

# **Committees versus individuals: an experimental analysis of monetary policy decision-making<sup>(1)</sup>**

*Clare Lombardelli, James Proudman and James Talbot,*  
Bank of England, Threadneedle Street, London, EC2R 8AH.<sup>(2)</sup>

Approximate Word Count (including tables etc.): 7800

**We report the results of an experimental analysis of monetary policy decision-making under uncertainty. We used a large sample of economically literate undergraduate and postgraduate students from the London School of Economics to play 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 group decision-making is superior. Our results show that some of the benefit is related to the ability of committees to strip out the effect of bad play in any given period. But there is a significant additional improvement, which we associate with the ability of committee members to share information and learn from each other by observing other members’ interest rate responses. One surprising result is that the superiority of committee decision-making does not appear to be related to the ability to discuss the interest rate decision.**

On 6 May 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%. This new framework replaced the previous system of a single individual – the Chancellor of the Exchequer – deciding on the appropriate level of UK base rates.

Why delegate monetary policy to a committee? The academic argument for central bank independence is well established (see, for example, Barro and Gordon (1983)). And in practice, there is strong evidence from across the world to suggest that committees are the preferred arrangement for setting monetary policy by central banks. For instance, a wide-ranging survey undertaken by Fry, Julius, Roger, Mahadeva and Sterne (1999) finds that 79 central banks out of a sample of 88 use some form of committee structure when setting monetary policy. By weight of

---

<sup>(1)</sup> 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 especially 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, Charlie Bean, Andrew Benito, Peter Chapman, Edward Dew, Phil Evans, Alex Golledge, Charles Goodhart, Jan Groen, Andrew Hauser, Jens Larsen, Lavan Mahadeva, Chris Mann, Frederic Mishkin, John Morgan, Katherine Neiss, Kalin Nikolov, Simon Price, Tom Sargent, Mohammed Sater, Gabriel Sterne, Jan Vlieghe and Chris Williams for helpful comments and suggestions. We are also grateful for the comments of three referees. All remaining errors are our own.

<sup>(2)</sup> E-mail: [clare.lombardelli@bankofengland.co.uk](mailto:clare.lombardelli@bankofengland.co.uk), [james.proudman@bankofengland.co.uk](mailto:james.proudman@bankofengland.co.uk), [james.talbot@bankofengland.co.uk](mailto:james.talbot@bankofengland.co.uk)

numbers, it appears to be accepted that setting interest rates by committee is superior. And the intuitive argument that committees make better decisions than individuals – because they allow decision-makers to pool judgement – also seems plausible.

With the exception of Gerlach-Kristen (2001), the theoretical economics literature has less 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 behaviour, 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 policy making in the context of a simple experiment.

Although experimental techniques are relatively new to monetary economics, they are well established in other branches of economics such as asset pricing, game theory and decision making under uncertainty.<sup>(3)</sup> In addition, psychologists have studied group behaviour for many years, and a series of experiments – for example, Hall (1971), Janis (1972) and Myers (1982) - have shown that group decisions are rarely equal to the sum of their parts. Group performance depends on the nature of the interaction between members and the task in hand, but the consensus view seems to be that for complex tasks, decisions taken by committee should be at least as good as the average of the individuals that comprise it.

This hypothesis was supported by the results of Blinder and Morgan (2000). In their experiment, groups made substantially better decisions on average than individuals. And, just as in real life, there were also disagreements between committee members over interest-rate decisions. But, contrary to their expectations, groups did not make decisions more slowly than individuals.

Examining whether groups make better decisions than individuals is the main focus of this paper. It describes a new experiment with students from the London School of Economics, which explored in more detail why groups are superior. One explanation is that majority voting helps to eliminate the poor decisions of a minority of members. But this experiment provided evidence that committees do more than just this, allowing members to pool information and learn more about the game they are playing.

---

<sup>(3)</sup> See Davis and Holt (1993) and Kagel and Roth (1995) for excellent surveys.

The experiment also explicitly tested whether the ability to exchange information through discussion improved performance. Such a finding would not be surprising if players held different prior views about the nature of the (unknown) model of the economy. So the experiment tried to examine such differences of opinion by means of a questionnaire designed to help establish the beliefs of players about the economy.

Section 1 describes the economic model used and the structure of the experiment; section 2 discusses the results; section 3 analyses them using a panel data framework; and in section 4 we conclude by trying to draw some inferences from our work for the design of monetary frameworks in the real world.

## 1. The experiment

### (i) *The model*

We asked participants to act as monetary policy makers 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 for example, Fuhrer and Moore (1995)). Where possible, it is calibrated to match UK macroeconomic data (see Bank of England (1999, 2001)) and is shown in equations (1) and (2) below:

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

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

Where  $y_t$  is log output,  $y^*$  is the natural rate of output,<sup>(4)</sup>  $\pi_t$  is inflation,  $R_t$  is the nominal interest rate and  $r^*$  is the neutral real interest rate (calibrated to 3% per annum).  $\bar{g}$  is a permanent shock,  $\eta_t$  and  $\nu_t$  are shocks corresponding to a random draw from a normal distribution  $\sim N(0, 0.01)$  in each period.

Equation (1) is an ‘IS curve’. The current output gap ( $y_t - y^*$ ) is a function of its one-period lag, and the deviation of the real interest rate from its neutral level in the current period ( $R_t - \pi_t - r^*$ ). The IS curve is also subject to two types of shock. The first,  $\bar{g}$ , is a permanent shock which occurs at random, and with equal likelihood, during one of the first five periods in each round; and the second –  $\eta_t$  – is white noise.

The structural shock  $\bar{g}$  takes the value +/- 0.5. It can be thought of as a permanent change in the equilibrium real interest rate. 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 adequately capture participants’ performance.

---

<sup>(4)</sup> In the model, this is arbitrarily calibrated to 5.

Equation (2) is a ‘Phillips Curve’. Inflation is a function of lagged values of itself and the current output gap. 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. The shock  $v_t$  is white noise.

The monetary authority’s decision rule for the short-term interest rate – as decided by the participants of the experiment – closes the model. It is possible to calculate the optimal rule under full information.<sup>(5)</sup> This is approximated by:

$$R_t = 1.6y_{t-1} + 0.27\pi_{t-1} + 0.115\pi_{t-2} + 2\bar{g} \quad (3)$$

In a backward-looking model of the type described in equations (1) and (2) above, optimal policy under partial information is the same as its full-information counterpart (see the introduction of Svensson and Woodford (2000)).<sup>(6)</sup> So the performance of this ‘optimal rule’ provides a useful benchmark against which to compare individual and group results. And we can attempt to assess whether the behaviour of players approaches this rule over time.

## (ii) *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 in their model. And this should be especially true of a committee where members have diverse backgrounds and specialities.

We posit that the differences of opinion observed in the Blinder and Morgan experiment reflected different subjective judgements about the structure of the model. So at the beginning of our experiment, players filled in a questionnaire that attempted to reveal their prior knowledge of the economy.<sup>(7)</sup> A set of ‘correct’ answers to this ‘priors’ questionnaire’ would yield the parameters of the model in question and therefore the structure of the optimal rule.

During the experiment, players should learn about the structure of the economy – just like real world policy-makers – 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

---

<sup>(5)</sup> In our experiment, the optimal rule does not correspond to a continuous function. So in order to derive equation (3), we approximate the scoring function as a linear quadratic, see Appendix 1.

<sup>(6)</sup> Again, under the assumption of a quadratic loss function.

<sup>(7)</sup> See Appendix 2.

attempted to capture this learning by asking participants to fill in the same questionnaire again at the end of the experiment.

### (iii) *Information flows and incentives for players*

To make the decision-problem of the players as similar to that of real-life policy-makers within the confines of a simple experiment, we also control carefully for their incentives and the information they receive.

Players received a clear mandate at the beginning of the experiment:<sup>(8)</sup> their objective was to maximise a ‘score’ function which penalised deviations of output and inflation from their target values of 5 and 2.5% respectively:

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

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:

$$Payoff = 10 + Average\ Score/10 \quad (5)$$

Where the maximum payoff was £20 for a perfect score, and was bounded from below at £10. In practice, students earned around £15-£16. We also offered top prizes of £100 for the best individual score the best committee.<sup>(9)</sup>

Just like actual policy-makers, participants in our experiment made decisions in an uncertain world, while observing only the evolution of the endogenous variables over time. 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 characterised the structure of the UK 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’*

And they were told that the economy was subject to other shocks in each period. This differs from Blinder and Morgan (2000), where subjects were told the probability laws governing the occurrence

---

<sup>(8)</sup> A copy of the oral and written instructions is available from the authors on request.

of the structural shock. We believe that our specification makes game play more typical of real-world policy making, where central bankers are unlikely to face shocks with a known distribution or size. The relative sizes of the three shocks were calibrated after testing the model on subjects within the Bank.

Some manipulation of equation (3) shows that a positive  $\bar{g}$  shock corresponds to a 1% increase in the neutral real interest rate to 4%, 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 behaviour accordingly in order to maximise their score.

**(iv) Outline of the experiment**

To analyse the effect of individual versus committee decision-making discussed above, we structured the experiment so that participants played the game under a number of different decision-making structures. The sequencing of the experiment can be summarised as follows:

---

**Table A**

**The structure of the monetary policy experiment**

---

Read instructions sheet		
Fill in ‘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	
Fill in ‘Priors Questionnaire’		
Students are paid according to their average score across the four stages		

---

After entering the laboratory, participants were allocated into groups of five. They were given a standard, short, oral briefing and were asked to read a set of instructions. Each player was asked to fill in the ‘priors’ questionnaire’ as a way of gauging his or her prior beliefs about the model. They were given about ten minutes to practise on their own with the actual version of the game used in the experiment before starting to play ‘for real’.

---

<sup>(9)</sup> These ‘bonus’ payments were instigated in order to try to ensure that players had an incentive not to exchange

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 in response to combinations of unobserved shocks. Players were scored according to equation (4), and the overall score for each round was taken as the average across the ten periods. It is these overall scores that we use in the analysis of Section 2.

In the first stage, the participants acted as individual policy makers, playing separate games on separate computers for four rounds. Beginning with round 1, the game started in period  $t = 1$  with inflation and output near the steady-state equilibrium ( $y = 5$ ,  $\pi = 2.5$ ).<sup>(10)</sup> In each round, inflation and output were observed with a one-period lag, so after viewing the level of output and inflation in period  $t = 0$ , players decided on the appropriate level of the interest rate for period 1 and entered this into the computer. The game then proceeded to the next period ( $t = 2$ ). The computer displayed output and inflation outturns for period 1, along with the score for that round and the interest rate decision. 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 a further 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 selected, and then set, the median vote for the group – as a proxy for a majority-voting rule – and participants observed this committee decision, as well as the response of output and inflation. They also saw the (unattributed) votes of their fellow committee members and overall score for the period and the round so far. Again each round lasted for ten periods. Stage 2 finished in round 8.

The committee phase was played in two stages – stage 2 and 3 in Table A above – each of which corresponded to a distinct scenario. The order of these two stages was randomised across committees in order to control for learning. Under scenario (i), discussion among members of the group 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 rates and scores – and entered their votes while sitting at separate computers, without talking with fellow players. In scenario (ii), participants were allowed to discuss their decisions in stage 2, and again, the computer would set the median interest rate of the group.<sup>(11)</sup> This discussion was not constrained in any way, and in practice could take many forms.

---

information with future participants outside the laboratory.

<sup>(10)</sup> The first observation at time  $t = 0$  would always be the steady state perturbed by a random shock to each equation of the model.

<sup>(11)</sup> 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.

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 the improvement in scores during the committee stages (rounds 5-12) was not just an extension of the learning trend observed in rounds 1-4.<sup>(12)</sup>

#### **(v) *The data***

The experiment was conducted on ten evenings between 12 November and 11 December 2001 at the London School of Economics. Participation in the experiment was voluntary, and the sample of students was entirely independent from the Bank of England. For the main experiment described above, 170 students participated in 34 independent experiments,<sup>(13)</sup> giving a cross-section of 34 committees with 16 time series observations for each. A further 15 students were used to play an alternative version of the experiment described in Section 2(iii) below. All participating students had taken at least one undergraduate-level economics course.

## **2. Results**

The main focus of the experiment was to provide evidence on the differences between group and individual policy making; and this is discussed in Section (ii) below. But because the nature of the experiment is one of decision-making under uncertainty with learning, we begin by discussing what we ‘learned about learning’ in Section (i). Section (iii) deals with the difference between discussion and no discussion committee scenarios.

#### **(i) *Learning***

This section is divided into two parts: evidence on learning about the structure of the model, and then on learning how to play the game. The two are inextricably linked.

##### **(i.i) *Priors***

Players’ answers to the ‘priors questionnaire’ give some insight into their initial beliefs about the structure of the economy. As noted in Section 1(ii), a set of ‘correct’ answers will reveal the key features of the model and the associated optimal rule. Participants also filled in the same questionnaire again at the end of the experiment; and from this we can judge whether their beliefs

---

<sup>(12)</sup> A better control for learning might have been to compare the results of a game where half the participants were randomised to a ‘committee scenario’ and the other half played only as individuals. We did in fact design such a purely ‘individual’ version of the game, but the students who participated were unwilling to play it, preferring instead to play the committee version.

<sup>(13)</sup> No student was allowed to play the game more than once.



converged on the actual parameters of the underlying model. One test of learning is the extent of convergence over the course of the game.

All answers to the questionnaire are in numeric form, allowing the calculation of the mean square error (MSE) of responses across questions. Overall, this statistic decreased from 0.17 in the initial questionnaire, to 0.15 at the end of the experiment, and this is significant ( $t = 3.4$ ). The standard deviation of responses to the questionnaire also narrowed significantly from 1.59 to 1.45 ( $t = 3.5$ ).

We can decompose this improvement further by computing the change in MSE for individual questions between the initial and final questionnaires. We find that participants learnt most about the lags in the transmission mechanism of monetary policy (Q2) and the weight they should attach to deviations of output from trend in their 'rule' (Q3). There was a significant decline in the MSE of their responses to both questions. Participants did less well at working out the appropriate degree of interest rate smoothing and the parameters of the model (Q1) and (Q4-8) respectively. Especially the long-run neutrality property of the model (Q8) and the impact of interest rate changes on output. The experiment may have been too short to learn much about these aspects.

#### **(i.ii) *Playing the game***

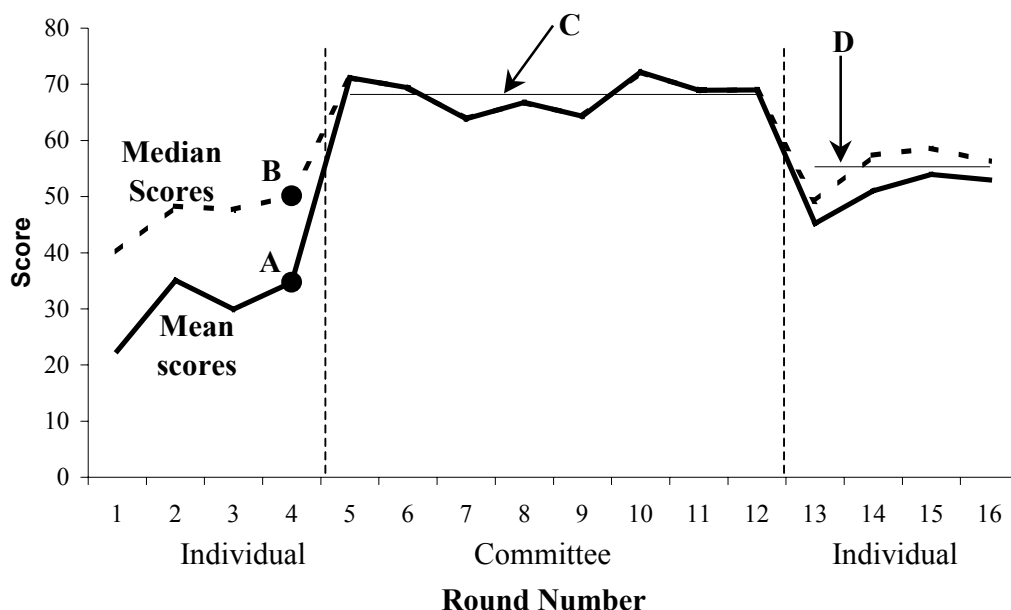
The results of the priors' questionnaire provide tentative evidence of learning about certain aspects of the model and the nature of the optimal rule, but did players actually get better at playing the game over time?

The solid line in chart 1 shows the mean scores attained by the 34 committees over time. This is broken down into the first set of individual play (rounds 1-4), committee play (rounds 5-12) and then individual play for a second time (rounds 13-16). For the individual rounds, the 'committee' score is taken to be the mean of the scores across the five individuals playing separately. For the committee rounds, this statistic is the mean score that each committee decision attracts.

There are three striking features of the data:

- (1) The significant upward trend in the results over time;
- (2) the large rise in scores when players moved to committee decision-making in Game 5; and
- (3) the large downward move in scores when participants returned to playing as individuals in Game 13.

**Chart 1: Scores for players over time**



Individuals' scores were higher in round 16 than in round 1, rising from a group average of 23 to 53. This increase is extremely significant ( $t = 5.12$ ), providing evidence of a significant learning effect during the game.

Within the individual rounds, there was strong evidence of learning. The average of the scores across each group was twelve points higher in round 4 than in round 1 and was eight points higher in round 16 than round 13. Both difference in mean tests are significant at the 1% level for a one-tailed test. These results suggest that learning occurred during the game, regardless of whether individuals were allowed to exchange information with others.

If we rank the players in each committee by their initial performance we find – perhaps unsurprisingly – that the worst players learnt most. This is consistent with the view that some players begin the game with the completely wrong 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. There is for example a significant positive correlation between the amount by which the worst players improve over the game, and their initial errors about the model when answering the questionnaire.

On average, the best players in each committee improved their scores during the experiment. This suggests that the improvement over time was not merely the result of the worst players learning from their better counterparts. These findings contrast with 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 Blinder and Morgan.

## (ii) *Groups vs individuals*

We found strong evidence that decisions taken by committees were superior to those of 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 the optimal rule under full information to calibrate the size of this improvement. The average score from simulating the game under this optimal rule is 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 the 'policy gap'.

How do we explain this improvement in committee performance? There are (at least) two distinct, competing hypotheses that can be used to explain why committee decisions are superior to those of the individuals that comprise it. We refer to these as Hypotheses 1 and 2:

**Hypothesis 1: A committee with 'majority' voting can neutralise 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 Chart 1 to give a visual representation of the contribution of these two hypotheses. The dashed line represents the average – over the 34 independent groups of five players – of the median player's score. The solid line is simply the mean score across all players in each committee.<sup>(14)</sup> Line C is the mean score over all the committee rounds and line D is the mean score over rounds 13-16 for the median players in each of those rounds. The overall improvement in performance – generated by setting interest rates by committee – is therefore measured as the distance between C and A: the difference between the average score in the final individual round and the committee rounds.

The chart decomposes this improvement into two distinct components. The difference between the score of the mean and median player in the individual rounds (represented by the distance B-A in Chart 1) 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. And this portion of the difference in means is significant ( $t = 3.7$ ). So we can not reject Hypothesis 1.

---

<sup>(14)</sup> Again, note that the mean score in the committee rounds is the score of the committee's interest rate decision.

The contribution of Hypothesis 2 should be represented by the residual, C-B (the portion of the committee improvement not explained by the move to majority voting). This difference is also significant ( $t = 2.8$ ), so we cannot reject Hypothesis 2 either.

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 over time, because the information set of the players must be at least as great in round 13 than it was before. We argue therefore that this residual effect stems from the ability of committees to pool judgement, expertise and skill. This is represented by the distance C-D in Chart 1 (12.9) and is also significant ( $t = 4.2$ ). In other words, there is ‘something special’ about committees beyond their ability to aid learning and to strip out the effects of ‘bad’ players.

The mean committee score (68) was also higher than that of the best individual (65) in each committee when playing alone ( $t = 1.51$ , significant at the 10% level). This provides further evidence that committees do more than just replicate the behaviour of their best individual.

### **(iii) *Discussion vs. No Discussion***

The experiment also included two different ways of organising committee decision-making: one where participants were allowed to discuss their views and another where no verbal communication was allowed. Perhaps the most surprising result was that the ability to discuss did not significantly improve committee performance.

This result was in contrast to earlier trials on Bank staff. So, in addition to the main experiment described above, a further small sample of students was asked to play a different version as a robustness check. This variant was designed so as to raise the implicit benefit of discussion: committee members were told - with a lag of up to two periods - that a shock had occurred, and the length of this information lag was allowed to vary across players. The ability to discuss was therefore more valuable because committee members with more timely information could share this with others more quickly by verbal communication. The average score for discussion was higher than non-discussion committees in this version of the game, although the small sample size – three committees – meant that the significance of this improvement could not be tested.

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. And it appears that for the main version of the game, and for this set of

students, discussion did not provide more information than could be acquired by observing others' votes.

There is also some evidence from the psychology literature that discussion may not always enhance group performance. The idea of 'group polarisation' – as proposed by Myers (1982) – suggests that discussion tends to polarise any initial tendency within the group. This is because people have an innate desire to compare themselves favourably with each other, and so take increasingly extreme positions in favour of the initial group proposition. One way around this problem is to ensure that a frank and open exchange of views takes place at the beginning of the discussion – as outlined in an earlier study by Hall (1971) who showed that groups who established a common consensus quickly were often less effective.

### 3. A Panel Data Approach

Can we bring together the stylised facts described above in an econometric model? A panel-data framework allows us to model the cross-sectional behaviour of committees over time as a function of a group of common variables. We use a 'fixed effects' approach. This captures the unobservable features of each committee – such as the innate ability of participants to play the game – within an intercept term which is allowed to exhibit cross-sectional variation.<sup>(15)</sup>

The 34 committees, each playing the game 16 times, group the data. This gives a fully rectangular panel of 544 observations. The time series dimension divides naturally into three different stages: the first set of individual rounds (numbered 1 to 4); the rounds played as a committee (5-12); and the final set of individual rounds (13-16). The individual rounds generated five independent observations for each period. For these, the 'committee observation' is taken to be the mean score of the five players, allowing us to model the evolution of the average performance of committee members over time. But this has implications for the structure of the matrix of regressors: for example, when we consider how actively interest rates are changed, the variable enters the regression separately for rounds played as individuals and rounds played by committee. In the committee stages, the standard deviation of the policy rate set by each committee in each game is included. For the individual games the mean of the standard deviations for the five individual players is used.

Estimation is by OLS, using a general-to-specific modelling strategy. The results for four candidate models given by equations (a) - (d) below are shown in Table B (t-statistics in brackets).

$$y_{it} = \alpha_i + \gamma_1 \ln(t)^{1-16} + \beta_3 s_{it} + \beta_4 \sigma_{it}^I + \beta_5 \sigma_{it}^C + u_{it} \quad (\mathbf{a})$$

---

<sup>(15)</sup> We reject the null hypothesis that the data should be modelled as a 'random effects' model (ie that there was no correlation between the individual effects and the independent variables). The Hausman test statistic is  $\chi^2 = 37.81$ , significant at the 0.01% level.

$$y_{it} = \alpha_i + \beta_1 \partial_1 + \beta_2 \partial_2 + \gamma_1 \ln(t)^{1-16} + \beta_3 s_{it} + \beta_4 \sigma_{it}^I + \beta_5 \sigma_{it}^C + u_{it} \quad (\mathbf{b})$$

$$y_{it} = \alpha_i + \beta_1 \partial_1 + \beta_2 \partial_2 + \gamma_2 \ln(t)^{1-4} + \gamma_3 \ln(t)^{5-12} + \gamma_4 \ln(t)^{13-16} + \beta_3 s_{it} + \beta_4 \sigma_{it}^I + \beta_5 \sigma_{it}^C + u_{it} \quad (\mathbf{c})$$

$$y_{it} = \alpha_i + \beta_1 \partial_1 + \beta_2 \partial_2 + \gamma_2 \ln(t)^{1-4} + \gamma_4 \ln(t)^{13-16} + \beta_3 s_{it} + \beta_4 \sigma_{it}^I + \beta_5 \sigma_{it}^C + u_{it} \quad (\mathbf{d})$$

**Table B: Panel data estimation results**

Coefficient on:	Model (a)	Model (b)	Model (c)	Model (d)
Committee dummy: rounds 5-12 ( $\beta_1$ )		25.89 (3.60)	39.08 (5.23)	36.49 (5.25)
Individual dummy: rounds 13-16 ( $\beta_2$ )		13.37 (2.58)	20.77 (4.95)	20.77 (4.96)
Period of structural shock ( $\beta_3$ )	3.49 (4.53)	2.51 (3.03)	2.34 (2.83)	2.42 (2.95)
Standard deviation of individual round decisions ( $\beta_4$ )	-29.69 (-9.24)	-18.02 (-3.24)	-18.25 (-3.29)	-18.23 (-3.28)
Standard deviation of committee round decisions ( $\beta_5$ )	-33.53 (-8.12)	-38.14 (-8.79)	-39.06 (-9.02)	-38.61 (-8.97)
Time trend for rounds 1 to 16 ( $\gamma_1$ )	22.33 (8.74)	8.66 (1.52)		
Time trend for rounds 1 to 4 ( $\gamma_2$ )			15.74 (2.19)	15.74 (2.19)
Time trend for rounds 5 to 12 ( $\gamma_3$ )			-3.82 (-0.94)	
Time trend for rounds 13 to 16 ( $\gamma_4$ )			14.76 (2.06)	14.76 (2.06)
Log likelihood	-2360.7	-2353.3	-2349.2	-2349.7
Akaike information criterion	-2398.7	-2393.3	-2391.2	-2390.7
Schwarz-Bayesian information criterion	-2412.7	-2408.0	-2406.6	-2405.7
R <sup>2</sup>	0.3595	0.3743	0.3797	0.3795

For models (a) to (d) ‘ $i$ ’ indexes committees  $i=1 \dots 34$  and ‘ $t$ ’ indexes the round number,  $t=1, \dots 16$ . So  $y_{it}$  is the score of committee ‘ $i$ ’ in round  $t$ ;  $\alpha_i$  is the constant for committee ‘ $i$ ’ (the unobserved, committee-specific, fixed effect); and the error  $u_{it}$  is assumed to be independent and identically distributed  $N(0, \sigma_u^2)$  and independent of the matrix of regressors.

To capture how the scores change over time we consider different time trends and dummy variables.  $\ln(t)^{i-j}$  is the time trend from round  $i$  to round  $j$ ,  $\partial_1$  is a dummy variable – representing an intercept shift – for the committee rounds (5 to 12) and  $\partial_2$  is for rounds 13 to 16. In practice the best fit is

achieved by fitting a separate logarithmic trend through the two sets of individual rounds<sup>(16)</sup>. The regressor set consists of a range of variables that attempt to capture both the structure of the game and the characteristics of policy-making.  $s_{it}$  indicates in which period of the round the structural shock occurs.<sup>(17)</sup> And  $\sigma_{it}^I$  and  $\sigma_{it}^C$  are the standard deviation of the interest rates set across the ten periods of each game played by individuals and committees respectively. A dummy variable for the discussion committee rounds is not significant across all models, consistent with the result that the ability to discuss does not have a significant effect on committee performance.

Model **(d)** is our preferred model. The coefficients on both dummy variables  $\partial_1$  and  $\partial_2$  are positive and we reject the null hypothesis that  $\beta_1$  equals  $\beta_2$  in favour of the alternative that  $\beta_1$  is greater (consistent with the evidence presented in Chart 3 of a positive, and significant, committee effect on mean scores).  $\beta_3$  is positive and significant: the later in the game the structural change occurred, the higher the score. Because of the unit root built into the Phillips curve, a shock occurring early in the game can lead to a substantial divergence in inflation and output from target, making the economy more difficult to control over the remainder of the round; particularly if it takes some time for the player to recognise that such a shock has occurred. Both  $\beta_4$  and  $\beta_5$  are negative: excessive interest rate movements during the game are associated with lower scores for both individuals and committees.

The panel data analysis reinforces the general conclusions from the rest of the paper: that committees were significantly better than individuals, and players' performance improves over time. But we also find evidence that higher interest rate activism – as measured by the standard deviation of the interest rate in each ten-period round – was associated with lower scores for both individuals and committees.

When the model described by equations **(1)** and **(2)** was simulated under the optimal rule **(3)**, interest rate movements were, on average, significantly less activist than those of the individuals, but not the committees, who played the game. Taken together with the results from the panel data analysis, one interpretation might be that committees allow players to learn the appropriate amount of activism more quickly by pooling information. 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 sub-optimal over time. 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 exchanging information.

---

<sup>(16)</sup> Using Akaike and Schwarz information criterion to select between models.

<sup>(17)</sup> Since players face shocks at different times when playing individually, this variable is only included for the games played as a committee.

#### **4. Conclusions**

In this paper we have undertaken an experimental analysis of monetary policy decision-making by individuals and committees. Our experiment suggests overwhelmingly that committees performed much better than the average of the individuals who compose them. And there is also evidence to suggest that committee performance was, on average, better than the performance of the best individual.

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 judgement and information (in whatever form) means that a group can be more than just the sum of its parts. And we present evidence to support the view that this pooling function has a significant role to play in explaining committee improvement. Perhaps surprisingly, committees that were able to discuss their decisions did not perform better than those who cannot. Our hypothesis is that – for this particular version of the game, and this set of students – participants were able to glean the same amount of information about the game from observing each other's play and therefore did not derive much extra benefit from discussion. In the real world, policy-making is undoubtedly a more complex affair, and the exchange of information and ideas is likely to be crucial for optimal monetary policy setting.

It is also possible to observe some evidence of learning within the experiment. The answers to the priors' questionnaire suggest that participants learnt a significant amount about certain aspects of the model during the game. And although only the worst two players in each committee demonstrated significant learning over time, even the best players improved somewhat. The econometric analysis concurs with the conclusion that committees gave a significant boost to the performance of the individuals that comprise them. And it also suggests that one way in which committees were able to do better in our experiment was by making less activist interest rate decisions over time.



## References

**Bank of England (1999)**, *Economic models at the Bank of England*.

**Bank of England (2000)**, *Economic models at the Bank of England*, September 2000 update.

**Barro, R J and Gordon, D B (1983)**, 'Rules, discretion and reputation in a model of monetary policy', *Journal of Monetary Economics*, Vol. 12, No. 1, pages 101-21.

**Blinder, A S and Morgan, J (2000)**, 'Are two heads better than one: an experimental analysis of group vs individual decision making', *NBER Working Paper*, No. 7909, September.

**Davis, D D and Holt, C A (1993)**, *Experimental economics*, Princeton University Press.

**Fry, M, Julius, D, Mahadeva, L, Roger, S and Sterne, G (1999)**, 'Key issues in the choice of monetary policy framework', in Mahadeva, L and Sterne, G (eds), *Monetary frameworks in a global context*, Routledge.

**Fuhrer, J C and Moore, G R (1995)**, 'Inflation persistence', *Quarterly Journal of Economics*, Vol. 110, No. 1, pages 127-59.

**Gerlach-Kristen, P (2001)**, 'Monetary policy committees and interest-rate setting', *mimeo*, University of Basel.

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

**Janis, I L (1972)**, *Victims of groupthink*, Boston: Houghton Mifflin.

**Kagel, J H and Roth, A E (1995)**, *The handbook of experimental economics*, Princeton University Press.

**Myers, D G (1982)**, 'Polarizing effects of social comparison', in Bransdatter, H, Davis, J H and Stocker-Kreichgauer, G (eds), *Group decision-making*, New York: Academic Press.

**Svensson, L E O and Woodford, M (2000)**, 'Indicator variables for optimal policy', *NBER Working Paper*, No. 7953, October.

## Appendix 1: Derivation and uses of the optimal rule for the monetary policy experiment

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

$$\begin{aligned} \text{Max}_{r_t} E_{t-1} \{S_t\} \quad \text{s.t.} \quad & \text{(1) } y_t = 0.8y_{t-1} - 0.5r_t + \bar{g} + \eta_t \quad \text{where } \eta_t \sim N(0, \sigma_\eta^2) \\ & \text{(2) } \pi_t = 0.7\pi_{t-1} + 0.3\pi_{t-2} + 0.2y_t + v_t \quad \text{where } v_t \sim N(0, \sigma_v^2) \end{aligned}$$

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

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

$$r_t = 1.6y_{t-1} + 0.27\pi_{t-1} + 0.115\pi_{t-2} + 2\bar{g} \quad \text{(4)}$$

Obviously, the distribution of  $\bar{g}$  is unknown to participants in the experiment, so (4) is the ‘certainty equivalence optimal rule’. Svensson and Woodford (2000) note that – under the assumption that the loss function is quadratic – the optimal policy rule under partial information is the same as its full-information counterpart. We use this optimal rule to conduct the simulations in Section 4(ii) and also to calibrate the correct responses to the priors’ questionnaire.

## 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 0 to 10 periods.

- 1) To *what extent should monetary policy makers respond cautiously to shocks* (ie if their interest rate reaction function includes the following expression  $i_t = \alpha i_{t-1} + \dots$ , what weight should they place on  $\alpha$ )?
- 2) After *how many quarters is the maximum impact of monetary policy on inflation felt*?
- 3) What *relative weight should monetary policy makers place on smoothing output compared with controlling inflation* (ie if their reaction function includes the following expression  $i_t = \alpha(y_t - Y) + (1 - \alpha)(\pi_t - \pi^*) + \dots$ , what weight should they place on  $\alpha$ )?
- 4) To *what extent are shocks to output persistent* (ie if the expression for output included the following term  $y_t = \alpha y_{t-1} + \dots$ , what weight do you think  $\alpha$  would take)?

- 5) How *sensitive is output to changes in interest rates* (ie if the expression for output included the following term  $y_t = \alpha i_t + \dots$ , what weight do you think  $\alpha$  would take)?
- 6) To *what extent are shocks to inflation persistent* (ie if the expression for inflation included the following term  $\pi_t = \alpha \pi_{t-1} + \dots$ , what weight do you think  $\alpha$  would take)?
- 7) To *what extent is inflation sensitive to deviations of output from trend in the short run* (ie if the expression for inflation included the following term  $\pi_t = \alpha(y_{t-1} - Y) + \dots$ , what weight do you think  $\alpha$  would take)?
- 8) To *what extent is inflation sensitive to deviations of output from trend in the long run*? Not at all sensitive (ie  $\alpha=0$ ) or highly sensitive (ie  $\alpha=1$ )?