

Does consultation improve decision-making?

Alessia Isopi · Daniele Nosenzo · Chris Starmer

Published online: 20 June 2014

© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract This paper reports an experiment designed to test whether prior consultation within a group affects subsequent individual decision-making in tasks where demonstrability of correct solutions is low. In our experiment, subjects considered two paintings created by two different artists and were asked to guess which artist made each painting. We observed answers given by individuals under two treatments: In one, subjects were allowed the opportunity to consult with other participants before making their private decisions; in the other, there was no such opportunity. Our primary findings are that subjects in the first treatment evaluate the opportunity to consult positively, but they perform significantly worse and earn significantly less.

Keywords Consultation · Decision-making · Group decisions · Individual decisions

JEL Classification C91 · C92 · D80

1 Introduction

Consultation is a key ingredient of many deliberative processes. In many walks of life, individuals consult with others before taking important decisions. Obvious exam-

A. Isopi
Department of Economics, University of Manchester,
Oxford Road, Manchester M13 9PL, UK
e-mail: alessia.isopi@manchester.ac.uk

D. Nosenzo · C. Starmer (✉)
School of Economics, University of Nottingham,
University Park, Nottingham NG7 2RD, UK
e-mail: chris.starmer@nottingham.ac.uk

D. Nosenzo
e-mail: daniele.nosenzo@nottingham.ac.uk

ples include investors consulting with financial advisors or individuals talking with health professionals when choosing between alternative medical interventions. With the growth of the internet, sources of ‘advice’ are expanding rapidly, while the costs of accessing them are often very low. But here, as in the examples mentioned above, the quality of advice obtained may be difficult to assess, raising interesting questions about the conditions under which consulting with others can be expected to improve (or worsen) individual decision-making.

One body of literature which might inform understanding of the influence of consultation on individual decisions is the extensive research examining the comparative success of decisions made by individuals versus decisions made by groups. There is now considerable evidence that groups can often ‘outperform’ individuals. The bulk of it comes from experiments in social psychology examining behavior in decision problems that have correct solutions and thus have a meaningful criterion for assessing decision accuracy. Within this literature, a widely reported finding is that groups are more likely to report the correct answer (see, e.g., [Hastie 1986](#); [Laughlin et al. 2003 October, 2006](#), and references therein). Economists have also compared individual and group decisions. A relatively small literature has investigated the incidence of preference ‘anomalies’ comparing groups and individuals, where one of the first contributions is [Bone et al. \(1999\)](#). A larger literature, starting with [Cason and Mui \(1997\)](#) and [Bornstein and Yaniv \(1998\)](#), has focused on interactive decisions where a common result is that groups’ decisions more closely track standard game theoretic predictions.

The fact that groups often perform better than individuals suggests the possibility that consultation with others, prior to an individual decision, might also have improving effects. There is, however, limited evidence of how deliberation within a group affects the quality of a later individual decision: [Maciejovsky et al. \(2013\)](#) report that subjects who have solved decision problems as part of a group subsequently perform better as individuals in similar decision tasks; [Charness et al. \(2010\)](#) find that group consultation mitigates some decision anomalies found in individual choice experiments. While these recent results chime with the broader literature comparing the success of individuals and groups, like many of the studies reviewed by [Hastie \(1986\)](#), both of them also share a design feature which may limit their scope: that feature is the use of tasks which have *demonstrably* correct solutions.¹

We will say that the solution to a decision problem is *fully demonstrable*, in a given context, when someone who knows it (or how to identify it) can convey that knowledge to other individuals facing the same decision.² In previous research, demonstrability has usually been implemented using tasks which have correct answers (e.g., multiplying two numbers together or finding the solution to a logical reasoning problem such as

¹ Also related is the paper by [Sausgruber and Tyran \(2011\)](#) who study voting on tax regimes in an experimental market. They examine whether deliberation mitigates a ‘tax-shifting bias’ whereby buyers/voters believe that taxing sellers comes at a lower economic cost than taxing themselves. They find that deliberation does not eliminate the bias. As in [Charness et al. \(2010\)](#) and [Maciejovsky et al. \(2013\)](#), also the task used in [Sausgruber and Tyran \(2011\)](#) has a solution which is in principle ‘demonstrable.’ Differently from these studies, however, the decision setting in [Sausgruber and Tyran \(2011\)](#) displays decision externalities within groups as subjects get paid as a function of the votes made in their group.

² A richer discussion of the concept of demonstrability is provided by [Laughlin and Ellis \(1986\)](#).

Wason's selection task). In such cases, while some individuals may not independently find the correct solution, the task is demonstrable if they will *recognize* the solution when presented with suitable arguments to identify it.

High demonstrability of solutions may be an important ingredient explaining the relative success of groups over individuals across a range of existing experimental findings.³ But it is not obvious that demonstrability is characteristic of most settings in the world where individuals seek advice from others. In fact, 'experts' or 'professional' advisors often encounter difficulties in providing compelling arguments in defense of their estimations of, say, the research publication potential of different candidates in an academic job market, the payoff from a particular model of corporate re-branding, the profitability of a proposed investment, or the risks associated with a new drug treatment. Indeed, such cases are often characterized by disagreements in the assessments of professionals.

In this paper, we examine the effect of consultation on individual decision-making in a task designed to have low demonstrability by comparing behavior across two treatments. In one treatment, before facing the decision, subjects discussed the task with other participants. We compare decisions made by these subjects with those made in a control group who had no opportunity to discuss the task with others. Notice that our study differs from the literature which compares the decisions of groups with those of single individuals because our study focuses on whether deliberating with others has an impact on subsequent *individual* decision-making.

Our primary findings are that subjects who had the opportunity to consult reported that it was helpful, but they actually performed worse and earned less than those who had no such opportunity. This effect is partly driven by a tendency for individuals to form consensus around uninformed opinion, a result which, as we discuss further below, has some resonance with the 'groupthink' phenomena reported widely in social psychology.

2 Experimental design and procedures

We use a decision task which is designed to have a correct solution known to the experimenter but for which demonstrability is low. All of the subjects in our experiment faced a pair of decision problems in which they were asked to consider two paintings. For each painting, subjects had to select which of two artists, Paul Klee or Wassily Kandinsky, had made the painting.⁴ Subjects received £1.50 for each correct answer (and nothing for any incorrect answer). Figure 1 shows the computer screen that subjects used to submit their answers.

³ This point is also made by, e.g., [Hastie \(1986\)](#), who reviews the social psychology literature comparing group and individual behavior in relation to tasks with varying degrees of solution demonstrability. Whereas groups clearly outperform individuals in tasks where solutions are easily demonstrated, the comparison is less clear in tasks where solutions are more difficult to demonstrate: Here groups seem to perform slightly better than the average individual, but usually below the level of the best performing individual.

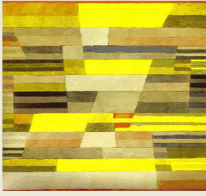
⁴ The two paintings were *Monument in Fertile Country*, 1929 by Paul Klee (painting A in the experiment), and *Weighing*, 1928 by Wassily Kandinsky (painting B). A similar task is also used by [Chen and Li \(2009\)](#) as part of a social identity manipulation.

For each painting on the right of the screen, you now have to select the artist who you think made the painting.
For each correct answer, you will be rewarded with € 1.50

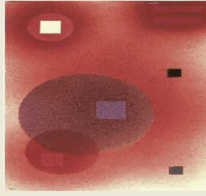
Painting A - My answer is: Klee
 Kandinsky

Painting B - My answer is: Klee
 Kandinsky

Painting A



Painting B



Once you have made your decision, press the SUBMIT button to submit your choices.

Fig. 1 The decision task

We contend that this task has relatively low demonstrability⁵ because the solutions to our painting tasks cannot be identified via the application of any system of reasoning that would be commonly understood by our subjects.⁶

Each subject took part in either the **Individual** or **Consultation** treatment. In the Consultation treatment, the decision task was preceded by a ‘group-discussion’ stage where subjects were randomly divided into three groups of six. After being assigned to a group, subjects took part in a computerized chat where they could discuss the task for 5 minutes with other group members before submitting their answers. Subjects knew that messages were only shared among the members of their own group. At the end of the 5 minutes, subjects individually submitted their answers. In the Individual treatment, by contrast, there was no ‘group-discussion’ stage before the decision task. Note that in both treatments subjects made private decisions as individuals. Thus, our study provides a controlled test of whether being able to discuss the task with others has an impact on subsequent individual performance as compared to a baseline situation (the Individual treatment) where group deliberation is not possible.

At the end of the experiment, there was a short questionnaire eliciting demographic and attitudinal information.⁷ This included self-assessment of risk attitudes (the SOEP general risk question) and trust attitudes (the WVS Trust question). In the data analysis, responses on these two questions enter as controls in a regression of subjects’ responses to the painting task.

⁵ Our task is similar in spirit to some of the low-demonstrability tasks used in the social psychology literature, such as world knowledge questions (e.g., ‘what is the capital city of Lithuania?’) or estimation tasks (e.g., ‘what is the population of China?’).

⁶ We do not rule out the possibility that there may be arguments that a community of art experts might recognize as identifying the correct responses to our task (perhaps involving, for example, references to the style of brush strokes used by the different artists). Our subjects, however, were not selected on the basis of, or expected to have, any particular expertise in art history.

⁷ The decision task reported here was the first part of a larger experiment and was followed by a one-shot sequential principal-agent game which is not related to the question examined in this paper.

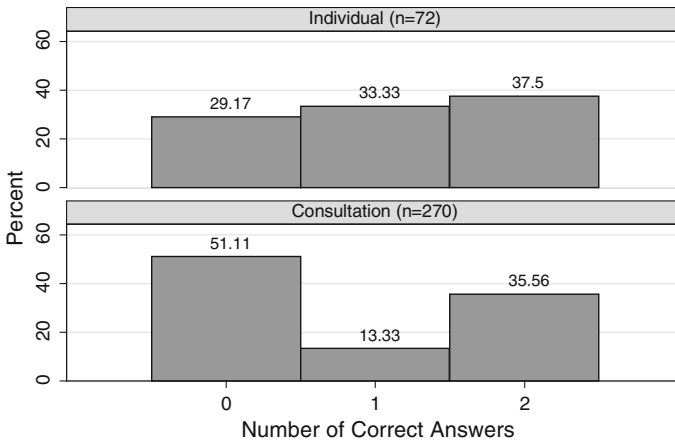


Fig. 2 Distribution of correct answers across treatments

A total of 342 subjects took part (270 in the Consultation treatment). Subjects were students recruited via ORSEE (Greiner 2004) had an average age 20.2 years, and 50 % were female. Subjects' earnings from the task ranged from £0.00 to £3.00, averaging £1.34. The experiment was conducted using the software z-Tree (Fischbacher 2007). The software and full instructions are available on request to the authors.

3 Results

3.1 Does consultation improve decision-making?

Figure 2 shows the distribution of correct answers in the Individual and Consultation treatments. In the Individual treatment (top panel) 38 % of the subjects answer correctly to both painting questions, 33 % answer correctly to one question, and 29 % submit two wrong answers.

In the Consultation treatment, the fraction of subjects correctly answering both questions is similar to that in the Individual treatment (36 % in Consultation). However, the two treatments differ markedly in the proportions of subjects with either zero or one correct answer. In Consultation, the proportion of subjects submitting two wrong answers is 51 %, while only 13 % of subjects submit one correct answer. We strongly reject the hypothesis that the distribution of correct answers is the same across the two treatments: $\chi^2(2^{df}) = 18.91$; $p < 0.001$. On average subjects in the Consultation treatment were *less* successful in identifying the correct painters and consequently they also earned less (earnings were 22 % lower than in the Individual treatment: $z = 2.14$; $p = 0.032$; two- sided Wilcoxon rank-sum test).

We note that the statistical tests reported above do not account for group interdependencies in the Consultation treatment; therefore, we proceed by analyzing the distribution of correct answers also using regression analysis. We use a generalized ordered logit regression model where the dependent variable records whether a subject

answers correctly to zero, one, or two questions.⁸ In Model I, the only independent variable is a treatment dummy (= 1 for the Consultation treatment). Model II adds controls for personal characteristics (gender, a dummy indicating whether a subject studies Humanities, and a self-assessment of the subject's risk and trust attitudes), and for session effects.⁹ Model III introduces interactions between the treatment dummy and the other regressors. Table 1 reports the regression results, displayed as factor changes in the odds of answering correctly. Note that a value >1 (resp. <1) implies a positive (resp. negative) effect on the odds of answering correctly.

Model I shows that being in the Consultation treatment reduces substantially (by a factor of 0.394) the odds of submitting at least one correct answer. In the Individual treatment (the benchmark condition), we expect to find approximately 2.43 subjects submitting at least one correct answer for every subject who submits no correct answer, in Consultation the same statistic falls to only ($0.394 \times 2.43 =$) 0.96. This effect is significant at the 1% level. Being in the Consultation treatment, however, has no significant impact on the odds of answering correctly to both questions (odds are reduced only by a factor of 0.919, $p = 0.810$). This is consistent with the intuitively plausible idea that those who know more are less likely to be swayed by the crowd.

These results are robust to the inclusion of controls for personal characteristics and session effects (Model II). In Model III, among the interaction terms, only the one between the treatment and gender dummies is significant (10% level). Interestingly, the model reveals that consultation is especially detrimental for females with the odds of at least one correct answer falling dramatically (by a factor of 0.088; $p = 0.007$). For male participants, the effect is smaller (the odds decrease by a factor of 0.204) and it is only marginally significant ($\chi^2(1^{df}) = 3.45$; $p = 0.063$). For female subjects, being in the Consultation treatment also reduces the odds of answering both questions correctly although the effect is only significant at the 10% level (the effect is insignificant for males).

3.2 The unrecognized curse of consensus

Why would the opportunity to consult with others have generated lower performance? A very striking feature of our data is a tendency for subjects in groups to give the same answers to the painting questions as those given by other members of their group. In approximately 84% of groups, an absolute majority of members submitted

⁸ The generalized ordered logit regression model allows relaxation of the 'parallel regression assumption' of the standard ordered logit model whereby the coefficients that describe the relationship between, e.g., submitting zero correct answers versus submitting one or more correct answers are the same as those that describe the relationship between submitting two correct answers versus submitting zero or one correct answers, etc. (see, e.g., Long and Freese 2006). This assumption is violated for the three regression models presented in Table 1 according to a Brant test (in all models $p < 0.05$). The test also showed that the largest violations were for the treatment dummy. Thus, we used the command `gologit2` in STATA 11 to estimate generalized ordered logit regressions where the parallel regression assumption was relaxed for the treatment dummy and maintained for other regressors.

⁹ The variable labeled 'Session' in Table 1 is an index of the order in which sessions were run (1, 2, 3, etc.). We include this to check whether there might have been some change in knowledge across the sequence of sessions. As we expected, the variable is insignificant.

Table 1 Regression analysis of the number of correct answers across treatments

	I		II		III	
	At least one correct	Both correct	At least one correct	Both correct	At least one correct	Both correct
Consultation treatment	0.394*** (0.007)	0.919 (0.810)	0.400** (0.022)	0.938 (0.873)	0.088*** (0.007)	0.228* (0.086)
1 if Male	-		0.862 (0.476)		0.453* (0.077)	
1 if Studies Humanities	-		0.836 (0.538)		0.853 (0.791)	
High Trust	-		0.863 (0.444)		0.808 (0.642)	
Risk Seeking	-		1.057 (0.182)		0.917 (0.401)	
Session	-		0.999 (0.983)		0.742 (0.103)	
Consultation * Male	-		-		2.312* (0.096)	
Odds of answering correctly in benchmark condition	2.43	0.60	2.09	0.51	11.58	2.56
N.	342		342		342	
Wald χ^2	11.47		15.12		20.84	
Prob > χ^2	0.003		0.034		0.053	
Pseudo R ²	0.025		0.029		0.039	

Generalized ordered logit regressions. The dependent variable is the number of correct answers submitted by a subject. The results are displayed as factor changes in the odds of answering correctly. *p* values are reported in parentheses. Standard errors are adjusted for intragroup correlation. A constant is included in all models, but omitted from the Table output. For the treatment dummy 'Consultation,' the parallel regression assumption is relaxed and the models report both the factor change in the odds of answering correctly to at least one question, and the factor change in the odds of answering correctly to both questions. Model III includes all possible interactions between the Consultation variable and the other regressors, but only interactions significant at the 10% level are reported in the Table. Significance levels: * 10%; ** 5%; *** 1%

identical answers to the two questions. About a third of the groups (14 out of 45) were *unanimous* in that all group members submitted the same answers; in another 24 groups, a *majority* (of four or five) submitted identical answers. This tendency arose even though participants submitted their answers *individually*, in private, and with no suggestion anywhere in the instructions that a group had to reach consensus.

Whether or not a group reached a consensus is strongly associated with subjects' evaluations of whether communicating with the other group members was a helpful input to the decision task. At the end of the experiment, but before being informed about the outcome of the decision task, subjects in the Consultation treatment were asked to rate how much they thought that communicating with other members of their group had helped them solve the two painting questions. They responded on a scale from 1 ('*not at all helpful*') to 10 ('*extremely helpful*').¹⁰ From these responses, we constructed a 'helpfulness index' as the mean of reported values for each group. For the 14 unanimous groups, the average helpfulness index is 6.56. This falls to 4.75 in the 24 groups where a majority of members submitted identical answers and to 2.40 in the 7 groups where no answer was submitted by an absolute majority. Both reductions are highly significant (two-sided Wilcoxon rank-sum tests give $p < 0.005$ for both comparisons).

The sting in the tail is that the poor performance in the Consultation treatment seems to be driven by those subjects who gave answers consistent with an absolute majority in their group. If we exclude from the Consultation data, all those subjects who formed part of a majority, we then find no significant difference between the average earnings of this subset (£1.438) and the average earnings of those in the Individual treatment (£1.625) (two-sided Wilcoxon rank-sum test: $z = 0.92$; $p = 0.355$). By contrast, average earnings in the Consultation treatment of those in majorities (£1.203) are significantly lower than those for the Individual treatment (two-sided Wilcoxon rank-sum test: $z = 2.43$; $p = 0.015$). One might wonder whether this inferior performance of majorities relative to subjects in the Individual treatment just reflects sorting according to knowledge, which results in relatively low representation of the better informed individuals among majorities. But while such sorting may account for some differential between subsets of subjects in two treatments, it cannot explain the *overall* difference in performance between treatments reported in the previous sub-section.

The face value interpretation of these data is that consensus makes you feel good and perform worse. There is of course considerable evidence from social psychology that individuals have a strong tendency to form consensus, even when there is no basis for it.¹¹ But an intriguing question is why majorities should have a tendency to coalesce around the *wrong* answers. We examine this in the next sub-section via analysis of the chat data from the Consultation treatment.

¹⁰ The average response was 4.95 (s.d. 3.17).

¹¹ For discussion of some early evidence see [Sherif \(1935\)](#).

Table 2 Percentage of correct suggestions about the paintings

	% of correct suggestions			
	Overall (%)	Initial suggestions (%)	Subsequent suggestions following a correct suggestion (%)	Subsequent suggestions following an incorrect suggestion (%)
Painting A	35	31	65	27
Painting B	50	44	57	37

3.3 The origins of consensus: insights from the chat data

In the Consultation treatment, subjects used a computer program to ‘chat’ with other group members. All but two of the 270 subjects sent at least one message, a total of 2,198 messages were exchanged (8.14 messages for the average individual). We coded messages and classified as ‘suggestions’ all cases where a subject explicitly suggested the artist of one of the paintings.¹² In total, 42 (about 6.5%) of messages were classified as containing suggestions.

Table 2 reports summary statistics for correct suggestions. Notice that suggestions made within a group tend to be correlated with initial suggestions influencing subsequent ones. In particular, for painting A, we find that when the first suggestion made in a group is correct, 65% of the subsequent suggestions are also correct. However, following an incorrect initial suggestion, only 27% of the subsequent suggestions are correct. Similarly, the probability of correctly identifying painting B is 57% if the initial suggestion in the group is correct and 37% if the initial suggestion is incorrect. Thus, there is a tendency among subjects in the Consultation treatment to develop a *consensus of opinions* around the answers that are first suggested in their group.

This consensus of opinions in the chat messages translates into a *consensus of actions* in the decision task. This is particularly evident in those groups where subjects receive homogeneous suggestions about which artists made which painting, i.e., in groups where all suggestions identify the same artist as the painter of a painting.¹³ In 23 groups, subjects received homogeneous suggestions about painting A. In six groups, the homogeneous suggestion is correct and 69% of subjects in these groups made the correct choice. In the remaining 17 groups, the homogeneous suggestion is incorrect and the probability of making the correct choice dropped to 12%. Similarly, for the 27 groups with homogeneous suggestions for painting B, the probability of

¹² Two examples of messages classified as suggestions are: ‘*i’d say the left one must’ve been made by klee*’ and ‘*i would go for Kandinsky for a*’. We do not consider as suggestions messages indicating agreement with suggestions made by others (e.g., ‘*I would agree Kadinsky A and Klee B*’; ‘*im going with klee for A too*’). We tested the robustness of our classification by having a research assistant independently code the messages.

¹³ This occurs in 66% of the groups where painting A is discussed, and in 84% of the groups where painting B is discussed. Note that homogeneous suggestions can occur either because there are different group members making the same suggestion about a painting (this is the case in 36% of groups with homogeneous suggestions), or because only one person makes a suggestion (64% of cases).

choosing the correct painter in the decision task is 82 % in the 11 groups with correct suggestions and 19 % in the 16 groups with incorrect suggestions.¹⁴

Overall, these findings reveal a strong tendency of subjects to form a consensus around the suggestions made in their group during the chat-discussion phase of the experiment. Suggestions initially made in a group influenced subsequent suggestions leading to a consensus of opinions. Partly as a result of this, in a large fraction of groups, subjects received homogeneous suggestions about the paintings. Choices made in these groups are strongly influenced by the suggestions observed in the group-discussion phase, i.e., we observe a consensus of actions around the suggestions made in a group. This may reflect something akin to the ‘drive toward consensus’ which Janis (1972) famously characterized as ‘groupthink.’ Several studies from social psychology, dating at least to Asch (1951), have illustrated how group pressures may lead individuals to conform to answers given by confederates, even when these are incorrect. (For a review see Esser 1998).

Perhaps more surprisingly, we find that subjects who are willing to make a suggestion typically make poor suggestions: About 60 % of the suggestions made in the chat-discussion phase are wrong. If we consider suggestions as a proxy for knowledge among those in the Consultation treatment who are willing to suggest their guess to others, and compare the proportion of correct suggestions with the proportion of correct decisions made in the Individual treatment, we find that these do not differ for painting B (50 % correct suggestions in Consultation vs. 44 % correct decisions in Individual, χ^2 (1df) = 0.39; $p=0.532$), whereas they are significantly different for painting A (35 vs. 64 %, χ^2 (1df) = 13.91; $p < 0.001$). This result holds if we only focus on initial suggestions made in a group ($p=0.002$ for painting A and $p=0.948$ for painting B).¹⁵

These results suggest a tendency in the Consultation treatment for the less informed individuals to be relatively more likely to promote their guess to others.¹⁶ A possible explanation for this may be the existence of a positive correlation between, on the one hand, an individual’s confidence that they know the answer and their willingness to suggest their guess to others and, on the other hand, between overconfidence and incompetence (e.g., Kruger and Dunning 1999). This, coupled with the strong tendency of subjects to form a consensus around the suggestions observed in their group regardless of their correctness, may suggest a possible mechanism at work in our experiment that induced subjects in groups to coalesce around wrong answers.

¹⁴ In 17 groups, subjects receive heterogeneous suggestions about the paintings. In these groups, choices are more diverse but most subjects tended to follow the suggestion made by at least 50 % of group members.

¹⁵ One may wonder whether subjects deliberately made wrong suggestions. This does not seem to be the case. The vast majority of subjects tended to follow the suggestion they made in the group, regardless of whether the suggestion was right or wrong.

¹⁶ Notice that this is quite distinct from the familiar ‘hidden profile problem’ (see for example, Lightle et al. 2009) in which individuals fail to share information which may inform an optimal decision. Our data, instead, point to active sharing of incorrect information.

4 Conclusion

We have reported an experiment designed to test the influence of consultation on individual decision-making. Our work is partly motivated by an extensive background literature which finds that groups often outperform individuals and we interpret our study as probing the *conditions* under which group interaction improves decision-making. As we noted, most of the evidence supporting conclusions of the form ‘teams make you smarter’ comes from experiments comparing decisions made by groups versus decisions made by individuals. However, we have argued that many interesting and important decisions where groups may play a role are better construed as *individual decisions* with an element of *consultation*, and part of our objective was to examine the extent to which the beneficial effects established for decisions made by groups extend to decisions made by individuals who consult.

A second distinguishing feature of our experiment was the use of a task with a (correct) solution which is low on *demonstrability*. This design feature had a number of connected motivations discussed in the introduction and reviewed here. The first stems from recognizing that the bulk of evidence pointing to beneficial effects of group decisions might be partly a by-product of experimental designs featuring tasks with high demonstrability. Indeed, when tasks have fully demonstrable solutions then, by definition, those who have knowledge of them can convey their knowledge to others. As such, adopting tasks with low demonstrability can be seen as providing a tougher test of the extent to which the knowledge possessed by some members of a group can be successfully transmitted to other members of it. That test is relevant, not least, because many interesting decisions in the world—and, in particular, many of those where consulting is commonplace—tend to have low demonstrability.

Our primary finding is that beneficial effects of group participation do not extend to our environment because of a tendency for individuals to follow the relatively poorly informed crowd, and this effect was particularly marked for females.

It is conceivable that consultation may lead to better outcomes in other settings, even when demonstrability is low. This may happen, for example, where there is substantial initial knowledge about the decision problem among individuals. Our findings, however, highlight that consultation does not always lead to better outcomes. A systematic analysis of the conditions under which consultation may positively or negatively effect decision-making seems a promising avenue for further research.

Acknowledgments We thank Spiros Bougheas, Niall Flynn, Jeroen Nieboer, Martin Sefton, Fangfang Tan and Jean–Robert Tyran, the Editors of the special issue and two anonymous referees for helpful comments. Andreas Aristidou provided excellent research assistance. Alessia Isopi acknowledges support from the Marie Curie Foundation (project TAAWE/220489). Daniele Nosenzo acknowledges support from the Leverhulme Trust (ECF/2010/0636). Chris Starmer acknowledges support from the Economic and Social Research Council via the Network for Integrated Behavioral Sciences (Award No. ES/K002201/1).

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Asch, S. E. (1951). *Effects of group pressure upon the modification and distortion of judgement. Groups, leadership and men*. Pittsburgh: Carnegie Press.
- Bone, J., Hey, J., & Suckling, J. (1999). Are groups more (or less) consistent than individuals? *Journal of Risk and Uncertainty*, 18(1), 63–81.
- Bornstein, G., & Yaniv, I. (1998). Individual and group behavior in the ultimatum game: Are groups more 'rational' players? *Experimental Economics*, 1(1), 101–108.
- Cason, T. N., & Mui, V. L. (1997). A laboratory study of group polarisation in the team dictator game. *Economic Journal*, 107(444), 1465–1483.
- Charness, G., Karni, E., & Levin, D. (2010). On the conjunction fallacy in probability judgment: New experimental evidence regarding Linda. *Games and Economic Behavior*, 68(2), 551–556.
- Chen, Y., & Li, S. X. (2009). Group identity and social preferences. *American Economic Review*, 99(1), 431–457.
- Esser, J. (1998). Alive and well after 25 years: A review of groupthink research. *Organizational behavior and human decision processes*, 73(2/3), 116–141.
- Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2), 171–178.
- Greiner, B. (2004). An online recruitment system for economic experiments. In *Forschung und wissenschaftliches Rechnen. GWDG Bericht 63*, ed by K. Kremer and V. Macho, 79–93. Göttingen: Ges. für Wiss. Datenverarbeitung.
- Hastie, R. (1986). Review essay: Experimental evidence on group accuracy. In G. Owen & B. Grofman (Eds.), *Information pooling and group accuracy* (pp. 129–157). Westport, CT: JAI Press.
- Janis, I. (1972). *Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes*. Mifflin: Houghton.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134.
- Laughlin, P. R., & Ellis, A. L. (1986). Demonstrability and social combination processes on mathematical intellectual tasks. *Journal of Experimental Social Psychology*, 22(3), 177–189.
- Laughlin, P. R., Hatch, E. C., Silver, J. S., & Boh, L. (2006). Groups perform better than the best individuals on letters-to-numbers problems: Effects of group size. *Journal of Personality and Social Psychology*, 90(4), 644–651.
- Laughlin, P. R., Zander, M. L., Knievel, E. M., & Tan, T. K. (2003 October). Groups perform better than the best individuals on letters-to-numbers problems: Informative equations and effective strategies. *Journal of Personality and Social Psychology*, 85(4), 684–694.
- Lightle, J. P., Kagel, J. H., & Arkes, H. R. (2009). Information exchange in group decision making: The hidden profile problem reconsidered. *Management Science*, 55(4), 568–581.
- Long, S. J., & Freese, J. (2006). *Regression models for categorical dependent variables using Stata* (2nd ed.). College Station, Texas: Stata Press.
- Maciejovsky, B., Sutter, M., Budescu, D. V., & Bernau, P. (2013). Teams make you smarter: How exposure to teams improves individual decisions in probability and reasoning tasks. *Management Science*, 59(6), 1255–1270.
- Sausgruber, R., & Tyran, J.-R. (2011). Are we taxing ourselves?: How deliberation and experience shape voting on taxes. *Journal of Public Economics*, 95(1–2), 164–176.
- Sherif, M. (1935). A study of some social factors in perception. *Archives of Psychology*, 187(27), 1–60.