Committees Versus Individuals: An Experimental Analysis of Monetary Policy Decision Making

Clare Lombardelli and James Proudman and James Talbot

8. February 2005

Committees Versus Individuals: An Experimental Analysis of Monetary Policy Decision Making*

Clare Lombardelli, James Proudman, and James Talbot
Bank of England

We report the results of an experimental analysis of monetary policy decision making under uncertainty. A large sample of economics students played a simple monetary policy game, both as individuals and in committees of five players. Our findings—that groups make better decisions than individuals—accord with previous work by Blinder and Morgan. We also attempt to establish why this is so. Some of the improvement is related to the ability of committees to strip out the effect of bad play, but there is a significant additional improvement, which we associate with players learning from each other's interest rate decisions.
JEL Codes: C91, C92, E5

On May 6, 1997, the Monetary Policy Committee (MPC) of the Bank of England was established and granted operational independence in setting short-term interest rates to achieve the government's inflation target of 2.5 percent. This new framework replaced the previous system of a single individual—the Chancellor of the Exchequer—deciding on the appropriate level of U.K. base rates.

*The views expressed are those of the authors and do not necessarily reflect those of the Bank of England. The authors would like to thank Mervyn King for providing the motivation for this paper. We are indebted to the help and assistance of the London School of Economics, and in particular Richard Jackman. We also thank, without implication, Peter Andrews, Kosuke Aoki, Jens Larsen, John Morgan, Simon Price, Tom Sargent, and Jonathan Temple for extensive help, comments, or suggestions. We are also grateful to seminar participants at the Bank of England, LSE Financial Markets Group, and the 2003 RES Conference for comments and suggestions. Chris Williams and Mohammed Sater provided excellent programming support; Edward Dew, Alex Golledge, and Richard Geare gave excellent research assistance. We are also grateful for the comments of four referees. All remaining errors are our own.

Why delegate monetary policy to a committee? A wide-ranging survey undertaken by Fry et al. (1999) found that seventy-nine central banks out of a sample of eighty-eight use some form of committee structure when setting monetary policy. So by weight of numbers, it appears that setting interest rates by committee is thought to be superior. And the intuitive argument that committees make better decisions than individuals—because they allow decision makers to pool judgment—also seems plausible.

With the exception of Gerlach-Kristen (2001) and Sibert (2003), the theoretical economics literature does not have much to say about the consequences of delegating responsibility to a committee. The hypothesis that groups make better monetary policy decisions is difficult to test, due to a lack of comparable empirical data. This problem motivated Blinder and Morgan (2000) to adopt a different approach: carrying out a “laboratory experiment” on a large sample of Princeton University students to test whether groups do indeed make monetary policy decisions differently.

In an experiment, the researcher can isolate the relative performance of individual and group behavior, controlling for differences in the abilities, incentives, and preferences of the decision makers, and of the environment in which they work. The main drawback is that it is artificial—it is not possible to replicate exactly the complexities of real world policymaking in the context of a simple experiment.

Psychologists have studied group behavior for many years. A series of experiments conducted by Stoner (1961) and Myers (1982) suggested that a group should not be expected to replicate the average of the individuals who compose it. Groups often polarize the initial tendency because participants are more likely to generate arguments in favor of the position initially favored by the majority of group members. And because of an innate desire to compare themselves favorably with their peers, individuals may often take increasingly extreme positions in favor of the group proposition, leading to what Stoner called a “risky shift.”

Janis’ (1972) study of U.S. foreign policymaking also found that where groups were highly cohesive and dominated by a strong leader, it was possible to make decisions that the majority did not agree with, especially when operating under time pressure. The best way to avoid this was by a frank and open exchange of views at the beginning of the meeting. Hall’s (1971) experiments also supported this hypothesis, finding that for complex tasks, the best performing groups were often those who were least consensual in the early stages of discussion.

But overall, the evidence suggests that, provided that groups do not reach decisions too easily or too quickly, their performance should be at least as good as the average of their individual members. This hypothesis was supported by the Blinder and Morgan (2000) experiment, where committees made substantially better decisions than individuals, although, contrary to expectations, groups were not more inertial.

This paper describes a new experiment using students from the London School of Economics (LSE). Like Blinder and Morgan (2000), we find that committees make better decisions than individuals, but we try to explore in more detail why this is the case. Our experiment suggests that committees do more than just eliminate the poor decisions of a minority of members, also allowing members to learn by observing the behavior of others. The experiment also explicitly tested whether the ability to exchange information through discussion improved performance; perhaps surprisingly, it did not.

Another question is, why do committee members' votes differ? One possibility could be that participants came to the experiment with different prior beliefs about the nature of the (unknown) model of the economy. So we designed a questionnaire to help establish these "priors," and asked participants to complete the same questionnaire again at the end of the experiment, in order to see how much they had learned about the model they were using.

The rest of the paper is organized as follows. Section 1 describes the economic model used, the structure of the experiment, how the priors questionnaire was used, the information flows and incentives facing players, and the data set. Section 2 discusses the results on learning over time, about both the model and how to play the game. Section 3 describes the scores in the individual and committee games and tries to explore why committee play was superior—including a discussion of why the ability to discuss did not improve committee performance. Section 4 looks at the relationship between activism and committee performance. Section 5 concludes.

1. The Experiment

1.1 The Model

We asked participants to act as monetary policymakers by attempting to "control" a simple macroeconomic model subject to shocks. We used a standard small-scale macro model of the type widely used for policy analysis in modern macroeconomics (see, e.g., Fuhrer and Moore 1995) and similar to that in Blinder and Morgan (2000). Where possible, it is calibrated to match U.K. macroeconomic data (see Bank of England 2000) and is shown in equations (1) and (2) below:

$$y_t - y^* = 0.8(y_{t-1} - y^*) - 0.5(R_t - \pi_t - r^*) + \bar{g} + \eta_t \tag{1}$$

$$\pi_t = 0.7\pi_{t-1} + 0.3\pi_{t-2} + 0.2(y_t - y^*) + \nu_t \tag{2}$$

where y_t is log output; y^* is the natural level of output, arbitrarily calibrated to 5; π_t is inflation; R_t is the nominal interest rate; and r^* is the neutral *real* interest rate (calibrated to 3 percent per annum). In equation (2) the coefficients on lagged inflation sum to one, implying that although a short-run trade-off between output and inflation exists, the Phillips curve is vertical in the long run.

Equation (1) is subject to two types of shock. The first, $\bar{g}$, which can be thought of as a permanent change in the equilibrium real interest rate, takes the value $+/-0.5$ and occurs once, and with equal likelihood, in any one of the first five periods in each round. This type of shock is attractive because it does not affect the inflation-output trade-off and therefore the ability of the score function outlined in equation (4) below to capture participants' performance adequately. Both the second shock in equation (1)—η_t—and ν_t in equation (2) are white noise, corresponding to a random draw from a normal distribution $\sim$N(0, 0.01) in each period.

The monetary authority's decision rule for the short-term interest rate—as decided by the participants of the experiment—closes the model.

1.2 Priors

An intriguing feature of Blinder and Morgan's (2000) results was that committee members frequently disagreed about their decisions, despite having identical loss functions and the same information set. But even without observing such differences in voting—whether experimentally or in real life—it seems entirely plausible that committee members can think differently about how to respond to shocks that are only indirectly observed via the response of the endogenous variables. This may be especially true of a committee where members have diverse backgrounds and specialties.

We posit that the differences of opinion observed in the Blinder and Morgan experiment reflected different subjective judgments about the structure of the (unknown) model. So at the beginning of our experiment, we asked the players to fill in a questionnaire that attempted to reveal their priors (see appendix 2 for details). A set of "correct" answers to this priors questionnaire would yield both the parameters of the underlying model and the structure of the optimal rule.

During the experiment, players should learn about the structure of the economy—just like real world policymakers—by observing the response of inflation and output to changes in interest rates, updating their priors, and changing their perception of the "correct" model accordingly. We attempted to capture the extent of this learning by asking participants to fill in the same questionnaire again at the end, in light of what they had learned during the experiment.

One way to calibrate the extent of learning is to compare both the "before" and "after" questionnaire responses with the parameters of the underlying model and the associated optimal rule. However, in our experiment, the optimal rule for the real interest rate, r_t, does not correspond to a continuous function. So instead, we calculated an approximation to the optimal rule, under full information and the assumption that the scoring function (see equation [4] below) is a linear quadratic.[1] This benchmark rule, derived in appendix 1, is given by

$$r_t = 1.6y_{t-1} + 0.27\pi_{t-1} + 0.115\pi_{t-2} + 2\bar{g}. \tag{3}$$

[1]Svensson and Woodford (2000) demonstrate that, in a backward-looking model, under the assumption of a quadratic loss function, optimal policy under partial information is the same as under full information because of the principal of certainty equivalence. However, equation (3) should only be thought of as an indicative benchmark for optimal policy setting.

1.3 Information Flows and Incentives for Players

Players received a clear mandate at the beginning of the experiment:[2] their objective was to maximize a "score" function that penalized deviations of output and inflation from their target values of 5 percent and 2.5 percent, respectively:

$$Score(t) = 100 - 40|Output(t) - 5| - 40|Inflation(t) - 2.5| \qquad (4)$$

As in Blinder and Morgan (2000), we chose a linear rather than quadratic loss function so that players could easily translate their (average) score into a final payoff. And at the end of the game, the participants were paid in pounds according to the following (known) formula:

$$\text{Payoff} = 10 + \text{Average Score}/10 \qquad (5)$$

where the maximum payoff was £20 for a perfect score and was bounded from below at £10. In practice, most students earned around £15–£16. We also offered top prizes of £100 for the best individual score and £100 to be shared equally among the best committee.[3]

Just like actual policymakers, participants did not know with certainty the exact structure of the economy, but they were told that the representative model was linear, learnable, and broadly characteristic of the U.K. economy. There was also uncertainty about the nature of the shocks hitting the economy. Players were informed that "... a structural change occurs at some point during each game. The key to playing successfully is to identify when the change has occurred and how best to respond to it."

They were also told that the economy was subject to other shocks in each period of play. In Blinder and Morgan (2000), subjects were told the probability laws governing the occurrence of the structural shock, but we believe that our specification makes game play more typical of real-world policymaking, where central bankers are unlikely to face shocks with a known distribution or size. The relative

[2] A copy of the oral and written instructions can be found in the Bank of England working paper version at www.bankofengland.co.uk.

[3] These bonus payments were instigated in order to try to ensure that players had an incentive not to exchange information with future participants outside the laboratory.

sizes of the three shocks were calibrated after testing the model on subjects within the Bank of England.

Some manipulation of equation (3) shows that a positive $\bar{g}$ shock corresponds to a 1 percent increase in the neutral real interest rate to 4 percent, and vice versa for a negative shock. So, for example, if players do not react to an upward shift in r^*, they risk accelerating inflation, and the model can quickly become unstable because of the unit root in inflation built into the Phillips curve. Players must therefore extract the signal from the noise and change their behavior accordingly in order to maximize their score.

1.4 Outline of the Experiment

To analyze the effect of individual versus committee decision making, we structured the experiment so that participants played the game under a number of different decision-making structures. The sequencing of the experiment is shown in table 1.

Table 1. The Structure of the Monetary Policy Experiment

Read instructions sheet		
Complete priors questionnaire		
Practice rounds	No score recorded	
Stage 1 (rounds 1–4)	Played as individuals	
Stage 2 (rounds 5–8)	Played as a group	(i): No discussion (ii): With discussion
Stage 3 (rounds 9–12)	Played as a group	(i): With discussion (ii): No discussion
Stage 4 (rounds 13–16)	Played as individuals	
Complete priors questionnaire again, using knowledge gained from participating in experiment		
Students are paid according to their average score across the four stages		

After entering the laboratory, participants were allocated randomly into groups of five; given a standard, short, oral briefing; and asked to read a set of instructions. Each player then filled in the priors questionnaire and, following that, was given about ten minutes to practice on their own with the actual version of the game used in the experiment before starting to play "for real."

The experiment itself comprised four stages. Each stage consisted of four rounds, with each round containing ten periods of play in which participants had to decide on what interest rate to set. Players were scored according to equation (4), with the overall score for each round being the average across the ten periods.

In the first stage, the participants acted as individual policymakers, playing separate games on separate computers for four rounds. Each experiment began at round 1, period $t = 1$, with inflation and output near the steady-state equilibrium ($y = 5$, $\pi = 2.5$).[4] In each period, inflation and output were observed with a one-period lag, and participants would then decide on the appropriate level of interest rates and enter this into the computer. The game then proceeded to the next period ($t = 2$). The computer displayed output and inflation outturns for period 1, the score for that period, the cumulative score so far, and the previous interest-rate decision(s). The same decision problem was repeated until the game reached $t = 10$. At this point, players were told their average score for round 1, the game was reset, and play continued, being repeated for another three rounds.

In stage 2—beginning in round 5—the group acted as a committee, with each member entering his or her own vote on their computer as before. But this time, in each period, the computer calculated and then implemented the group's median vote—as a proxy for a majority-voting rule[5]—and participants observed this committee decision, as well as the response of output and inflation to it. They also saw the (unattributed) votes of their fellow committee members and the committee score for the period and the round so

[4]The first observation at time $t = 0$ would always be the steady state perturbed by a random shock to each equation of the model.

[5]Not all monetary policy committees make interest rate decisions in this way. But in practical terms, one appealing feature of our rule is its ease of implementation during the "no discussion" committee stages, when an alternative rule, such as unanimity, would be difficult to implement.

far. Again, each round lasted for ten periods, and stage 2 finished at the end of round 8.

The committee phase was played over two stages—stage 2 and 3 as shown in table 1 above—each of which corresponded to a distinct scenario. The order of these two stages was randomized across committees in order to control for learning. Under scenario (i), discussion among members of the committee was not allowed in stage 2. The five players observed the same information in each period—the level of output and inflation of the previous period(s) as well as the history of interest rate decisions and scores—and entered their votes while sitting at separate computers, without talking to their fellow players. Participants were then allowed to discuss their decisions in stage 3, and again, the computer would set the median interest rate of the group.[6] This discussion was not constrained, and in practice could take many forms. Under scenario (ii), stages 2 and 3 were reversed.

Stage 4 (rounds 13–16) served as another control, to ensure that the comparison between individual and committee play was not biased by the fact that participants had had four (or more) individual rounds to learn before entering the committee stage. By returning to individual play at the end of the experiment, it was possible to verify that not all of the improvement in scores during the committee stages (rounds 5–12) was an extension of the learning trend observed in rounds 1–4.

When the experiment was over, participants were asked to fill in the questionnaire again in light of what they had learned about the economy from playing the game.[7] Each experiment lasted between 90 and 150 minutes.

1.5 The Data

The experiment was conducted on ten evenings between November 12 and December 11, 2001. Participation in the experiment was voluntary. The sample was independent of the authors: participants were students at the London School of Economics. A total of 170

[6]Participants were again asked to sit at their own computers to enter their votes: during testing we observed that if the committee gathered around one computer, this created a bias towards the decision of a chairperson who entered the votes.

[7]The wording at the beginning of the priors questionnaire was changed to reflect this focus.

students participated in thirty-four independent experiments,[8] giving a cross-section of thirty-four committees with sixteen time series observations for each. All participating students had taken at least one undergraduate-level economics course.

2. Learning Results

The main focus of the experiment was to provide evidence on the differences between group and individual policymaking; this is discussed in section 3 below. But because the nature of the experiment is one of decision making under uncertainty, we begin by discussing the results on learning about both the model and how to play the game.

2.1 Learning About the Model

Players' answers to the priors questionnaire give some insight into their initial beliefs about the structure of the economy. Participants were also asked to fill in the questionnaire again at the end in light of what they had learned during the experiment, and from this we can judge whether their beliefs had converged on the actual parameters of the underlying model. Summary statistics for the questionnaire responses are shown in table 2.

For all questions apart from question 1 (Q1), the dispersion of responses to the final questionnaire is lower, although this decline is not significant. For each question, more than half of the participants improved in the final questionnaire.

The mean square error (MSE) of responses for each individual question was also calculated with reference to the underlying model and benchmark rule. We can use paired sample tests to examine the changes in players' responses after participating in the experiment. Across all questions, the MSE of responses fell from 0.17 in the initial questionnaire to 0.15 at the end of the experiment, which is significant at the 1 percent level ($t = 3.4$). We can also examine whether this learning is concentrated on certain aspects of the model.

The biggest improvements were in answers to Q2 and Q3. This shows that participants learned most about the lags in the

[8] No student was allowed to play the game more than once.

Table 2. Summary Statistics for the Distribution of Answers to the Priors Questions

Question Number	Correct Answer	Mean Response Before	Mean Response After	MSE of Responses Before	MSE of Responses After	Percentage of Players Who Improve	t-Statistic (Difference in MSE)	Wilcoxon Signed Rank Test
1	0	0.626	0.591	0.426	0.403	59%	0.74	<u>−0.92</u>
2	0.1	0.427	0.298	0.140	0.057	87%	6.74*	<u>−7.16*</u>
3	0.8	0.438	0.569	0.171	0.094	81%	5.38*	<u>−5.59*</u>
4	0.8	0.555	0.550	0.105	0.109	62%	−0.37	−0.22
5	0.5	0.499	0.649	0.037	0.058	53%	−4.43*	−4.13*
6	0.7	0.582	0.586	0.055	0.053	68%	0.33	<u>−0.68</u>
7	0.2	0.485	0.502	0.126	0.129	63%	−0.29	<u>−0.19</u>
8	0	0.490	0.531	0.310	0.320	68%	−0.51	−0.13
Over All Questions	n/a	n/a	n/a	0.170	0.150	n/a	3.4*	<u>−3.63</u>

Note: Starred values indicate statistical significance for a two-tailed test at the 1 percent level. Underlined entries indicate where the Wilcoxon signed rank test is supportive of the hypothesis that the MSE of responses to the questionnaire has fallen during the experiment. Starred values indicate a significant difference (in either direction).

transmission mechanism of monetary policy (Q2) and the weight they should attach to deviations of output from trend in their rule (Q3). The responses to Q5, which considers the responsiveness of output to interest rate changes, are interesting: this is the only question where responses got significantly worse. On average, players started the experiment with surprisingly accurate beliefs about this parameter, but after playing the game they believed that output was more sensitive to interest rate changes than it actually is. There are several reasons why this might be the case: first, the short-run policy multiplier is larger for output than inflation, and second, output responds much more to the structural shock in the short run than inflation.

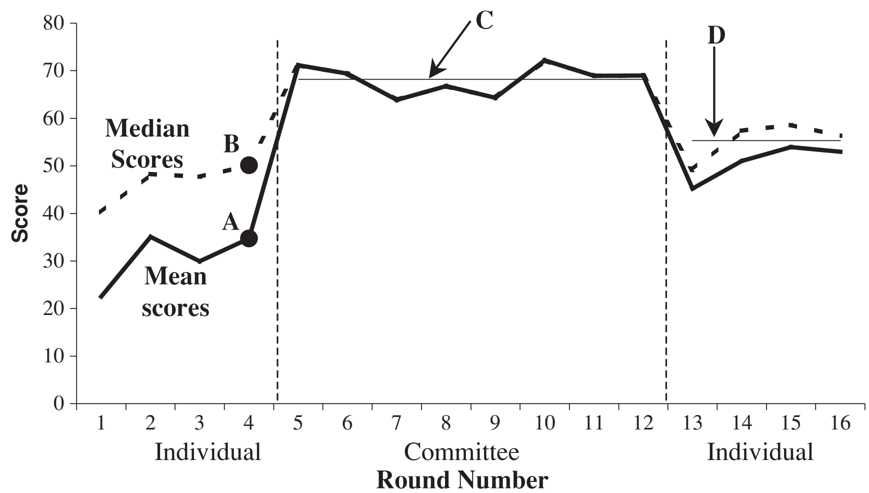
For the other five questions, beliefs changed by much less, and t-tests show that the changes in means are not significant. Participants did not learn much, if anything, about the appropriate degree of interest rate smoothing and the parameters of the model (Q1 and Q4–Q8, respectively). But this may not be too surprising: although participants played the game a number of times, each individual round lasted only ten periods and may have been too short for players to deduce much about the long-run properties of the model. We cross-checked these results with Wilcoxon signed rank tests, which support the conclusions of the t-tests and are also reported in table 2.

Which players learned most? There is a positive and significant (at the 1 percent level, $t = 10.7$) correlation between players whose MSE was largest at the beginning and those who learned most about the parameters of the model. There is also a positive and significant (at the 1 percent level, $t = 3.66$) correlation between the fall in the MSE of responses to the priors questionnaire and performance in rounds 13–16. So those who learned most about the model during the game—whether through their own experience or by playing with others—did better in the final set of individual rounds.

2.2 Learning About Playing the Game

Did players actually get better at playing the game over time? The solid line in figure 1 shows the mean score attained by the 170 individuals in each of rounds 1–4 and 13–16, and the mean score achieved

Figure 1. Scores for Players Over Time



by the thirty-four committees during rounds 5–12.[9] There are three striking features:

1. the significant upward trend in the results over time
2. the large rise in scores when players moved to committee decision making
3. the large downward move in scores when participants returned to playing as individuals

Individuals' scores were higher in round 16 than in round 1: the mean score rose from 23 to 53. This increase is significant at the 1 percent level ($t = 5.12$), providing evidence of a significant learning effect during the game.

Within the individual rounds, there was also some evidence of learning. The average individual score was twelve points higher in round 4 than in round 1 and was eight points higher in round 16 than in round 13. Both differences in mean tests are significant at the 1 percent level for a one-tailed test ($t = 2.38$ and

[9]It is not possible to calculate an individual score for each person in the committee rounds, as the evolution of the round and therefore the score are determined by the decision of the committee as a whole.

$t = 2.61$). These results suggest that it is possible to get better at playing the game, even when not exchanging information with others.[10]

If we rank the players in each committee by their initial performance, we find—perhaps unsurprisingly—that the worst players improved most. Although even the best players in each committee improved somewhat, the improvement in scores was only significant for the worst two players in each committee. This is consistent with the view that some players begin the game with a totally incorrect model in their head, and so their decisions attract a very low score initially relative to others with more accurate priors. As they learn that their priors do not accord with the truth—through playing the game and observing their scores—they update their beliefs and their performance improves accordingly.

These findings contrast with those of Blinder and Morgan (2000), where there was less evidence of learning. One reason for this might be that our model is slightly simpler. For example, participants have to learn fewer parameters in our game: five as compared with seven in the Blinder and Morgan experiment.

3. Groups Versus Individuals

We found strong evidence that decisions made by committees were superior to those made by individuals. The average score during the committee rounds was nearly two-thirds better (68 compared with 41), and significantly higher ($t = 7.4$) than the average for the individual rounds.

We can also use our benchmark rule to calibrate the size of this improvement. The average score from simulating the game under this rule was 85, much higher than the *best* individual player's score (71), but only slightly better than the *best* committee (83). On average, moving from individual decision making to a committee structure closed nearly two-thirds of this "policy gap."

[10]Because we use a "within subject" rather than a "between subject" experimental design, it is difficult to draw firm conclusions about how much learning in rounds 13–16 is related to earlier group feedback. However, the upward trend over the final four individual rounds is consistent with at least *some* learning being related to individual play.

3.1 Explaining the Improvement

How do we explain the improvement in performance of committees over individuals? There are (at least) two distinct, competing hypotheses that can be used to explain why committee decisions are superior:

Hypothesis 1: A committee with majority voting can neutralize the impact of some members playing badly in any given game.

Hypothesis 2: Committees allow members to improve performance by sharing information and learning from each other.

We can use figure 1 to give a visual representation of the contribution of these two hypotheses. We first took the scores of the five players in each committee during the individual rounds and calculated that of the median player (the third best performer in each committee across all the individual rounds). So the dashed line in figure 1 represents the mean of the median players' scores in each of rounds 1–4 and 13–16. The solid line is the mean score across all 170 players in the individual rounds, and the mean score across the thirty-four committees during each of rounds 5–12. Line C is the mean score across rounds 5–12 taken together, and line D is the average of the median players' scores over rounds 13–16 taken together.

The overall improvement in performance—generated by setting interest rates by committee—is taken to be C − A: the difference between the mean individual score in round 4 and the mean committee score across rounds 5–12.

Figure 1 decomposes this improvement into two distinct components. One component is the difference between the score of the mean and median player in the individual rounds (represented by the distance B − A in figure 1), which should be equal to the adverse effect of a minority of poor performers on the mean individual score. This is therefore the extent of improvement under hypothesis 1 described above. This portion of the difference in means is significant at the 1 percent level (t = 3.7), so we cannot reject hypothesis 1. The contribution of hypothesis 2 should be represented by the second component—the residual, or C − B (the portion of the committee improvement not explained by the move to majority voting). This difference is also significant at the 1 percent level ($t = 2.8$), so we cannot reject hypothesis 2 either.

The mean committee score (68) was also higher ($t = 1.51$, significant at the 10 percent level) than that of the best individual (65—calculated as the person with the highest average score across all the individual rounds) in each committee when playing alone. This provides further evidence that committees do more than just replicate the behavior of their best individual.

The significant decline in scores as participants move back to individual play is a striking feature of both our results and those of Blinder and Morgan (2000). By definition, this "residual" component of the committee improvement cannot be associated with learning about the game, because the information set of the players must be at least as great in round 13 as it was before. We argue therefore that at least some of this residual effect stems from the ability of committees to pool judgment, expertise, and skill. The distance C – D in figure 1 represents the difference between the average committee score and that of the median individual player in each committee. At 12.9 points, it is also significant at the 1 percent level ($t = 4.2$). This evidence also supports the view that there is something special about committees, beyond their ability to strip out the effects of bad players.

In order to try to abstract from learning effects in the initial rounds of the game, we can also compare committee scores with those achieved by the first, second, and third best players—as judged by their performance in rounds 13–16 of the game. Though committees do not outperform their best player (as judged by this criterion), scores are significantly higher for all other members (for example, for the second best player, $t = 2.91$, significant at the 1 percent level) when they set monetary policy together. This is a striking result because it suggests that even after players have had repeated opportunities to learn about the game, and from each other, most still perform better when they make collective decisions. Furthermore, given that it may be very difficult in practice to judge ex-ante who would be the best decision maker, this result shows that on balance it is better to make decisions by committee.

Figure 2. Distribution of Scores Across Players

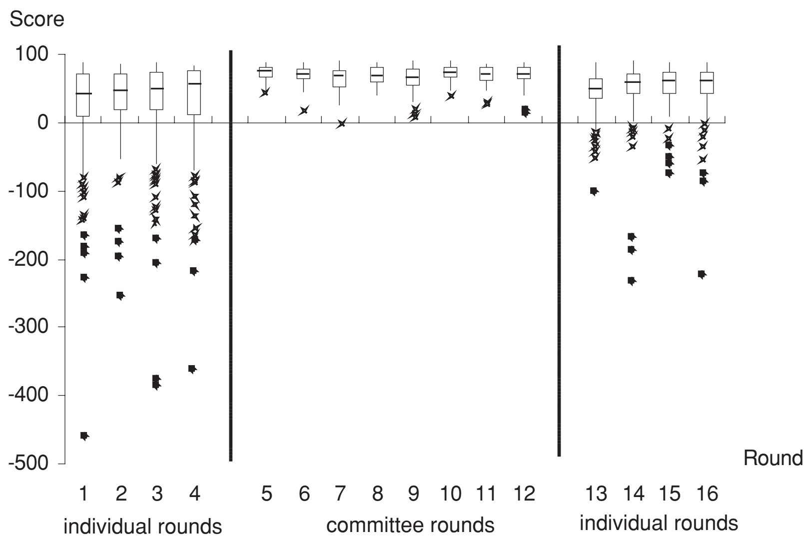


Figure 2 shows a Tukey box-plot[11] of both the scores across the 170 individual players in rounds 1–4 and 13–16 and across the thirty-four committees in rounds 5–8. For the box-plots, the height of the box represents the interquartile range for each round, with the position of the median score shown by the horizontal line inside the box. The lines running vertically from the bottom and top edges of the boxes are "whiskers," which extend to the furthest observation within one and a half interquartile ranges of the bottom and top of the box. Stars mark out "minor outliers" (outside the range of the whiskers), and dots mark "major outliers," which lie greater than three interquartile ranges from the bottom of the box.

Several features are immediately striking. First, the range of scores is much wider for the individual than for the committee rounds, both in terms of the interquartile range of the distribution and the number of outliers. Individual scores are more often negative, and some particularly poor players achieve very low (negative) scores. For the committee games, only one outlier achieves a negative score (of -1.8 in round 7). Second, the individual scores

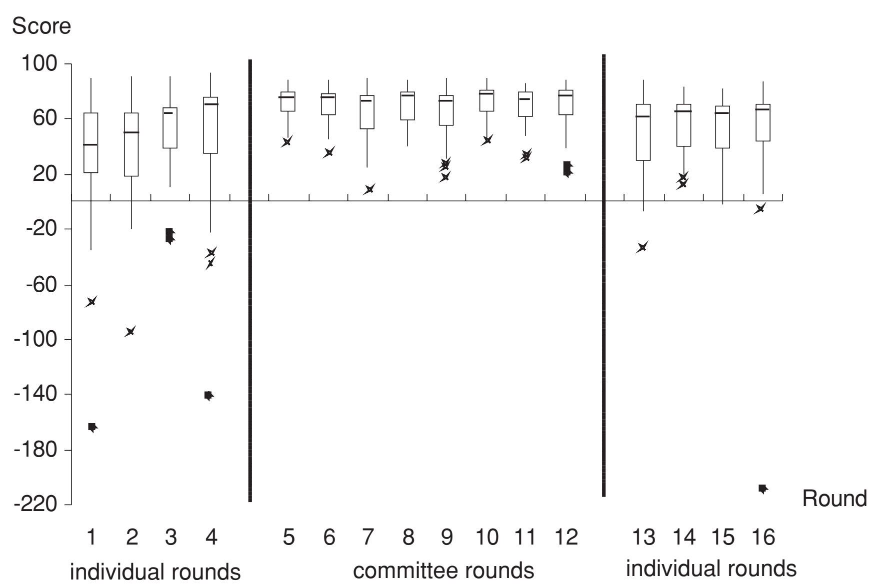
[11]See Tukey (1977).

show a much larger negative skew: the median is much closer to the upper than lower quartile and there are long tails of negative scores.

Comparing the distribution of scores between the first and second individual stages may also give an insight into what players learn from being in a committee. The whole distribution is a lot narrower in stage 4, and although there is only a moderate shift up in the second and third quartiles, the first quartile is much higher, the very long tails in rounds 1–4 are absent, and the number of outliers falls somewhat. This is consistent with the hypothesis suggested above that the worst players learn a lot from being part of a committee.

We can also examine a breakdown of the performance of an "artificial committee" over time, using the median player in each individual round as a proxy for the committee (see figure 3). Again we see a much narrower distribution during the committee rounds—demonstrated by the smaller boxes (the interquartile range in the committee rounds is half the size on average) and shorter whiskers. And although the performance improvements are greatest for the

Figure 3. Distribution of Score for "Artificial Committee"



lower quartile, it is definitely possible to discern an upward shift in the positions of the boxes in the figure during rounds 5–12. The first, second, and third quartiles increase by twenty-eight, sixteen, and ten points, respectively.

3.2 Does Discussion Help?

The experiment also included two different ways of organizing committee decision making: one where participants were allowed to discuss their views and another where no verbal communication was allowed. Although the mean score of committees who were allowed to discuss their decisions (67) was significantly higher ($t = 7.9$, significant at the 1 percent level) than the mean individual score in rounds 13–16 of the experiment (51), perhaps the most surprising result was that the ability to discuss did not significantly improve committee performance.[12] The mean score of committees who did not discuss their decisions was 69, which was actually slightly (but not significantly, $t = 1.39$) higher than those committees who were able to discuss their decisions.

So our committees must have pooled information in other ways. The benefits of different forms of communication are likely to depend on the nature of the game, as well as the individuals taking part. There are many games—for example, snooker or chess—that may be easier to learn by watching, rather than discussion.

Our results are also supportive of those from the social psychology literature discussed above, which suggest that discussion may not always enhance group performance. And although it appears that for this experiment, and for this set of students, discussion did not provide more information than could be acquired by observing others' votes, we cannot tell for sure how much of this result is being driven by the particular specification of our experiment. Real-world policymaking is undoubtedly a much more complex affair, and the benefits of discussing decisions with fellow committee members are likely to be high.

[12]This contrasted with earlier trials on Bank of England staff. Observation of the LSE experiment in progress suggested that the quality of the discussion varied across committees—some would just talk about what interest rate to set, whereas others talked about what the structure of the model might be.

4. Policy Activism and Performance

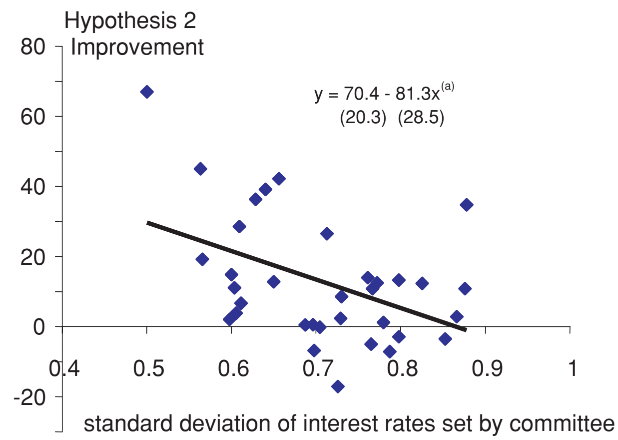
When the model described by equations (1) and (2) was simulated under the benchmark rule (3), interest rate movements were, on average, significantly less activist than those of the individuals ($t = 15.0$, significant at the 1 percent level), but not the committees ($t = -7.08$, significant at the 1 percent level), who played the game. If we calculate interest rate activism as the standard deviation of the interest rate during each ten-period round, 92 percent of individuals were more active than the benchmark rule, whereas only 9 percent of committees were. One interpretation might be that committees allow players to learn the appropriate amount of activism more quickly by pooling information.[13] In other words, they can work out that aggressive movements in interest rates make the economy more difficult to control, and players understand this property of the model more quickly by either exchanging information or watching the votes of others.

Some supporting evidence is provided by figure 4, which shows a significant, negative relationship between hypothesis 2 improvement—that is, the portion of the committee improvement not explained by the move to majority voting—and policy activism (as measured by the standard deviation of committee interest rate decisions over time). However, figure 4 tells us nothing about causation.

A further test would be whether individuals are less active in rounds 13–16 than rounds 1–4. Although the standard deviation of rate changes falls slightly, the change is not significant, so another interpretation might be that committees naturally make more gradual decisions. This is not formally tested in our experiment, but Blinder and Morgan (2000) found that, on average, committees did not take longer to change their views than individuals. This was in direct contrast to what Alan Blinder expected in light of his experience sitting on the U.S. Federal Open Market Committee.

[13] Even if the causation were to run in the opposite direction—that is to say that bad players need to vary interest rates more because the economy is further away from target—it is still the case that good players can learn that this strategy is suboptimal over time.

Figure 4. Hypothesis 2 Improvement and Committee Activism



(a) standard errors in brackets

5. Conclusions and Suggestions for Further Work

In this paper, we have undertaken an experimental analysis of monetary policy decision making by individuals and committees. An experiment like this has obvious limitations: it can only be indicative of what might happen in the real world, where policymaking is a much more complex affair. Because of the need to design a relatively simple experiment, and the focus on testing the proposition that committees are better than individuals, our research has nothing to say on other aspects of optimal committee design such as committee size, composition, or how discussion should be structured.

Our experiment does suggest overwhelmingly that committees performed much better than the average of the individuals who composed them. There is also evidence to suggest that committees perform significantly better than all but their best member, and given that it might be difficult to discern ex-ante who is the best person, on balance it would seem advisable to make decisions by committee.

We argue that, while some of the improvement associated with group play reflects the averaging of errors across members, the ability of committees to allow the pooling of judgment and information (in whatever form) also has a significant role to play in explaining why committees do better. Perhaps surprisingly, committees who were able to discuss their decisions did not perform better than those who could not. In our experiment, participants were able to glean the same amount of information about the game from observing each other's play as from discussion. These results certainly merit further investigation, perhaps using a more complicated or more lifelike setup than ours.

It is also possible to observe some evidence of learning within the experiment. The answers to the priors questionnaire suggest that participants learned a significant amount about certain aspects of the model during the game, although the experiment may not have been long enough for participants to learn much about the structural parameters. In addition, the distribution of scores across players narrowed during the course of the game—with the worst players learning most.

Appendix 1. Derivation of the Benchmark Rule

Assuming that players attempt to maximize their score (S_t) in each period of the game, the decision problem can be written as

$$\underset{r_t}{Max}\ E_{t-1}\{S_t\} \ \text{s.t.}\ (1)\ y_t = 0.8y_{t-1} - 0.5r_t + \bar{g} + \eta_t \quad \text{where } \eta_t \sim N(0, \sigma_\eta^2)$$

$$(2)\ \pi_t = 0.7\pi_{t-1} + 0.3\pi_{t-2} + 0.2y_t + \nu_t \quad \text{where } \nu_t \sim N(0, \sigma_v^2)$$

$$\text{where} \quad (3)\ S_t = 100 - 40|y_t - y^*| - 40|\pi_t - \pi^*|$$

Equations (1) and (2) are written in deviation from equilibrium form, and r_t is the real interest rate.

Approximating (3) as a linear quadratic, we derive the benchmark rule by substituting in the constraints (1) and (2) and differentiating with respect to r_t, to give

$$r_t = 1.6y_{t-1} + 0.27\pi_{t-1} + 0.115\pi_{t-2} + 2\bar{g} \tag{4}$$

Obviously, the distribution of $\bar{g}$ is unknown to participants in the experiment, so (4) is the "certainty equivalence benchmark rule." If our loss function were quadratic, the optimal policy rule under partial information would be the same as its full-information counterpart, according to Svensson and Woodford (2000).

We use this benchmark rule to conduct the simulations in section 3 and also to calibrate the responses to the "before" and "after" priors questionnaires.

Appendix 2. Priors Questionnaire

Players were asked to give numeric responses to the following questions. They could choose any value from $\alpha = 0$ to $\alpha = 1$, with intervals of 0.1, apart from question (2), where the options ranged from zero to ten periods. Although all players would have taken at least one university-level macroeconomics course, it is likely they may not have been familiar with the type of model used for the experiment, or the "correct" calibration of the parameter values in the questions below. Participants were told that if they did not understand the jargon in brackets, they should instead make a guess governed by how much they agreed or disagreed with the statement in the question.[14]

(1) To what extent should monetary policymakers respond cautiously to shocks (i.e., if their interest rate reaction function includes the expression $i_t = \alpha i_{t-1} + \ldots$, what weight should they place on α)?

(2) After how many quarters is the maximum impact of monetary policy on inflation felt?

(3) What relative weight should monetary policymakers place on smoothing output compared with controlling inflation (i.e., if their reaction function includes the expression $i_t = \alpha(y_t - Y)$ $+(1 - \alpha)(\pi_t - \pi^*) + \ldots$, what weight should they place on α)?

[14] A complete copy of the questionnaire can be found in the Bank of England working paper version at www.bankofengland.co.uk.

(4) To what extent are shocks to output persistent (i.e., if the expression for output included the term $y_t = \alpha y_{t-1} + \ldots$, what weight do you think α would take)?

(5) How sensitive is output to changes in interest rates (i.e., if the expression for output included the term $y_t = \alpha i_t + \ldots$, what weight do you think α would take)?

(6) To what extent are shocks to inflation persistent (i.e., if the expression for inflation included the term $\pi_t = \alpha \pi_{t-1} + \ldots$, what weight do you think α would take)?

(7) To what extent is inflation sensitive to deviations of output from trend in the short run (i.e., if the expression for inflation included the term $\pi_t = \alpha(y_{t-1} - Y) + \ldots$, what weight do you think α would take)?

(8) To what extent is inflation sensitive to deviations of output from trend in the long run? Not at all sensitive (i.e., $\alpha = 0$) or highly sensitive (i.e., $\alpha = 1$)?

References

Bank of England. 2000. *Economic Models at the Bank of England.* September 2000 update. London: Bank of England.

Blinder, Alan S., and John Morgan. 2000. "Are Two Heads Better Than One?: An Experimental Analysis of Group vs. Individual Decisionmaking." NBER Working Paper No. 7909 (September).

Fry, Maxwell, DeAnne Julius, Lavan Mahadeva, Sandra Roger, and Gabriel Sterne. 1999. "Key Issues in the Choice of Monetary Policy Framework." In *Monetary Frameworks in a Global Context*, ed. Lavan Mahadeva and Gabriel Sterne. London: Routledge.

Fuhrer, Jeffrey C., and George R. Moore. 1995. "Inflation Persistence." *Quarterly Journal of Economics* 110 (1): 127–59.

Gerlach-Kristen, Petra. 2001. "Monetary Policy Committees and Interest-Rate Setting." Mimeo, University of Basel.

Hall, J. 1971. "Decisions, Decisions, Decisions." *Psychology Today* (November).

Janis, Irving L. 1972. *Victims of Groupthink.* Boston: Houghton Mifflin.

Myers, David G. 1982. "Polarizing Effects of Social Comparison." In *Group Decision-Making*, ed. Hermann Brandstätter, James H. Davis, and Gisela Stocker-Kreichgauer. New York: Academic Press.

Sibert, Anne. 2003. "Monetary Policy Committees: Individual and Collective Reputations." *Review of Economic Studies* 70 (3): 649–65.

Stoner, J. A. F. 1961. "A Comparison of Individual and Group Decisions Involving Risk." Master's thesis, Massachusetts Institute of Technology.

Svensson, Lars E.O., and Michael Woodford. 2000. "Indicator Variables for Optimal Policy." NBER Working Paper No. 7953 (October).

Tukey, John W. 1977. *Exploratory Data Analysis.* Reading, MA: Addison-Wesley.