group guesses should be lower than that of individual guesses. We find indeed that the standard deviation of pool guess dates is smaller than the standard deviation of individual guess dates by approximately one day (see Table 1). This result is not driven by differences in sample sizes, since the sample of individual guesses is much larger than the sample of pool guesses. As a consequence, the differences in standard deviations are also statistically significant. The p-value in a variance ratio test of the hypothesis that the standard deviation of individual guess dates is greater than the standard deviation of pool guess dates is less than 0.0001. Thus, group guesses appear to be more moderate than individual guesses.

To characterize the betting strategies in more detail, we analyse the frequency distribution of guesses. We focus primarily on guess dates, since the distributions for guess times, as measured in minutes, are less smooth which complicates the analysis. Figure 2 compares the frequency distribution of pool guesses to individual guesses. While pools place very similar weights on the three main historical spikes, individuals appear to place too much weight on the first spike (30 April) and too little on the third (8 May). Thus, the distribution of guesses made by groups appears to conform more to the distribution of historical break-up dates than the distribution of guesses made by individual bettors. This suggests that individual bettors rely relatively less on historical data than pools.

This pattern is not driven by individuals placing more weight on recent historical information. A similar pattern emerges if we use only the most recent 42 years (50% of the sample) of break-up history (results not reported).[28] It also appears that individuals placed less weight on current environmental information. According to the ice measurements available on the Nenana Ice Classic website, the average ice thickness of the Tanana River in 2002 was higher in the two full months during which tickets could be purchased, February (47″) and March (57.26″), than in 2001 (36.83″ and 39.72″, respectively). Since the break-up day in 2001 occurred on 8 May, four days after the historical mean break-up day, this was at least suggestive that the break-up day in 2002 might also be after 4 May, rather than before (as was indeed the case, since the ice broke on 7 May).

Result 3 of the model suggests that groups should be less likely to guess very early or very late dates than individuals. Consistent with this prediction, we find that the cumulative frequency of individual guesses of dates prior to the earliest recorded break-up day of 20 April (0.0156) is larger than the respective number for pools (0.0088). The cumulative frequency of

27. In most of our analysis, we do not emphasize the concept of statistical significance. There are two reasons for this. First, our sample comes very close to covering the entire relevant population. Thus, it is unclear what the gains from applying sample theory are in such cases. Second, with more than 290,000 observations, almost any difference we encounter will be highly statistically significant, even though it may not be economically significant.

28. In an article that appeared in *Science*, Sagarin and Micheli (2001) argue that the trend in historical ice break-up dates recorded by the Nenana Ice Classic indicates the presence of global warming. However, the article raised some controversy (Daly, 2001; O'Ronain, 2002), because, for example, it did not include the 2001 break-up day, which occurred four days after the mean historical break-up day of 4 May, in its analysis. Thus, the distribution of historical break-ups days until 2002 does not show clear evidence that betting earlier was a better strategy.

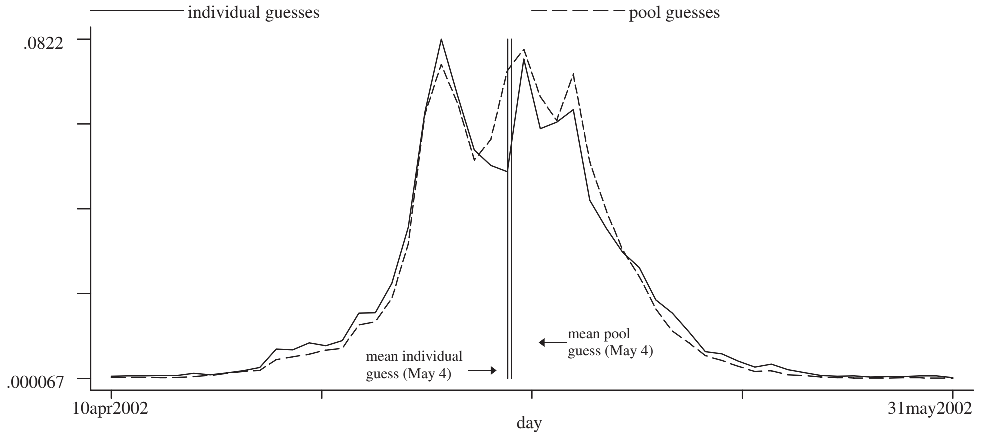


FIGURE 2

Comparison of frequency distribution of days of individual guesses to days of pool guesses in 2002 (sample of individual guesses: 215,713 observations; sample of pool guesses: 74,332 observations; only guesses from 10 April to 31 May displayed)

individual guesses of dates after the last recorded break-up day of 20 May (0.0015) is also larger than for pools (0.0008).

The fact that individuals are more likely to guess dates both earlier than 20 April and later than 20 May is consistent with our hypothesis that groups tend to make less extreme decisions than individuals. Figure 2 provides further evidence in favour of the idea that groups are more moderate. As is clear from the figure, individuals place more weight on the tails of the distribution than groups do. The two distributions cross only twice. To the left and right of the crossing points, the individual distribution always lies above the pool distribution, whereas the pool distribution lies above the individual distribution in between the crossing points. The cumulative frequency of individual (pool) guesses to the left of the earliest crossing point (on 3 May) is 0.417 (0.377); the cumulative frequency of individual (pool) guesses to the right of the latest crossing point (11 May) is 0.102 (0.082). Furthermore, the cutoffs for 5%, 10% and 20% of the individual guess distribution are in each case a day earlier than the corresponding cutoffs for the pool guesses. The cutoffs for 95% and 90% of the individual guess distribution are both a day later than for the pool guess distribution (the 80% cutoff is the same for both distributions).

Our strategy of estimating the probability density functions of guesses with histograms has potential pitfalls. In particular, the excessively "jagged" appearance of the functions displayed in Figures 1 and 2 suggests that one could improve the precision of the estimates by using a non-parametric smoothing procedure. In Figure 3, we display kernel density estimates of the individual and pool guess distributions using the Epanechnikov kernel function with a bandwidth of two days. This choice of bandwidth is just sufficient to make the distributions look "single-peaked". The key feature of Figure 3 is that it highlights, even more strongly, the fact that individuals tend to place more weight on the tails of the distribution. Our results, therefore, do not appear to be driven by undersmoothing.[29] If anything, the differences between

29. Because we do not use information about the guessed minutes in Figures 1 and 2, our graphs of frequency distributions also represent smoothed versions of the true frequency distribution of guesses. The same is true of the historical break-up frequency.

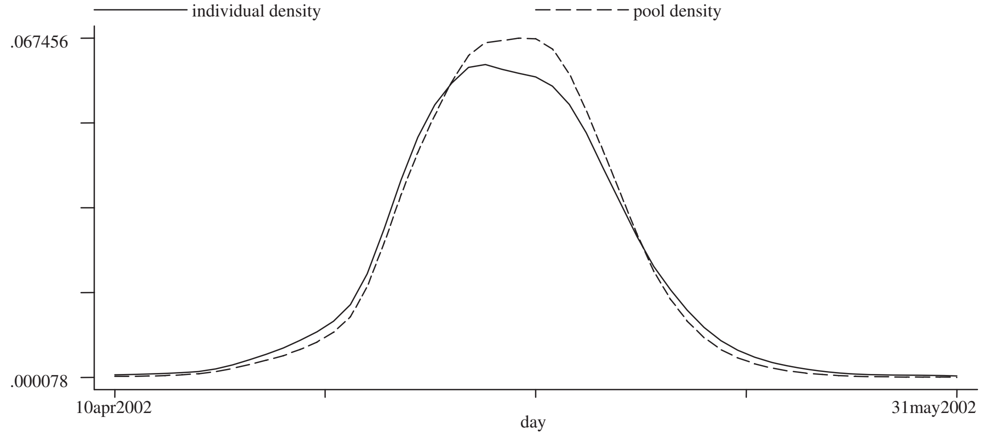


FIGURE 3

Kernel density estimates of days of individual and pool guess distributions in 2002 (sample of individual guesses: 215,713 observations; sample of pool guesses: 74,332 observations; only guesses from 10 April to 31 May displayed. Densities estimated using the Epanechnikov kernel function with a bandwidth of 2 days)

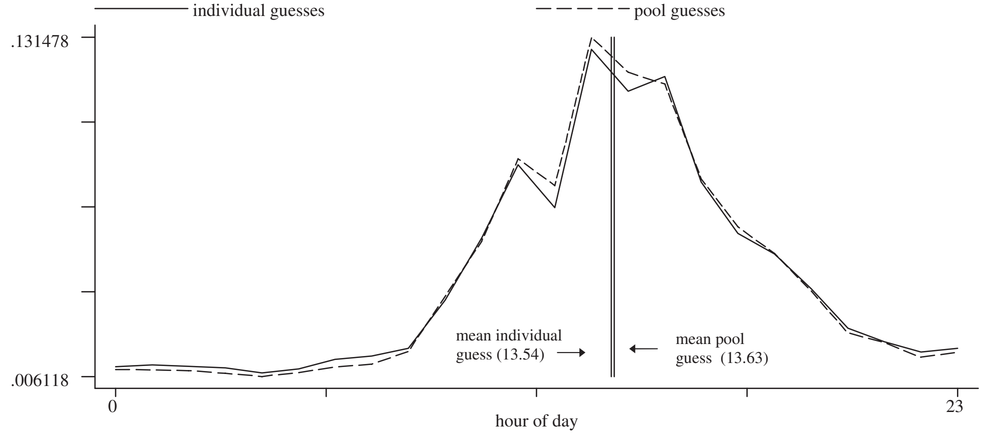


FIGURE 4

Comparison of frequency distribution of hours of individual guesses to hours of pool guesses in 2002 (sample of data on hours of individual guesses: 208,765 observations; sample of data on hours of pool guesses: 72,247 observations)

the weights which groups and individuals assign to the tails appear to be much more pronounced when we use kernel estimation techniques.

From Table 1, we see that the standard deviation of individual guess times is 12 minutes larger than the standard deviation of pool guess times (this difference is statistically significant), which is consistent with the pattern we find for the guess days. In Figure 4, we plot the histograms for individuals and pools based on the hours of the guessed times. A similar pattern as in Figure 2 emerges: individuals are more likely to guess both earlier and later hours of the day than pools. The differences between the two histograms are much less pronounced, however, with the distributions crossing each other six times, and overall behaving very similarly. We conclude that the evidence for the hours and minutes of the day also suggests that pools make less extreme guesses, although the differences between the guess time distributions are less pronounced than for the guess dates.

6. SELECTION, MEASUREMENT ERRORS AND ADDITIONAL EVIDENCE

So far, we can only interpret our evidence in terms of the *joint hypothesis* that groups make more moderate decisions either because they require compromise among their members or because individual characteristics that are correlated with betting behaviour affect group membership (or both). But, in order to obtain a clearer picture of the differences between group and individual behaviours, it would be nice to know whether our results are completely driven by selection, for reasons other than the ones we have discussed in the theory section.

Our hypothesis is that we observe more moderate guesses in groups either because group members compromise or because individuals choose not to be group members to avoid compromising (or equivalently that groups terminate membership of individuals not willing to compromise). Our model clarifies the role of actual or potential compromise for membership decisions. However, it could also be that personal characteristics (and in particular characteristics other than betting preferences) bring similar people together, so that groups may appear more moderate despite the absence of actual or potential compromise. In order to address this possibility, in this section we attempt to control for potential selection on personal characteristics.

There are two main types of selection mechanisms that are relevant for our analysis. First, individuals choose whether or not to participate in the game by making a guess. Second, conditional on participation, individuals choose whether to make guesses individually, in groups, or both.

The first selection problem is virtually impossible to address. This is not necessarily a problem, because it is still an interesting question to compare the betting behaviour of groups versus individuals conditional on the decision to participate. Of course, if the characteristics that affect participation also affect the behavioural differences between groups and individuals, then it would be difficult to generalize our results to different settings.[30]

It is also difficult to examine the second selection issue, because it is not clear which types of individuals self-select into groups. Barber *et al.* (2003) face similar problems in their study of stock clubs, but argue that their experimental evidence suggests that selection is not likely to be a major issue. Hamilton, Nickerson and Owan (2003) provide evidence that more productive workers self-select into teams. However, Kocher, Strauß and Sutter (2006) show the opposite in the context of a beauty-contest game. Individuals self-selected into groups hoping to improve performance, while individuals who opted to stay alone appeared to value autonomy in decision making. Nevertheless, we use three complementary methods to examine whether selection alone can explain our results.

1. *Individual fixed effects*: we use individual fixed effects in a subsample of individuals who appeared to make guesses both as individuals and as part of a pool. This analysis suggests

30. Because the Ice Classic has a longstanding tradition in Nenana and Alaska, selection due to the decision to participate may not be as important a concern as in other games. In our data, we identify 31,563 different individual bettors. The 2000 US census estimated the populations of Alaska and Nenana to be 626,932 and 402, respectively. Even though people outside Alaska can bet, it is more difficult because they must contact the Ice Classic office directly. If we assume that most bettors are from Alaska, this suggests that roughly 5% of the population placed a bet. This is a substantial fraction of the population. Because accounts of the Ice Classic picture it as a community affair that provides part-time employment for roughly 100 residents, it is likely that most Nenana residents bet.

that individual characteristics other than betting preferences do not fully explain the difference between group and individual guesses.

2. *Family pools*: we identify groups that appeared to be formed on the basis of family ties. Such groups are less likely to be formed as a consequence of the similarity of opinions of their members, although family pools are also possibly more homogeneous since family members may have similar opinions.

3. *Pool size*: we identify a subsample of pools for which we can infer group size. We then analyse the effect of group size on the variability of guesses.

None of these approaches rules out the importance of selection, but together they provide evidence that selection alone is not sufficient to explain our results. Because family pools and pools for which we can infer the number of members are less likely to be misclassified as pools, this analysis also acts as a robustness check that measurement error in our pool classification is not driving the results.

6.1. *Individual fixed effects*

It is possible that individuals with more moderate predictions will be more likely to join a betting pool. One reason is the membership effect we discussed in Section 3: people with more extreme predictions may choose not to join groups if they expect groups to compromise (or similarly groups may refuse membership to individuals with extreme predictions). Another possibility is that omitted personal characteristics may be correlated both with moderation in betting preferences and with group membership decisions. If the latter is true, then groups may appear more moderate even if there is no (actual or expected) compromise effect.

To address this issue, we construct a dataset of guesses placed by individuals who also bet in groups. We then use individual fixed effects to identify within-individual variation in betting strategy that can be attributed to betting in a group. For example, if we define b_{ijk} to be a measure of betting strategy for guessed date k of individual i who also guesses as part of pool j (where pools include pools with multiple members and the extreme "pool" of betting alone), POOL_j to be a dummy variable which is equal to 1 if POOL_j has more than one member and 0 otherwise, γ_i to be individual fixed effects and ε_{ijk} to be error terms, then *for the same individual* we can identify the effect of betting in a group on betting strategy by estimating the following regression:

$$b_{ijk} = \alpha + \beta\text{POOL}_j + \gamma_i + \varepsilon_{ijk}. \tag{17}$$

While ex ante it seems difficult to identify individuals who also bet in groups in our dataset, we often observe what appear to be married couples betting together as well as individually. For example, we observe guesses placed by "Ray & Ruth Birky", as well as guesses placed by "Ray Birky" and "Ruth Birky". We also observe individuals who appear to be betting with a pool, for example, "Andrea F. Staniforth & Co" and "Helen Lazeration Pool" as well as on their own, for example, "Andrea F. Staniforth" and "Helen Lazeration". Under the assumption that no other individual with exactly the same name is betting, for example, there are not two different bettors called "Andrea F. Staniforth", and that guesses associated with such individuals' names are truly individual guesses, we can construct a dataset of guesses placed by individuals who also bet in pools. While these assumptions are strong, we believe the second assumption would simply bias us against finding a result. If guesses associated with an individual's name who also bets in a pool are also pool guesses, then β should not be significantly different from 0. It is less clear how the first assumption would affect our results.

If the guesses we associate with, for example, “Andrea F. Staniforth”, are individual guesses but are not those belonging to the Andrea F. Staniforth of “Andrea F. Staniforth & Co”, then we are still comparing guesses of individuals to guesses of pools but in this case the individual fixed effect only captures the effect of having the same name and is essentially meaningless. We believe it unlikely that many bettors with *exactly* the same names will be present in our data,[31] but our results should still be interpreted with care given that we cannot cross-check the identity of the individuals.

To construct our sample for this section, we examined all pools that contain the names of individuals and searched for other occurrences of the individual’s name. We were able to identify 351 individuals whose name appeared alone as well as in the name of a pool. Our sample of guesses consists of all guesses made by individuals whose name also occurred in the name of a pool as well as those of the associated pools.

We examine three measures of an individual’s betting strategy: the guess, the absolute value of the distance of the guess to the sample mean guess and the square of this absolute value. We focus primarily on the latter two measures of guess dispersion.

In Table 2, we report the results of estimating the above equation for both guess dates and guess times (minutes of the day).[32] From columns I and IV, we see that groups guessed earlier dates but later times. However, these differences are not large: mean pool and individual guesses are on the same day, and pool guesses are on average 17 minutes later than individual guesses. More important are the results for the variability measures. Columns II, III, V and VI all show that both the absolute value and the square of the deviation from the mean are lower for pool guesses, both for the guess day and for the guess time. Consider, for example, column III. The coefficient on the constant can be interpreted as the variance of individual guesses, so the implied standard deviation of individual guesses is 6.72 days, a number that is slightly lower than in Table 1 (6.86 days). The implied standard deviation for pools is 6.29 days, which is higher than in Table 1 (5.95 days). If we consider the guess time instead, the implied standard deviation for individuals is 278.86 minutes, which is higher than in Table 1 (263.89 minutes), while the implied standard deviation for pools is 238.87, which is lower than in Table 1 (251.23).

The results show that group guesses display lower standard deviations than individual guesses, for both guess days and guess times, in a subsample of guesses by individuals who made guesses both individually and as part of a group. Thus, individual characteristics may explain some of the difference between the standard deviations of the two groups, but not all of it. These results appear even more striking if one recalls that all regressions include 351 individual dummy variables.

Why would an individual who chooses to make multiple guesses make his more extreme guesses alone? Either groups choose a betting rule that leads to compromise or individuals choose not to make their extreme guesses in groups.[33] Similarly, the group members betting together with such an individual may refuse to be part of a betting pool that places an extreme

31. As robustness checks, we created measures of “common” names, like “Johnson”, and “uncommon” names, like “Arrants”. To help in our classification, we used the number of unique individuals with the same last name as a guide. For example, there were roughly 233 individual bettors with the last name “Johnson”, but only one bettor with the last name “Arrants”. Thus, “Johnson” is common, while “Arrants” is uncommon. We redid our analysis after excluding the common names, since for the uncommon names it was less likely that another individual with exactly the same name was betting. Naturally, this procedure reduces our sample substantially, so our results are statistically weaker. Nevertheless, the signs of our coefficient estimates are the same as when we use our full subsample.

32. Reported constants are averages over all individual effects.

33. Of course, this does not rule out the possibility that individuals expect groups to compromise even when they do not.

TABLE 2
Individual fixed effect results for individuals betting alone and in groups

	I	II	III	IV	V	VI
	Guess date (in days)			Guess time (in minutes)		
	Guess	Abs. distance to mean guess	Square distance to mean guess	Guess	Abs. distance to mean guess	Square of distance to mean guess
Pool dummy	−0.712***	−0.242*	−5.657**	17.582*	−29.381***	−20,704.424***
	[3.50]	[1.84]	[2.09]	[1.89]	[4.86]	[4.96]
Constant	124.200***	5.110***	45.168***	836.586***	210.030***	77,761.704***
	[1003.41]	[62.61]	[26.15]	[139.80]	[54.21]	[28.49]
Observations		5431			5268	

Notes: The sample consists of observations on individual and group guesses for 351 individuals whom we classified as betting on their own and in a group. We exclude six guesses of days prior to the date guesses could be made by buying tickets (1 February). An individual was considered to bet on her own and in a group if her name appeared both on its own and as part of a group name in the list of bettors, for example, "Lois Swanberg" and "Lois Swanberg & Bob Hager". Columns I–III show individual fixed effect regressions of the guessed day, the absolute value of the distance to the sample mean guessed day and the square of that distance on a group dummy. This dummy is defined to be equal to one if the bettor's name suggests it is a group with more than one member, for example, "Lois Swanberg & Bob Hager" and is 0 otherwise. The guess day is measured in days of the year since 31 December. Columns IV–VI replicate the regressions in columns I–III using the guessed time. The guess time is measured in minutes of the day. All standard errors are corrected for heteroscedasticity. The absolute values of the robust t-statistics are in brackets. *, **, *** indicate significance at the 10%, 5% and 1% level, respectively.

bet. All these possibilities are plausibly related to betting preferences and the expectation of compromising in groups. However, what the fixed effects results rule out is the possibility that moderation in groups is fully driven by personal characteristics (other than betting preferences) that determine the match between group members.[34]

6.2. *Family pools*

To address the possibility that our results may be driven by the fact that individuals self-select into groups based on similarities in betting preferences, we try to identify pools that are formed for reasons unrelated to betting preferences. Family pools are a natural example of pools formed for "exogenous" reasons, because families are likely to bet together even if they disagree about dates. While we cannot control for initial selection (whether families choose to participate at all) or family member inaction (some family members may not participate in the decision making process), we can be reasonably confident that family composition is not driven by betting preferences. However, it is still possible that family pools are more homogeneous in their opinions than non-family pools.

34. The fixed-effect analysis does not hold the characteristics of other group members constant. However, we can control for partner effects to a certain extent by restricting our sample to individuals who bet alone and with members of the opposite sex. For example, we can restrict our sample to women who bet alone and in groups with only one other person who is male, for example, Barbara Bluekens versus Barbara & Thomas Bluekens. In this restricted sample, we find a coefficient of −0.449 (significant at the 10% level) on the pool dummy in an individual fixed effect regression with the absolute value of the distance to the mean guess as the dependent variable. This is at least suggestive that our results are not fully driven by characteristics of partners that may determine the match between group members.

TABLE 3
Comparison of individual guesses to guesses of family pools

	I	II	III	IV	V	VI
	Guess date (in days)			Guess time (in minutes)		
	Guess	Abs. distance to mean guess	Square distance to mean guess	Guess	Abs. distance to mean guess	Square of distance to mean guess
Family pool	−0.322***	−0.159***	−6.485***	8.364***	−9.762***	−9154.122***
	[5.13]	[3.90]	[5.61]	[3.38]	[6.04]	[8.66]
Constant	124.029***	4.988***	46.787***	838.444***	196.225***	69,636.798***
	[8421.59]	[495.03]	[64.81]	[1451.72]	[508.12]	[264.35]
Observations		226,511			219,202	

Notes: The sample consists of observations on guesses made by individuals and 965 family pools. We exclude six guesses of days prior to the date guesses could be made by buying tickets (1 February). A pool is classified as a family pool if it is clear that the members of the pool have the same last name, e.g. the bettor's name is "Ray & Ruth Birky", or are members of the same family, for example, "John Pickett & Sons". Columns I–III show regressions of the guessed day, the absolute value of the distance to the sample mean guess day and the square of that distance on a "Family pool" dummy. This dummy is defined to be equal to one if the bettor is classified as family pool and is 0 if the bettor is classified as an individual. The guess day is measured in days of the year since 31 December. Columns IV–VI show the same regressions using the guess time. The guess time is measured in minutes of the day. All standard errors are corrected for heteroscedasticity. The absolute values of robust t-statistics are in brackets.
*, **, *** indicate significance at the 10%, 5% and 1% level, respectively.

Our subsample for this section is the set of observations on guesses made by individuals and family pools. We classify a pool as a family pool if it is clear that the members of the pool have the same last name, e.g. the bettor's name is "Ray & Ruth Birky", or are members of the same family, for example, "John Pickett & Sons". Of the 3093 pools in our sample, 965 (31.12%) are family pools.

Table 3 shows the results of similar regressions as in Table 2 (without fixed effects). The results are similar to those for the full sample. For the guess days, the implied standard deviation for individuals is 6.86 days, which is identical to the number reported in Table 1 (as it should, because the sample of individuals is not restricted), while the implied standard deviation for pools is 6.35 days (higher than the number in Table 1: 5.95). For the guess times, the implied standard deviation for individuals is 263.89 minutes, while the implied standard deviation for pools is 245.93 days (lower than the number in Table 1: 251.23).

We conclude that family pools exhibit less extreme betting behaviour than individual bettors. These results also help to address concerns about misclassification, as family pools (or pools that use family names or family identifiers) are straightforward to classify as pools.

6.3. *Variation in pool size*

Our model implies that guess variance should decrease with pool size if membership is exogenously determined (see Result 2).[35] Thus, we could reject the exogenous membership hypothesis if we find evidence that the variance is invariant or increasing in pool size. We investigate this issue in a subsample of observations on guesses placed by pools whose names

35. By exogenous membership we mean that betting preferences are not the reason groups are formed, not that groups are formed randomly.

TABLE 4
Variation in guesses according to pool size

	I	II	III	IV	V	VI
	Guess date (in days)			Guess time (in minutes)		
	Guess	Abs. distance to mean guess	Square distance to mean guess	Guess	Abs. distance to mean guess	Square of distance to mean guess
Pool size	0.002	−0.052***	−0.796***	1.014***	−1.054***	−562.344***
	[0.95]	[36.79]	[8.68]	[8.64]	[13.07]	[11.97]
Constant	123.868***	4.540***	40.849***	825.852***	219.411***	84,446.079***
	[2268.21]	[119.87]	[13.15]	[333.56]	[134.40]	[74.37]
Observations		18,964			18,434	

Notes: The sample consists of observations on guesses made by 1244 pools whose names could be used to proxy for their sizes. We exclude six guesses of days prior to the date guesses could be placed by buying tickets (1 February). A pool's name was considered to reflect its size if its name consisted of a list of individual names, for example, "Brett Miller/Bruce Atkinson" (a size of 2) or its name suggested the number of members, for example, "7 Lucky Ladies" (a size of 7). To control for numbers that were potentially introduced into a pool's name by scanning error, we restrict ourselves to pools whose name occurs more than once as a bettor. Columns I–III show regressions of the guessed day, the absolute value of the distance to the sample mean guess day and the square of that distance on pool size. The guess day is measured in days of the year since 31 December. Columns IV–VI show the same regressions using the guessed time. The guess time is measured in minutes of the day. All standard errors are corrected for heteroscedasticity. The absolute values of the robust t-statistics are in brackets.
*, **, *** indicate significance at the 10%, 5% and 1% level, respectively.

can be used to proxy for their sizes. We assume a pool's name reflects its size if its name consists of a list of individual names, for example, "Brett Miller/Bruce Atkinson" (a size of 2) or its name suggests the number of members, for example, "7 Lucky Ladies" (a size of 7). We were able to assign a pool size to 40.22% of pools (1244) because a large number of pools consists of couples (1149). Average pool size is 2.53 with a maximum of 50. To control for numbers that were potentially introduced into a pool's name by scanning error, we restrict ourselves to pools whose name occurs more than once in the list of bettors.

In Table 4, we report the results of the same types of regressions as in Table 3. For both dates and times, guess variability decreases with pool size. Our results imply that a pool of size 2 would have a standard deviation of guess dates of 6.27 days, while a pool of size 7 would display a standard deviation of 5.93 days. The analogous effect for guess times is smaller, with a change in the standard deviation of roughly 6 minutes.

The evidence from this table suggests that group size is negatively related to the variability of guesses. Thus, larger groups may be characterized by more moderation in decisions, or larger groups attract individuals with more moderate opinions, or both. Since we find similar results when we compare pools of different sizes as when we compare pools and individuals, misclassification of bettors into individuals and pools does not appear to be the primary explanation of our results.

7. THE 2008 DATA: ADDITIONAL EVIDENCE AND ROBUSTNESS

Although we discussed potential measurement error problems and addressed them in different ways in the previous section, in this section we reexamine the measurement error problem in order to get a sense of its magnitude. As we discussed earlier, there are several reasons to believe that our results are not generated by measurement error. Instead, it is more likely that measurement error introduces noise that will cause attenuation bias, making it harder for us to find that pools are more moderate. The question is, how large will this bias be?

To examine this issue, we tried to obtain a list of pools that registered their members with the Nenana Ice Classic office in 2002.[36] For these pools, there can be no doubt that they are pools, so there is no measurement error associated with the assignment of their guesses. Unfortunately, we were unable to obtain this information for 2002. However, we were able to obtain this information for 2008 along with the Nenana Ice Classic's *List of Guesses* for 2008. We define an *official pool* to be a pool that registered its name with the Nenana Ice Classic office. We followed a similar procedure for scanning the data as for the 2002 *List of Guesses* and used a similar pool classification method to assign guesses to pools (other than the official ones) and individuals as in 2002. However, we used only a simplified cleaning procedure, because the full cleaning procedure we employed for the 2002 data was too time intensive. This limits the usefulness of the 2008 data (e.g. we do not have enough detailed data to replicate the analysis from Section 6 fully). However, the availability of registered pool data is a major advantage.

The 2008 *List of Guesses* contains 239,579 guesses. After scanning and cleaning, we are able to classify 238,684 guess dates placed by 41,977 different bettors and 238,307 guess times according to whether they were placed by individuals or pools.[37] We identify a total of 4998 pools, 150 of which are official pools. These 150 pools account for 27.56% of all pool date guesses (49,336) and 27.54% of pool time guesses (49,262). This suggests that pools are more likely to register their members with the Nenana Ice Classic office if they place many guesses. Indeed, official pools placed on average 83.7 more guesses than other pools, a difference which is significant at lower than the 1% level.

As a first step towards assessing the measurement error problem, we examine the list of official pools to see whether our method of assigning bettors to the pool category in 2002 is reasonable. Two pool names contain the identifier "&", eight names contain one of the words "bunch", "clan", "committee", "company", "pack", "partners", "kids" and "sisters", two names contain the word "gang", two names contain the word "crew", two names contain the word "group", two names contain the word "staff", two names contain the word "team", five names contain the word "family", 15 names contain a number, for example, "6 R US" and "The Three Amigos", and 81 names contain the word "pool". The remaining names are names such as "CHILLY DOGS" and "Northwest Airlines" that do not contain a clear pool identifier, but that also do not contain the name of an individual in it. Thus, these are names that we would not have assigned to the individual category based on our merge with census data. Instead, we would have classified these as pools by hand. We conclude that we would have correctly identified each official pool as a pool based on our 2002 classification method. We also note that no official pool is registered under the name of an individual without an accompanying pool identifier. This is at least suggestive that bettors who bet under individual names are not likely to be pools.

Next we use the 2008 data to compare the betting behaviour of pools and individuals.[38] This serves as a robustness check that our findings are not specific to a given year. We replicate our previous figures using the entire sample of guesses. The figures are similar to those in 2002.

36. The Nenana Ice Classic encourages pools to register their members to avoid disputes.

37. In 2008, ten guesses were for dates prior to the opening of the ticket selling season, none of which were placed by pools. We exclude these bets from our analysis, since it is irrational to guess a date that has already passed.

38. We have information on 511 non-missing guess dates and 510 non-missing guess times placed for dates between 1 February and 5 April, which is the period during which guesses could be made by buying tickets. Of these, 66 date and time guesses were made by pools. These numbers are comparable to the numbers for 2002.

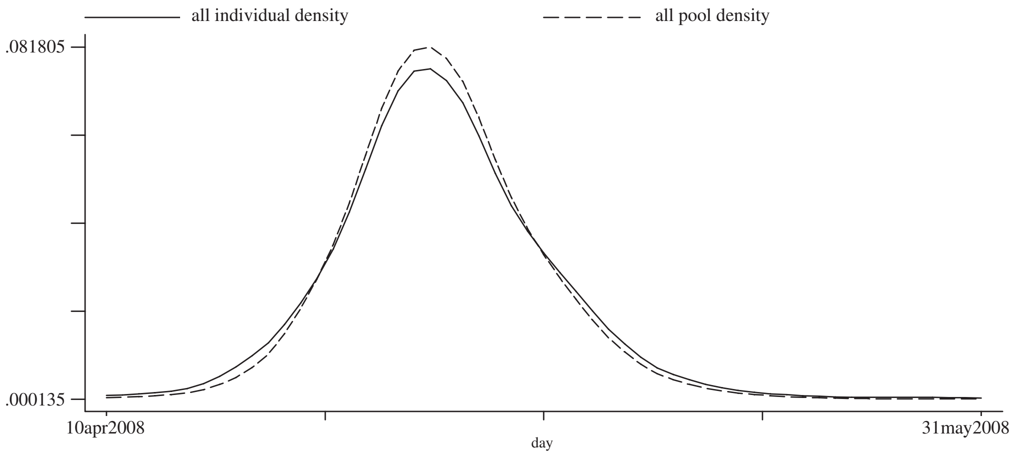


FIGURE 5

Kernel density estimates of days of individual and pool guess distributions in 2008 (sample of individual guesses: 189,348 observations; sample of pool guesses: 49,336 observations; only guesses from 10 April to 31 May displayed. Densities estimated using the Epanechnikov kernel function with a bandwidth of 2 days)

For the sake of brevity, we provide only the replication of Figure 3 in Figure 5. From Figure 5, it is clear that in 2008 individuals also tend to place more weight on the tails of the distribution than pools.

Finally, as an alternative method of addressing potential measurement error problems, we compare the behaviour of *official* pools to the behaviour of individuals. Because we are certain that official pools are indeed pools, this comparison will be subject to measurement error problems only if pools bet under the names of individuals. The official pool classification thus represents the cleanest measure of pools available and we can use it both to verify that our previous results are not an artifact of measurement error and to assess the magnitude of the measurement error problem.

Because the official pools represent a subsample of all pools, we first compare official pools to a random subsample of individuals with roughly the same amount of guesses as the official pools. We construct the sample of individuals by sorting them alphabetically and truncating the list of names when the cumulative number of individual guesses is at least as large as the number of official pool guesses (13,598 guess dates and 13,568 guess times). This gives us a sample of 2612 individuals with 13,617 guess dates and 2607 individuals with 13,571 guess times. To ensure that our choice of random sample is not biasing the results, we also compare official pools to the sample of all individuals.

We report the results in Table 5. We first regress the guessed date, the absolute value of the distance of the guess date to the sample mean guess, and the square of this absolute value on a pool dummy variable. "Pool dummy" is defined to be one if a bettor is classified as a pool using our 2002 methodology. We show the results for three different sample types: the full sample, official pools plus random individuals, and official pools plus all individuals. The top panel of Table 5 shows the results. The bottom panel of Table 5 shows the same regressions using the guess time.

Overall, we find that the 2008 data deliver results that are qualitatively identical to the 2002 results (i.e. groups are more moderate). The pool dummy enters with a negative sign in

TABLE 5
Variation in guesses in 2008

	Full sample			Random individuals + official pools			All individuals + official pools		
	I	II	III	IV	V	VI	VII	VIII	IX
	Guess	Abs. distance to mean guess	Square distance to mean guess	Guess	Abs. distance to mean guess	Square distance to mean guess	Guess	Abs. distance to mean guess	Square distance to mean guess
					Guess date (in days)				
Pool dummy	−0.045	−0.538***	−11.493***	−0.672***	−0.932***	−25.752***	−0.675***	−0.775***	−16.590***
	[1.50]	[25.07]	[8.05]	[8.77]	[16.31]	[4.70]	[14.35]	[23.33]	[6.72]
Constant	121.294***	4.525***	43.966***	121.292***	4.683***	53.127***	121.294***	4.525***	43.966***
	[7959.99]	[406.34]	[58.76]	[1941.77]	[97.84]	[10.72]	[7959.98]	[406.34]	[58.76]
Observations		238,684			27,215			202,946	
					Guess time (in minutes)				
Pool dummy	3.041**	−4.436***	−4,849.930***	9.282***	−6.515***	−6,249.972***	3.413*	−8.194***	−8,033.334***
	[2.50]	[5.44]	[8.79]	[3.21]	[3.37]	[4.83]	[1.65]	[5.98]	[8.99]
Constant	867.780***	181.822***	61,628.443***	861.911***	180.142***	59,845.072***	867.780***	181.822***	61,628.443***
	[1519.85]	[467.71]	[229.55]	[410.56]	[126.78]	[61.54]	[1519.85]	[467.71]	[229.55]
Observations		238,307			27,138			202,613	

Notes: The full sample consists of observations on guesses in 2008. We exclude 10 guesses of days prior to the date guesses could be made by buying tickets (1 February).The top panel shows regressions of the guess day, the absolute value of the distance to the sample mean guess day and the square of that distance on the pool dummy in three different samples: the full sample, official pools plus random individuals, and official pools plus all individuals. The bottom panel shows the same regressions using the guess time. "Pool dummy" is defined to be one if a bettor's name contains a pool identifier, such as the word "pool", "group", "family", "and", or a symbol such as "&" or "/", or is an official pool. An official pool is one of 150 pools that officially registered its members with the Nenana Ice Classic office. The "official pools plus random individuals" sample includes only official pools and a random subsample of 2612 (in the top panel) or 2607 (in the bottom panel) individuals with roughly the same amount of date guesses (13,617 in the top panel) or minute guesses (13,571 in the bottom panel) as official pools. The "official pools plus all individuals" sample includes only official pools and the subsample of all individuals. The guess day is measured in days of the year since 31 December. The guess time is measured in minutes of the day. All standard errors are corrected for heteroscedasticity. The absolute values of the robust t-statistics are in brackets.
*, **, *** indicate significance at the 10%, 5% and 1% level, respectively.

all regressions using absolute and square distances as dependent variables.[39] The difference in variances between groups and individuals is similar to the one found in the 2002 data. For example, from Column III we find that the standard deviation for individuals is 0.93 day larger than the one for pools (in 2002, this difference was 0.9 day; see Table 1). Once we use the official pool classification compared to the full sample of individuals (Column IX), this difference increases to 1.4 days. Thus, our results suggest that potential measurement errors may have an effect of 0.5 day on the difference in standard deviations.[40]

Our conclusions from this section are as follows. First, as in 2002, the 2008 data unambiguously show that the distribution of pool guesses displays lower variability than the distribution of individual guesses, both for guess dates and for guess times. Thus, our results are robust to the choice of year. Second, our 2002 classification procedure correctly classifies all official pools in 2008 as pools. Third, at least in the 2008 data, consistent with an attenuation bias, potential measurement error in the classification of pools seems to reduce the differences in variances.

8. FINAL REMARKS

The main contribution of this paper is to compare individual and group outcomes in a setting in which group formation is endogenous. We use a unique dataset on guess portfolios to compare the decisions made by individual bettors to those by betting pools. We find that the decisions made by pools are more moderate than the decisions made by individuals. On average, individuals appear to rely less on historical information than groups. They also place more weight on outlying days and times than groups do. Both these pieces of evidence are consistent with the idea that group decisions are more moderate, either due to the fact that groups have to reach a compromise when their members disagree, or because individuals with extreme opinions are less likely to be part of a group, or both.

Our results may be special to situations in which participants behave in a risk or fun seeking way, and situations in which individuals can choose to act alone or with a group. But they are an important complement to the existing (usually experimental) literature on group decision making. Motivated by some of the findings in this literature, economists have recently displayed renewed interest in group decision-making biases such as risky shifts, group polarization and groupthink (see e.g. Eliaz, Ray and Razin, 2006; Glaeser and Sunstein, 2007; Bénabou, 2008). Consistent with the modern psychology literature, our results highlight that cognitive biases such as group polarization, risky shifts, and groupthink may not always be present, or may not always dominate other characteristics of group decision making that can lead to moderation in groups.

APPENDIX: DATA CLEANING FOR 2002 DATA

We labelled name observations as possibly problematic if they were unique observations. For example, there were a series of guesses by BOB JOHNSON and a single guess by 8OB JOHNSON. This single guess was identified as problematic. These problem observations were checked manually one-by-one. Observations with wrong entries were fixed when the correct name was clear (as in 8OB JOHNSON).

39. In 2008, pools guessed on average earlier dates but later times of the day. As in 2002, these differences are economically small (again, on average pools and individuals guess the same day).

40. The measurement error in 2008 (except when using the official pools) is likely to be larger than in 2002, because we did not apply the same cleaning procedure nor cleaned as many observations by hand as we did in 2002 (due to the amount of time it would have taken).

For the final phase of the cleaning process we used a program to clean the names as follows. Each name was compared to every other name by comparing each character in the same position in each word; the number of character matches was saved. For instance, a comparison of BOB JOHNSON to 8OB JOHNSON results in ten matches, as all but the first characters match. These two words are then given a matching percentage of 91 (total matches = 10)/(length of word = 11). All words are compared and the highest matching percentage is recorded so that each word is associated with the word which it best matches. However, scanning errors sometimes change the length of the names, which creates a shifting problem for these comparisons. A common scanning error occurred when the character "M" scanned as "14". In a comparison of a name like MARY to 14ARY, the scanning error would not be picked up, because the program would make the following comparisons. First, "M" would be compared to "1", which is not a match. Then, "A" would be compared to "4", "R" to "A", and "Y" to "R", yielding zero matches. To help remedy this problem, we amended the program to also check for matches starting from the back of the words. Whichever matching percentage was higher between the forward check and the backward check was used as the matching percentage for those names. In order to increase the program's speed during this checking process, we only ran the backward check if the words were of different lengths. We also skipped checking words that differed in length by three or more letters, as it was highly unlikely that scanning errors could have accounted for such a large difference.

The next step was to determine whether we wanted to replace a name by its best match. If the matching name occurred more frequently than the original name in the list of guesses, it was less likely the product of scanning error than the original name. We therefore replaced the original name with its match only if the match percentage was greater than 85% and the matching name occurred more frequently than the original name.

One final issue was the possibility of an original name being replaced by a match name, which itself was replaced by its match name, as in the following example:

Observation	Name	Frequency in list of bets
1	Bob Johnso	2
2	Bob Johnson	10
3	8ob Johnso	1

In this case, the program will match Observation 1 with Observation 2, Observation 2 with Observation 1, and Observation 3 will match Observation 1. We will end up with the following:

Observation	Name	Frequency	Match name	Match percentage
1	Bob Johnso	2	Bob Johnson	1.0 (10 matches/10 length)
2	Bob Johnson	10	Bob Johnso	0.91 (10 matches/11 length)
3	8ob Johnso	1	Bob Johnso	0.9 (9 matches/10 length)

The program would then replace the name in Observation 1 with the name in Observation 2, as they have a match percentage above 85 and Observation 2 occurs more frequently than Observation 1. Observation 2 will not be replaced as there is no name that occurs more frequently. The problem arises when Observation 3 is replaced with Observation 1, which is itself an incorrect name. We want Observation 3 to be replaced by Observation 1 only if Observation 1 is correct. Otherwise, we want Observation 3 to be replaced by the same replacement as Observation 1. The final portion of the code implements a chaining procedure, which solves this by replacing an original name with the final match of its match name.

After running the program, we determined that removing all spaces from the names would increase the efficiency and accuracy of the program. For example, two names that are clearly the same, such as BOB JO HNSO N and BOBJOHNSON would not match because their percentages were below the threshold due to errant spaces input by the scanner. We therefore reran the program after removing spaces in the names to obtain the final output of the cleaning program.

Acknowledgements. We benefited from comments on previous versions by Heitor Almeida, Jordi Blanes-i-Vidal, Tore Ellingsen, Thomas Kittsteiner, Don Morgan, Marco Ottaviani, Maitreesh Ghatak and Bruno Biais (the editors), and especially two anonymous referees. We thank Cherrie Forness, Manager of the Nenana Ice Classic, for providing us with copies of the List of Guesses and Amanda Wubbold for her assistance with the 2008 data. For excellent research assistance, we thank Elise Aatenir and especially Kyle Lewis.

REFERENCES

ADAMS, R., ALMEIDA, H. and FERREIRA, D. (2005), "Powerful CEOs and their Impact on Corporate Performance", *Review of Financial Studies,* **18**, 1403–1432.

Arctic Science Journeys (1997), "Nenana Ice Classic", Radioscript, http://www.uaf.edu/seagrant/NewsMedia/97ASJ/04.29.97_IceClassic.html.

BÄR, M., KEMPF, A. and RUENZI, S. (2005), "Team Management and Mutual Funds" (CFR Working Paper No. 05-10, Cologne).

BARBER, B. M., HEATH, C. and ODEAN, T. (2003), "Good Reasons Sell: Reason-Based Choice Among Group and Individual Investors in the Stock market", *Management Science,* **49**, 1636–1652.

BÉNABOU, R. (2008), "Groupthink: Collective Delusions in Organizations and Markets" (Working Paper, Princeton University).

BLINDER, A. S. and MORGAN J. (2005), "Are Two Heads Better than One? Monetary Policy by Committee", *Journal of Money, Credit and Banking,* **37**, 789–812.

BONE, J., HEY, J. and SUCKLING, J. (1999), "Are Groups More (or Less) Consistent than Individuals?" *Journal of Risk and Uncertainty,* **18**, 63–81.

CASON, T. N. and MUI, V. L. (1997), "A Laboratory Study of Group Polarization in the Team Dictator Game", *Economic Journal,* **107**, 1465–1483.

DALY, J. (2001), "The Nenana Ice Classic: Betting on Warming", Web Document, http://www.john-daly.com/nenana.htm.

ELIAZ, K., RAY, D. and RAZIN, R. (2006), "A Decision-Theoretic Basis for Choice Shifts in Groups", *American Economic Review,* **96**, 1321–1332.

FARRELL, L., HARTLEY, R., LANOT, G. and WALKER, I. (2000), "The Demand for Lotto: The Role of Conscious Selection", *Journal of Business and Economic Statistics,* **18**, 228–241.

FINKEL, M. (1998), "This Break-up Is the Talk of the Town: When the River Thaws in Nenana, Alaska, a Pile of Money Changes Hands", *Sports Illustrated Magazine,* **88**, R12.

FORBES, D. and MILLIKEN, F. (1999), "Cognition and Corporate Governance: Understanding Boards of Directors as Strategic Decision-Making Groups", *Academy of Management Review,* **24**, 489–505.

FORREST, D., SIMMONS, R. and CHESTERS, N. (2002), "Buying a Dream: Alternative Models of Demand for Lotto", *Economic Inquiry,* **40**, 485–496.

GLAESER, E. L. and SUNSTEIN, C. R. (2007), "Extremism and Social Learning" *Journal of Legal Studies,* (forthcoming).

HAMILTON, B. H., NICKERSON, J. A. and OWAN, H. (2003), "Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation", *Journal of Political Economy,* **111**, 465–497.

JANIS, I. (1982), *Groupthink: Psychological Studies of Policy Decisions and Fiascoes* (Boston: Houghton Mifflin).

KERR, N. (1992), "Group Decision Making at a Multialternative Task: Extremity, Interfaction Distance, Pluralities, and Issue Importance", *Organizational Behavior and Human Decision Processes,* **52**, 64–95.

KOCHER, M., STRAUß, S. and SUTTER, M. (2006), "Individual or Team Decision-Making–Causes and Consequences of Self-Selection", *Games and Economic Behavior,* **56**, 259–270.

KOCHER, M. G. and SUTTER, M. (2005), "The Decision Maker Matters: Individual versus Group Behavior in Experimental Beauty-Contest Games", *Economic Journal,* **115**, 200–223.

KOGAN, N. and WALLACH, M. (1966), "Modification of Judgmental Style Through Group Interaction", *Journal of Personality and Social Psychology,* **4**, 165–174.

LEVINE, J. M. and MORELAND, R. L. (1998), "Small Groups", in Gilbert, D. T., Fiske, S. T. and Lindzey, G. (eds.) *The Handbook of Social Psychology*, 4th edn, vol 2 (New York: McGraw-Hill) 415–469.

MOSCOVICI, S. and ZAVALLONI, M. (1969), "The Group as a Polarizer of Attitudes" *Journal of Personality and Social Psychology,* **12**, 125–135.

Nenana Ice Classic (2002), *List of Guesses* (Alaska: Nenana).

OTTAVIANI, M. and SORENSEN, P. N. (2006), "The Strategy of Professional Forecasting", *Journal of Financial Economics,* **81**, 441–466.

O'RONAIN, M. (2002), "The Nenana Ice Classic 2002 or How to Lose Money Gambling on Global Warming", Web Document, http://www.numberwatch.co.uk/nenana_ice_classic_2002.htm.

PRATHER, L. J. and MIDDLETON, K. L. (2002), "Are N+1 Heads Better than One? The Case of Mutual Fund Managers", *Journal of Economic Behavior and Organization,* **47**, 103–120.

RICHARDS, B. (1995), "Forget the Calendar, Alaskans Bet They Can Spot Spring: This Is the Season for Wagers On When the Tanana Ice–and Winter–Will Break", *The Wall Street Journal,* 20 April, A1.

ROCKENBACH, B., SADRIEH, A. and MATHAUSCHEK, B. (2007), "Teams Take the Better Risks", *Journal of Economic Behavior and Organization,* **63**, 412–422.

SAGARIN, R. and MICHELI, F. (2001), "Climate Change in Nontraditional Data Sets", *Science,* **294**, 811.

SAH, R. K. and STIGLITZ, J. (1986), "The Architecture of Economic Systems: Hierarchies and Polyarchies", *American Economic Review,* **76**, 716–727.

SAH, R. K. and STIGLITZ, J. (1991), "The Quality of Managers in Centralized versus Decentralized Organizations", *Quarterly Journal of Economics,* **106**, 289–295.

SAUER, R. (1998), "The Economics of Wagering Markets", *Journal of Economic Literature,* **36**, 2021–2064.

STONER, J. F. A. (1968), "Risky and Cautious Shifts in Group Decisions" *Journal of Experimental Social Psychology,* **4**, 442–459.

SUNSTEIN, C. R. (2002), "The Law of Group Polarization", *Journal of Political Philosophy,* **10**, 175–195.

SUROWIECKI, J. (2005), *The Wisdom of Crowds* (New York: Anchor Books).

SUTTER, M. (2005), "Are Four Heads Better Than Two? An Experimental Beauty-Contest Game with Teams of Different Size", *Economics Letters,* **88**, 41–46.

THALER, R. and ZIEMBA, W. (1988), "Anomalies: Parimutuel Betting Markets: Racetracks and Lotteries", *Journal of Economic Perspectives,* **2**, 161–174.

THE SEATTLE TIMES (1986), "All Bets are on Ice in Quest for Alaska Cash", *The Seattle Times,* 4 May, E1.

WALLACH, M. and KOGAN, N. (1965), "The Roles of Information, Discussion, and Consensus in Group Risk Taking", *Journal of Experimental Social Psychology,* **1**, 1–19.

WILLIAMS, L. (1999), "Information Efficiency in Betting Markets: A Survey", *Bulletin of Economic Research,* **51**, 1–30.

The Economic Journal, **118** (*January*), 258–283. © The Author(s). Journal compilation © Royal Economic Society 2008. Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

RISK PREFERENCE DIFFERENTIALS OF SMALL GROUPS AND INDIVIDUALS*

Robert S. Shupp and Arlington W. Williams

This research compares lottery valuation decisions made by individuals with similar decisions made by small groups. There is an extensive social psychology literature addressing group versus individual decision making but few studies explore this issue in economic contexts with cash rewards. Willingness-to-pay data elicited from independent samples of individuals and three-person groups in a repeated-measures experimental design reveal that: the variance of risk preferences is generally smaller for groups than individuals and the average group is *more* risk averse than the average individual in high-risk situations, but groups tend to be *less* risk averse in low-risk situations.

Do small groups reveal systematically different risk preferences than individuals and, if so, how do the risk preferences of the individual group members aggregate into a group risk preference? The research reported here explores these topics, motivated by both its importance as a methodological issue in experimental economics and by the long-standing effort in economics and social psychology to describe valuation decisions over risky prospects. Normative models of economic behaviour typically utilise a single objective function that implicitly treats all decisions as individual decisions, even those likely to be made through group interaction in the context of families, committees, management teams, etc. This distinction is important, since laboratory experiments almost exclusively elicit decisions from isolated individuals in spite of the fact that group discussion and problem solving is commonplace in many economic environments. Thus, the external validity of results from many experimental studies hinges partially on whether isolated individuals make decisions that are significantly different from groups when faced with identical information about uncertain outcomes.

In the research reported here, this important methodological issue is addressed using data from an experimental design focused on the potential existence of a group versus individual risk preference differential. This experiment assessed the risk preferences of three-person groups and individuals in the context of revealed certainty equivalents for dichotomous lotteries. Nine maximum willingness-to-pay (WTP) decisions for the right to play each of nine different lotteries were elicited in a non-sequential repeated-measures experimental design. The lotteries ranged from a 10% chance of winning \$20 per person (\$60 per group) to a 90% chance of winning \$20 per person. Losing lotteries paid \$0. Subjects made cash-motivated decisions as either an individual or group member, not both, yielding statistically independent individual and group samples for each of the nine lotteries. After confirming the existence of significant differences in group versus individual decisions using independent samples, a

* The authors gratefully acknowledge constructive comments from Ronald Baker, Timothy Cason, Charles Holt, Steven Kachelmeier, Susan Laury, David Schmidt, Pravin Trivedi, Rusty Tchernis, Konstantin Tyurin, James Walker, this Journal's editors Leonardo Felli and David de Meza, and anonymous referees. The research has also benefited from comments by participants at the regional meetings of the Economic Science Association (September 2000), the Purdue University Economic Theory Workshop (April 2001), and Indiana University students enrolled in the E427 and E627 Seminar in Experimental Economics.

follow-up experiment explored how individual WTP decisions are aggregated into a group WTP decision. In this two-phase sequenced experimental design, participants made individual decisions and were then randomly assigned to a three-person group.

The WTP data from both the independent-samples and sequenced experiments reveal that groups were significantly more risk averse than individuals in the higher-risk lotteries. In the lowest-risk lotteries the independent-samples data suggest that groups tend to be less risk averse than individuals but this finding was not confirmed in the sequenced experiments. Also, group WTP decisions exhibited significantly smaller variance than individual decisions in seven of nine lotteries.

1. Previous Research Examining Group Versus Individual Decisions

There is a long history of social psychologists investigating differences in group versus individual decisions. Studies dating back to the early 1960s found that groups tend to make riskier decisions than individuals in the context of responses to choice dilemma questionnaires (CDQs). This 'risky-shift' finding led to many other CDQ-based studies – see Isenberg (1986) for a more detailed review of this literature – where participants chose actions in hypothetical situations involving risk but without a salient response-contingent reward structure. Further research confirmed that groups tend to make decisions that differ from those of individuals, however, groups do not always make riskier decisions. Thus, 'choice shift' due to 'group polarisation', where individual biases are amplified by group discussion, are now the prevailing terms used to summarise the CDQ-based research. Alternatives to the CDQ methodology are also prominent in the more recent social psychology literature.

A comprehensive review by Kerr *et al.* (1996) concluded 'there are several demonstrations that group discussion can attenuate, amplify, or simply reproduce the judgmental biases of individuals' and 'research conducted to date indicates that there is unlikely to be any simple, global answer to the question' (p. 693). Group discussion of a task appears to improve performance only when there is a 'demonstrably correct' normative solution to the problem under consideration. For the decision task confronted by subjects in the present study, the fact that truthfully revealing WTP for each lottery was the best strategy may have a high degree of demonstrability (especially since this was explicitly stated in the instructions) but the correctness of any particular WTP was likely to be less demonstrable since risk preferences are subjective. It is possible, however, that the expected value of each lottery (discussed briefly in the instructions) could emerge as the objective, demonstrably correct, 'statistically rational' solution to the WTP problem in the sense that it maximises expected monetary earnings. This focal-point logic suggests the possibility of a systematic choice shift where groups submit a decision that is more consistent with risk neutrality than is the average of the individual group members. Alternatively, the social psychology literature suggests the possibility of a choice shift rooted in group polarisation that leads risk-averse (or risk-preferring) individuals to a group decision that amplifies their risk preferences and moves the group away from the risk-neutral benchmark relative to the average of the individual decisions.

In the absence of a choice shift that alters individual preferences, group discussion can be viewed as unstructured bargaining between heterogeneous agents with fixed preferences that is likely to result in a group WTP decision that simply reflects the average preference of the individuals in the group. Even if a group does not consciously focus on calculating the mean or median of the individual decisions, it is quite possible that an averaging process will emerge naturally from group discussions.

The data analysis presented in Section 3.4 evaluates the validity of specific research conjectures based on this logic. The *group-averaging null conjecture* states that there will be no systematic difference between the group's WTP decision and the average of the group members' individual decisions for a specific lottery. The *group-shift alternative conjecture* states that a systematic difference between the group and average individual decision will be observed. For lotteries where the null conjecture is rejected, the existence of a choice shift toward the risk-neutral focal point is evaluated relative to the alternative of a choice shift away from risk neutrality. Whether the group decision-making process leads to an averaging of individual decisions or to a choice shift via some other preference aggregation process, the variance of group decisions in a given lottery is expected to be smaller than that of individual decisions.

Given the additional complexity and data acquisition costs of implementing experiments with group decisions, it is not surprising that there are so few studies in the experimental economics literature on group versus individual decision making. Also, given economists' natural concern with the use of meaningful incentives, it is not surprising that the few studies conducted by economists utilise salient response-contingent cash rewards.

In general, this literature focuses on two possible consequences of having a group rather than an individual make a decision. The first is whether groups make more rational decisions in the sense that their decisions are more in line with the game-theoretic prediction for a given task. The conclusions are mixed on this account. Three studies found evidence that support the conjecture that group decisions are closer to the game-theoretic prediction while two do not. Bornstein and Yaniv's (1998) ultimatum bargaining study, Cox's (2002) study involving trust games and, to a certain extent, Kocher and Sutter's (2005) beauty contest experiments all found that groups do, although not always initially, play closer to the game-theoretic prediction. In contrast, Cason and Mui's (1997) dictator game study and Cox and Hayne's (2006) first-price sealed-bid common-value auction experiments found that groups tend to be less rational (or, in some cases, no less irrational) than individuals in the sense that group decisions lie further (or just as far) from the game-theoretic prediction than the decisions made by individuals. The second consequence involves Expected Utility Theory (EUT). Three studies (Bateman and Munro, 2005; Bone *et al.*, 1999; Rockenbach *et al.*, 2007) investigate whether groups make choices that are more in line with EUT. None of the studies found evidence that EUT describes group decisions under risk any better than it describes individual decisions. Two recent working papers by Harrison *et al.* (2005) and Baker *et al.* (2006) explicitly focus on individual versus group risk preference measurement and are thus more closely related to the research reported here. Neither study reports significant risk preference differences between individuals and three-person groups as measured by the count of 'safe' choices using the lottery-choice mechanism developed by Holt and Laury (2002). For a more extensive description of many of these studies see the excellent review of literature in Kocher and Sutter (2005).

2. Experimental Design and Procedures

Risk preference differentials are investigated here using certainty equivalents elicited via a lottery valuation task. There are many studies in the economics literature involving similar lottery valuation experiments. While there is considerable variation in the specific decision tasks, experimental procedures and reward levels utilised, the ratio of the certainty equivalent to the expected value of a dichotomous lottery tends to be higher with low monetary prizes, low probabilities of winning and the use of minimum selling price (rather than maximum demand price) certainty-equivalent elicitation techniques.

One of these studies in particular influenced the experimental design. Kachelmeier and Shehata (K&S, 1992) used a lottery valuation experiment in an interesting study of risk preferences using participants from The People's Republic of China, the US, and Canada. The primary focus of their research was the effect on Chinese students' risk preferences of using lotteries with very high monetary rewards. K&S initially relied on selling prices to elicit certainty equivalents. They found 'a marked downward curvilinear trend from highly risk-seeking preferences to risk-neutral or slightly risk-averse preferences as the win percentages increase' (p. 1124). In two subsequent trials, K&S utilised a maximum demand price method to elicit certainty equivalents and found significantly lower numbers. This is consistent with several earlier studies (Kahneman *et al.*, 1990), suggesting that the use of the selling price method was a likely reason for the absence of risk aversion in their low win-percentage lotteries. This result is noteworthy, since risk aversion over these lotteries is predicted by prospect theory and supported by data reported by Tversky and Kahneman (1992). A maximum demand price method was used in the research reported here in order to avoid the so-called 'endowment effect' associated with the minimum selling price method.

A total of 100 participants were recruited from undergraduate economics classes at Indiana University, Bloomington. As mentioned, two experimental designs were implemented. Design I was used to investigate the existence of a statistically significant difference between individual and group risk preferences. In this design, each participant made decisions as either an individual or a member of a three-person group, thus maintaining strict independence between the individual and group samples for a given lottery. Sixteen participants were used as individual decision makers and 48 other participants were randomly assigned to three-person groups. Design II used 36 other participants first making decisions as an individual then as a member of a randomly assigned three-person group. This two-phase design probes beyond the existence-of-difference issue addressed in Design I by allowing an initial exploration of how the risk preferences of specific individuals are aggregated into a group risk preference.

2.1. *Design I Procedures*

In this design, decision-making units (individuals or groups) were spatially separated in a large classroom and no communication between units was permitted. The experiment instructions (Appendix A for individuals, Appendix B for groups) were read aloud while the participants followed along on their personal copies. Participants had few questions and did not appear to have difficulty understanding the experimental procedures. After completing the instructions a practice run through the full set of procedures was conducted without monetary rewards. This was followed immediately by a second run for cash rewards. Six separate experimental sessions were completed (two with individual decisions and four with three-person group decisions). Each session lasted less than one hour and the average payout was $21.98 per participant.

Appendices A and B also contain the text of an overhead shown to subjects outlining the post-instruction sequence of events comprising a run through the experiment. A detailed explanation and discussion of each step follows.

Step 1 – Entry of Lottery Bid Decisions. Each individual or group decision-making unit entered on a record sheet (also in the appendices) nine 'bids' corresponding to nine different lotteries with a chance of winning equal to 10% to 90% in 10% increments. In the sessions with bids submitted by individuals, all participants were endowed with $20 and all lotteries paid either $20 or $0. In the sessions with groups, all groups were endowed with $60 and all lotteries paid either $60 or $0. Each group member was paid an equal one-third share of total group earnings. Groups were given a maximum of twenty minutes to discuss the problem and agree on the bids to be entered. If there was disagreement when time expired, participants were informed that each group member would independently submit a bid for each lottery and the mean of the three bids would be entered as the group's bid for that lottery. All groups were able to reach unanimous agreement in considerably less than the allotted time.

Step 2 – Random Choice of One Lottery. After all bids were entered on the record sheets, eight of the nine lotteries were randomly eliminated from further consideration. This was accomplished by having each decision-making unit blindly draw a poker chip from a vessel containing nine chips labelled 1 to 9 (corresponding to the 10% to 90% chance-of-winning lotteries, as shown on the record sheet). Whatever chip was drawn, the corresponding lottery was the only lottery utilised in the remaining steps of the experiment.

Step 3 – Random Choice of Lottery Purchase Price. After focusing on a single lottery, a random purchase price was determined in the range from $0 to $19.99 for individuals and $0 to $59.99 for groups. The four digits comprising the lottery purchase price were established by having participants blindly choose numbered poker chips. To save time, a single purchase price was applied to all lotteries in an experimental session. Decision-making units with bids greater than or equal to the random purchase price bought the right to play the lottery and paid the random purchase price. All others did not play the lottery and thus received a final cash payment equal to their endowment. The instructions carefully explained, and the experimenters verbally emphasised, that submitting a bid equal to maximum willingness to pay was the best strategy in this game.

Step 4 – Lottery Outcome Determination. Finally, for those individuals and groups who purchased a lottery ticket, the lottery outcome was determined by having a participant blindly draw one of ten poker chips numbered 0 to 9. For chip draw D, all lotteries with a chance of winning greater than $(10 \times D)\%$ were declared winners and all others were declared losers. Thus, all lotteries were winners if a zero was drawn and all lotteries were losers if a nine was drawn. Final cash payments for those who played a lottery were equal to the cash endowment-purchase price + lottery earnings.

2.2. *Design II Procedures*

In this design, each participant first made a set of decisions as an individual, then as a member of a randomly assigned three-person group. It is important to note that the participants were not informed that they would be making a second set of decisions as a member of a group until after collection of their individual decisions, therefore the Design I and Design II procedures are identical up to this point. The participants in Design II sessions were spatially separated in a large classroom and the experiment instructions for making an individual decision (Appendix A) were read aloud while the participants followed along on their personal copies. After completing these instructions a practice run through the full set of individual decision procedures was conducted without monetary rewards. The participants were asked to enter their lottery bid decisions (See step 1 above) knowing these decisions were for cash. After collection of these individual decisions, the participants were told that they were now going to make a similar set of decisions as a member of a previously determined and randomly chosen three-person group. They were reassured that the experimenters would return to their individual decisions and play out all lotteries. Before being given a group record sheet, an overhead (Appendix C) was used to explain how the group decision situation differed. After answering any questions, they were instructed to make their group decisions. Upon collection of the group decisions, steps 2 to 4 from above were performed for the individual and then the group decisions. Three separate experimental sessions were completed (each involving twelve participants and thus four groups). As in Design I, all of the Design II groups were able to come to unanimous agreement on their nine bids within the allotted time. Each session lasted less than one and a half hours and the average payout was $44.72 per participant.

3. Experimental Results

The analysis presented below focuses on a decision maker's certainty-equivalent ratio (CER), defined as maximum willingness to pay for a lottery divided by the lottery's expected value. A CER equal to unity is consistent with risk-neutral preferences, a CER greater than unity is consistent with risk-seeking preferences, and a CER less than unity is consistent with risk-averse preferences.[1] The reporting of results will initially focus on the existence of an individual versus group risk preference difference by aggregating the individual data from both Design I ($N = 16$) and Design II ($N = 36$) and comparing them to the group data from Design I ($N = 16$). Since the individual decisions made in Design II sessions occurred first, and without subject knowledge of the subsequent group decision, statistical independence of individual and group decisions is maintained. The presentation of experimental results begins with a simple graphical overview of CERs across lotteries, followed by estimation of a regression model, followed by supporting paired-comparison tests focusing on individual lotteries. The CER

[1] In the Sections that follow, for CERs less than one, smaller (larger) CERs are assumed to correspond to more (less) risk-averse preferences over a specific lottery. Of course, it is possible that either decision errors or motivations other than maximisation of expected utility from lottery payoffs can influence willingness-to-pay bid choices. For example, participants might derive nonmonetary utility from the excitement of playing out a lottery or from submitting bids that they feel will either please or displease the experimenter. Furthermore, such anomalous effects might not be symmetric across individuals and groups.

data are then evaluated in the context of risk-aversion coefficients calculated assuming a constant relative risk averse utility function over lottery payoffs. Finally, the Design II individual and group CER data are used to investigate how the risk preferences of specific individuals aggregate into a group risk preference.

3.1. *Graphical Overview*

Figures 1, 2 and 3 report the CER mean, median and standard deviation for individuals and three-person groups in each of the nine lotteries. Several informal observations emerge from studying these figures. The analysis presented in the next subsection addresses the statistical validity of these informal point-estimate comparisons.

Observation 1. In the 10% to 40% lotteries, using either mean or median CERs, groups exhibit more risk aversion than individuals. This differential is substantially smaller using median CERs primarily due to the individual median CER being lower than the individual mean CER. A few highly risk-seeking individuals have a large impact on the individual mean.

Observation 2. In the 70% to 90% lotteries, using either mean or median CERs, individuals exhibit more risk aversion than groups. Both groups and individuals in these high win-percentage lotteries are closer to the risk-neutral benchmark than in the low win-percentage lotteries.

Observation 3. CER dispersion is smaller for groups than individuals in all lotteries, with the dispersion for individuals being substantially greater in the low win-percentage lotteries. Both groups and individuals exhibit smaller CER dispersion in the high win-percentage lotteries.

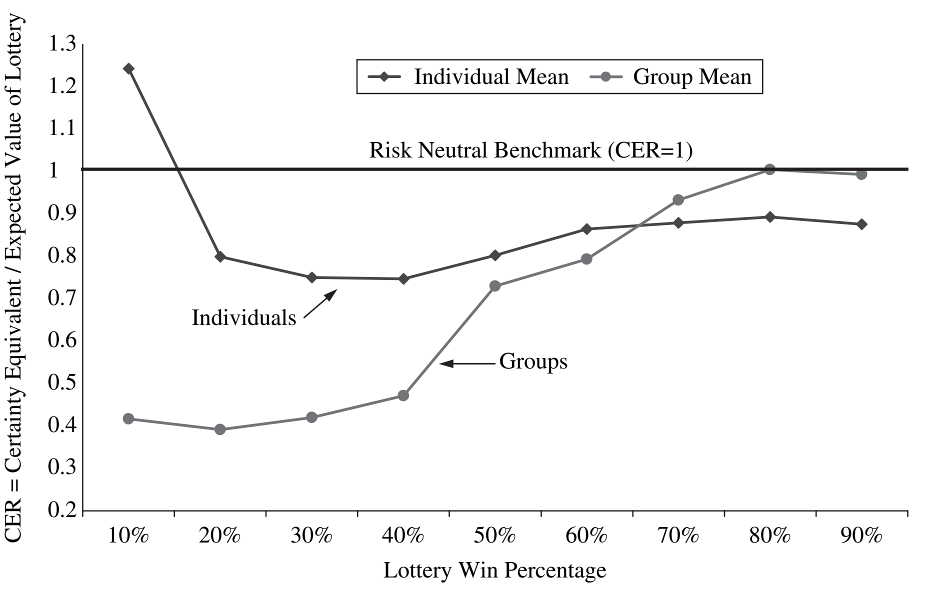


Fig. 1. *Mean Certainty-Equivalent Ratio (CER) Comparison*

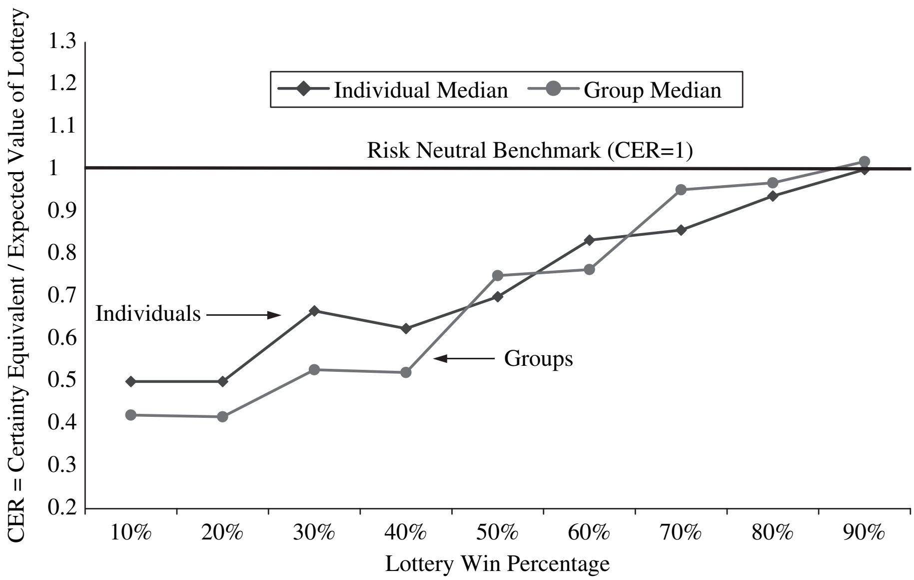


Fig. 2. *Median Certainty-Equivalent Ratio (CER) Comparison*

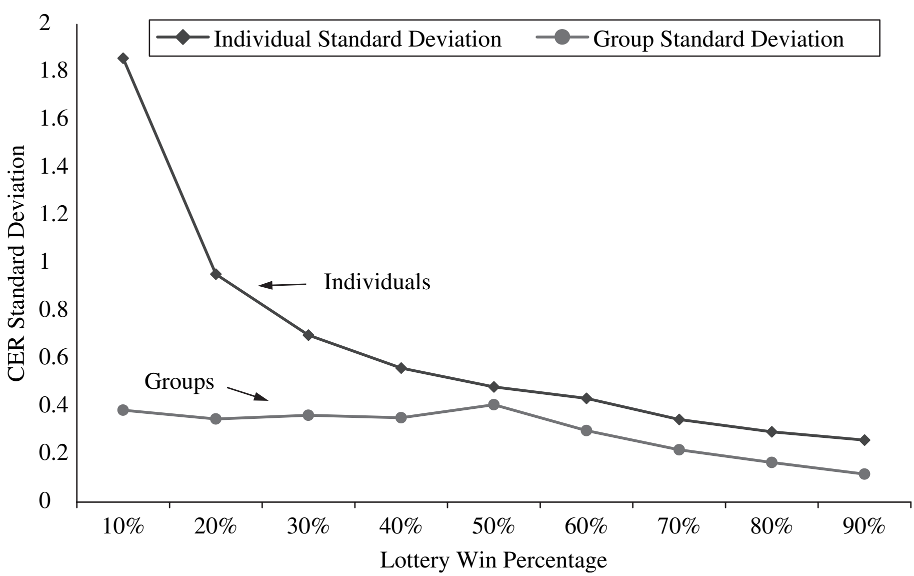


Fig. 3. *Certainty-Equivalent Ratio (CER) Standard Deviation Comparison*

3.2. *Statistical Analysis*

The analysis begins with a regression model estimated using all 612 observations (9 lotteries × 68 decision-making units) where CER is the dependent variable and the independent variables are: a group-decision dummy variable (*GRP*), a set of eight lottery dummy variables (*LOTi*, $I = 20, 30, \ldots, 90$, where *i* corresponds to the lottery win percentage), and eight $GRP \times LOTi$ interaction terms. Inclusion of the interaction terms in the model is critical since it is clear from Figure 1 that both the sign and magnitude of mean CER differences due to the group-vs-individual effect vary systematically across the range of lottery win percentages. The constant term gives the predicted CER for individuals in the 10% lottery. To account for lack of independence across the nine CERs elicited from each individual and group, robust standard errors are utilised where the data are clustered by these within-subject observations.[2]

Table 1 displays the regression coefficient point estimates, robust clustered standard errors, and two-tailed significance tests of the coefficients. A Wald test of the null hypothesis that the lottery dummy coefficients are jointly equal to zero is rejected ($p = 0.000$), as is the null hypothesis that the interaction coefficients are jointly equal to zero ($p = 0.012$). Of primary interest here is the significance of lottery-specific effects of group vs. individual decision making. This is evaluated using Wald tests of the null hypotheses $(GRP + GRP \times LOTi) = 0$ for the 20%–90% lotteries, and $GRP = 0$ for the 10% lottery (where names of independent variables imply the corresponding point estimates of coefficients). The p-values for these tests are included in Figure 4 (discussed further below) – the null hypothesis of no lottery-specific effect is rejected ($p < 0.10$, two-tailed test) in the 10%, 20%, 30%, and 40% lotteries, where the mean group CER is below the mean individual CER, and in the 80% and 90% lotteries, where the mean group CER is above the mean individual CER.[3]

These formal regression-based results are generally consistent with the informal observations derived from visual inspection of the mean CER data in Figure 1. To examine the validity of these observations further, the analysis now turns to two-sample tests focusing on the significance of CER differences for each lottery. Figure 4 summarises the results of the regression-based Wald tests referred to above, as well as three additional tests comparing the group and individual CERs for each of the nine lotteries: t-tests and Mann-Whitney U-tests for central tendency equality, and Levene (1960) F-tests for variance equality. The Mann-Whitney and Levene tests do not rely on the underlying populations being normally distributed. Consistent with the Wald tests, the null hypothesis of equal population means is rejected ($p < 0.10$) using two-sample (two-tailed) t-tests for the 10%, 20%, 30%, 40%, 80% and 90% lotteries. The

[2] For a detailed discussion of the heteroscedasticity-robust Huber/White sandwich estimator of variance in clustered samples see, for example, Cameron and Trivedi (2005, ch. 24, sect. 24.5). The specific implementation utilised here is documented in Rogers (1993). The results reported below are nearly identical to GLS estimation of a cluster-specific random effects model.

[3] Approximately 11% of the CER observations occur at the fixed lower boundary of the decision space (bid of zero) and 8% of the CER observations occur at the variable upper boundary of the decision space (maximum bid/expected value of lottery). These boundary observations are asymmetrically distributed across lotteries – 95% of the CERs at the lower bound occur in the 10%–40% lotteries, and 88% of the CERs at the upper bound occur in the 60–90% lotteries. The fact that bids occur at both corners of the decision space suggests the use of a two-limit censored-normal (Tobit) regression model. This is not a conventional example of censoring, however, since the bounds on the random purchase price mean that bids below (above) the lower (upper) bound of the decision space are equivalent to bids equal to the lower (upper) bound. In spite of the fact that the censored-normal model is rooted in very strong distributional assumptions, Wald tests similar to those reported in Figure 4 (with slightly higher p-values) are obtained if a lottery win-percentage variable is used to model heteroscedasticity in the error term. Without this specification of the conditional variance, the lottery-specific effect of group decision making is not significant in the 80% and 90% lotteries.

Table 1

Regression Model Dependent Variable: CER = (Bid/Expected Value of Lottery)

Independent Variable	Coefficient Estimate	Robust Standard Error	Ho: Coefficient = 0	
			t	2-tailed p-value
CONSTANT	1.243	0.261	4.770	0.000
GRP	−0.826	0.278	−2.970	0.004
*LOT*20	−0.443	0.134	−3.300	0.002
*LOT*30	−0.492	0.181	−2.720	0.008
*LOT*40	−0.495	0.206	−2.400	0.019
*LOT*50	−0.440	0.222	−1.980	0.051
*LOT*60	−0.378	0.236	−1.600	0.115
*LOT*70	−0.363	0.242	−1.500	0.138
*LOT*80	−0.350	0.249	−1.400	0.165
*LOT*90	−0.367	0.254	−1.440	0.153
GRP × *LOT*20	0.418	0.141	2.970	0.004
GRP × *LOT*30	0.495	0.189	2.610	0.011
GRP × *LOT*40	0.549	0.213	2.580	0.012
GRP × *LOT*50	0.753	0.243	3.100	0.003
GRP × *LOT*60	0.754	0.254	2.970	0.004
GRP × *LOT*70	0.879	0.267	3.300	0.002
GRP × *LOT*80	0.937	0.272	3.440	0.001
GRP × *LOT*90	0.942	0.275	3.420	0.001

Total number of observations = 612 = 68 clusters of 9 observations.
Model: F (17,67) = 7.31, p = 0.000; R-squared = 0.066.

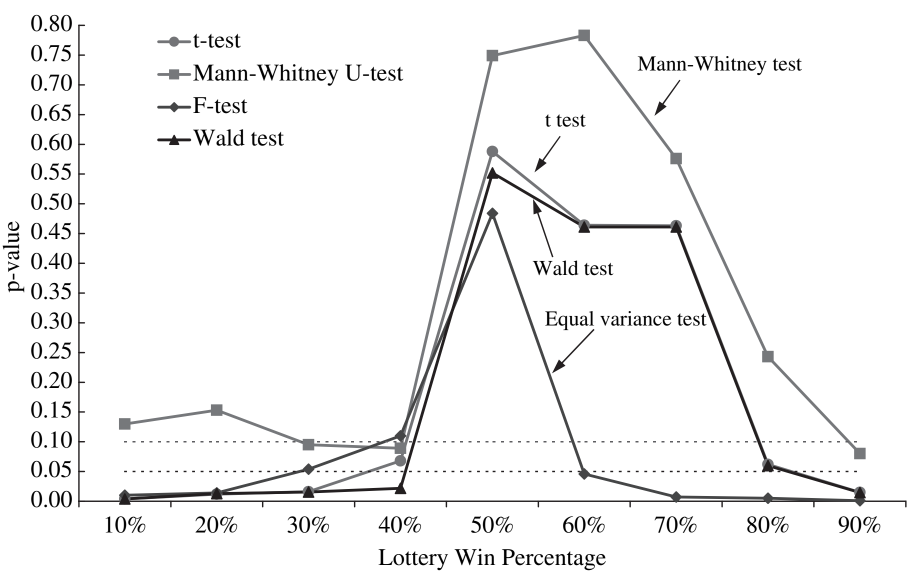


Fig. 4. *Probability of Type-I Error: Groups vs. Individuals*

Wald and t-tests are nearly identical in all except the 40% and 50% lotteries where the equal variance null hypothesis is not rejected ($p > 0.10$) and thus the degrees of freedom in the t-tests are not adjusted to account for unequal sample variances. Non-parametric Mann-Whitney U-tests also reject ($p < 0.10$) central tendency equality for the 30%, 40% and 90% lotteries. In summary, these two-sample tests support the existence of a significant interaction between the lottery win-percentage and the effect of group decision making. Groups tend to be more risk averse than individuals in the higher-risk (lower win-percentage) lotteries and less risk averse (approaching risk neutral) in the lowest-risk (high win-percentage) lotteries.[4] Section 3.4 will explore the robustness of these results when a specific mapping of individual to group decisions is observed.

Since risk neutrality is a natural point of reference for risk preference measures, the null hypothesis that the population mean CER = 1 is evaluated for individuals and groups in each lottery using a one-sample (two-tailed) t-test. For the individual data, the null hypothesis is rejected ($p < 0.10$) for all lotteries except the high-variance 10% and 20% lotteries.[5] For the group data, the null hypothesis is rejected ($p < 0.10$) in all lotteries except the 70%, 80%, and 90% lotteries. Similar results are obtained using nonparametric tests except that, for individuals, the null is also rejected in the 20% lottery.

3.3. *Calculation of Risk-Aversion Coefficients*

Commenting on the paper by Kachelmeier and Shehata (K&S, 1992), Ortona (1994) makes the point that CERs will approach unity monotonically as the lottery win percentage increases, if one assumes a stable exponential utility function over lottery payoffs with a constant risk-preference coefficient. Thus, the K&S observation that mean CERs fall toward unity as the lottery win percentage increases (interpreted by K&S as revealing less risk-seeking behaviour) could be consistent with a fixed risk coefficient utility function. While this observation appears to be roughly consistent with their CER data, K&S's (1994) reply shows that the risk preferences implied by their CER data are unstable. For their high-prize condition, the mean risk coefficients presented by K&S tend towards less risk-seeking, and eventually risk-averse, choices as the lottery win percentage increases.

[4] The existence of a risk preference gender effect was also investigated. Neither two-sample t-tests nor Mann-Whitney tests allow rejection ($p < 0.10$) of the null hypothesis of equal central tendency for the male and female certainty equivalents in any of the nine lotteries. A regression model employing gender as a between-subjects effect supports this conclusion. Also, the null of homogeneous population variances is rejected in only the 90% lottery, with the female sub-sample having higher variance than the male sub-sample. Given the results of these tests and the fact that the volunteer participants (approximately 70% male and 30% female) were randomly assigned to three-person groups, it is very unlikely that the gender composition of groups had a significant impact on the group-versus-individual results.

[5] Data collected during demonstration experiments using graduate and advanced undergraduate student subjects enrolled in the Seminar in Experimental Economics at Indiana University suggest that 'highly sophisticated' subjects with formal training in statistics, economic theory and behavioural economics are far more likely to submit individual CERs near the risk-neutral benchmark than the first-year and second-year undergraduate economics students used in this research. Generalisations from the informal classroom experiments are very difficult, however, since the data were collected under either hypothetical or very low monetary incentives relative to those used in the research reported here. From Holt and Laury (2002), low and hypothetical monetary incentives are predicted to result in less risk-averse decisions. Despite this confounding of effects, it is likely that the subject-sophistication effect is significant.

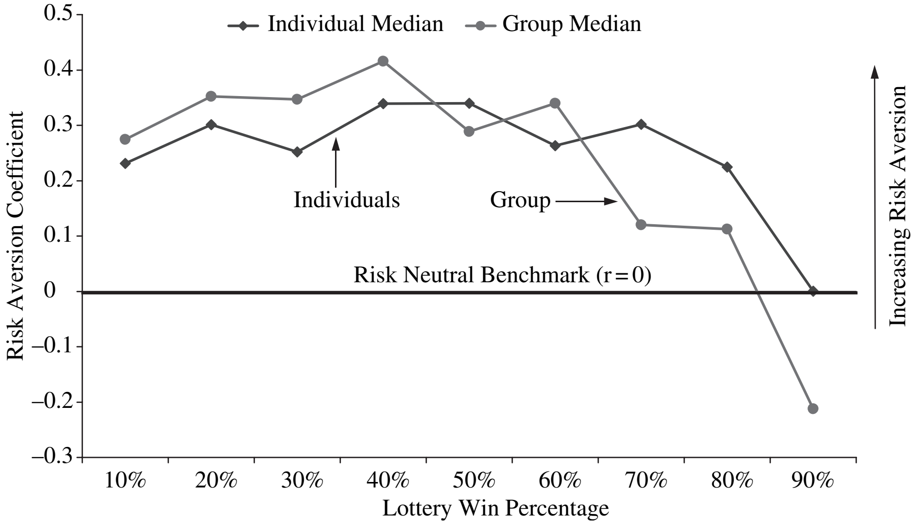


Fig. 5. *Median Coefficient of Risk Aversion Comparison*

Figure 5 displays median risk-aversion coefficients calculated for both groups and individuals in each of the nine lotteries, using a constant relative risk-averse utility function.[6] The Figure illustrates fairly stable median risk-aversion coefficients for both groups and individuals over the 10% to 60% lotteries. As the win percentage increases beyond this range, groups reveal less risk aversion in the 70% to 90% lotteries and individuals follow this pattern in the 80% and 90% lotteries. Thus, the risk-aversion coefficients obtained from the CER data reported here are much more stable than those reported by K&S (1994). There is, however, movement toward the risk-neutral benchmark in the highest win-percentage lotteries, which is similar to the pattern seen in the K&S minimum selling price CER data.

3.4. *Aggregation of Individual Decisions into a Group Decision*

The objective of this section is to use the Design II individual ($N = 36$) and group ($N = 12$) data to explore how individual group members' CERs are aggregated into a group CER, and, in doing so, to evaluate the robustness of the basic results from the independent-samples data presented in the previous Sections. The null hypothesis that the Design I and Design II group-decision samples are drawn from populations with identical central tendency cannot be rejected ($p > 0.20$ in Mann-Whitney tests for

[6] Letting X represent the elicited certainty equivalent, P the probability of winning Y dollars, and $(1 - P)$ the probability of winning zero dollars, a constant relative risk-averse utility function over lottery payoffs implies $X^{1-r} = P(Y^{1-r})$, $r \neq 1$. Thus, $X = P^{1/(1-r)}\ Y$ and the $CER = X/(PY) = P^{r/(1-r)}$, where $r = 0$ implies risk neutrality, $r > 0$ implies risk aversion, and $r < 0$ implies risk-seeking preferences. Solving for r, the coefficient of risk aversion, yields $r = \ln\ CER/(\ln\ P + \ln\ CER)$. Of the 612 CER observations (468 from individuals and 144 from groups), 66 are equal to zero (46 from individuals and 20 from groups) reflecting a willingness-to-pay of $0 in low win-percentage lotteries. For these instances where r is undefined, r is set equal to one when calculating the medians shown in Figure 5. Also, the Figure focuses on the median r, rather than the mean, due to a few extreme negative r estimates in high win-percentage lotteries that have a large effect on the mean.

each of the nine lotteries). Thus, it appears that group decisions are not, on average, significantly affected by the fact that Design II participants had previously submitted bids as individuals (but had not seen the outcome from those decisions).[7]

The empirical validity of *a priori* conjectures about how individual decisions aggregate into a group decision is evaluated using descriptive statistics and Wilcoxon matched-pairs tests. For each lottery the *group-averaging null conjecture*, that the group CER does not systematically differ from the average of the individual group members' CER decisions (hereafter, average individual CER), is evaluated relative to the *group-shift alternative conjecture*. The *group focal-point conjecture*, that group discussion tends to move the group CER toward the risk-neutral benchmark relative to the average individual CER, is also considered as a specific type of choice shift. The results presented in the previous Sections suggest that the group-averaging null conjecture will be rejected in the low (10%–40%) win-percentage lotteries, where groups appear to be more risk averse than individuals, and the highest (80%–90%) win-percentage lotteries, where groups appear to be less risk-averse than individuals. The independent-samples analysis suggests that the group focal-point conjecture will be rejected in all but perhaps the two highest (80%–90%) win-percentage lotteries.

Table 2 displays descriptive statistics and Wilcoxon matched-pairs tests comparing, for each lottery, each group's CER with the mean CER of the three group members. The Table reveals that, in the high-risk (low win-percentage) lotteries, the group CER tends to be substantially less than the mean of the group members' individual CERs. This difference is significant ($p < 0.10$) in the 10% to 40% lotteries. These rejections of the group averaging null in the low win-percentage lotteries are consistent with the

Table 2

Comparison of Group CER and Average of Three Group Members' Individual CERs

Lottery Win Percentage	Mean CER Individuals	Mean CER Groups	Mean Difference Group CER minus the Group Members' Mean Individual CER	Wilcoxon Matched-Pairs Test Ho: Group CER = Members' Mean CER 2-tailed p-value ($N = 12$)
10%	1.288	0.354	−0.934	0.013
20%	0.835	0.365	−0.471	0.021
30%	0.829	0.456	−0.374	0.028
40%	0.851	0.510	−0.340	0.041
50%	0.912	0.826	−0.086	0.326
60%	0.995	0.924	−0.071	0.209
70%	0.987	0.954	−0.033	0.814
80%	0.989	0.996	0.007	0.638
90%	0.961	0.978	0.018	0.456

[7] While this small-sample finding is suggestive, it does not eliminate the possibility that the Design II group decisions reflect a confounding of a group-discussion effect and a pure decision-sequence effect. The problem is that individual CERs are not observed in the second phase of the experiment immediately before group discussion occurs. It is assumed that there is no systematic change in individual CERs from the first to the second phase, which are separated by only a few minutes. Sessions with sequential individual decisions could be used to test for the existence of learning or order effects that are independent of the effects of group discussion. Focusing the Design II subject payment funds on the individual-then-group sequence allowed a larger sample to be collected in this important treatment, which most directly addresses a referee's suggestion to check the robustness of the (Design I) independent-samples results in a sequenced environment where group decisions can be compared to the specific individual decisions of group members. A careful investigation of decision-sequence effects in this setting remains a topic for future research.

prior independent-samples analysis. However, unlike the independent-samples analysis, the group-averaging null cannot be rejected for either the 80% or 90% lotteries. The data reveal that this is in part due to the fact that the individuals in the Design II sample are, on average, less risk averse in the high win-percentage lotteries than the individuals in the Design I sample.[8]

Further study of the Design II individual and group data in the high-risk lotteries supports the conclusion that group discussion leads to more risk-averse decisions. In each of the four highest-risk lotteries 10 of the 12 group CER observations, 83.3%, are more risk-averse than the mean individual CER. In contrast, over the five lowest-risk lotteries 48.3% of the group CERs are more risk-averse than the mean individual CER. Part of the significant choice shift in the four highest-risk lotteries is consistent with an averaging process that underweights the upper extremum of the individual CERs. But 25% of the group CERs that are less than the mean individual CER are also less than the *most* risk-averse group member's individual CER, and are thus inconsistent with an averaging process. The lower extremum of the individual CERs in the four highest-risk lotteries is adopted as the group CER in 39.6% of cases, and is thus a weak focal point outcome for the choice shift toward increased risk aversion. In the five lowest-risk lotteries this outcome occurs in 8.3% of cases. The Design II data are thus consistent with the prior independent-samples analysis with regard to decisions in the highest risk lotteries and confirm, albeit with a small sample, that group discussion tends to induce a choice shift toward increased risk aversion in high-risk situations.

As a refinement of the group-shift alternative conjecture, the group focal-point conjecture implies that group CERs will be closer to the risk-neutral benchmark than the corresponding individual mean CER. This is true for only 33.3% (12 of 36), 30.6% (11 of 36), and 41.7% (15 of 36) of the paired observations over the three highest, medium, and lowest risk lotteries, respectively. This evidence, in combination with the finding that the group averaging null conjecture can not be rejected in the 50% to 90% lotteries, implies that there is not general support for the group focal-point conjecture in the Design II data.

4. Summary and Discussion

The research presented here documents an economic decision-making environment using salient monetary rewards where statistically significant risk-preference differentials are observed for group versus individual decision makers. Use of a maximum willingness-to-pay procedure eliminates the preponderance of seemingly risk-seeking choices observed by Kachelmeier and Shehata (1992) using a minimum compensation-demanded procedure and is generally consistent with the levels of risk aversion found in recent research by Holt and Laury (2002). A regression model and two-sample tests using independent group ($N = 16$) versus individual ($N = 52$) willingness-to-pay certainty equivalents support the following general conclusions.

[8] The individual median CER was also considered in the evaluation of the group averaging null conjecture. Results using the median are qualitatively similar, but the null is rejected ($p < 0.10$) in only the 10% to 30% lotteries.

Conclusion 1. Certainty equivalent ratios vary significantly across lotteries as the win percentage varies from 10% to 90%. For both groups and individuals, the median CER tends to increase toward the risk-neutral benchmark as the lottery win percentage increases. The corresponding median coefficients of constant relative risk aversion are fairly stable across the 10% to 60% lotteries but eventually move toward the risk-neutral benchmark over the 70% to 90% lotteries.

Conclusion 2. For higher-risk lotteries with a win percentage of 40% or lower, the average group is significantly more risk averse than the average individual.

Conclusion 3. For the lowest-risk lotteries with a win percentage of 80% and 90%, the average group is approximately risk neutral and significantly less risk averse than the average individual.

Conclusion 4. For lotteries with a win percentage of 50%, 60% and 70%, the average group and the average individual are both risk averse and are not significantly different.

Conclusion 5. The variance of CERs is lower for groups than for individuals in all lotteries. This difference is significant for all except the 40% and 50% lotteries. In general, CER variance tends to decrease as the lottery win-percentage increases.

The effect on risk preference measurement of using three-person groups instead of individuals as decision makers is more complex than was anticipated prior to conducting this research. Rather than a simple uniform shift of certainty equivalent ratios across different win-percentage lotteries, the magnitude and possibly the direction of the group effect appears to depend on the inherent riskiness of the property right being considered for acquisition. Analysis of the smaller sample of Design II data (36 individual decisions followed by 12 three-person group decisions) supports the conclusion that group discussion leads to a significant shift toward more risk aversion in the four highest risk lotteries but does not reveal significant separation of the group CER and the mean individual CER in any of the five lowest risk lotteries.

For both groups and individuals, as the win percentage increases:

(1) the median CER trends upward toward unity (see Figure 2),
(2) CER dispersion is reduced (see Figure 3), and
(3) the coefficient of relative risk aversion eventually decreases toward the risk-neutral benchmark (Figure 5).

Why do both groups and individuals reveal lower variance and more risk-neutral bids as the risk associated with the lottery decreases? One possible explanation focuses on differentials across lotteries in the monetary incentive to submit statistically rational risk-neutral bids that maximise expected earnings in this experimental setting. For each of the nine \$20 lotteries, Figure 6 shows expected earnings as a function of the bid.[9] Each function's peak occurs at the risk-neutral bid for that lottery, so bids

[9] If a winning lottery pays \$20, a losing lottery pays \$0, decision makers receive a \$20 endowment, and the random purchase price is uniformly distributed with supports at \$0 and \$19.99, then expected earnings in a specific lottery = $\{P_{BUY}$ [\$20 – E(*Price* | *Bid* $\geq$ *Price*) + (P_{WIN} \$20)]} + {[(1 − P_{BUY}) \$20]}, where: P_{BUY} = *Bid*/19.99 is the probability of buying a lottery conditional on the submitted bid, E(*Price* | *Bid* $\geq$ *Price*) = *Bid*/2 is the expected purchase price conditional on having bought the lottery, and P_{WIN} is the probability of winning the lottery. For example, a risk-neutral bid of \$10 in the 50% lottery yields expected earnings of [0.50025 (\$20 − \$5 + \$10)] + (0.49975 \$20) = \$22.50, rounding to the penny. Prior to the procedural step where only one lottery is selected for further play (see Section 2.1), each lottery has a 1/9 probability of being chosen. At this stage a risk-neutral bidder has expected earnings of \$23.17 over all nine lotteries.

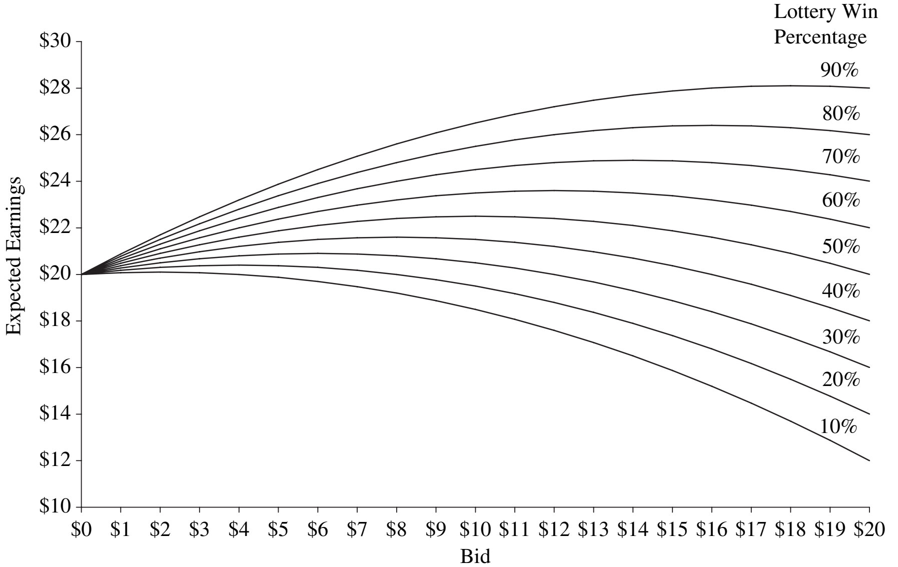


Fig. 6. *Expected Earnings Functions*

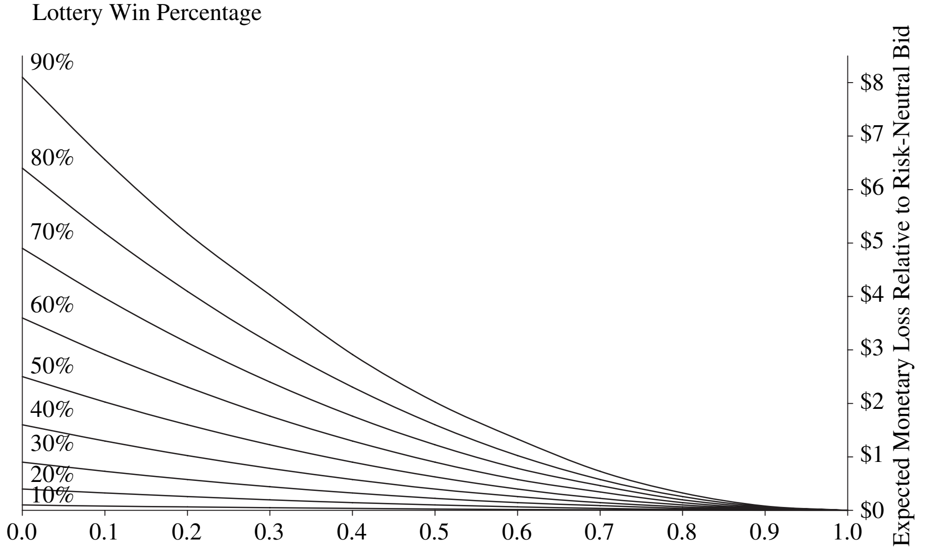


Fig. 7. *Expected Monetary Loss Functions*

consistent with either risk aversion or risk seeking lead to a loss in expected earnings relative to the risk-neutral bid. Figure 7 translates the expected earnings curves shown in Figure 6 into expected monetary loss (relative to the risk-neutral benchmark) as a function of the CER.[10] The expected loss functions reveal that, using CER as the bid

[10] Only the risk-averse CER range is shown in Figure 7 since the functions are symmetric around CER = 1 and the majority of the observations are consistent with risk aversion.

metric, the monetary cost of deviating from the risk-neutral bid increases as the lottery win percentage increases. For example, in the 20% lottery, submitting a bid that generates a CER of 0.5 results in an expected loss of \$0.10 relative to risk-neutral expected earnings of \$20.40. In the 80% lottery, a bid generating a CER of 0.5 results in an expected loss of \$1.60 relative to risk-neutral expected earnings of \$26.40. The figures also reveal that the earnings differential over the interval from CER = 1 to CER = 0 ranges from \$8.10 (\$28.10−\$20.00) in the 90% lottery to \$0.10 (\$20.10−\$20.00) in the 10% lottery.

Given the relatively small incentive to bid precisely in the higher-risk lotteries, the accuracy of bids as a measure of true certainty equivalents is perhaps reduced relative to the lower-risk lotteries. While previous research suggests that this reduced incentive is not expected to affect the central tendency of the sample bid distributions, it is likely to increase their variance (Smith and Walker, 1993). Both individuals and groups are seemingly responsive to the magnitude of the expected monetary losses associated with deviations from expected-value bidding, as evidenced by the reduced variance of bids and the trend toward expected-value bidding as the lottery win percentage increases. The independent-samples analysis indicates that group CERs are on average significantly closer to the risk-neutral benchmark than individual CERs only in the highest win-percentage lotteries. While this finding is not present in the Design II matched-pairs comparison of group versus mean individual CERs, it suggests that group discussion is more likely to facilitate outcomes that are consistent with risk neutrality when the decision costs required to reach this solution are offset by a sufficiently large expected monetary gain.

Appendices

A. Instructions, Record Sheet, and Overhead for Sessions with Individuals

Experiments Instructions

This is an experiment in behavioural economics focusing on individual valuation of events (called lotteries) that have uncertain outcomes. You are starting with a money endowment of \$20. This \$20 can be used to purchase the right to play a lottery (i.e. buy a lottery ticket). Since the outcome of a lottery is uncertain, you might win more money or lose some of your \$20 endowment through participation in the experiment. At the conclusion of the experiment you will be paid your final earnings privately in cash.

Before getting into the details of the experimental procedures and the decisions you will make, it is important that you understand exactly what is meant by the term 'lottery'.

What is a lottery?

In this experiment, a lottery is a chance to win a monetary prize of \$20. There are nine possible lotteries in today's experiment. The nine lotteries are associated with the following chances of winning: 90%, 80%, 70%, 60%, 50%, 40%, 30%, 20%, 10%. For example, consider the lottery with a 70% chance of winning. If this lottery was run many times, the player would win (receive \$20) 70% of the time and lose (receive \$0) 30% of the time. Using formal statistical terminology, the 'expected value' or average payout from this lottery is $0.7 \times \$20 = \$14$. This is the average payout after many repetitions, in any one lottery the payout is either $20 or $0. [Are there any questions?]

How is the outcome of a lottery determined?

To determine whether a lottery pays out $20 or $0, a number between 0 and 9 will be randomly drawn. The random number will be determined using 10 poker chips numbered 0 to 9 drawn from a bucket. If the number times 10 is less than the stated chance of winning $20, the lottery pays $20. If the number times 10 is greater than or equal to the stated chance of winning, the lottery pays $0. For example, the 70% lottery pays $20 if the random number drawn is 0, 1, 2, 3, 4, 5 or 6. This lottery pays $0 if the random number drawn is 7, 8, or 9. [Are there any questions?]

How do I purchase the right to play a lottery?

Now that you understand how the lotteries work, it is time to explain how you can purchase the right to play one of the nine lotteries. There are three phases to this process: a bid decision phase, a choice of lottery phase, and a purchase price determination phase.

Bid decision phase

In the bid decision phase you will enter (on the attached Experiment Record Sheet) the *maximum amount* that you are willing to pay for the right to play each of the nine lotteries. These nine maximum willingness-to-pay decisions are your 'bids' for the nine lotteries. The minimum bid is $0 and the maximum is $19.99. After your bids are recorded, only one of the lotteries will be randomly chosen for further use in the experiment. The other eight will be eliminated. [Are there any questions?]

Choice of lottery phase

Which lottery is chosen for further use in the experiment will be determined randomly by drawing a poker chip from a bucket containing nine chips numbered 1 to 9. If chip 1 is drawn, the 10% chance-of-winning lottery is chosen. If chip 2 is drawn, the 20% chance-of-winning lottery is chosen. Similarly, chips 3 to 9 correspond to the 30% to 90% chance-of-winning lotteries. [Are there any questions?]

Purchase price determination phase

The purchase price for the right to play the lottery will be determined randomly in the range from $0 to $19.99 using 10 poker chips numbered 0 to 9 drawn from a bucket. There will be four draws-the first draw (using only the two chips numbered 0 and 1) will determine the first digit and the second though fourth draws (using all 10 chips) will determine the other three digits. [Are there any questions?]

If your bid (maximum willingness to pay) is greater than or equal to the purchase price, you pay the purchase price and play the lottery. If your bid is less than the purchase price, you do not play the lottery.

It is important to understand that, if you play the lottery, *you pay the randomly determined purchase price rather than your bid price* (unless your bid is exactly equal to the purchase price). This is why your best strategy is to submit a bid equal to your maximum willingness to pay for the right to play a lottery.

After the purchase price is determined, a random number will be drawn to determine which lotteries pay $20 and which lotteries pay $0. [Are there any questions?]

How is my final cash payment determined?

If you play the lottery, your final cash payment = your $20 endowment + your lottery winnings ($20 or $0) – the lottery purchase price. For example, suppose your bid for the chosen lottery is $10. If the random purchase price is $8.75 then your bid is greater than the purchase price, so you have bought the right to play the lottery and the price you pay is $8.75 (not your $10 bid price). If you win the lottery your final cash payment = $20 endowment + $20 lottery winnings − $8.75 purchase price = $31.25. If you do not win the lottery your final cash payment = $20 endowment + $0 lottery winnings − $8.75 purchase price = $11.25.

If you do not play the lottery, your final cash earnings = $20 endowment.
[Are there any questions?]

What is the largest final cash payment possible?

The largest possible cash payment is $40. This outcome will occur only if the random purchase price for the chosen lottery is $0 and you win the lottery. Your final payment would thus be $20 (endowment) + $20 (lottery winnings) − $0 (purchase price) = $40. Note that there is a 1 in 2,000 (0.05%) chance that any one purchase price in the range from $0 to $19.99 will be drawn.

What is the smallest final cash payment possible?

The smallest possible cash payment is $0.01. This outcome will occur only if the random purchase price for the chosen lottery is $19.99, your bid for this lottery is $19.99, and you lose the lottery. Your final payment would thus be $20 (endowment) + $0 (lottery winnings) − $19.99 (purchase price) = $0.01.

What is the exact sequence of events in the experiment?

1 Everyone enters their nine bids associated with the nine lotteries on their Experiment Record Sheet. Remember, your bid price is the maximum price that you are willing to pay to play a specific lottery. You do not actually pay what you bid unless the randomly determined purchase price is exactly equal to the bid price. In all other cases, if you play a lottery, the purchase price is less than the bid.

[Enter nine practice bids on your Practice Record Sheet now.]

2 When everyone is finished entering their bids, the experiment monitor will collect all of the record sheets. The monitor will then visit each participant and each will draw a poker chip to determine which one of the nine lotteries will be utilised for that individual in the remainder of the experiment.

[Demonstration of lottery choice via chip draw.]

[In this practice exercise, enter this number yourself on the Practice Record Sheet after 'Lottery Chosen'.]

3 The random purchase price ($00.00–19.99) will then be determined via four poker chip draws. This purchase price will be displayed to everyone and will apply to all lotteries.

[Demonstration of purchase price determination via four chip draws.]

[Enter this amount on the Practice Record Sheet after 'Purchase Price Chosen'.]

4 Everyone can now determine if they purchased the right to play the lottery. If your bid is greater than or equal to the purchase price, you pay the purchase price and play the lottery. If your bid is less than the purchase price, you pay nothing and do not play the lottery.
[Circle 'YES (play lottery)' or 'NO (Final Cash Payment = $20)' on the Practice Record Sheet as appropriate.]
5 The random number (0–9) used to determine the lottery outcome will then be chosen via a poker chip draw. This number will be displayed to everyone and will apply to all lotteries.
[Demonstration of lottery outcome determination via chip draw.]
[If you played the lottery, enter this number on the Practice Record Sheet after '# drawn'.]
6 Each individual that played a lottery can now determine if they won the $20 lottery. If the number times 10 is less than the stated chance of winning, the lottery pays $20. Otherwise, the lottery pays $0.
[After 'WIN $20?' on your Practice Record Sheet circle 'YES' or 'NO' as appropriate, then calculate your final cash payment on your Practice Record Sheet.]
7 At the end of the experiment, the monitor will call each participant to the front of the room one at a time. Please remain seated until you are called. In your presence, the monitor will determine your final cash payment using the procedures described previously. You will be paid this amount privately in cash and must sign a payment sheet for our financial records.

This is the end of the instructions. Are there any final questions? If not, it is time to begin the actual experiment to determine your cash earnings. Good luck to everyone!
[Display large wad of cash.]

Experiment Record Sheet

Session ________________ Participant # ______

	Lottery	Your Bid to Purchase Lottery ($0–$19.99)
1)	10% chance of winning $20	$__________________
2)	20% chance of winning $20	$__________________
3)	30% chance of winning $20	$__________________
4)	40% chance of winning $20	$__________________
5)	50% chance of winning $20	$__________________
6)	60% chance of winning $20	$__________________
7)	70% chance of winning $20	$__________________
8)	80% chance of winning $20	$__________________
9)	90% chance of winning $20	$__________________

Do not write below this line.

Lottery Chosen: ______ Purchase Price Chosen: $______________
Bid $\geq$ Purchase Price? YES (play lottery) NO (Final Cash Payment = $20)
Lottery Outcome: # drawn: ______ WIN $20? YES NO
Final Cash Payment: $20 Endowment + $__________ Lottery Prize ($20 or $0) − $__________ Purchase Price = $__________

Overhead Outlining the Sequence of Events

Step 1.	Enter a bid (maximum willingness to pay) for each of the 9 lotteries. (9 bids in the \$0–\$19.99 range)
Step 2.	Randomly choose 1 of the 9 lotteries (1–9) for each participant.
Step 3.	Randomly choose the purchase price (\$0–19.99): \$ __________
Step 4.	Is your bid greater than or equal to the purchase price?
Result a.	NO – you do not play the lottery. Your final cash payment is \$20.
Result b.	YES – you play the lottery and pay the purchase price from step 3.
Step 5.	Randomly choose the lottery outcome (0–9): __________
Step 6.	Is the step 5 number less than the lottery number chosen in step 2?
Result a.	NO – you do not win the \$20 lottery. Your final cash payment is \$20-purchase price.
Result b.	YES – you win the \$20 lottery. Your final cash payment is \$20 + \$20-purchase price.
Step 7.	When your name is called, come to the desk at the front of the room to receive your final cash payment.

B. Instructions, Record Sheet, and Overhead for Sessions with Three-Person Groups

Experiments Instructions

This is an experiment in behavioural economics focusing on group valuation of events (called lotteries) that have uncertain outcomes. Your three-person group is starting with a money endowment of \$60. This \$60 can be used to purchase the right to play a lottery (i.e. buy a lottery ticket). Since the outcome of a lottery is uncertain, your group might win more money or lose some of the \$60 endowment through participation in the experiment. At the conclusion of the experiment your group will be paid its final earnings privately in cash. Each member of the group will receive a one-third share of the group earnings.

Before getting into the details of the experimental procedures and the decisions you will make, it is important that you understand exactly what is meant by the term 'lottery'.

What is a lottery?

In this experiment, a lottery is a chance to win a monetary prize of \$60. There are nine possible lotteries in today's experiment. The nine lotteries are associated with the following chances of winning: 90%, 80%, 70%, 60%, 50%, 40%, 30%, 20%, 10%. For example, consider the lottery with a 70% chance of winning. If this lottery was run many times, the player would win (receive \$60) 70% of the time and lose (receive \$0) 30% of the time. Using formal statistical terminology, the 'expected value' or average payout from this lottery is $0.7 \times \$60 = \$42$. This is the average payout after many repetitions, in any one lottery the payout is either \$60 or \$0. [Are there any questions?]

How is the outcome of a lottery determined?

To determine whether a lottery pays out \$60 or \$0, a number between 0 and 9 will be randomly drawn. The random number will be determined using 10 poker chips numbered 0 to 9 drawn from a bucket. If the number times 10 is less than the stated chance of winning \$60, the lottery pays \$60. If the number times 10 is greater than or equal to the stated chance of winning, the lottery pays \$0. For example, the 70% lottery pays \$60 if the random number drawn is 0, 1, 2, 3, 4, 5 or 6. This lottery pays \$0 if the random number drawn is 7, 8, or 9. [Are there any questions?]

How do I purchase the right to play a lottery?

Now that you understand how the lotteries work, it is time to explain how your group can purchase the right to play one of the nine lotteries. There are three phases to this process: a bid decision phase, a choice of lottery phase, and a purchase price determination phase.

Bid decision phase

In the bid decision phase your group will enter (on the attached Experiment Record Sheet) the *maximum amount* that it is willing to pay for the right to play each of the nine lotteries. These nine maximum willingness-to-pay decisions are your group's 'bids' for the nine lotteries. The minimum bid is $0 and the maximum is $59.99. You have a maximum of 20 minutes to agree on your group's bid decisions. If you can not unanimously agree on your group's bid for a specific lottery, each group member will privately submit an individual bid and your group's bid will be the average of the three individual bids.

After your group's bids are recorded, only one of the lotteries will be randomly chosen for further use in the experiment. The other eight will be eliminated. [Are there any questions?]

Choice of lottery phase

Which lottery is chosen for further use in the experiment will be determined randomly by drawing a poker chip from a bucket containing nine chips numbered 1 to 9. If chip 1 is drawn, the 10% chance-of-winning lottery is chosen. If chip 2 is drawn, the 20% chance-of-winning lottery is chosen. Similarly, chips 3 to 9 correspond to the 30% to 90% chance-of-winning lotteries. [Are there any questions?]

Purchase price determination phase

The purchase price for the right to play the lottery will be determined randomly in the range from $0 to $59.99 using 10 poker chips numbered 0 to 9 drawn from a bucket. There will be four draws-the first draw (using only the six chips numbered 0, 1, 2, 3, 4, and 5) will determine the first digit and the second though fourth draws (using all 10 chips) will determine the other three digits. [Are there any questions?]

If your group's bid (maximum willingness to pay) is greater than or equal to the purchase price, your group pays the purchase price and plays the lottery. If your group's bid is less than the purchase price, your group does not play the lottery.

It is important to understand that, if your group plays the lottery, *the group pays the randomly determined purchase price rather than the bid price* (unless the bid is exactly equal to the purchase price). This is why your best strategy is to submit a bid equal to your group's maximum willingness to pay for the right to play a lottery.

After the purchase price is determined, a random number will be drawn to determine which lotteries pay $60 and which lotteries pay $0. [Are there any questions?]

How is my final cash payment determined?

If your group plays the lottery, your group's final cash payment = your $60 endowment + your lottery winnings ($60 or $0) – the lottery purchase price. For example, suppose your group's bid for the chosen lottery is $30. If the random purchase price is $26.25 then your group's bid is greater than the purchase price, so your group has bought the right to play the lottery and the price paid is $26.25 (not the $30 bid price). If your group wins the lottery, your group's final cash payment = $60 endowment + $60 lottery winnings − $26.25 purchase price = $93.75. If your group does not win the lottery, your group's final cash payment = \$60 endowment + \$0 lottery winnings − \$26.25 purchase price = \$33.75.

If your group does not play the lottery, your group's final cash earnings = \$60 endowment. [Are there any questions?]

What is the largest final cash payment possible?

The largest possible cash payment to a group is \$120. This outcome will occur only if the random purchase price for the chosen lottery is \$0 and your group wins the lottery. Your group's final payment would thus be \$60 (endowment) + \$60 (lottery winnings) − \$0 (purchase price) = \$120. Note that there is a 1 in 6,000 (0.017%) chance that any one purchase price in the range from \$0 to \$59.99 will be drawn.

What is the smallest final cash payment possible?

The smallest possible cash payment is \$0.01. This outcome will occur only if the random purchase price for the chosen lottery is \$59.99, your group's bid for this lottery is \$59.99, and your group loses the lottery. Your group's final payment would thus be \$60 (endowment) + \$0 (lottery winnings) −\$59.99 (purchase price) = \$0.01.

What is the exact sequence of events in the experiment?

1 Every group enters their nine bids associated with the nine lotteries on their Experiment Record Sheet. Remember, your group's bid price is the maximum price that it is willing to pay to play a specific lottery. Your group does not actually pay what it bids unless the randomly determined purchase price is exactly equal to the bid price. In all other cases, if your group plays a lottery, the purchase price is less than the bid.

[Enter nine practice bids on your group's Practice Record Sheet now.]

2 When every group is finished entering their bids, the experiment monitor will collect all of the record sheets. The monitor will then visit each group and each will draw a poker chip to determine which one of the nine lotteries will be utilised for that group in the remainder of the experiment.

[Demonstration of lottery choice via chip draw.]

[In this practice exercise, enter this number yourself on the Practice Record Sheet after 'Lottery Chosen'.]

3 The random purchase price (\$00.00−\$59.99) will then be determined via four poker chip draws. This purchase price will be displayed to everyone and will apply to all lotteries.

[Demonstration of purchase price determination via four chip draws.]

[Enter this amount on the Practice Record Sheet after 'Purchase Price Chosen'.]

4 Each group can now determine if it purchased the right to play the lottery. If the group's bid is greater than or equal to the purchase price, it pays the purchase price and plays the lottery. If the group's bid is less than the purchase price, it pays nothing and does not play the lottery.

[Circle 'YES (play lottery)' or 'NO (Final Cash Payment = \$60)' on the Practice Record Sheet as appropriate.]

5 The random number (0–9) used to determine the lottery outcome will then be chosen via a poker chip draw. This number will be displayed to everyone and will apply to all lotteries.

[Demonstration of lottery outcome determination via chip draw.]

[If your group played the lottery, enter this number on the Practice Record Sheet after '# drawn'.]

6 Each group that played a lottery can now determine if it won the $60 lottery. If the number times 10 is less than the stated chance of winning, the lottery pays $60. Otherwise, the lottery pays $0.
[After 'WIN $60?' on your Practice Record Sheet circle 'YES' or 'NO' as appropriate, then calculate the group's final cash payment on your Practice Record Sheet.]

7 At the end of the experiment, the monitor will call each group to the front of the room one at a time. Please remain seated until your group is called. In your presence, the monitor will determine your group's final cash payment using the procedures described previously. Each member of the group will be paid a one-third share of this amount privately in cash and must sign a payment sheet for our financial records.

This is the end of the instructions. Are there any final questions? If not, it is time to begin the actual experiment to determine your cash earnings. Good luck to everyone!

[Display large wad of cash.]

Experiment Record Sheet

Session ________________ Group # ______

	Lottery	Your Group's Bid to Purchase Lottery ($0–$59.99)
1)	10% chance of winning $60	$________________
2)	20% chance of winning $60	$________________
3)	30% chance of winning $60	$________________
4)	40% chance of winning $60	$________________
5)	50% chance of winning $60	$________________
6)	60% chance of winning $60	$________________
7)	70% chance of winning $60	$________________
8)	80% chance of winning $60	$________________
9)	90% chance of winning $60	$________________

Do not write below this line.

Lottery Chosen:______ Purchase Price Chosen: $______________

Bid $\geq$ Purchase Price? YES (play lottery) NO (Final Cash Payment = $60)

Lottery Outcome: # drawn:______ WIN $60? YES NO

Final Cash Payment: $60 Endowment +$__________ Lottery Prize ($60 or $0) − $__________ Purchase Price = $__________

Overhead Outlining the Sequence of Events

Step 1.	Enter a bid (maximum willingness to pay) for each of the 9 lotteries. (9 bids in the $0–$59.99 range)
Step 2.	Randomly choose 1 of the 9 lotteries (1–9) for each group.
Step 3.	Randomly choose the purchase price ($0–59.99): $__________
Step 4.	Is your group's bid greater than or equal to the purchase price?
Result a.	NO – your group does not play the lottery. Your group's final cash payment is $60.
Result b.	YES – your group plays the lottery and pays the purchase price.
Step 5.	Randomly choose the lottery outcome (0–9):_________
Step 6.	Is the step 5 number less than the lottery number chosen in step 2?
Result a.	NO – your group does not win the $60 lottery. Your group's final cash payment is $60-purchase price.
Result b.	YES – your group wins the $60 lottery. Your group's final cash payment is $60 + $60-purchase price.
Step 7.	When your group's number is called, come to the desk at the front of the room to receive your final cash payment.

C. Overhead for Design II Transition from Individual to Group Decisions

Group lottery decision

You will now make bid decisions as a member of a randomly chosen three-person group.

The bid and lottery determination processes will be identical to those that have already been described WITH THE FOLLOWING EXCEPTIONS:

1. Your three-person group will start with a money endowment of $60.
2. A lottery will be the chance to win $60.
3. Each member of the group will receive a one-third share of group earnings.
4. For each of the lotteries the maximum bid will now be $59.99 instead of $19.99.
5. You will have a maximum of 20 minutes to agree on your group's bid decisions. If you can not unanimously agree on your group's bid for a specific lottery, each group member will privately submit an individual bid and your group's bid will be the average of the three individual bids.
6. The purchase price will be randomly determined in the range from $0 to $59.99. As such, the first draw will now use six chips numbered 0, 1, 2, 3, 4, 5.
7. The largest cash payment to a group will be $120. The smallest cash payment will still be $0.01.

Any there any questions?

Michigan State University
Indiana University

Submitted: 18 December 2001
Accepted: 17 January 2007

References

Baker, R.J., Laury, S.K. and Williams, A.W. (2006). 'Comparing group and individual behavior in lottery-choice experiments', unpublished manuscript, Indiana University.

Bateman, I. and Munro, A. (2005). 'An experiment on risky choice amongst households', ECONOMIC JOURNAL, vol. 115 (March), pp. C176–89.

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 behavior in the ultimatum game: are groups more rational players', *Experimental Economics*, vol. 1, pp. 101–8.

Cameron, A.C. and Trivedi, P.K. (2005). *Microeconometrics: Methods and Applications*, New York: Cambridge University Press.

Cason, T.N. and Mui, V. (1997). 'A laboratory study of group polarisation in the dictator game', ECONOMIC JOURNAL, vol. 107 (September), pp. 1465–83.

Cox, J.C. (2002), 'Trust, reciprocity, and other regarding preferences: group vs. individual and male vs. female', 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. (2006), 'Barking up the right tree: are small groups rational agents?', *Experimental Economics*, vol. 9(3), pp. 209–22.

Harrison, G.W., Lau, M.I., Rutström, E.E. and Tarazona-Gómez, M. (2005). 'Preferences over social risk', unpublished manuscript, University of Central Florida.

Holt, C.A. and Laury, S.K (2002). 'Risk aversion and incentive effects', *American Economic Review*, vol. 92 (December), pp. 1644–55.

Isenberg, D.J. (1986). 'Group polarization: a critical review and meta analysis', *Journal of Personality and Psychology*, vol. 50, pp. 1141–51.

Kachelmeier, S.J. and Shehata, M. (1992). 'Examining risk preferences under high monetary incentives: experimental evidence from the Peoples's Republic of China', *American Economic Review*, vol. 82 (December), pp. 1120–41.

Kachelmeier, S.J. and Shehata, M. (1994). 'Examining risk preferences under high monetary incentives: reply', *American Economic Review*, vol. 84 (September), p. 1105.

Kahneman, D., Knetsch, J.L. and Thaler, R.H. (1990). 'Experimental tests of the endowment effect and the coase theorem', *Journal of Political Economy*, vol. 98 (December), pp. 1325–48.

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

Kocher, M.G. and Sutter, M. (2005). 'The decision maker matters: individual versus group behaviour in experimental beauty-contest games', Economic Journal, vol. 115 (January), pp. 200–23.

Levene, H.(1960). 'Robust tests for the equality of variance', in (I. Olkin, ed.), *Contributions to Probability and Statistics*, pp. 278–92, Palo Alto, CA: Stanford University Press.

Ortona, G. (1994). 'Examining risk preferences under high monetary incentives: comment', *American Economic Review*, vol. 84 (September), p. 1104.

Rockenbach, B., Sadrieh, A. and Matauschek, B. (2007). 'Teams take the better risks', *Journal of Economic Behavior and Organization*, vol. 63 (July), pp. 412–22.

Rogers, W.H. (1993). 'Regression standard errors in clustered samples', *Stata Technical Bulletin Reprints*, vol. 3, pp. 88–94.

Smith, V.L. and Walker J.M. (1993). 'Monetary rewards and decision costs in experimental economics', *Economic Inquiry*, vol. 31 (April), pp. 245–61.

Tversky, A. and Kahneman, D. (1992). 'Advances in prospect theory: cumulative representation of uncertainty', *Journal of Risk and Uncertainty*, vol. 5, pp. 297–323.

American Economic Review 2009, 99:5, 2247–2257
http://www.aeaweb.org/articles.php?doi=10.1257/aer.99.5.2247

Individual Behavior and Group Membership: Comment

By Matthias Sutter*

The influence of group membership on individual behavior has attracted attention recently, because it questions economic theories that regard individual behavior as solely determined at the individual level. Largely inspired by George A. Akerlof's and Rachel E. Kranton's (2000) model on the effects of identity on economic outcomes, experimental economists have become increasingly interested in examining the effects of group identity on individual behavior. Gary Charness, Luca Rigotti, and Aldo Rustichini (2007, hereafter CRR) have found that salient group membership makes subjects more aggressive in coordination and prisoner's dilemma games. Saliency has been induced by letting other group members observe individual behavior or by using payoff commonality. The latter means that an individual's decision has consequences for the payoffs of other group members, even though the other members cannot influence the individual's decision. CRR argue that salient group membership lets individuals shift their decisions toward those that are more favorable for the group as a whole, meaning that individuals take into account the payoffs of other group members. CRR conclude as their paper's first, and foremost, lesson that "groups profoundly affect individual behavior in social situations" (1350).[1]

The main goal of this comment is to relate the effects of group membership on individual behavior to team decision making. I find in an investment experiment that individual decisions with salient group membership are indistinguishable in the aggregate from those made by unitary teams. This result bridges the gap between the emerging literature on group membership effects (e.g., Goette et al. 2006; CRR; Charness and Matthew O. Jackson 2009; Chen and Li 2009) and the literature on team decision making (e.g., David J. Cooper and John H. Kagel 2005; Charness, Edi Karni, and Dan Levin 2008) and provides further insights into the determinants of team decision making. A secondary aim of this paper is to show that the findings of CRR also apply to nonstrategic decisions (as has already been documented by Chen and Li 2009), even in situations where there is no outgroup at all. Whereas all former studies on group membership effects have considered a setting where an outgroup exists, this is the first study showing that it is the mere fact of being in a group that changes individual behavior, regardless of whether an outgroup exists. This finding makes the hitherto documented effects of group membership on

* Department of Public Finance, University of Innsbruck, Universitaetsstrasse 15, A-6020 Innsbruck, Austria, and Department of Economics, University of Gothenburg (e-mail: matthias.sutter@uibk.ac.at). I thank two anonymous referees for excellent comments and Cornelia Horn and Felix Oettl for valuable research assistance. Financial support from the Max Planck Society, the Austrian Science Foundation (FWF-Project P16617), the Tyrolean Science Foundation (TWF), and the Center of Experimental Economics at the University of Innsbruck (sponsored by Raiffeisen Landesbank Tirol) is gratefully acknowledged.

[1] Their statement is supported by several other recent studies that typically examine individual behavior toward members of an ingroup or an outgroup. Lorenz Goette, David Huffman, and Stephan Meier (2006), for instance, report that officers of the Swiss Army are more cooperative in a prisoner's dilemma game toward members of their own platoon than those of other platoons. Yan Chen and Xin Li (2009)—who provide an excellent survey of the literature on group identity—show that the degree of other-regarding social preferences in an allocation task is stronger toward members of an ingroup than toward outgroup members. Ernst Fehr, Helen Bernhard, and Bettina Rockenbach (2008) study the behavior of three- to eight-year-old children in a simple allocation task and find that the differences in behavior toward ingroup or outgroup members develop in this life span.

individual behavior even more widely relevant, since they obviously do not depend on the existence of an outgroup.[2]

In contrast to individual decision making under group membership, team decision making requires several subjects to reach a joint decision, where there is typically no internal conflict in terms of payoffs among team members.[3] The literature on team decision making has captured a lot of interest in recent years, because many economic decisions are made by teams, such as families, company boards, management teams, committees, or central bank boards. Cooper and Kagel (2005) provide a thorough survey of the relevant literature, documenting that team decisions are typically closer to standard game theoretic predictions than individual decisions are. For example, teams send and accept smaller transfers in the ultimatum game (Gary Bornstein and Ilan Yaniv 1998), send or return smaller amounts in the trust game (James C. Cox 2002, Tamar Kugler et al. 2007), and are more selfish in dictator games (Wolfgang J. Luhan, Martin G. Kocher, and Sutter 2009).[4] Teams exit the centipede game at earlier stages (Bornstein, Kugler, and Anthony Ziegelmeyer 2004), and they outperform individuals in beauty-contest games because they converge more quickly to the equilibrium (Kocher and Sutter 2005). Concerning nonstrategic tasks, teams take more rational decisions in intellective tasks, such as the Wason selection task (Boris Maciejovsky and David V. Budescu 2007), they are more forward-looking in a noninteractive common-pool-resource game (Joris Gillet, Arthur Schram, and Joep Sonnemans 2009), and they achieve a higher payoff/risk ratio in a portfolio selection task (Rockenbach, Abdolkarim Sadrieh, and Barbara Mathauschek 2007). Charness, Karni, and Levin (2007) also show that teams violate the principles of Bayesian updating less often than individuals do. Summarizing the evidence, teams can be considered more "rational players" (Bornstein and Yaniv 1998) in a broad variety of strategic and nonstrategic tasks.

Of course, an important question is why differences between team decisions and individual decisions occur. Opening the "black box" of team decision making, Cooper and Kagel (2005) have analyzed the content of team members' dialogues. In their experiment, teams play a limit pricing game against another team. In this signaling game, a market incumbent can signal his cost type to a potential entrant by choosing a particular output level. Teams are found to act more strategically than individuals by choosing higher output levels in order to signal that market entry is not profitable. The team dialogues reveal that one particular reason for this behavior is that teams often put themselves into the shoes of their competitors, meaning that they view the game from the competitor's perspective before making their own decisions. This perspective taking lets teams act more strategically and more competitively. Studying team behavior in experimental tournaments, Sutter and Christina Strassmair (2009) find in their analysis of team members' dialogues that members often refer to the payoffs of their colleagues to support a more competitive strategy. Members frequently raise the goal of beating the competitors in order to get more money for each team member. Sutter and Strassmair (2009) can show that team members

[2] One implication of my finding is that it potentially challenges some well-known and widely accepted theories in social psychology that rely on a distinction between ingroup and outgroup to explain individual behavior in groups (see, for example, the optimal distinctiveness model of Marilynn B. Brewer 1991, or the social identity theory of Henri Tajfel and John Turner 1979). Exploring this implication in more detail is beyond the scope of this comment, though.

[3] Team decision making is often referred to as "group decision making" in the literature. However, in order to separate more clearly between group membership in the spirit of CRR and team decisions, I will use the term "team" for situations in which several subjects have to agree on a joint decision, and the term "group membership" for situations in which individuals make decisions independently of others, but are somehow related to others, for example, by being observed by others or by the prevalence of payoff commonality.

[4] Since the seminal paper of Timothy N. Cason and Vai-Lam Mui (1997), some researchers have considered team behavior in the dictator game to be more generous than individual behavior. Cason and Mui (1997), however, report only that team choices are more other-regarding than individual choices for teams that made unequal individual choices.

increase their efforts in the tournament significantly if such a concern for the other members' payoffs is raised within a team. In sum, the evidence from team decision making experiments suggests that team decisions are strongly influenced by the intention to increase not only one's own payoff, but also the payoffs of the other team members.

This latter aspect of team decision making is remarkably similar to the findings of CRR, who have shown that salient group membership changes individual behavior in a direction that yields more favorable outcomes for the other group members. Given this similarity, it seems straightforward to ask whether salient group membership influences individuals in such a way that individual decisions become similar to—and possibly indistinguishable from—decisions taken by teams. Neither CRR nor the existing literature on team decision making has examined this question, even though answering it will provide a link between these two hitherto unrelated strands of literature.

I will examine the relation between team decision making and individual decisions under salient group membership by running a simple investment experiment. The experiment is based on one treatment of a paper by Uri Gneezy and Jan Potters (1997) in which they have investigated the effects of myopic loss aversion. My paper is not about myopic loss aversion, though.[5] Nevertheless, the design of Gneezy and Potters (1997) provides a very suitable framework to study the effects of group membership on individual decisions and compare them in a straightforward way to the outcome of team decisions. Since the experiment is nonstrategic, using this design also has the advantage that comparing individual decisions under group membership with team decisions cannot be confounded by any possible interaction effect of strategic interaction with individual or team decisions. Using a nonstrategic game also provides a robustness check of the results of CRR for a completely different decision task. The investment task used here does not even require the existence of an outgroup, which distinguishes this paper also from Chen and Li (2009).

My experimental results yield an affirmative answer to the question posed above. Individual decisions under salient group membership and decisions made by teams are largely similar. One implication of this finding is that team decision making and individual decision making under salient group membership can be considered substitutes. This insight promotes a better understanding of the characteristics and determinants of team decision making, as it documents that decision making in teams changes individual behavior in the same way as individual decisions are influenced when payoff commonality applies. Although previous analyses of team dialogues in Cooper and Kagel (2005) or Sutter and Strassmair (2009) have shown that team behavior is influenced by team members referring to joint payoffs, it has remained an open question whether payoff commonality *itself*—when subjects do *not* become members of a team—changes individual behavior or whether the mere fact of being member of a team adds something "extra" to the differences in individual and team behavior. My results suggest that payoff commonality itself yields the difference. Finally, since I also find a profound effect of salient group membership on individual decisions in a nonstrategic task without any outgroup, the results of CRR are obviously applicable beyond the domain of strategic games or tasks where an outgroup is involved.

The rest of the paper is organized as follows. Section I describes the basic experimental design. Section II presents the experimental treatments and results and is divided into three subsections.

[5] In a nutshell, myopic loss aversion assumes that people are myopic in evaluating outcomes over time, and are more sensitive to losses than to gains. In the context of financial decision making, myopic loss aversion implies that subjects invest less in risky assets the more frequently their returns are evaluated and the more often subjects can change their investment decision (see Gneezy, Arie Kapteyn, and Potters 2003; or Michael S. Haigh and John A. List 2005, for more details and for evidence that myopic loss aversion affects also professional traders and experimental markets). In Sutter (2007), I have shown that team decision making is also prone to myopic loss aversion. The two experimental treatments presented in Section IIA use data from the SHORT-condition in Sutter (2007). All other treatments presented in this paper are novel.

Subsection A examines whether individuals and teams make different decisions in the experimental task. Subsection B tests the effects of salient group membership on individual decisions and compares them to those of team decision making. Subsection C then addresses the question whether team membership (i.e., not only group membership) also has an impact on individual decisions. Section III concludes the paper by summarizing the main findings and discussing their implications.

I. Basic Experimental Design

All treatments reported below rely on the basic design of Gneezy and Potters (1997). Subjects receive an endowment of 100 euro-cents (i.e., €1) in each out of nine rounds. Then they have to choose in each round how much to invest in a lottery with the following properties. With a probability of $1/3$ the lottery returns two and a half times the invested amount X in addition to the initial endowment, yielding a round payoff of $100 + 2.5X$ euro-cents. With a probability of $2/3$ the invested amount is lost, yielding $100 - X$ euro-cents as payoff. Such a lottery yields the highest expected value (of 116.67 euro-cents) in case of a maximum investment of $X = 100$ euro-cents. Subjects are informed at the end of each round about the lottery's outcome, the resulting payoff in this round, and the accumulated payoffs up to the present round. Note that the maximum investment in each round is 100. That means that an endowment not invested in previous rounds cannot be carried over to be invested in later rounds.

Experimental sessions were run with z-Tree (Urs Fischbacher 2007)[6] and conducted at the Max Planck Institute of Economics in Jena. Sessions lasted on average 40 minutes. A total of 358 students from the University of Jena participated in the experiment (using the software ORSEE by Ben Greiner (2004) for recruitment). No subject was allowed to participate in more than one session and all subjects were randomly assigned to any of the treatments described in the following section. Participants earned on average €12.2 (including a show-up fee of €2).

II. Experimental Treatments and Results

A. *Individual versus Team Decisions*

Treatments.—The treatment variable considered first is the type of decision maker being either an individual (treatment INDIVIDUALS; $N = 64$ subjects) or a team of three subjects who can communicate and discuss their decisions (treatment TEAMS; $N = 84$ subjects, yielding 28 teams). Teams are randomly assigned at the beginning of the experiment. Each team is seated in a room of its own and stays together for the whole experiment. The experimental instructions are identical in both treatments, with two exceptions.[7] First, teams have to agree on a joint decision that is binding for all team members. Second, *each* of the three team members gets paid the full amount earned by the team over the nine rounds. The latter procedure holds the per capita payoffs and marginal incentives constant across both treatments.

[6] In order to check whether the determination of the lottery's outcome via computer influenced behavior, I also ran some control sessions with paper and pen where the lottery's outcome was determined by drawings from an urn. There was never any significant difference between computerized sessions and those with paper and pen (given a particular treatment), allowing the pooling of data.

[7] All experimental instructions are available as additional material on the *AER* Web site (available at http://www.aeaweb.org/articles.php?doi=10.1257/aer.99.5.2247).

Results.—Figure 1 shows the average investments in single rounds. Starting from the very first period, there is a clear difference between both treatments. Overall average investments are 39.4 in INDIVIDUALS, and 55.7 in TEAMS ($p < 0.05$; $N = 92$; Mann-Whitney U-test[8]). Average earnings are €12.0 in INDIVIDUALS, respectively €12.6 in TEAMS.[9] I summarize as a first finding:

RESULT 1: *Teams invest significantly higher amounts than individuals do.*

Result 1 can be further qualified by checking whether the higher investments of teams are due to fewer individuals investing anything (meaning that it could be caused by a larger fraction of individuals shying away from positive investments) or whether those individuals with positive amounts also invest less than teams. It turns out that both explanations are valid. About 17.5 percent of individual choices are zero investments ($X = 0$), compared to 4.4 percent of team choices ($p < 0.05$; χ^2 test). However, those individuals who invest positive amounts also invest less than teams. If we consider only the positive investments, we find individuals investing 47.8, but teams 58.2 on average ($p < 0.1$; Mann-Whitney U-test).

B. *Individual Behavior and Group Membership: Payoff Commonality and Exchange of Messages*

This subsection introduces two treatments that investigate individual behavior under group membership. Both treatments share the characteristic that decisions are taken by individuals, but individuals are members of groups. Saliency of group membership is induced in two steps. First, I use payoff commonality.[10] This means that individual decisions have payoff consequences for the other group members, even though any individual is free to choose according to own preferences. Second, I add the opportunity of sending nonbinding messages to other group members. Both factors—payoff commonality and the exchange of messages—are of obvious importance in team decision making, where team members typically have the same payoffs from a given decision and where team members can communicate with each other by exchanging messages. If both factors influenced individual decisions under group membership such that they were no longer different from team decisions, this would establish a firm link between team decision making and individual behavior under group membership, and it would support the notion that both types of decision making can be considered substitutes.

Treatments.—In treatment PAY-COMM ($N = 54$ subjects), groups of three subjects are formed. Subjects get labels as member 1, member 2, or member 3. Decisions are made subsequently and independently, with each member being responsible for three rounds,[11] i.e., member 1 decides in rounds 1–3. The other two members are informed about the decisions and the outcome of the lottery after each round, and they earn the same amount as member 1. Member

[8] All tests reported in this paper are two-sided. When teams are considered, the team decision of three subjects is treated as one independent observation.

[9] Note that the relative ranking of earnings corresponds to the relative levels of investment here, but that earnings are not significantly different. The latter is due to a high variance in earnings because they depend heavily on the outcomes of the lottery in each round and on whether high investments coincide with positive lottery outcomes.

[10] CRR show that payoff commonality makes group membership salient. They also report that observation and feedback can make group membership salient. For the purpose of this paper—comparing individual behavior under group membership to team decisions—payoff commonality is most appropriate to induce saliency.

[11] The feature of members making decisions for three rounds only is motivated by letting each member be responsible for one-third of the decisions that real teams of three subjects make in TEAMS.

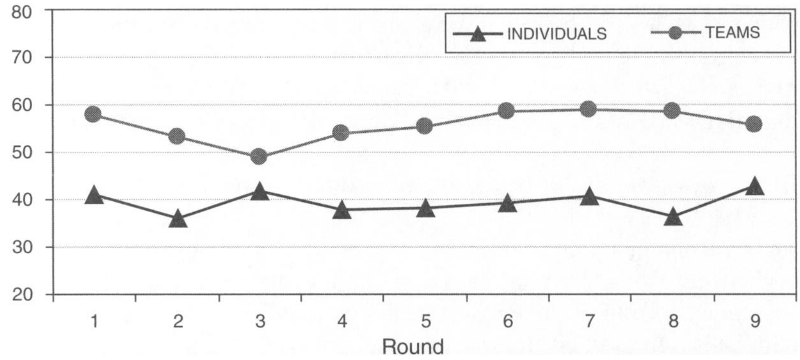


FIGURE 1. INVESTMENTS IN INDIVIDUALS AND TEAMS

2 decides for rounds 4–6, and member 3 for rounds 7–9, with the same information and payoff conditions as in rounds 1–3. Other than the payoff commonality, there is no interaction between the linked members, and members remain anonymous. The whole procedure is common knowledge to all members before member 1 starts making decisions.

Treatment MESSAGE ($N = 72$ subjects) is identical to PAY-COMM, but adds the opportunity of sending nonbinding messages. Members can write on a sheet of paper suggestions for investments or any other message to their predecessors (i.e., members deciding earlier) or their successors (i.e., members deciding later). Thus, member 1 receives two separate sheets of paper with messages from member 2, respectively member 3, before member 1 can make decisions for rounds 1–3. After round 3, members 1 and 3 can send messages to member 2 who can then decide for rounds 4–6. Finally, member 3 gets messages from members 1 and 2 before making decisions in rounds 7–9. Information conditions concerning the lottery's outcome are as in PAY-COMM, i.e., all linked members get to know the outcome as soon as a given round is over. Note that anonymity is preserved in treatment MESSAGE by forbidding subjects to send messages that might reveal their identity. In case a subject had violated this rule, he or she would not receive any payment. Yet, all subjects adhered to preserving anonymity.

Results.—Figure 2 shows the average investments in PAY-COMM and MESSAGE, where individuals make decisions as members of groups. Figure 2 also includes the investments in INDIVIDUALS as a benchmark, where individuals are isolated decision makers. It turns out that salient group membership through payoff commonality (PAY-COMM) induces higher investments than in INDIVIDUALS, though the difference needs three rounds to evolve clearly (overall investments are 50.3 in PAY-COMM versus 39.4 in INDIVIDUALS; $p < 0.05$; $N = 82$; Mann-Whitney U-test). Adding the opportunity of sending messages increases investment levels even further (with 61.4 in MESSAGE versus 50.3 in PAY-COMM; $p < 0.1$; $N = 42$). Earnings are €12.4 in PAY-COMM, respectively €12.3 in MESSAGE. It is important to stress that investment levels both in PAY-COMM and in MESSAGE are not significantly different from those in TEAMS (with 55.7 on average; $p > 0.3$ in any comparison). Hence, salient group membership and team decision making can be considered substitutes in their influence on individual behavior. This leads to:

RESULT 2: *Individual behavior is also strongly affected by salient group membership in a non-strategic task (without any outgroup). If group membership entails two factors that are important in team decision making—payoff commonality and the exchange of messages—individual decisions are, in the aggregate, no longer different from team decisions.*

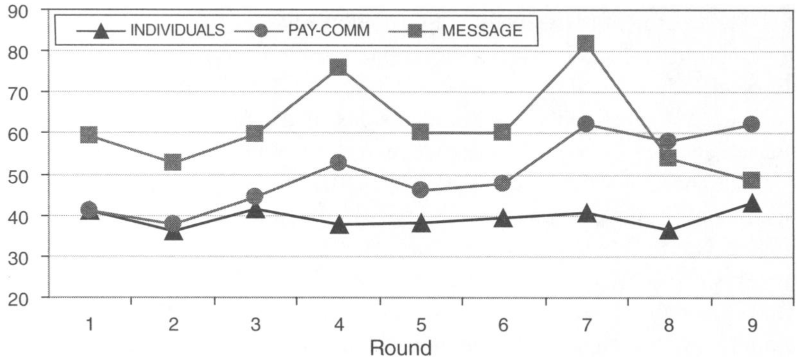


FIGURE 2. INVESTMENTS IN INDIVIDUALS, PAY-COMM, AND MESSAGE

Treatment MESSAGE offers an opportunity to classify the type of messages sent back and forth and how they affect investment levels. The coding has been done independently by two research assistants who later jointly clarified diverging assessments. The three most frequently occurring messages are the following. Message M1 proposes to make high investments, because the expected payoff for all group members is maximized with the maximum investment of 100. This message can be found on 30 percent of the sheets used for sending messages. Receivers of message M1 seem to respond to it, since the average investment level of a receiver is significantly positively correlated with message M1 ($r = 0.23$; $p < 0.05$, Pearson correlation). Message M2 suggests investing little, because the probability of losing in a single round is double that of winning in the lottery. This message is included in 18 percent of sheets, but it is not significantly correlated with actual investment levels chosen by the recipients of these messages ($r = -0.08$; $p = 0.50$). Message M3 recommends high investments, because the group can reasonably expect to win on average in three out of nine rounds. It is used on 15 percent of sheets, but has no significant correlation with investment levels either ($r = 0.09$; $p = 0.47$). All other types of messages are rare. Some sheets also contain only a suggested investment level, without providing any reasoning for it. In sum, the most frequently used messages try to give information on what to do and to influence the recipient in the sender's favorite direction, and appeal to joint payoffs in the group.[12]

C. *The Influence of Team Decision Making on Individual Decisions*

The previous subsection has shown that group membership has an effect on individual decisions. Now I investigate whether the experience of team decision making (contrary to mere group membership where decisions are still taken independently from other subjects) can also affect individual behavior. If this is to be found and the effects are similar to those reported in the previous subsection, this would corroborate the finding that both group membership—where individual decisions are taken independently from other subjects—and team decision making—which requires a joint decision of several team members—have largely similar effects on individual behavior.

[12] Of course, it would be very interesting to compare the messages used in MESSAGE with the content of communication in treatment TEAMS. Unfortunately, the communication in TEAMS has not been recorded (since teams sat in separate rooms). It seems clear, though, that messages M1 and M3 target the issue of how to maximize expected earnings from the experiment. This is also what has been found in an analysis of the video-protocols of team communication in a signaling game (Sutter 2009), providing indirect support for the conjecture that in the experiment reported here the messages used in MESSAGE were probably similar to the arguments exchanged in TEAMS.

Treatment.—The treatment MIXED ($N = 84$ subjects) intends to examine how the experience of team decision making affects individual decision making. In rounds 1–3, each subject decides independently of all other members. Payoff commonality does not apply in these rounds, and group members are not informed about the decisions of the other members. For rounds 4–6, however, three subjects are linked together to form a team. They are then connected via an electronic chat in which they can exchange any messages (that do not reveal their identity) in order to reach a team decision. Team decisions are valid only if all team members enter the same decision on their computer. Naturally, this means that in rounds 4–6 all group members earn the same amount of money since they make a team decision. Participants are not informed at the beginning of the experiment about the need to make team decisions in rounds 4–6, but this is revealed only after round 3.[13] After round 6, it is announced that a final phase of individual decision making in rounds 7–9 (identical to rounds 1–3) completes the session. This final phase is important to examine the effects of team decision making on subsequent individual decisions.

Results.—Figure 3 compares the investments in MIXED to the benchmark of INDIVIDUALS. The dotted lines in MIXED indicate the transitions from individual to team decision making (after round 3) and from team to individual decision making (after round 6). As one would have expected from the results in Section IIA, investments increase from rounds 1–3 to rounds 4–6 (45.3 versus 53.9 on average; $p = 0.08$; $N = 28$; Wilcoxon signed-ranks test)[14]. Individual investments in rounds 7–9 are not significantly lower than the investments of teams in rounds 4–6, though ($p > 0.6$; $N = 28$; Wilcoxon signed-ranks test), but they are higher than the individual investments in rounds 1–3 (50.7 versus 45.3 on average; $p = 0.08$; $N = 84$; overall average earnings are €11.9). The development of investment levels suggests that the experience of team decision making has an impact on individual behavior as well. It also seems noteworthy that the average investments in rounds 7–9 in MIXED are remarkably close to the overall average investments in PAY-COMM (50.3), which indicates that the effects of group membership and those of experiencing team decision making are very similar. These findings are summarized in:

RESULT 3: *Subjects increase their investments when they switch from individual to team decision making, but they do not decrease investments significantly when switching back. Hence, the experience of team decision making also affects individual behavior.*

Examining the content of communication between team members in rounds 4–6 reveals that the same messages that have been used most frequently in MESSAGE have been exchanged in MIXED. In 32 percent of teams, message M1, which proposes high investments in order to maximize the expected payoff, is exchanged. It is strongly correlated with the average investment ($r = 0.62$; $p < 0.01$, Pearson correlation). Message M2—proposing small investments due to the high probability of losing—is voiced in 21 percent of teams, but has no significant effect

[13] This approach was taken to avoid that individual decisions in rounds 1–3 might be influenced by the prospect of deciding in a team later on. Note that participants in MIXED are informed in the initial instructions (before round 1 starts) that the way in which decisions have to be made might change in the course of the experiment. Hence, there is no deception of subjects involved here.

[14] I use a conservative measure for testing here, because I match the investments of a team in rounds 4–6 with the average investments of the three members in rounds 1–3. Hence, each team of three members constitutes one independent unit of observation. Using a random effects panel regression (with clustering on the individual decision maker) as a less conservative test of team effects yields an estimated coefficient of 6.45 ($p < 0.05$) for the dummy of team decision making in rounds 4–6.

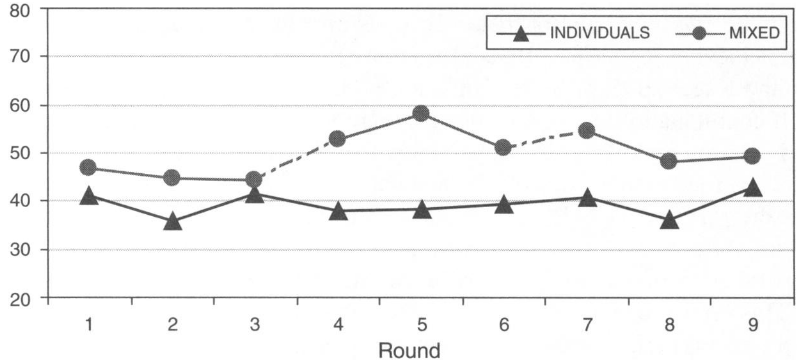


FIGURE 3. INVESTMENTS IN INDIVIDUALS AND MIXED

on the investment levels. Neither has message M3, which advocates positive investments due to an expected number of three wins and which is found in 11 percent of teams.

III. Conclusion

In this comment I have used a nonstrategic investment task to explore further the effects of salient group membership on individual behavior. CRR have shown that group membership changes individual behavior, making it more competitive in coordination games and prisoner's dilemma games. This paper adds to their findings in three important ways.

The first lesson from this paper is that the effects of salient group membership on individual behavior prevail also in a nonstrategic task that has no outgroup. Consequently, for the effects of group membership to show up requires neither strategic interaction between different groups nor the existence of an outgroup. Group membership in itself is important, regardless of whether other groups exist or not. This lesson implies that the findings of CRR or Chen and Li (2009) are robust to variations in the type of tasks and the institutional structure. A recent paper by Charness, Karni, and Levin (2008) provides further evidence for the effects of group membership on individual behavior, also in a context without an outgroup. They have found that subjects make fewer errors in probability judgments, i.e., they are less prone to the conjunction fallacy when they can consult with others before making a decision.

The second, and more important, lesson from this paper is that team decision making has the same effects as salient group membership on individual decisions. This result is by no means trivial, because team decision making requires deliberation, compromise, and consensus among team members, whereas salient group membership leaves the decision-making power unconditionally with a single individual. Individual behavior under group membership does not require compromise or any other form of coordination with other group members. Hence, although both forms of decision making—in teams as well as individually with group membership—have distinctly different institutional structures, they yield decisions that are largely the same in the aggregate. An important qualification to this statement is to acknowledge the—in my view most interesting—finding of CRR that group membership has to be salient to affect individual behavior. Payoff commonality obviously satisfies this condition. In the strategic games of CRR and in my nonstrategic task payoff commonality has shifted individual decisions in a direction that is more favorable for the group as a whole, meaning that individuals wish to take actions that are expected to be good not only for themselves, but also for the other group members.[15,16] The same goal of achieving higher payoffs is also often invoked in the dialogues of team members (see Cooper and Kagel 2005, or Sutter 2009, for evidence from signaling games, for instance). Hence, payoff commonality seems to be one driving force for individual behavior in groups, but also for team decision making. The second lesson therefore links two hitherto unrelated strands of literature, i.e., group membership effects and team decision making.

A third lesson from this paper is that the experience of team decision making also affects individual behavior. After having made decisions in a team, individual decisions are much closer to the previously taken team decisions than to the decisions that the same individuals have taken as individuals. This finding can be considered a robustness check for the second lesson. Therefore, the bottom line of this paper is that both salient group membership—where individual decisions are taken independently from other subjects—and team decision making—which requires a joint decision of several subjects—have largely the same effects on individual behavior.

REFERENCES

Akerlof, George A., and Rachel E. Kranton. 2000. "Economics and Identity." *Quarterly Journal of Economics*, 115(3): 715–53.

Bornstein, Gary, Tamar Kugler, and Anthony Ziegelmeyer. 2004. "Individual and Group Decisions in the Centipede Game: Are Groups More 'Rational' Players?" *Journal of Experimental Social Psychology*, 40(5): 599–605.

Bornstein, Gary, and Ilan Yaniv. 1998. "Individual and Group Behavior in the Ultimatum Game: Are Groups More 'Rational' Players?" *Experimental Economics*, 1(1): 101–08.

Brewer, Marilynn B. 1991. "The Social Self: On Being the Same and Different at the Same Time." *Personality and Social Psychology Bulletin*, 17(5): 475–82.

Cason, Timothy N., and Vai-Lam Mui. 1997. "A Laboratory Study of Group Polarisation in the Team Dictator Game." *Economic Journal*, 107(444): 1465–83.

Charness, Gary, and Matthew O. Jackson. 2009. "The Role of Responsibility in Strategic Risk-Taking." *Journal of Economic Behavior and Organization*, 69(3): 241–47.

Charness, Gary, Edi Karni, and Dan Levin. 2007. "Individual and Group Decision Making under Risk: An Experimental Study of Bayesian Updating and Violations of First–Order Stochastic Dominance." *Journal of Risk and Uncertainty*, 35(2): 129–48.

Charness, Gary, Edi Karni, and Dan Levin. 2008. "On the Conjunction Fallacy in Probability Judgment: New Experimental Evidence." University of California–Santa Barbara Economics Working Paper 14–08.

Charness, Gary, Luca Rigotti, and Aldo Rustichini. 2007. "Individual Behavior and Group Membership." *American Economic Review*, 97(4): 1340–52.

Chen, Yan, and Sherry Xin Li. 2009. "Group Identity and Social Preferences." *American Economic Review*, 99(1): 431–57.

Cooper, David J., and John H. Kagel. 2005. "Are Two Heads Better Than One? Team Versus Individual Play in Signaling Games." *American Economic Review*, 95(3): 477–509.

[15] If this wish determined individual behavior under salient group membership, then investing the full amount in order to maximize the expected payoff not only for oneself, but also for the linked members, should be more frequent in PAY-COMM and MESSAGE than in INDIVIDUALS. In fact, this is what I find, since the relative frequency of investing the full endowment ($X = 100$) is significantly higher both in PAY-COMM (18.5 percent) and in MESSAGE (36.6 percent) than in INDIVIDUALS (12.5 percent; $p < 0.05$ in both comparisons; χ^2 tests).

[16] Maximizing expected payoffs requires higher investments—and thus more exposure to risk—in my experiment. Charness and Jackson (2009) find in a Stag Hunt game (which is a two-player coordination game) that individuals make less risky decisions when payoff commonality applies. Though this might seem conflicting evidence at first sight, both findings are compatible when considering the expected payoffs in the Stag Hunt game under the assumption that the opponent player chooses randomly. Taking the safe option "Hare" yields a sure payoff of 8 in the experiment of Charness and Jackson (2009), irrespective of the other's choice. However, the risky option "Stag" has an expected payoff of 5 only (getting either 9 if the other player also chooses "Stag," or 1 if "Hare" is chosen by the other player). Thus, the results of Charness and Jackson (2009) can be interpreted as subjects maximizing the expected payoff in the face of strategic uncertainty about the other player's behavior.

Cox, James C. 2002. "Trust, Reciprocity, and Other-Regarding Preferences: Groups vs. Individuals and Males vs. Females." In *Experimental Business Research*, ed. Rami Zwick and Amnon Rapoport, 331–50. Dordrecht: Kluwer Academic.

Fehr, Ernst, Helen Bernhard, and Bettina Rockenbach. 2008. "Egalitarianism in Young Children." *Nature*, 454(28 August): 1079–84.

Fischbacher, Urs. 2007. "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics*, 10(2): 171–78.

Gillet, Joris, Arthur Schram, and Joep Sonnemans. 2009. "The Tragedy of the Commons Revisited: The Importance of Group Decision-Making." *Journal of Public Economics*, 33(5–6): 785–97.

Gneezy, Uri, Arie Kapteyn, and Jan Potters. 2003. "Evaluation Periods and Asset Prices in a Market Experiment." *Journal of Finance*, 58(2): 821–37.

Gneezy, Uri, and Jan Potters. 1997. "An Experiment on Risk Taking and Evaluation Periods." *Quarterly Journal of Economics*, 112(2): 631–45.

Goette, Lorenz, David Huffman, and Stephan Meier. 2006. "The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups." *American Economic Review*, 96(2): 212–16.

Greiner, Ben. 2004. "An Online Recruitment System for Economic Experiments." In *Forschung und wissenschaftliches Rechnen 2003*, ed. Kurt Kremer and Volker Macho, 79–93. Göttingen: Gesellschaft für Wissenschaftliche Datenverarbeitung Göttingen.

Haigh, Michael S., and John A. List. 2005. "Do Professional Traders Exhibit Myopic Loss Aversion? An Experimental Analysis." *Journal of Finance*, 60(1): 523–34.

Kocher, Martin G., and Matthias Sutter. 2005. "The Decision Maker Matters: Individual versus Group Behaviour in Experimental Beauty-Contest Games." *Economic Journal*, 115(500): 200–23.

Kugler, Tamar, Gary Bornstein, Martin G. Kocher, and Matthias Sutter. 2007. "Trust between Individuals and Groups: Groups Are Less Trusting Than Individuals but Just as Trustworthy." *Journal of Economic Psychology*, 28(6): 646–57.

Luhan, Wolfgang J., Martin G. Kocher, and Matthias Sutter. 2009. "Group Polarization in the Team Dictator Game Reconsidered." *Experimental Economics*, 12(1): 26–41.

Maciejovsky, Boris, and David V. Budescu. 2007. "Collective Induction Without Cooperation? Learning and Knowledge Transfer in Cooperative Groups and Competitive Auctions." *Journal of Personality and Social Psychology*, 92(5): 854–70.

Rockenbach, Bettina, Abdolkarim Sadrieh, and Barbara Mathauschek. 2007. "Teams Take the Better Risks." *Journal of Economic Behavior and Organization*, 63(3): 412–22.

Sutter, Matthias. 2007. "Are Teams Prone to Myopic Loss Aversion? An Experimental Study on Individual Versus Team Investment Behavior." *Economics Letters*, 97(2): 128–32.

Sutter, Matthias. 2009. "Deception through Telling the Truth?! Experimental Evidence from Individuals and Teams." *Economic Journal*, 119(534): 47–60.

Sutter, Matthias, and Christina Strassmair. 2009. "Communication, Cooperation and Collusion in Team Tournaments–an Experimental Study." *Games and Economic Behavior*, 66(1): 506–25.

Tajfel, Henri, and John Turner. 1979. "An Integrative Theory of Intergroup Conflict." In *The Social Psychology of Intergroup Relations*, ed. William Austin and Stephen Worchel, 33–47. Monterey, CA: Brooks/Cole.