Committees versus individuals: an experimental analysis of monetary policy decision-making

Clare Lombardelli
James Proudman
and
James Talbot

Working Paper no. 165

E-mail: clare.lombardelli@bankofengland.co.uk
james.proudman@bankofengland.co.uk
james.talbot@bankofengland.co.uk

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, in making the experiment possible. Special thanks go to Chris Williams who wrote the visual basic code used for the model, Mohammed Sater for programming support, and Edward Dew for excellent research assistance. We also thank, without implication, Peter Andrews, Kosuke Aoki, Charlie Bean, Andrew Benito, Peter Chapman, 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, Gabriel Sterne and Jan Vlieghe for helpful comments and suggestions. We are also grateful for the comments of three referees. Finally, we would like to thank many other members of the Bank of England staff for help in trialling the experiment, and – most of all – the LSE students for taking part. All remaining errors are our own.

Copies of working papers may be obtained from Publications Group, Bank of England, Threadneedle Street, London, EC2R 8AH; telephone 020 7601 4030, fax 020 7601 3298, e-mail mapublications@bankofengland.co.uk

Working papers are also available at www.bankofengland.co.uk/wp/index.html

The Bank of England's working paper series is externally refereed.

ISSN 1368-5562

Contents

Abstract

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.

Key words: Monetary policy, experimental economics, central banking, uncertainty.

JEL classification: C91, C92, E5.

Summary

Evidence from around the world suggests that the majority of central banks take monetary policy decisions by committee rather than through a single individual. Despite this observation, there is little direct empirical or theoretical evidence on the relative merits of monetary policy decision-making by committees versus individuals. Recent work by Blinder and Morgan has sought to shed light on this question by taking an experimental approach, the main result being that the decisions of committees were superior to those of individuals. Although the results of our paper support this conclusion, we attempt to extend their work by examining and testing several hypotheses as to why this improvement might come about.

To this end, we asked a large sample of economically literate undergraduate and postgraduate students from the London School of Economics to play a simple monetary policy game. Participants acted as monetary policy makers, setting interest rates to 'control' a simple macroeconomic model calibrated to match UK data and subject to an unknown combination of shocks. Each participant acted as both individual decision-maker and as part of a committee of five players. All players faced an identical incentive structure: performance was judged according to a score function that penalises deviations of output and inflation from their target values; and they were paid according to their performance.

Just like actual policy-makers, participants in our experiment were forced to make decisions in an uncertain world, while observing only the evolution of the endogenous variables over time. As in real life, these monetary policy makers did not know with certainty the exact structure of the economy they were attempting to analyse. To the extent that players came to the experiment with different prior beliefs about the structure of the model, they may have responded differently to the same set of shocks. So we modelled these differences of opinion by asking participants to fill in a questionnaire that attempted to reveal these prior beliefs. By asking players to fill in the same questionnaire at the end of the game we were able to discern some evidence of players learning about the underlying model of the economy over time. And for the 'worst' players, their improvement in scores over time was positively and significantly related to the extent of their learning about the underlying model.

Like Blinder and Morgan, we find that committees performed significantly better than the individuals who composed them. There are several competing hypotheses as to why. Our results suggest two reasons why committees make better decisions. First, collective decision-making appears to give more weight to the better and less weight to the worse committee members – as judged by their scores when playing the game as individuals – than would be implied by taking the mean of their individual performance. Second, we find evidence that committees do more than this, enabling all members to improve their performance by sharing information and learning from each other. For example, the performance of the committee was on average better than that of its 'best' policy-maker when playing alone.

In our experiment, we also explicitly tested whether the ability to discuss a decision drives the observed improvement in performance. In practice, this did not appear to be the case: in our simple monetary policy game, participants were able to share enough information by simply observing each other's behaviour. But we were able to illustrate how the relative importance of different types of communication depends upon the nature of the decision problem in a variant of the game in which we slightly altered the structure so as to raise the relative importance of discussion. When we did so, committees that discuss performed better.

1. Introduction

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 (eg Barro and Gordon (1983)). And in practice, there is strong evidence from across the world to suggest that committees are the preferred vehicle for setting monetary policy by central banks. For instance, a wide-ranging survey undertaken by Fry *et al* (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 numbers, it seems that it is accepted that setting interest rates by committee is superior. The intuitive argument that committees make better decisions than individuals – because they allow decision-makers to share information and pool judgment – also seems plausible.

With the exception of Gerlach-Kristen (2001), the theoretical economics literature has less to say about the consequences of delegating interest-rate setting responsibility to a committee. The hypothesis that groups make better monetary policy decisions is also 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 groups of Princeton students to test whether groups do indeed make monetary policy decisions differently.

A laboratory experiment is one way to overcome the problem of a lack of observable empirical data when testing a hypothesis. 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-maker, and of the environment in which they work. The main drawback is that it is artificial – we cannot hope to exactly replicate the complexities of real-world policy making in the context of a simple experiment.

Blinder and Morgan (2000) uncovered (at least) two striking results. First, and contrary to their expectations, groups did not make decisions more slowly than individuals. Second, they found that groups made substantially better decisions on average than individuals. The second result is the main focus of this paper. We first examine its robustness to a different game and a different sample; and then explore in more detail the reasons behind it. There are two competing hypotheses. One is that pooling information and judgment among the group does help committees to make better decisions. The other is that majority voting helps to eliminate poor decisions of a minority of members.

In our experiment, we used a large sample of economically literate[1] undergraduate and postgraduate students at the London School of Economics to play a simple monetary policy game both as individuals and in committees of five players. We find that committees make substantially better monetary policy decisions than individual decision-makers. When individuals come together as a committee, they close nearly two-thirds of the gap between the score associated with the optimal policy rule and the average they score as individuals.

We found evidence that some of the improvement associated with committees is due to their ability to offset (through majority voting) the impact of a minority of poor performers. But we also found strong evidence that committees do more than just this, allowing members to pool information and – through communicating with each other – learn more about the game they are playing.

Somewhat counter-intuitively, we found that, in the main variant of the game, the ability of individuals to discuss their ideas aloud did not improve the performance of a committee. We argue that the benefits of different forms of communication depend on the nature of the game as well as the individuals. In some games – for example, snooker or chess – it may be easier to learn by watching someone else's play, while in others, it is easiest to learn by talking about it to others. We illustrate how the relative importance of different types of communication depends upon the nature of the decision-problem in a variant of the game in which we slightly alter the structure so as to raise the relative importance of discussion. When we do so, committees that do discuss do better.

The structure of the paper is as follows. In Section 2, we discuss the related literature. In Section 3 we describe the economic model we use and the structure of the experiment. In Section 4 we discuss our results and in Section 5 we conclude.

2. Related literature

The idea that monetary policy committees generate better outcomes for complex decision-problems because they increase the flow of information between members is closely linked to the debate on monetary policy under uncertainty. In the absence of uncertainty – over both the nature of shocks hitting the economy and its underlying structure – and given a well-defined loss function, monetary policy would be essentially mechanical.

What causes difficulty is the pervasive uncertainty faced by policy-makers, and understanding the implications of this is one of the growth fields of theoretical monetary economics. But how does it affect the operation of optimal monetary policy in practice? Using a simple model, Brainard (1967) showed that 'uncertain monetary policy makers should do less.' And although this result

[1] We are extremely grateful to all the students who took part in the experiment, and to the staff at the LSE who assisted us in running the experiment, in particular Richard Jackman, Gill Wedlake and Paul Jackson. As in Blinder and Morgan (2000), all participants had taken at least one undergraduate-level economics course.

was derived under quite restrictive assumptions, this general rule appears to have become ingrained in modern central bank thought, see Blinder (1998). More recently, Aoki (1998, 2000), Sargent (1999), Svensson and Woodford (2001, 2002) among others have explored the implications of uncertainty for optimal policy.

Theoreticians have also begun to explore how learning about the underlying structure of the economy affects policy decisions. But learning dynamics are inherently complex – see Evans and Honkapohja (2001) for a comprehensive summary. Wieland (2000) outlines an optimal learning strategy within the context of a policy-maker learning about a model with two unknown parameters – considerably simpler than the model in our experiment. In this context, there is a trade-off between control and estimation.

So how much do policy-makers learn by setting policy? Do they learn more or less in a committee? The main objective of this paper is to answer these types of questions. There is little empirical evidence on monetary policy under uncertainty, and even less – with the exception of Blinder and Morgan (2000) – on how this is affected by group dynamics.

Psychologists however have long studied the impact of groups on individual behaviour, and this literature may contain some important lessons for monetary policy makers. To the extent that setting interest rates contains a technical aspect that can be thought of as solving an optimal control problem, a committee can benefit by selecting from a pool of different analyses. Under this scenario, performance should be equal to that of the best individual: in other words, the one who can solve the (essentially mathematical) problem. But we argue that the monetary policy decision-making process is more complicated and subtle than this. When setting interest rates, decisions are not usually black or white. And an exchange of both subjective and objective views in a committee setting – what we characterise as members pooling information and learning from each other – means that, on average, we might expect committees to outperform their best individual.

So it may be possible for committee performance to be better than that of its best individual. But how can this achieved in practice? Hall (1971) conducted a series of laboratory experiments where people were asked to solve a complicated problem. He noticed that the best-performing groups were those which were least consensual in the early stages of discussion – exploring all possible avenues and ideas. This idea of exchanging ideas and arguing about key points could be one reason why groups might be more than just the sum of their parts. Hall's experiment also showed that groups who establish a common consensus quickly were often much less effective.

Other studies have shown that too much agreement can be detrimental: a classic example being Janis' (1972) study of US foreign policy. Janis' work suggests that highly cohesive groups, isolated from outside influence, dominated by a strong leader and attempting to take decisions under time pressure, may be prone to an extreme form of group polarisation called 'groupthink'.[2] The best way to avoid this is for the chairman to encourage freedom of expression, and for disagreements among committee members to be fully aired and discussed.

The idea of group polarisation was proposed by Myers (1982) and is now well founded in the social psychology literature. It suggests that we should not expect groups to replicate the average performance of the individuals that compose them. This phenomenon was first uncovered in a series of experiments by Stoner (1961), who suggested that groups encouraged increased risk taking, which he labelled the 'risky shift' phenomenon. But later work – eg Myers (1982) – suggested that discussion in a group just tends to polarise the initial tendency, whatever that may be.

There are two main explanations for this polarisation. The first is 'persuasive arguments': group discussion generates arguments in favour of the position favoured by the majority of group members. And as more arguments stack up in support of the proposition, reactions become even more extreme in favour of it. The second is 'social comparison theory' whereby polarisation occurs because of people's innate desire to compare themselves favourably with each other. This encourages individuals to take increasingly extreme positions during the course of a discussion in favour of the group proposition. This is done so as to show support for the consensus and to distinguish themselves favourably from the arguments that have gone before.

So the evidence from the social psychology literature seems to suggest that complex decisions taken by committee should be at least as good as the average of those individuals that comprise it. And if group synergy can be achieved – perhaps through a frank and open exchange of views – performance may even exceed that of the best individual. But for committees who discuss their decisions, 'group polarisation' may cause the decision to veer away from the optimum, and in our experiment we explicitly test whether discussion affects committee performance for better or worse.

Because there is little empirical evidence on how uncertainty affects monetary policy decision-making, an experimental analysis – like that of Blinder and Morgan (2000) – is attractive. Although this approach is relatively new to monetary economics, it is well established in other branches of economics: asset pricing, game theory and decision-making under uncertainty for example; see Davis and Holt (1993) and Kagel and Roth (1995) for excellent surveys. And in order to ensure best experimental practice, economists have developed a series of 'protocols' – see Friedman and Sunder (1994) for a good summary – which we have adhered to in our experiment as far as possible (see Section 3 below).

An experimental approach has both advantages and disadvantages. The first advantage is replicability – we could, for example, design an experiment to be identical to that of Blinder and Morgan in order to independently verify their results, just as any other researcher could replicate our work. The second chief benefit is control, or the capacity to manipulate the experiment so as

[2] Janis' (1972) classic example of this was the US government's decision to invade Cuba in 1961.

to exactly evaluate alternative theories and hypotheses. In fact, one of the main drawbacks of trying to evaluate our hypothesis using actual data would be this lack of control – eg for other features of monetary policy design across countries. Our experiment allows us to isolate the exact benefit of groups over individuals by controlling for all other factors within a strictly controlled laboratory environment.

Our paper also aims to extend the work of Blinder and Morgan in (at least) four ways:

(i) We explicitly test several alternative hypotheses as to why committees might make better decisions than individuals. And we study what forms of communication matter – concluding that in our simple experiment committee members could learn as much from observing each other's votes as they could from discussing their decisions.

(ii) We use panel data techniques to study which features of committee behaviour may be associated with their improvement in scores.

(iii) We try to explicitly model the differences of opinion among committee members by means of a questionnaire designed to help establish players' priors about the (unknown) model they are using.

(iv) By asking players to fill in the priors' questionnaire again at the end of the experiment, we can also judge how much they learned about the underlying model during the experiment, in addition to the implicit improvement seen in their improvement in scores over time.

3. 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 that is 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:

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

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

Where y_t is log output, y^* is the natural rate of output,[3] π_t is inflation, R_t is the nominal interest rate and r^* is the neutral <u>real</u> interest rate (calibrated to 3% per annum). $\bar{g}$ is a permanent shock, η_t and ν_t are shocks corresponding to a random draw from a normal distribution ~ N(0, 0.01) in each period. The main difference between our model and a standard new Keynesian model is the

[3] In the model, this is arbitrarily calibrated to 5.

absence of forward-looking expectations (which we were unable to build into our analysis given the nature of the game).

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 - π_t - r*). The IS curve is also subject to two types of shock. The first, $\overline{g}$, 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 – η_t – is white noise.

The structural shock $\overline{g}$ takes the value +/- 0.5. As in Blinder and Morgan (2000) this shock can be though 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.

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, reflecting the fact that although a short-run trade-off between output and inflation exists, the Phillips curve is vertical in the long run. The shock ν_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.(4) This is approximated by:

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

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)).(5) 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

(4) 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. We describe the derivation of this optimal rule in more detail in Appendix 1.
(5) Again, under the assumption of a quadratic loss function.

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 prior beliefs 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.[6] 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 attempted to capture this learning by asking participants to fill in the same questionnaire again at the end of the experiment.

Given that the specification of the optimal rule depends crucially on the parameters of the model, subjective judgment about its structure can also differ. This provides a rationale for differences in responses to shocks across committee members.

(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: 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:

$$Score(t) = 100 - 40\left|Output(t) - 5\right| - 40\left|Inflation(t) - 2.5\right| \qquad \textbf{(4)}$$

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:

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

[6] See Appendix 4 for a copy of the questionnaire.

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 a top prize of £100 for the best individual score and another of £100 for the best committee.(7)

Just like actual policy-makers, participants in our experiment were forced to make decisions in an uncertain world, while observing only the evolution of the endogenous variables over time. As in real life, the participants did not know with certainty the exact structure of the economy they were attempting to analyse. 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 is slightly different than Blinder and Morgan (2000), where subjects were told the probability laws governing the occurrence 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.

Some manipulation of equation **(3)** shows that a positive $\overline{g}$ 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. But because players may come to the model with different prior beliefs about the nature of the structural change, it can be rational for their responses to be different. As noted above, if players update their beliefs with information obtained while playing the game, these differences should diminish over time.

(7) These 'bonus' payments were instigated in order to try to ensure that, wherever possible, players had an incentive not to exchange information with future participants outside the laboratory. At first sight, it might seem that they could encourage risk-taking behaviour. But in fact they do not affect the structure of the optimal rule and therefore the optimal interest-rate setting strategy: the maximum probability of winning the prize still occurs when a player sets policy based on her 'best guess' of the parameters of the model. This is because during the individual rounds, each participant plays an independent game (the shocks are a random draw from a normal distribution and the structural shock can occur in different periods – which makes each player's observation of the evolution of the endogenous variables different). So under this scenario there is no incentive to deviate from the optimal strategy as outlined above. In a situation where all players observe exactly the same information set in each game, it is possible that a 'tournament' incentive might be created by the prize structure of our game. Under this 'common information' scenario, players may wish to deviate from their perceived optimal strategy in order to differentiate themselves from other participants so as to stand a better chance of winning this individual 'prize'. But in our experiment, the independence of the individual games should rule out such 'risk-taking' behaviour.

The relative sizes of the three shocks were calibrated after testing the model on subjects within the Bank. We found that this specification of shocks made the game not too easy and not too difficult for most participants.

(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 (see Appendix 2 for both). 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'.

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 4.

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 ($y = 5$, $\pi = 2.5$).[8] In each round, inflation and output were observed with a one-period lag, so after viewing the level of output and

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

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.[9] This discussion was not constrained in any way, and in practice could take many forms. Sometimes groups would discuss their views on the underlying properties of the model and its shocks before taking decisions, and on other occasions they immediately proceeded to discussion of the interest rate which was to be set. Whichever committee scenario had been played in stage 2, the other was played in stage 3 (rounds 9-12).

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.[10]

---

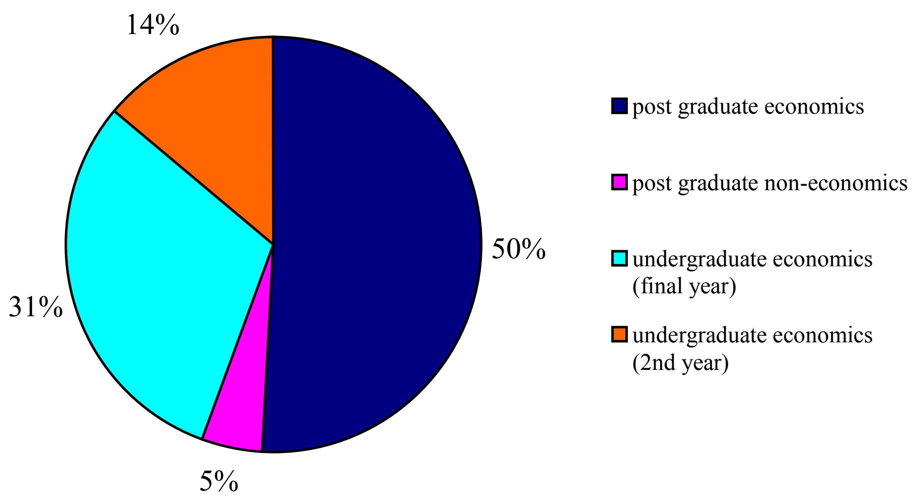
[9] But participants were again asked to sit at their own computers to enter their votes: during testing we observed that gathering round one computer to set rates created an artificial bias towards a committee structure where one person would act as chairperson by entering all the votes. This sometimes appeared to influence group decisions, limiting the spread of votes across committee members and giving undue influence to the chairperson.

[10] 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.

(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,[(11)] 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 4(iii) below.

Chart 1: Breakdown of players by course studied



The game was (in principle) open to postgraduate as well as 2nd and 3rd year undergraduate students studying an economics-related discipline. Chart 1 above shows a breakdown of the 185 participants by course studies: half of students were postgraduate economists. And although a small minority (5%) was not currently studying an economics-related discipline, all students had taken at least one undergraduate-level economics course.

4. 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). Sections (iii)-(iv) deal with what aspects of committee play appear to be associated with an improvement over the individual rounds, and Section (v) reports the results of a panel data analysis of the data.

(11) No student was allowed to play the game more than once. As noted above, in order to discourage students from passing information on to future participants, we offered a £100 cash prize for the best individual performance, and a further £100 for the best group. Importantly, there was also no evidence of scores improving over successive days.

(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*

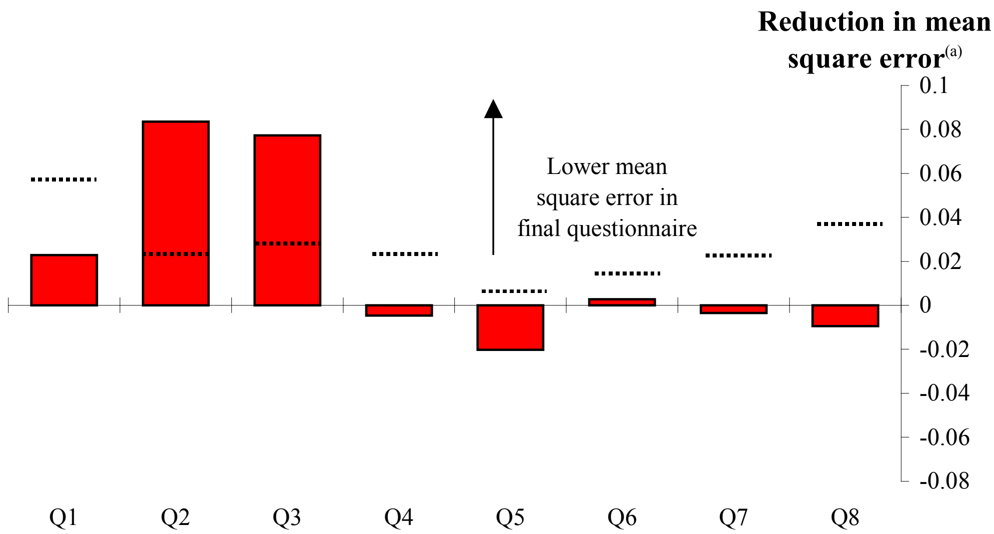
Players' answers to the 'priors questionnaire' give some insight into their initial beliefs about the structure of the economy. As noted in Section 3(ii), a set of 'correct' answers will reveal the key features of the model and the associated optimal rule (see Appendix 1 for a fuller discussion). Participants also filled in the same questionnaire again at the end of the experiment; and from this we can judge whether their beliefs converged on the actual parameters of the underlying model. One test of learning is the extent of convergence over the course of the game.(12)

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).

But we can decompose this improvement further? Chart 2 shows the change in MSE for individual questions: the dashed lines represent the reduction in error required for a significant improvement in response to each of the questions. This implies 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). The change in response to the other questions was more mixed. Participants did less well at working out the parameters of the model (Q4-8) – particularly how much impact interest rate changes had on output (Q5) and the long-run impact of output on inflation (Q8). But each game may be too short to learn much about these aspects – especially the long-run neutrality property of the model. There was also a fall in the MSE of responses to the interest-rate smoothing question (Q1), but this is not significant.

(12) We note that players were not explicitly paid according to their performance in filling in the priors' questionnaire. But they were asked to fill it in 'honestly', and observation suggested that they took it seriously.

Chart 2: Reduction in mean square error of responses between the initial and the final questionnaire



(a) dotted lines indicate significance at the 5% level

(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?

Chart 3 below shows a summary of the mean scores attained by the 34 committees over time.[13] 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:

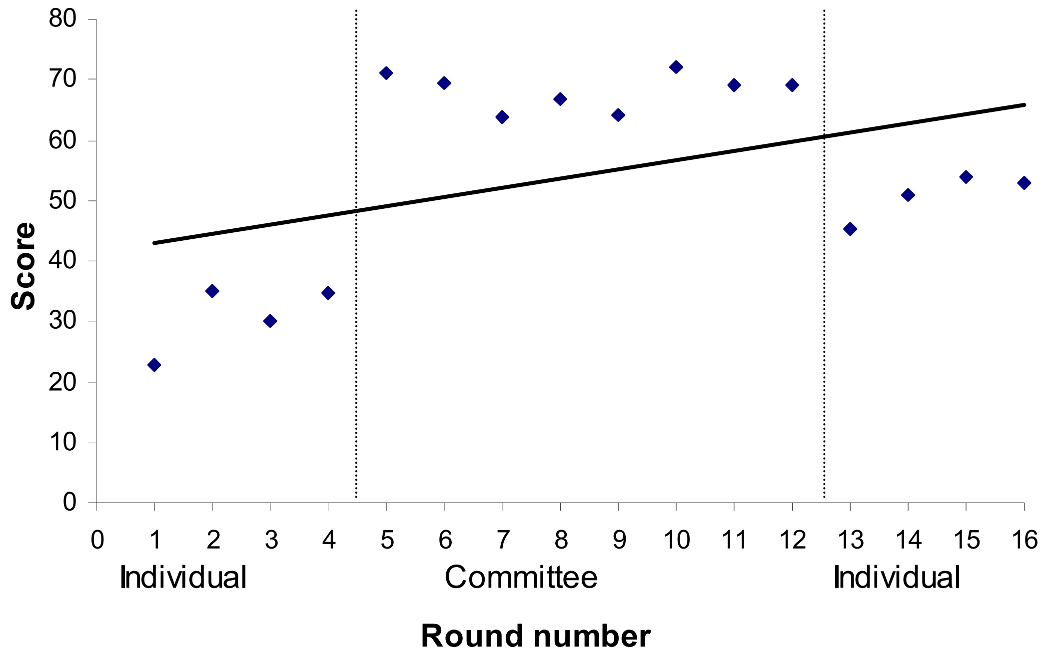
(1) The significant upward trend in the results over time;

(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.

[13] In this section, we are interested in learning over time, so we aggregate the results of the committee rounds, rather than partitioning them by 'discussion' or 'no discussion'. This distinction is dealt with in Section 4(iii) below where we discuss how committees communicate with each other.

Chart 3: Average committee scores over time



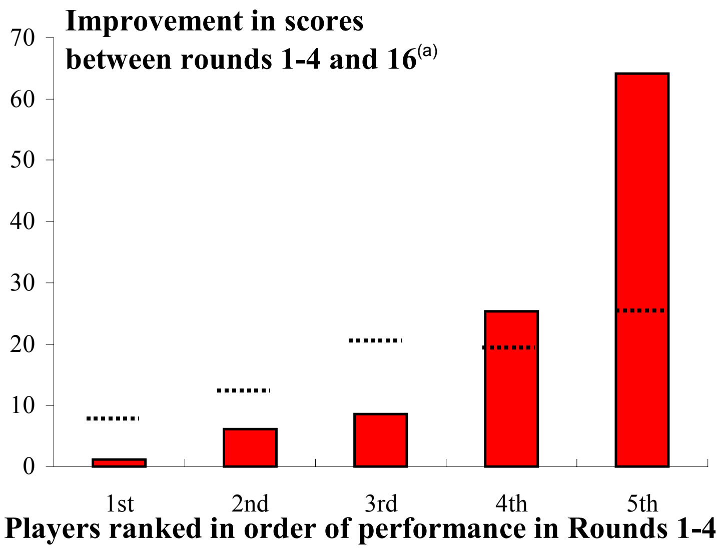
Individuals' scores were higher in round 16 than in round 1, rising from a group average of 23 to 53. This increase is extremely significant (t = 5.12), providing evidence of a significant learning effect during the game.

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.

The standard deviation of scores in any given round more than halved during the game from 76 in round 1 to 35 in round 16. This suggests that the worst players learnt most about the game: those who performed poorly in the first rounds got disproportionately better.

But it is not just the worst players who learnt during the course of the experiment. Chart 4 shows evidence that if we rank the players in each committee by their initial performance, even the best players improved their scores somewhat by the end; although only the worst two players made a significant improvement (again the dashed lines represent significance levels). This suggests that the improvement over time – measured as the score in the final round minus the average score in the first four rounds – was not merely the result of the worst players learning from their better counterparts. It also contrasts 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.

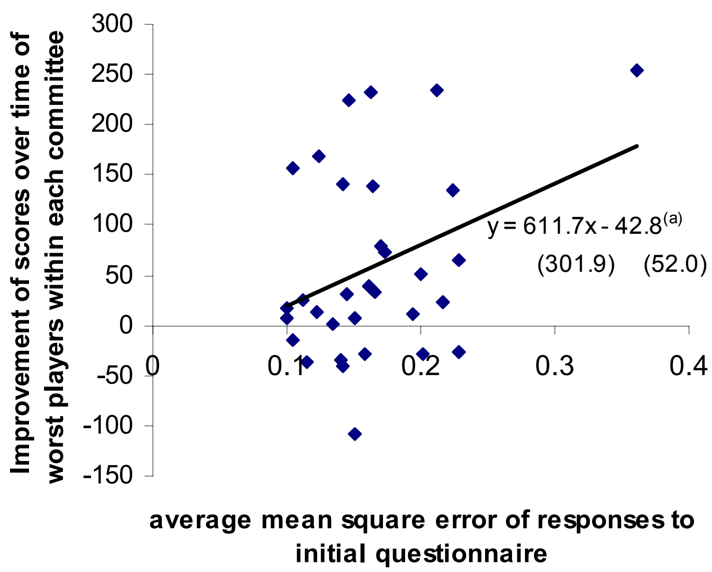
Chart 4: Improvement in scores for players ranked by initial performance



(a) dotted lines indicate significance at the 5% level

As we would expect, the worst players learnt most. This is consistent with the view that some players may 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. And Chart 5 provides evidence that the extent of this learning was positively, and significantly, related to the errors in their responses to the priors' questionnaire for the worst committee members.

Chart 5: Improvement of the worst player per committee and the MSE of their responses to the initial questionnaire



(a) standard errors in brackets

(ii) *Groups vs individuals*

We found strong evidence that decisions taken by committees were superior to those of individuals. The average committee score was nearly two-thirds better (68 compared with 41). A difference in means test between scores in the committee rounds (rounds 5-12 in Chart 3 above), and the individual rounds (1-4 and 13-16 in Chart 3) shows that committee scores were significantly higher ($t = 7.4$).

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:

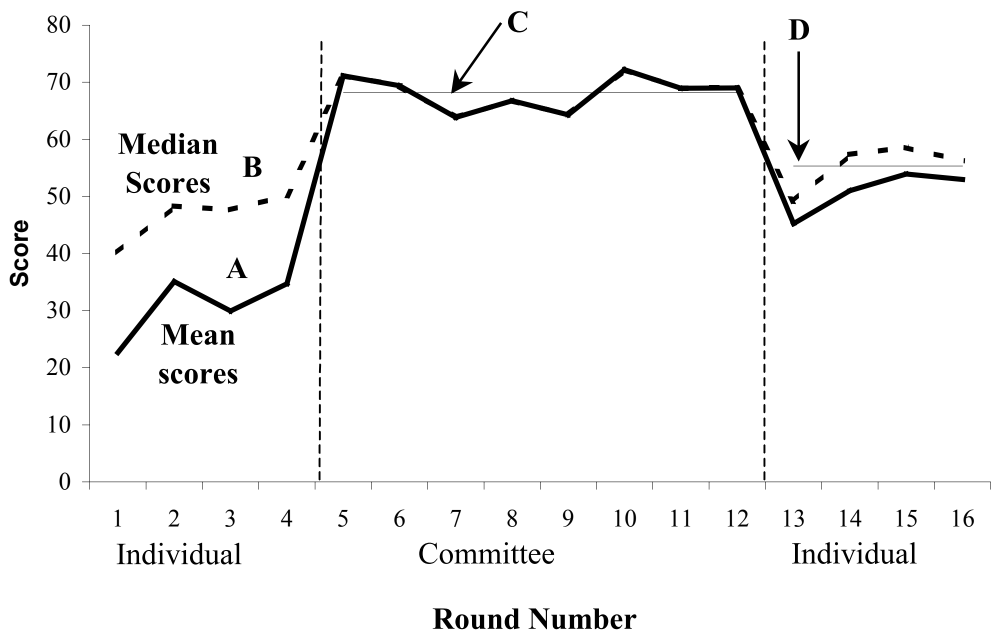
Hypothesis 1: A committee with 'majority' voting can neutralise the impact of some members playing badly in any given game.

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

Chart 6 shows a visual representation of the contribution of these two hypotheses to the improvement of committees over individuals. 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.[14] 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.

[14] Again, note that the mean score in the committee rounds is the score of the committee's interest rate decision.

Chart 6: Mean and median scores for committee members



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 6) 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.

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 – what we have called the 'pooling improvement' – stems from the ability of committees to pool judgment, expertise and skill. This is represented by the distance C-D in Chart 6 (12.9) and is also significant (t = 4.2). In other words, there is 'something special' about committees apart from their ability to aid learning and to strip out the effects of 'bad' players.

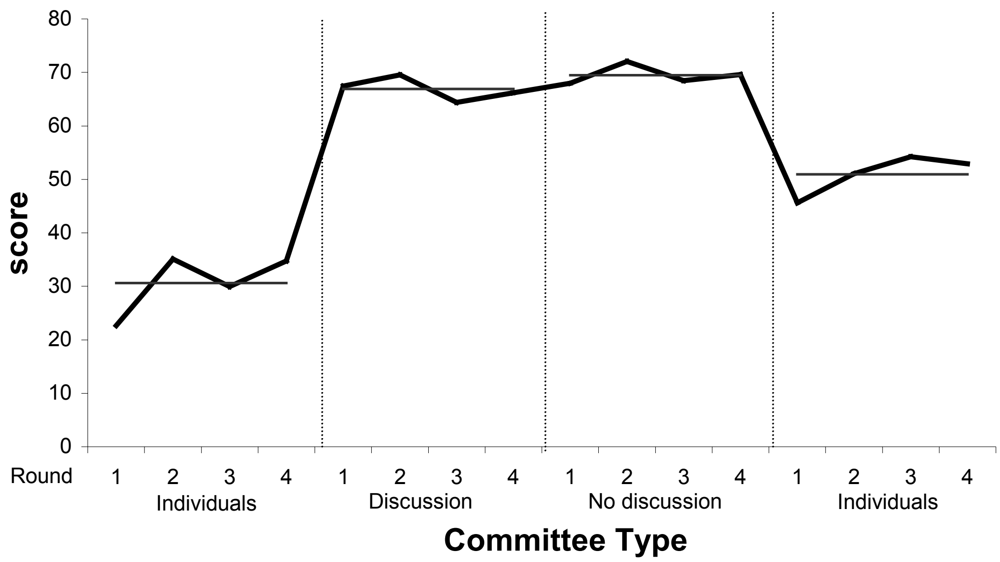
Further evidence that a committee is more than just the sum of its parts is shown by the following test: was performance in the committee stages better than the mean score of the best individual in each committee when playing alone? The mean committee score (68) was somewhat higher than that of the best individual (65) (t = 1.51, significant at the 10% level), providing evidence that committees do more than just replicate the behaviour of their best individual.

(iii) ***How do committees communicate?***

If committees improve decision-making by exploiting their members' ability to pool information and knowledge, and to learn from each other – communication must be key. Our strong prior had been that preventing committees from discussing their decisions would significantly reduce their performance, and one element of our experiment was designed to test this hypothesis.

As discussed earlier, we conducted the experiment under two different committee scenarios: one where participants were allowed to discuss their views and another where no verbal communication was allowed. As shown in Chart 7, perhaps the most surprising result is that the ability to discuss did not significantly improve committee performance: there was no significant difference between the scores in the 'discussion' and 'no discussion' stages.[15] This result was in contrast to our trialling on Bank staff. How can we explain this puzzle?

Chart 7: Breakdown of results by committee type



We argue that there are lots of different ways in which people can communicate: observation and talking to name but two. And 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 where it is easy to learn by watching, but which you would be hard pressed to pick up by talking about. Our hypothesis is that – for this particular version of the game, and this set of students – discussion did not provide more information than they acquired by watching each other.

[15] The order of the discussion and no-discussion stages was randomised across committees to control for any learning effects.

We attempted to verify this hypothesis by playing a different version of the game,(16) altering the format so as to raise the implicit benefit derived from discussion. In this variant, committee members observed their own noisy indicator of when the structural change occurred. This signal could be received with a lag of up to two periods, and the length of the lag was randomised across players. The ability to discuss becomes more valuable in this context because committee members with more timely information can share this with others more quickly by verbal communication. And the average score of discussion committees was higher (73.4) than non-discussion (65.2) in this version of the game.(17)

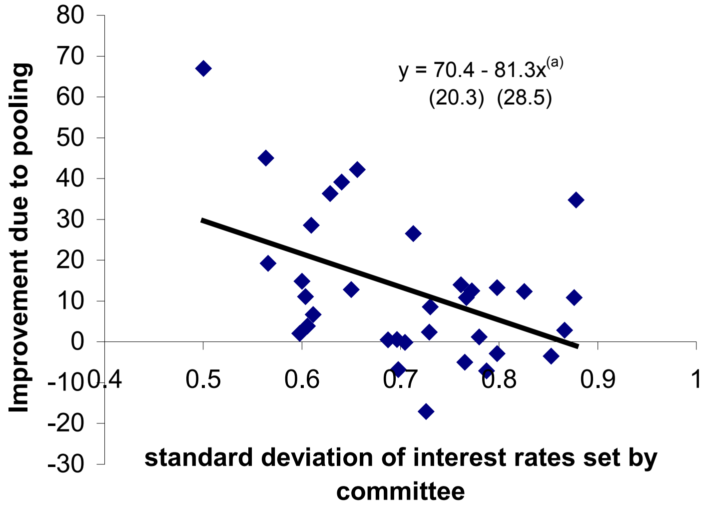
(iv) *What makes a good committee?*

Charts 8-10 attempt to shed some light on this question by showing the relationship between 'pooling improvement' and the nature of the decisions taken by each committee. We use this residual improvement measure because it strips out any 'median voting' effect on the scores. And there appears to be some evidence that the committees who improved most by this measure are those where decision-making was least activist.

Chart 8 shows a significant, negative relationship between 'pooling improvement' and policy activism (measured as the standard deviation of committee interest rate decisions over time). Although the chart tells us nothing about causation, one interpretation might be that committees scored more highly because the optimal strategy(18) in our game was to be 'cautious' – and good committees can learn this more quickly by pooling information.(19)

(16) We are very grateful to Tom Sargent for suggesting this version of the game.
(17) We cannot formally test the significance of this result, as the sample size – 15 players, in three committees – is too small.
(18) When the model was simulated under the optimal rule, interest rate movements were, on average, significantly less activist than those of the individuals, but not the committees, who played the game.
(19) Even if 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.

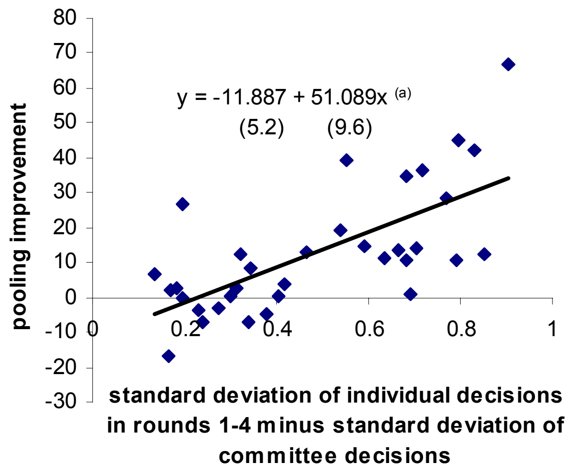
Chart 8: Pooling improvement and policy activism



(a) standard errors in brackets

Chart 9 below provides further evidence that better committees enabled participants to learn more about the optimal strategy in the game. It shows a positive and significant relationship between pooling improvement and the extent to which committee decisions were less activist than those of their individual players in the first four rounds of the experiment.

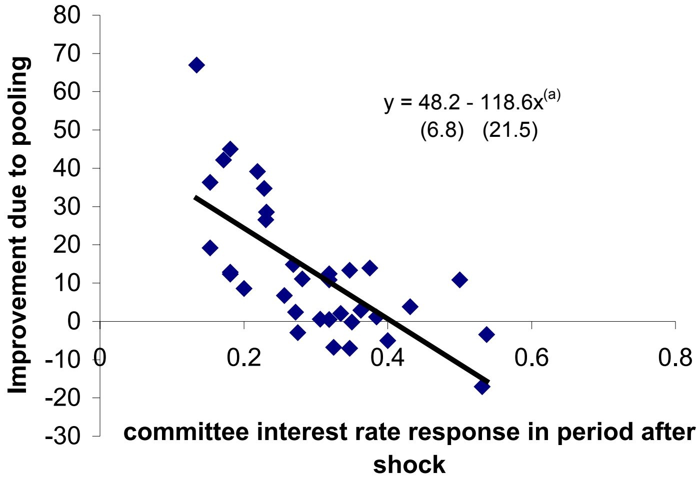
Chart 9: Pooling improvement and the fall in activism in the committee rounds as compared with individual rounds



(a) standard errors in brackets

An alternative interpretation is that the decisions of less activist committees inject less noise into the data and therefore these groups learned more quickly about the model.[20] Chart 10 offers some support for this alternative hypothesis; it uses another measure of activism – the absolute size of the interest rate response in the period following the structural shock. The same significant, negative relationship with 'pooling improvement' is again observed, suggesting that once again, committees who were less active did better. But – in contrast to the test above – the optimal rule suggests a one-for-one response to the structural shock: much bigger than the responses observed in the actual experiment.

Chart 10: Pooling improvement and absolute interest rate response in the period after the structural change



(a) standard errors in brackets

From this analysis, it would appear that there is a reasonably robust negative relationship between the extent of interest-rate activism and the improvement gained by committees through pooling their resources. We note also that this improvement cannot be attributed to median voting (because it has been stripped out) and that there is some evidence to suggest that this result is not entirely driven by the specific features of our model and the associated optimal rule. This has interesting ramifications for optimal monetary-policy setting, and it would be interesting to see if this result were true for a wider class of models.

(v) *A panel data approach*

Can we bring together the stylised facts described above in an econometric model? One way to look at the data generated from our experiment is within a panel-data framework. This allows us

[20] This appears to be at odds with the results in Wieland (2000). When model parameters are unknown, a fully-optimising policy-maker should conduct 'experiments' – eg tightening monetary policy more than is necessary – in order to learn more about the relevant policy multipliers. But within the context of our experiment, it is impossible to distinguish between active and passive learning strategies; more active committees may be simply introducing more noise into their decisions as suggested above.

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 sort of model is able to capture 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.[(21)]

The 34 committees, each playing the game 16 times, group the data. This gives a fully rectangular panel of 544 observations. For games 1 to 4 and 13 to 16, the individuals acted alone as policy-makers, generating five independent observations in each period. So for these eight games, the 'committee observation' is taken to be the mean score of the five players. This is appropriate because it allows us to model the evolution of the average performance of committee members over time (our metric of committee improvement). The panel is specified as:

$$y_{it} = \alpha_i + X_{it}'\beta + u_{it} \qquad \textbf{(6)}$$

where 'i' indexes committees i=1…34 and 't' indexes the game number, t=1,…16. y_{it} is the score of committee 'i' in game t; α_i is the constant for committee 'i' (the unobserved, committee-specific, fixed effect); and X_{it} is a matrix of regressors. The specification of the X_{it} matrix is given in more detail in equation **(7)** below. In this sort of 'fixed-effects' model, it is possible – and indeed we have good reason to expect that – the constant may be correlated with the regressors. The error u_{it} is assumed to be independent and identically distributed $N(0, \sigma_u^2)$, and is independent of X_{it}.

The time series dimension divides naturally into three different stages: the first set of individual games (numbered 1 to 4); the games played as a committee (5-12); and the final set of individual games (13-16). This has some implications for the structure of the X_{it} matrix. For example, when we consider how actively interest rates are changed – by measuring the standard deviation of interest rates set in each period across the ten periods that make up each game – this 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 this standard deviation over the five individual players is used.

Estimation is by OLS, using a general-to-specific modeling strategy. The results are shown in Table B (t-statistics shown in brackets). Equation **(7)** – or 'Model (4)' in the table is our preferred model:

$$y_{it} = \alpha_i + \beta_1\partial_1 + \beta_2\partial_2 + \gamma_1 \ln(t)^1 + \gamma_2 \ln(t)^2 + \beta_3 s_{it} + \beta_4\sigma_{it}^I + \beta_5\sigma_{it}^C + u_{it} \qquad \textbf{(7)}$$

(21) 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 χ^2 = 37.81, significant at the 0.01% level.

where α_i is again the committee-specific intercept term. To capture how the scores change over time we include a logarithmic time trend for the individual stages $\ln(t)$, and in practice the best fit is achieved by fitting a separate trend through the two sets of individual rounds (we discuss this further below). ∂_1 is a dummy variable – representing an intercept shift – for the committee games (5 to 12) and ∂_2 is for games 13 to 16. 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 game the structural shock occurs.[22] And σ_{it}^{I} and σ_{it}^{C} are the standard deviation of the interest rates set in the ten periods of each game played by individuals and committees respectively.

Table B: Specification of the preferred panel data model

Coefficient on:	**Coefficient value (t-stat in brackets)**
Committee dummy: rounds 5-12 (β_1)	36.49 (5.25)
Individual dummy: rounds 13 to 16 (β_2)	20.77 (4.96)
Period of structural shock (β_3)	2.42 (2.95)
Standard deviation of individual round decisions (β_4)	-18.23 (-3.28)
Standard deviation of committee round decisions (β_5)	-38.61 (-8.97)
Time trend for rounds 1 to 4 (γ_1)	15.74 (2.19)
Time trend for rounds 13 to 16 (γ_2)	14.76 (2.06)
R^2	0.3795

All coefficients in the 'Model 4' regression are significant. The two dummy variables ∂_1 and ∂_2 represent a common shift in the intercept of the regression at the beginning and the end of the committee stages. Their coefficients are both positive and we can reject the null hypothesis that β_1 equals β_2 in favour of the alternative that β_1 is greater (this is also consistent with the evidence presented in Chart 3 of a positive, and significant, committee effect on mean scores). β_3 is positive – the later in the game the structural change occurred, the higher the score. Because of

[22] Since players face shocks at different times when playing individually, this variable is only included for the games played as a committee.

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, if the appropriate policy action is not taken. Both β_4 and β_5 are negative: excessive interest rate movements during the game are associated with lower scores for both individuals and committees (the latter result is consistent with the analysis in the Section 4(iv)). A dummy variable for the discussion committee rounds is not significant, consistent with the result in Section 4(iii) that the ability to discuss does not seem to have a significant effect on committee performance.

After accounting for the shift in the intercept caused by the move to a committee structure, our preferred model includes a logarithmic time trend to capture learning in the individual rounds only. A logarithmic trend gives the best fit and also accords with the observation from the data that scores improved more quickly at the beginning of the game, eventually seeming to asymptote towards some upper bound. We also estimated several other versions of the panel data regression (see Appendix 3 for estimation details).

Model (1) uses a log time trend, without intercept dummies, to capture the changes in score over time. But such a model ignores the clear intercept shift in rounds 5-12 shown in Chart 3. So we fitted model (2) instead, which includes dummies to model these treatment effects. Both dummies are significant. The reported Akaike and Schwarz-Bayesian information criteria suggest an improvement on Model (1), but the time trend across the experiment as a whole is now insignificant.

Model (2) assumes a constant, logarithmic time trend throughout the 16 periods. A Wald Test of this null hypothesis against the alternative that the time trend differs across different committee structures show that this assumption is rejected at the 2% level. Model (3) allows the time trend to be different across each of the three stages of the game (rounds 1-4, rounds 5-12 and rounds 13-16 respectively). Again the information criteria indicate that this model provides a closer representation of the data.

But when we split the time trend across treatments in this way, we see that the time trend in the committee stages is actually insignificant. Removing this variable gives us Model (4) – our 'preferred model'. This seems to imply that learning over time only occurred within the individual rounds. But it is difficult to decompose the initial increase in scores in the committee stages – which is modelled better in the panel by an intercept shift – from the upward trend in scores we observe between rounds 4 and 12. The fact that the shift up in scores between rounds 4 and 5 is significantly greater than the jump down in scores after round 12 is evidence of a significant and positive committee learning effect.

Taken together, the panel data analysis reinforces the conclusions from the rest of the paper: that committees were significantly better than individuals, that less activist committees did better and that there is some evidence that participants learnt about the game over time.

5. 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 judgment 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 panel data analysis in Section 4(v) 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

Aoki, K (1998), 'On the optimal monetary policy response to noisy indicators', *Working Paper*, Princeton University.

Aoki, K (2000), 'Optimal commitment policy under noisy information', *Working Paper*, Kobe University.

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 (1998), *Central banking in theory and practice*, MIT Press.

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.

Brainard, W (1967), 'Uncertainty and the effectiveness of monetary policy', *American Economic Review*, Vol. 57, No. 2, pages 411-25.

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

Evans, G W and Honkapohja, S (2001), *Learning and expectations in macroeconomics*, Princeton University Press.

Friedman, D and Sunder, J (1994), *Experimental methods: a primer for economists*, Cambridge 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.

Sargent, T J (1999), 'Comment on Ball (1999)', in Taylor, J B (ed), *Monetary policy rules*, pages 144-54.

Stoner, J A F (1961), 'A comparison of individual and group decisions involving risk', unpublished master's thesis, MIT.

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

Svensson, L E O and Woodford, M (2001), 'Indicator variables for monetary policy under asymmetric information', *NBER Working Paper*, No. 8255, April.

Wieland, V (2000), 'Learning by doing and the value of optimal experimentation', *Journal of Economic Dynamics and Control*, Vol. 24, Issue 4, April.

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:

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

$$\textbf{(2)}\ \pi_t = 0.7\pi_{t-1} + 0.3\pi_{t-2} + 0.2y_t + \nu_t \ \text{ where } \nu_t \sim N\left(0, \sigma_\nu^2\right)$$

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

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:

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

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.

The correct responses to the priors' questionnaire and the distribution of the answers given both before and after participants played the game are contained in the table below. The 'correct' answers to questions 1 and 3 come directly from the form of the optimal rule **(4)** derived above. There is no interest rate smoothing term in that rule, so the correct answer to question 1 is zero. The relative weight monetary policy makers should place on smoothing output as compared to inflation is 0.8: this is calculated by calculating the relative weight on output to the sum of the weights on lagged inflation in the optimal rule.

The correct answer to question 2 is that the maximum impact of monetary policy on inflation is felt after one period. This is derived from simulating the model under the full information optimal rule.

The ideal answers to questions 4 and 5 are taken directly from the parameters of the IS curve **(1)**, and those to questions 6, 7 and 8 are taken from the Phillips curve **(2)**. The correct answer to Q8 follows from the coefficients on lagged inflation summing to one in equation **(2)** – which gives the model its long-run neutrality property.

From the responses given by players, we can see that in all cases except question one, the variance of the answers given by participants in response to each question fell after they had played the game. For half of the questions, players' final answers were closer to the correct answers derived from the model than their initial responses. And these changes in view were significant in the case of questions 2 and 3 (see Section 4(i.i) for a more detailed discussion).

Table A1: Summary statistics for the distribution of answers to the priors' questions

Question number	Correct answer	Mean response before	Mean response after	SD of responses before	SD of responses after
1	0	0.626	0.591	0.192	0.241
2	0.1	0.427	0.298	0.199	0.141
3	0.8	0.438	0.569	0.202	0.194
4	0.8	0.555	0.550	0.216	0.208
5	0.5	0.499	0.649	0.195	0.184
6	0.7	0.582	0.586	0.207	0.192
7	0.2	0.485	0.502	0.212	0.197
8	0	0.490	0.531	0.265	0.210

Appendix 2a: Oral briefing

Thank you for coming today. In a moment we are going to ask you to play a simple monetary policy game using a linear and learnable small macroeconomic model that we believe is representative of the structure of the UK economy. We will give you 15 minutes to read the instruction sheet that I'm about to hand out and to have a practice before playing the game for real. During the experiment it is very important that you do not speak to each other, except at the times indicated by the instructions sheet. If you have any questions, please ask us, not the person sitting next to you. [The candidates were then talked through an example screen and shown how to play the game.]

Appendix 2b: Written instructions

These were handed out to each participant. There were two versions:

1) The version shown here.

2) A second version for the game where the 'discussion' round was played before the 'no discussion' round (bullet points 3 and 4 below are reversed).

Monetary policy game

Congratulations – you are now a monetary policy maker!
Your task is to vary the short-term nominal interest rate to keep inflation as close as possible to the government's inflation target of 2.5% while minimising the deviation of output from its 'natural' level of 5.

To measure your performance, you will be scored in each round as follows:[23]

$$Score = 100 - 40\left|Output - 5\right| - 40\left|Inflation - 2.5\right|$$

Each game consists of ten rounds in which you must decide what interest rate to set. And the overall score for each game is the average across these ten rounds, with the minimum possible score set at zero.

Before each round begins, different types of shock may hit the economy. Like real policy-makers, you will not observe those shocks – only the response of certain economic variables: in this case, output and inflation. To set rates as well as possible, you will have to try to work out what shock has hit the economy and when. Only over time will you learn how the economy responds to your interest rate decisions. To make things more exciting, a structural change will occur at some point during each game. The key to playing successfully is to identify

(23) $|x|$ refers to the absolute value of a number, eg $|-3| = 3$.

when the change has occurred and how best to respond to it. So try to think creatively – are the facts you observe consistent with one or more hypotheses? That is what real policy-makers do!

You will be asked to play the game both on your own and under two different 'committee' scenarios. The game consists of six stages:

1. Fill in a questionnaire about your economic 'views'. Then ten minutes practice to get a feel for the model.
2. Four games setting interest rates as 'governor' on your own.
3. Four games as a 'committee' of five people, but without discussing your vote with other committee members. For this stage of the game each person votes and the interest rate set will be the median. The game cannot proceed on to the next round until all committee members have voted.
4. Four further games as a committee of five. But now you are allowed to discuss your vote with other committee members. Again, the median interest rate will be set.
5. Repeat stage 2.
6. Fill in the 'views' questionnaire again.

We will be there to help with any problems during the game – good luck!!

Appendix 3: Panel data estimation results

The table below reports the estimation results of our attempts to model the time trend in the panel regression (the figures in brackets are t-statistics).

Coefficient on:	Model **(1)**	Model **(2)**	Model **(3)**	Model **(4)**
Committee dummy: rounds 5-12 (β_1)		25.89 (3.60)	39.08 (5.23)	36.49 (5.25)
Individual dummy: rounds 13 to 16 (β_2)		13.37 (2.58)	20.77 (4.95)	20.77 (4.96)
Period of structural shock (β_3)	3.49 (4.53)	2.51 (3.03)	2.34 (2.83)	2.42 (2.95)
Standard deviation of individual round decisions (β_4)	-29.69 (-9.24)	-18.02 (-3.24)	-18.25 (-3.29)	-18.23 (-3.28)
Standard deviation of committee round decisions (β_5)	-33.53 (-8.12)	-38.14 (-8.79)	-39.06 (-9.02)	-38.61 (-8.97)
Time trend for rounds 1 to 16	22.33 (8.74)	8.66 (1.52)		
Time trend for rounds 1 to 4			15.74 (2.19)	15.74 (2.19)
Time trend for rounds 5 to 12			-3.82 (-0.94)	
Time trend for rounds 13 to 16			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^2	0.3595	0.3743	0.3797	0.3795

The four models estimated are specified as follows:

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

$$y_{it} = \alpha_i + \beta_1 \partial_1 + \beta_2 \partial_2 + \gamma_1 \ln(t)^{1-16} + \beta_3 s_{it} + \beta_4 \sigma_{it}^I + \beta_5 \sigma_{it}^C + u_{it} \quad \textbf{(2)}$$

$$y_{it} = \alpha_i + \beta_1 \partial_1 + \beta_2 \partial_2 + \gamma_1 \ln(t)^{1-4} + \gamma_2 \ln(t)^{5-12} + \gamma_3 \ln(t)^{13-16} + \beta_3 s_{it} + \beta_4 \sigma_{it}^I + \beta_5 \sigma_{it}^C + u_{it} \quad \textbf{(3)}$$

$$y_{it} = \alpha_i + \beta_1 \partial_1 + \beta_2 \partial_2 + \gamma_1 \ln(t)^{1-4} + \gamma_2 \ln(t)^{13-16} + \beta_3 s_{it} + \beta_4 \sigma_{it}^I + \beta_5 \sigma_{it}^C + u_{it} \quad \textbf{(4)}$$

Appendix 4: Priors' questionnaire

Date:	Group:

Please spend a few minutes filling in this questionnaire, concentrating in particular on the questions in italics. It doesn't matter if you are not familiar with the jargon in brackets: this is merely to help us calibrate your response.

BE AS HONEST AS YOU CAN – THERE ARE NO RIGHT OR WRONG ANSWERS!

What is your player number?

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

Not at all cautiously (ie α = 0) Very cautiously (ie α = 1)

0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1

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

0	1	2	3	4	5	6	7	8	9	10

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 $i_t = \alpha(y_t - Y) + (1-\alpha)(\pi_t - \pi^*) + \ldots.$, what weight should they place on α)?

None (ie α = 0) All (ie α = 1)

0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1

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

Not at all persistent (ie α = 0) Completely persistent (ie α = 1)

0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1

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

Not at all sensitive (ie α = 0) Very sensitive (ie α = 1)

0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1

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

Not at all persistent (ie α = 0) Completely persistent (ie α = 1)

0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1

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 $\pi_t = \alpha(y_{t-1} - Y) + ...$, what weight do you think α would take)?

Not at all sensitive (ie α = 0) Highly sensitive (ie α = 1)

0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1

8) To *what extent is inflation sensitive to deviations of output from trend in the long run*?

Not at all sensitive (ie α = 0) Highly sensitive (ie α = 1)

0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1

9) What course are you studying?

……
……….

10) Are you….

Undergraduate: 2nd Year	*Undergraduate: 3rd Year*	*Graduate Student*