ALAN S. BLINDER
JOHN MORGAN

Are Two Heads Better Than One? Monetary Policy by Committee

Two experiments were conducted to test the common hypothesis that groups make decisions more slowly than individuals. One of these experiments imitates real-life monetary policy decisions. In both cases, the hypothesis is found wanting: groups are *not* slower than individuals. In both experiments, we also find that group decisions are on average *better* than individual decisions. This holds regardless of whether the groups make decisions by unanimity or majority rule. Simple mechanical theories of group decisionmaking—that the group follows its average player, median player, or best player—do not explain the results. Group interactions seem to matter.

JEL codes: C91, C92, E52, E58
Keywords: monetary policy; group decision making.

THE OLD SAW "it works in practice, now let's see if it works in theory" has a direct application to the design of monetary policy institutions. In recent years, central banking practice has exhibited a notable shift from individual to group decisionmaking, that is, toward more monetary policy committees (MPCs). For example, J.P. Morgan's "Guide to Central Bank Watching" (J.P. Morgan & Co. 2000, p. 4) noted that "One of the most notable developments of the past few years has been the shift of monetary policy decision-making to meetings of central bank policy boards." Two of the best-known examples of this institutional change are the Bank of England and the Bank of Japan, which (roughly) switched from individual to group decisionmaking in 1997 and 1999, respectively. The Governing Council of

We gratefully acknowledge financial support from Princeton's Center for Economic Policy Studies, thank Felix Vardy for excellent and extensive research assistance, and thank the editor and referees of this *Journal* for helpful comments.

ALAN S. BLINDER *is Gordon S. Rentschler Memorial Professor of Economics at the Department of Economics at Princeton University* (*E-mail:* blinder@princeton.edu). JOHN MORGAN *is a Professor at the Haas School of Business and Department of Economics at the University of California, Berkeley* (*E-mail:* morgon@haas.berkeley.edu).

Received January 29, 2003; and accepted in revised form October 15, 2003.

Journal of Money, Credit, and Banking, Vol. 37, No. 5 (October 2005)

the European System of Central Banks, patterned loosely on the Federal Open Market Committee, also opened for business in 1999, although the Bundesbank Council had made decisions as a group long before that.

Nevertheless, economic theory has had precious little to say on the pros and cons of making monetary policy individually or in groups.[1] Nor is there substantial empirical evidence on the relative merits of the two types of monetary policy decisionmaking processes.

In fact, this chasm between theory and practice generalizes. While economics has been characterized as the science of *choice*, almost all the choices economists analyze are modeled as *individual* decisions. A consumer with a utility function and a budget constraint decides what to purchase. A firm, modeled as an individual decisionmaker, decides what will maximize its profits. A unitary central banker with a well-defined loss function selects the optimal interest rate. In many instances, these modeling choices abstract from the fact that a group is in fact making the decision, presumably on the grounds that the group members have the same objective and share the same information. The question, "Do decisions made by groups differ systematically from the decisions of the individuals who comprise them?" is infrequently asked.[2]

But we all know that many decisions in real societies—including some quite important ones—are made by groups. Legislatures, of course, make the laws. The Supreme Court is a committee, as are all juries. Some business decisions, e.g., in partnerships or management committees, are made collectively, rather than dictatorially. And, as just noted, monetary policy in most countries these days is made by a committee rather than by an individual. While one of us served as Vice Chairman of the Federal Reserve Board, he came to believe that economic models might be missing something important by treating monetary policy decisions as if they were made by a single individual minimizing a well-defined loss function. As he subsequently wrote:

> While serving on the FOMC, I was vividly reminded of a few things all of us probably know about committees: that they laboriously aggregate individual preferences; that they need to be led; that they tend to adopt compromise positions on difficult questions; and—perhaps because of all of the above—that they tend to be inertial. (Blinder 1998, p. 20)

This sentiment reflects what is probably a widely-held view: that groups make decisions more slowly than individuals.[3] One of the major questions for this paper is: is it true?

But there is a deeper question: why are so many important decisions entrusted to groups? Presumably because of some belief in collective wisdom. In a complicated

1. See, among other sources, Goodfriend (2000), Gerlach-Kristen (2003), and Gersbach and Hahn (2001) for recent discussions of the pros and cons of group versus committee decisionmaking on monetary policy. Blinder (2004) contains a reasonably comprehensive survey of the (limited) literature on the question.

2. The most prominent and famous exception is surely Arrow (1963).

3. In the case of monetary policy, slowness is reflected in the amount of *data* the central bank feels compelled to accumulate before coming to a decision rather than in the length of time a *meeting* takes. As will be clear shortly, we designed our experiment to reflect that aspect of reality.

world, where no one knows the true model or even all the facts, where data may be hard to process or to interpret, and where value judgments may influence decisions, it may be beneficial to bring more than one mind to bear on a question. While it has been said that nothing good was ever written by a committee, could it be that committees actually make better decisions than individuals—at least when committees are organized, orderly, and cohesive?

So these are the two central questions for this paper: do groups such as monetary policy committees reach decisions more slowly than individuals do? (We have never heard the opposite suggested.) And are group decisions, on average, better or worse than individual decisions?

Since neither the theoretical nor the empirical literature offers much guidance on these questions, our approach is experimental. We created two laboratory experiments in which literally everything was held equal except the nature of the decisionmaking body—an individual or a group. Even the identities of the individuals were the same, since each experimental group consisted of five people who also participated as individuals. We therefore had automatic, laboratory controls for what are normally called "individual effects." The experimental setting also allowed us to define an objective function—which was known to the subjects—that distinguished better decisions from worse ones with a clarity that is rarely attainable in the real world. That is one huge advantage of the laboratory approach. The artificiality is, of course, its principal drawback.

The two experiments, one of which is described in detail below, were very different.[4] In the simpler setup, which is sketched briefly in Section 2, we created a purely statistical problem devoid of any economic content, but designed (as will be explained) to mimic certain aspects of monetary policymaking. In the more complex and interesting setup, which we discuss at some length in Section 1, we placed subjects explicitly in the shoes of monetary policymakers: subjects were asked to steer an (electronic model of an) economy by manipulating the interest rate.

The results were strikingly consistent across the two experiments. Neither experiment supported the commonly-held belief that groups are more inertial than individuals. In fact, the groups required no more data than the individuals before reaching a decision. That came as a big surprise to us; our priors were like seemingly everyone else's. Despite the fact that both groups and individuals were operating with similar amounts of information, both experiments found that groups, on average, made better decisions than individuals. (Here our priors were much more diffuse.) Moreover, and strikingly, groups outperformed individuals by almost exactly the same margins in the two experiments.

In addition, the experiments unearthed one other surprising result: there were practically no differences between group decisions made by majority rule and group decisions made under a unanimity requirement. This finding, which also conflicted with our priors, is also highly relevant to monetary policy. The ESCB, for example, claims to reach decisions unanimously while, e.g., the Bank of England relies

4. The data, program code, and instructions for both experiments are available on request.

on majority vote. (The Fed is somewhere in between, but probably closer to the unanimity principle.)

Before proceeding, a few words on the experimental literature on individual versus group choices may be useful. Much of it comes from psychology and centers on how individual biases are reflected in group decisions. The evidence on whether groups or individuals make "better" decisions in this framework is mixed. A metastudy of this literature (Kerr, MacCoun, and Kramer, 1996) concluded that there is no general answer to the question. Other studies (e.g., Wallach, Kogan, and Bem 1964) have found that group decisions can lead to excessive risk taking—the so-called *risky shift*, which we will discuss later.

In the economics literature, most individual-versus-group experiments are in game-theoretic settings, rather than the decision-theoretic settings of our experiments (Bornstein and Yaniv, 1998, Cox and Hayne, 1998, Kocher and Sutter, 2000). Methodologically, our paper is closest to Cason and Mui (1997), who explored individual and group decisions with objective payoffs in a decision-theoretic setting. Their substantive concerns, however, were completely different from ours. In particular, the decisionmaking task in their experiment was straightforward, but there were potentially strong differences of opinion about the allocation of rewards. Our experiments were structured in precisely the opposite way: the decisionmaking task was complex, but there was no room for dispute over the division of the spoils. Thus, our focus was mainly on differences between groups and individuals in what psychologists refer to as *intellective* tasks rather than *judgmental* tasks.

The remainder of the paper is organized into four sections. Section 1 describes the monetary policy experiment and what we found. Section 2 summarizes the purely statistical experiment very briefly. Section 3 reports on some attempts to model the group decisionmaking process, and Section 4 is a brief summary.

1. THE MONETARY POLICY EXPERIMENT

1.1 Description of the Monetary Policy Experiment

Our main experiment asked subjects to assume the role of monetary policymaker. For this reason, we imposed a prerequisite in recruiting student subjects: they had to have taken at least one course in macroeconomics. We brought students into the laboratory in groups of five, telling them that they would be playing a monetary policy game. Specifically, we programmed each computer with a simple two-equation macroeconomic model that approximates a canonical model made popular in the recent theoretical literature on monetary policy (Ball, 1997, Rudebusch and Svensson, 1999), choosing (not estimating) parameter values that resemble the U.S. economy:

$$U_t - 5 = 0.6(U_{t-1} - 5) + 0.3(i_{t-1} - \pi_{t-1} - 5) - G_t + e_t \tag{1}$$

$$\pi_t = 0.4\pi_{t-1} + 0.3\pi_{t-2} + 0.2\pi_{t-3} + 0.1\pi_{t-4} - 0.5(U_{t-1} - 5) + w_t \,. \tag{2}$$

Equation (1) can be thought of as a reduced form combining an IS curve with Okun's Law. Specifically, U is the unemployment rate, and the assumed "natural rate" is 5%. Since i is the nominal interest rate and π is the rate of inflation, the term $i - \pi - 5$ connotes the deviation of the *real* interest rate from its equilibrium or "neutral" value, which is also set at 5%.[5] Higher (lower) real interest rates will push unemployment up (down), but only gradually. But our experimental subjects, like the Federal Reserve, controlled only the *nominal* interest rate, not the *real* interest rate.

The G_t term connotes the effect of fiscal actions on unemployment and is the random event that our experimental monetary policymakers were supposed to recognize and react to. G starts at zero and randomly changes *permanently* to either $+0.3$ or -0.3 sometime within the first 10 periods. As is clear from Equation (1), this event changes unemployment by precisely that amount, but in the opposite direction, on impact. Subsequently, of course, the model economy converges back to the natural rate. Prior to the shock, the model's steady-state equilibrium is $U = 5$, $i - \pi = 5$. Because the long-run Phillips curve is vertical, any constant inflation rate can be a steady state. But we always began the experiment with inflation at 2%—which is the target rate. As is apparent from the coefficients in Equation (1), the fiscal shock changes the "neutral" real interest rate to either 6% or 4% *permanently.* Our subjects were supposed to react to this change, presumably with a lag, by raising or lowering the nominal interest rate accordingly. The length of the lag, of course, is one of our primary interests.

Equation (2) is a standard accelerationist Phillips curve. Inflation depends on the lagged unemployment rate and on its own four lagged values, with weights summing to one. While the weighted average of past inflation rates can be thought of as representing expected inflation, the model does not demand this interpretation. The coefficient on the unemployment rate was chosen to (roughly) match empirically-estimated Phillips curves for the United States.

Finally, the two stochastic shocks, e_t and w_t, were drawn from uniform distributions on the interval $[-.25, +.25]$.[6] Their standard deviations are approximately 0.14, or about half the size of the G shock. This parameter controls the "signal to noise" ratio in the experiment. We sized the fiscal shock to make it easy to detect, but not "too easy."[7]

Monetary policy affects inflation only indirectly in this model, and with a distributed lag that begins two periods later. All of our subjects understood that higher interest rates reduce inflation and raise unemployment with a lag, and that lower interest rates do the reverse.[8] But they did not know any details of the model's specification,

5. The neutral real interest rate is defined as the real rate at which inflation is neither rising nor falling. See Blinder (1998, pp. 31–33).

6. The distributions were iid and uniform, rather than normal, for programming convenience.

7. This is a probabilistic statement. It is possible, for example, that a two-standard-deviation e or w shock in the opposite direction obscures the G shock for a while.

8. Remember, all our subjects had at least some exposure to basic macroeconomics. Lest they had forgotten, the instructions (available on request) reminded them that raising the rate of interest would lower inflation and raise unemployment, while lowering the rate of interest would do the opposite.

coefficients, or lag structure. Nor did they know when the shock occurred. But they did know the probability law that governed the shock—which was a uniform distribution across periods 1 through 10.

While the model looks trivial, stabilizing such a system can be rather tricky in practice because the model will diverge from equilibrium when perturbed by a G shock—unless it is stabilized by monetary policy. One useful way to think about this dynamic instability is as follows. Start the system at equilibrium with $U = 5$, $\pi = 2$, and $i = 7$, as we did. Now suppose G rises to 0.3. By Equation (1), the neutral real rate of interest increases to 6%. So the initial real rate, which is 5%, is now below neutral—and hence expansionary. With a lag, inflation begins to rise. If the central bank fails to raise the nominal interest rate, the real rate falls further—stimulating the economy even more. But the lags make the divergence from equilibrium pretty gradual.

Each play of the game proceeded as follows. We started the system in steady state equilibrium with $G = 0$, lagged U set at 5%, all lags of π at 2%, and current and lagged nominal interest rates at 7% (reflecting a 5% real rate and a 2% inflation rate). The computer then selected values for the two random shocks and displayed the first-period values, U_1 and π_1, on the screen for the subjects to see. Normally, these numbers were quite close to the optimal values of $U = 5\%$ and $\pi = 2\%$. In each subsequent period, new random values of $\mathrm{e_t}$ and $\mathrm{w_t}$ were drawn, thereby creating statistical noise, and the lagged variables that appear in Equations (1) and (2) were updated. The computer would calculate U_t and π_t and display them on the screen, along with all past values. Subjects were then asked to choose an interest rate for the next period, and the game continued.

At some period chosen at random from a uniform distribution between $t = 1$ and $t = 10$, G_t was either raised to $+0.3$ or lowered to -0.3. (Whether G rose or fell was also decided randomly, with a 50% probability for either choice.) Students were not told when G changed, nor in which direction. Even though our primary interest was in the decision lag, that is, the lag between the change in G and the *first* change in the interest rate, we did not stop the game when the interest rate was first changed because this seemed artificial in a monetary-policy context. Instead, each play of the game continued for 20 periods. (Subjects were told to think of each period as a quarter.)

It is important to note that no time pressure was applied; subjects were permitted to take as much clock time as they pleased to make each decision. In comparing the speed with which decisions are taken by individuals and groups, we were *not* concerned with clock time—just as we are not concerned with the number of hours an MPC meeting takes. Some readers of earlier drafts were confused on this point, so let us be explicit. There are certainly examples of real world decisions in which speed, in the literal sense of clock time, is of the essence—think of skiing down a narrow slope, shooting the rapids, or auto racing. But very few *economic* decisions are of this character. (Real-time bond and commodity trading may be an exception.) No one much cares if a consumer takes five minutes or five hours to decide on her consumption bundle, nor if a firm takes an hour or a day to decide how much labor to hire.

Certainly in the context of monetary policy committee meetings, clock time is irrelevant. Nobody cares whether the FOMC deliberates for two hours or four hours. What *is* relevant, and what is measured here as the *decision lag*, is the amount of *data* that the decisionmakers insist on seeing before they change interest rates. In the real world, this data flow corresponds, albeit unevenly, to calendar time—e.g., the Fed may see five relevant pieces of data one week, three the next, etc. At some point, it decides that it has accumulated enough information to warrant changing interest rates. In our experiment, data on unemployment and inflation flow evenly—one new observation on each variable per period. So when we say that one type of decisionmaking process "takes longer" than another, we mean that more *data* (not more *minutes*) are required before the decision is made.

To evaluate the *quality* of the decisions, our other main interest, we need a loss function. While quadratic loss functions are the rule in the academic literature, they are rather too difficult for subjects to calculate in their heads. So we used an absolute-value function instead. Specifically, subjects were told that their score for each quarter would be:

$$s_t = 100 - 10\left|U_t - 5\right| - 10\left|\pi_t - 2\right|, \tag{3}$$

and the score for the entire game (henceforth, *S*) would be the (unweighted) average of s_t over the 20 quarters. The coefficients in Equation (3) scale the scores into percentages—giving them a ready, intuitive interpretation. Equal weights on unemployment deviations and inflation deviations were chosen to facilitate mental calculations: every miss of 0.1 cost one point. Thus, for example, missing the unemployment target by 0.5 (in either direction) and the inflation target by 0.7 would result in a score of 100 - 12 = 88 for that period. At the end of the entire session, scores were converted into money at the rate of 25 cents for each point. Subjects typically earned about \$21–\$22 out of a theoretical maximum of \$25.[9]

Finally, we "charged" subjects a fixed cost of 10 points each time they changed the rate of interest, regardless of the size of the change.[10] The reason is as follows. The random shocks, e_t and w_t, were an essential part of the experimental design because, without them, changes in G_t would be trivial to observe: no variable would ever change until *G* did. After some experimentation, we decided that random shocks with standard deviations about half the size of the *G* shock made it neither too easy nor too difficult to discern the G_t "news" amidst the e_t and w_t "noise."

But this decision created an inference problem: our subjects might receive several false signals before *G* actually changed. For example, a two-standard-deviation *e*

9. It may seem "natural" to compare subjects' scores to the theoretical optimum computed as the solution to the optimal control problem. We did not do this, however, because this problem is *not* linear-quadratic. Because of the absolute value objective function (Equation 3) and the fixed cost and integer constraints mentioned in the following paragraph, the optimal control solution would (a) be virtually impossible to derive analytically and, more importantly, (b) have "bang-bang" features that we did not expect our students to emulate (nor did they).

10. To keep things simple, only integer interest rates were allowed.

shock appears just like a negative *G* shock, except that the latter is permanent while the former is transitory. (The random shocks were iid.) Furthermore, subjects knew neither the size of the *G* shock nor the standard deviations of *e* and *w*; so they had no way of knowing that a two-standard-deviation disturbance would look (at first) like a *G* shock.

In some early trials designed to test the experimental apparatus, we observed students moving the interest rate up and down quite frequently—sometimes every period.[11] Such behavior would make it virtually impossible to measure (or even to define) the decision lag in monetary policy. So we instituted a small, 10-point charge for each interest rate change. Ten points is not much of a penalty; averaged over a 20-period game, it amounts to just 0.5% of total score. But we found it was large enough to deter most of the excessive fiddling with interest rates. It also had the collateral benefit of making behavior much more realistic.[12] The Fed does not jigger the interest rate around every quarter, presumably because it perceives some cost of doing so that is not captured in Equation (3).[13]

The game was played as follows. Each session had five subjects, mostly Princeton undergraduates. Subjects were read detailed instructions, which they were also given in writing, and then allowed to practice with the computer apparatus for about five minutes—during which time they could ask any questions they wished. Scores during those practice rounds were displayed for feedback, but not recorded. At the end of the practice period, all machines were reinitialized, and each student was instructed to play 10 rounds of the game *alone*—without communicating in any way with the other students. Subjects were allowed to proceed at their own pace; clock time was irrelevant. When all five subjects had completed 10 rounds, the experimenter called a halt to Part One of the experiment.

In Part Two, the five students gathered around a single computer to play the same game 10 times *as a group*. The rules were exactly the same, except that students were now permitted to communicate freely with one another—as much as they pleased. During group play, all five students received the group's common scores. Thus, since everyone in the group had the same objective function and the same information, there was no incentive to engage in self-interested behavior.[14]

We ran 20 sessions in all, involving 100 distinct subjects. (No subject took part in more than one session.) In half of the sessions, decisions in Part Two were made by *majority rule*: the experimenter told the group that he would do nothing until he had instructions from at least three of the five students. In the other half, decisions

11. These early trials are *not* included in the experimental data reported below. Nor were any of the students used as subjects in these trials also used as subjects in the actual experiment.

12. With one exception: since the game terminated after 20 periods, students generally concluded that it was not worth paying 10 points to change the rate of interest in one of the last few periods. We pay virtually no attention to this end-game data.

13. Empirically estimated reaction functions for central banks typically include a Δi term that is rationalized by some sort of cost of changing interest rates.

14. For essentially these reasons, the literature on information aggregation in groups is mostly irrelevant to our experiment.

were made *unanimously*: the experimenter told the subjects that he would do nothing until all five agreed.

After 10 rounds of group play, the subjects returned to their individual machines for Part Three, in which they played the game another 10 times alone. Following that, they returned to the group computer for Part Four, in which decisions were now made *unanimously* if they had been by majority rule in Part Two, or by *majority rule* if they had previously been under unanimity. Table 1 summarizes the flow of each session.

A typical session (of 40 plays of the game) lasted about 90 minutes. Each of the 20 sessions generated 20 individual observations per subject, or 2000 in all, and 20 group observations, or 400 in all. We would have liked to have taken longer and generated more observations, but it was unrealistic to ask subjects to commit more than two hours of their time,[15] and 40 plays of the game were about all we could count on finishing within that time frame.

1.2 The Three Main Hypotheses

While several subsidiary questions will be considered below, our interest focused on the three main hypotheses mentioned in the introduction, especially the first two:

H_1: *Group decisions are more inertial than individual decisions.*

The main idea that motivated this study was our prior belief that groups are inertial—that is, they need to accumulate more data before coming to a decision. We repeat once again that we measured the decision lag in *number of periods* that elapsed between the change in G and the *first* change in the interest rate—that is, the amount of *information* required before a decision was reached to change monetary policy. We did not measure elapsed clock time, which, in the context of monetary policy, seemed irrelevant.[16] The decision lag, L, could be positive, as was true in 84.8% of the cases, or negative (if subjects changed the interest rate before G changed).[17]

Specifically, let L_i be the average lag for the i-th individual in the group (i = 1,..., 5) when he or she plays the game *alone*, and let L_G be the average lag for those same five people when making decisions *as a group*. Under the null hypothesis of no group interaction, the group's mean lag would equal the average of the five individual mean lags. Furthermore, under this null, and assuming independence across observations, a simple t-test for difference in means is the appropriate test.[18]

The typical lags in the monetary policy game were short, averaging just over 2.4 "quarters" across the 2400 observations. In fact, a number of subjects "jumped the gun" by moving interest rates *before* G had changed. (As just noted, this happened

15. Although sessions normally took closer to 1.5 hours, we insisted that subjects agree to commit two hours, since the premature departure of even one subject would ruin an entire session.

16. It was clear from observing the experiments, however, that groups took more clock time.

17. In defining L, we used only the *first* change in the rate of interest. Subsequent changes in the interest rate affected the students' scores, S, but not the measured decision lag, L.

18. We thank Alan Krueger for reminding us of this simple consequence of the Neymann—Pearson lemma.

TABLE 1

The Flow of the Monetary Policy Experiment

Instructions
Practice rounds (no scores recorded)
Part One: 10 rounds played as individuals
Part Two: 10 rounds played as a group under majority rule (alternatively, under unanimity)
Part Three: 10 rounds played as individuals
Part Four: 10 rounds played as a group under unanimity (alternatively, under majority rule)
Students are paid in cash, fill out a short questionnaire, and leave

in 15.2% of all cases.) Surprisingly, the groups actually made decisions slightly *faster* than the individuals on average, with a mean lag of just 2.30 periods (with standard deviation 2.75) versus 2.45 periods (with standard deviation 3.50) for the individual decisions. While this scant 0.15 difference goes in the direction opposite to the null hypothesis, it does not come close to statistical significance at conventional levels ($t = 0.91$, $p = 0.18$ in a one-tailed test).[19] Histograms for the variable L for individuals and groups look strikingly alike (see Figure 1). In fact, the left-hand panel resembles a mean-preserving spread of the right-hand panel. In a word, we find no experimental support for the commonsense belief that groups are more inertial than individuals in their decision making.

Before examining this counterintuitive finding further, an important methodological issue must be addressed. The t-test that we just employed treats each play of the game as an independent observation. But in presenting this work to several audiences, we found that many people insisted that strong individual effects (e.g., person i is inherently slower than person j) meant that the 40 observations on the behavior of one person in a single session were far from independent. We have two answers to that criticism. First, Section 3 will present strong empirical evidence that individual effects were actually quite *unimportant* in our experiments.

Second, our findings *do not* rest on the independence assumption. We demonstrate the latter in two different ways. First, we will present tests based on an extremely conservative view of the data that treats each *session* as a *single observation.* Doing that collapses our 2400 observations into just 20 matched pairs of individual and group lags. When treating the data this way, we use a nonparametric test: the Wilcoxon signed-ranks test. Second, we can take a less extreme view that is not so wasteful of the data by using a heteroskedastic and autocorrelation consistent (HAC) estimator for the standard error (Newey and West 1987).

In the case of the differential lag that we are now discussing, the Wilcoxon test detects no significant difference between the groups and the individuals ($z = 0.5$ in a one-tailed test).[20] And the HAC estimator is identical to our original estimator

19. Throughout, all t-tests use unequal variances.

20. Of course, no test will have much power with only 20 observations. The Wilcoxon test results are more interesting when we try to *demonstrate* differences rather than *deny* them, as we do just below.

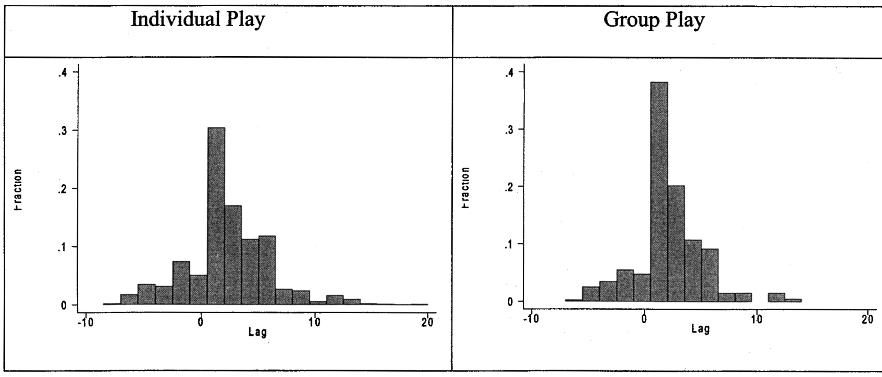

FIG. 1. Histograms of Lag in Monetary Policy Experiment

($t = 0.91$).[21] Thus all three estimation techniques lead to the same conclusion: there is no statistically significant difference in lag times between individuals and groups.

A more important question is: could the counterintuitive finding of (statistically) equal lags be an artifact of the experiment? One possibility relates to learning, about which we will have much more to say later. In both our experiments, subjects always began by making decisions as individuals. Suppose the typical student was still learning how to play the game in these early rounds of individual play—even though he or she had been given the opportunity to practice before the start of the game. In that case, learning effects could mask the fact that individuals are "really" faster than groups once they learn how to play the game—thus biasing our results toward showing no significant differences in average lags between individuals and groups.

We examined this possibility as follows. Discarding the data from Part One (rounds 1–10 of individual play), we compared *individual* decisions in Part Three (rounds 21–30) with *group* decisions made just before (in Part Two, rounds 11–20) and just after (in Part Four, rounds 31–40). Interestingly, in 12 of the 20 sessions, the average individual lag in Part Three was actually *longer than* the average group lag in Part Two. The difference is significant at conventional levels, regardless of whether we treat an individual decision as the unit of observation ($t = 2.2$) or the session as the unit of observation ($z = 1.9$). Comparing individual lags in Part Three with group lags in Part Four, however, shows a mixed picture ($t = 1.9$, $z = 0.6$). To summarize, there is no evidence to support the notion that learning accounts for our finding that group and individual lags are similar.

Another possible explanation for why we did not see any differences between individual and group lags is the phenomenon labeled the "risky shift" in the psychology literature. The risky shift is the observation that the diffusion of responsibility in groups leads them to take on more risk than they would as individuals. In our

21. The HAC estimator comes from a regression in which we allowed for first-order autocorrelation. The regression also included dummy variables for each session.

setting, one way in which groups could implement a riskier strategy would be to make decisions *sooner* (that is based on less information)—and this effect just might compensate for the group inertia that we expect to see (but do not find).

Fortunately, there is a second way in which groups could take more risk in our experiment: by making *larger* interest rate changes when they decide to move. Thus, if the risky shift is empirically important in our experiment, we should find that the *average size* of the initial interest rate change is larger for groups than for individuals. This hypothesis is directly testable with our data. In fact, the mean absolute value of the initial move is almost identical between the groups and the individuals. Moreover, the tiny difference is not close to being statistically significant (t = 0.11). This ancillary evidence leads us to doubt that group decisions were strongly influenced by risky shift considerations.

In sum, neither alternative statistical approaches nor learning effects nor the risky shift seem to explain away our counterintuitive finding that groups make decisions as quickly as individuals do.

H_2: Groups make better decisions than individuals.

A quite different hypothesis concerns the *quality* of decisionmaking, rather than the *speed*. Do groups make better decisions than individuals? Recall that in both our experimental setups, every subject has the same objective function and receives the same information. So, were our subjects to behave like *homo economicus*, they would all make the same decisions.

In reality, we all know that different people placed in identical situations often make different decisions. Furthermore, as we observed in the introduction and motivation section, many important economic and social decisions in the real world are assigned to groups rather than to individuals. Presumably, there is a reason. In any case, the hypothesis that groups outperform individuals is strongly supported by our experimental data.

Remember, we designed the experiment to yield an unambiguous measure of the quality of the decision: S ("score"), as defined by Equation (3). We scored (and paid) our *faux* monetary policymakers according to how well they kept unemployment near 5% and inflation near 2% over the entire 20-quarter game. As mentioned earlier, average scores were quite high—almost 86%. (We designed the experiment this way.) But the groups did significantly better than the individuals. The mean score over the 400 group observations was 88.3% (with standard deviation 4.7%), versus only 85.3% (standard deviation 10.1%) over the 2000 individual observations. This difference is both large enough to be economically meaningful and highly significant statistically ($t = 9.3$, if we treat each round as an independent observation; $t = 8.3$ with the HAC estimator that allows for autocorrelation and heteroskedasticity; or $z = 3.8$ using a Wilcoxon signed-ranks test, which treats each session as an observation).

Allowing for learning by comparing group play in Part Three with individual play in Parts Two and Four reiterates the same message: the groups outperform the individuals. We obtain t-statistics of 2.9 and 5.0, respectively, when treating an individual decision as the unit of observation, and z-statistics of 2.1 and 2.8, respectively, when treating the session as the unit of observation. All these results are highly significant.

Thus, in a nutshell, we find that group decisions are superior to individual decisions without being slower—which suggests that group decisions dominate individual decisions in this setting. Maybe two heads—or, in this case, five—really are better than one.

For subsequent comparison with the purely statistical experiment that we will discuss in Section 2, we also constructed a variable that indicates *directional accuracy.* Specifically, when G rises, subjects are supposed to *increase* interest rates; and when G falls, subjects are supposed to *decrease* interest rates. So define the dummy variable C ("correct") as 1 if the first interest rate change is made in the *same* direction as the G change, and 0 if it is made in the *opposite* direction. While the variable C does not enter the loss function *directly,* we certainly expect subjects to attain higher scores if their first move is in the right direction.[22]

Here, once again, groups outperformed individuals by a notable margin. The average value of C was 0.843 for individuals but 0.905 for groups. This difference is highly significant statistically ($t = 4.3$ when each individual decision is treated as an independent observation; $t = 4.2$ with the HAC estimator; $z = 3.5$ when we treat each session as an observation). Economically, it is even more noteworthy. When playing as individuals, our ersatz monetary policymakers moved interest rates in the *wrong* direction 15.7% of the time. When acting as a group, however, these same people got the direction wrong only 9.5% of the time. Looked at in this way, the "error rate" was reduced by about 40% when groups made decisions instead of individuals.

H_3: Decisions by majority rule are less inertial than decisions under a unanimity requirement.

Before we ran the experiment, we believed that requiring unanimous agreement would slow down the group decisionmaking process relative to using majority rule. But observing the subjects interacting face-to-face in real time showed something quite different. If you watched the game without having heard the instructions, it was hard to tell whether the game was being played under the unanimity principle or under majority rule. Perhaps it was peer group pressure, or perhaps it was simply a desire to be cooperative.[23] But for whatever reasons, majority decisions quickly evolved into unanimous decisions. In almost all cases, once three or four subjects agreed on a course of action, the remaining one or two fell in line immediately.[24]

Observationally, it was hard to tell whether groups were using majority voting or unanimous agreement to make decisions. Statistically, the mean lag under unanimity was indeed slightly longer than under majority rule—2.4 periods versus 2.2

22. And they do. The simple correlation between moving in the right direction initially and the final score is +0.37.

23. Students typically did not know one another prior to the experiment, though in some cases, purely by chance, they did.

24. One student noted that her group *unanimously* agreed to decide by *majority* vote.

periods—in conformity with H_3. However, the difference did not come close to statistical significance ($t = 0.9$ with either the conventional means test or the HAC estimator). When it came to average scores, the two decision rules finished in what was essentially a dead heat: 88.0% under majority rule, and 88.6% under unanimity. Hence, we pool observations from the majority-rule and unanimity treatments. The data support such pooling.

1.3 Other Findings

Learning. Having mentioned the issue of learning several times, we now turn to it explicitly. Because the dynamics of the monetary policy game are rather tricky, we suspected that there would be learning effects, at least in the early rounds: subjects would get better at the game as they played it more (up to a point). That is why we began each experimental session with a practice period in which subjects could familiarize themselves with the apparatus. Still, it is entirely possible that many students were not fully comfortable with the game when play started "for real."

While we performed a variety of simple statistical tests for learning, Figure 2 probably displays the results better than any regressions or *t*-tests. To construct this graph, we partitioned the data by round, reflecting the chronological order of play. There are 40 rounds in each session—20 played as individuals and 20 played as groups (see Table 1). So, for example, we have 100 observations (20 sessions times five individuals in each) on each of the first 10 rounds, 20 observations on each of rounds 11–20 (the 20 groups), and so on. Figure 2 charts the mean score by round. Vertical lines indicate the points where subjects switched from individual to group decisionmaking, or vice-versa, and horizontal lines indicate the means for each part of the experiment.

If there were continuous learning effects, scores should improve as we progress through the rounds. Figure 2 does give that rough impression—until you look closely. But careful inspection shows a rather different pattern. There is no indication whatsoever of any learning *within* any part of the experiment consisting of 10 rounds of play. However, the first experience with group play (rounds 11–20) not only yields better performance, but appears to make the individuals better monetary policymakers when they go back to playing the game alone (in rounds 21–30). Nonetheless, within that second batch of 10 rounds of individual play, average performance is inferior to what it was in the preceding ten rounds of group play. The pattern repeats itself when we compare rounds 21–30 of individual play with rounds 31–40 of group play.

So the conclusion seems clear: there is little or no evidence of learning, but overwhelming evidence for the superiority of groups over individuals. What learning there is appears to be *learning from others*, not *learning by doing*.

t-tests verify these graphical impressions. Looking first at individual play, the increase in mean score from Part One (rounds 1—10) to Part Three (rounds 21—30) is notable (3.2%) and extremely significant ($t = 6.1$). The standard deviation also drops markedly. All this suggests that substantial learning took place during Part Two. Learning effects were minor across the two rounds of group play—the

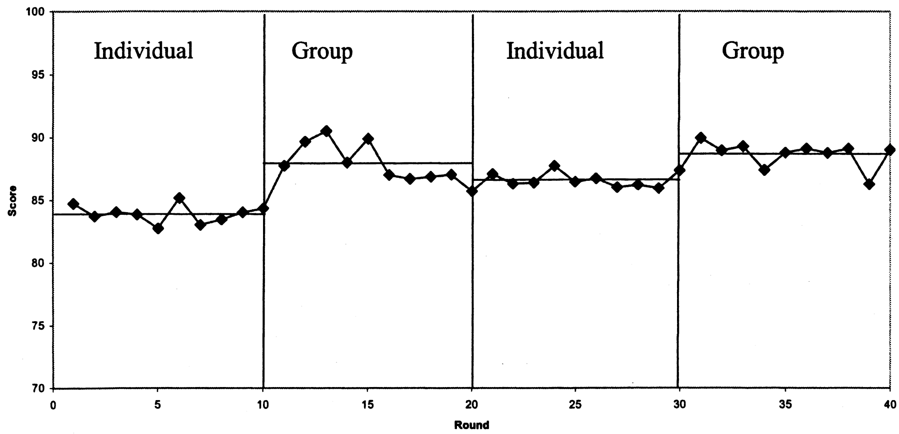

FIG. 2. Mean Score by Round in Monetary Policy Experiment

mean score in Part Four was just 0.9% higher than the mean score in Part Two. This improvement is not statistically significant ($t = 1.6$, $p = 0.12$).

Experimental order. In any experimental design, there is always a danger that results may be affected by the ordering of parts of the experiment. That is precisely why we arranged the parts of the experiment as we did: to have group play both precede and follow individual play, and to have unanimity both precede and follow majority rule. Nonetheless, the question remains: does ordering matter? Fortunately, the answer appears to be: no. Neither the scores from group play in Part Four nor the scores from individual play in Part Three appear to be affected by whether the subjects' first participation in group decisionmaking (in Part Two) was under majority rule or a unanimity requirement.

2. THE PURELY STATISTICAL EXPERIMENT[25]

2.1 Description of the Urn Experiment

Our second experiment placed subjects in a probabilistic environment devoid of any economic content, but designed to evoke the nature of monetary policy decisionmaking wherever possible. While such content-free problem solving may be of limited practical relevance, our motives were twofold. First, we wanted to create an experimental setting into which students would carry little or no prior intellectual baggage, so as to isolate the pure effect of individual versus group decisionmaking. Second, we wanted to see if our monetary policy results would be replicated in a quite different setting.[26]

25. This extremely short section omits much. For the details, see our NBER Working Paper No. 7909 (September 2000) of the same title.

26. Actually, we conducted the urn experiment first. We report them in reverse order because the monetary policy experiment is more interesting substantively.

Specifically, the problem was a variant of the classic "urn problem" in which subjects sample from an urn and then are asked to estimate its composition. In our application, groups of five students were placed in front of computers which were programmed with electronic "urns" consisting, initially, of 50% "blue balls" and 50% "red balls." They were told that the composition of the urn would change to either 70% blue or 70% red at some randomly-selected point (corresponding to the change in G in the monetary policy experiment). Once again, subjects were not told when the change would take place, nor in which direction. But they did know the probability law that governed the timing of the color change: uniform over the first 10 draws. Their task was to guess the post-change color composition of the urn—the equivalent of making a monetary policy decision. And they could make their guess whenever they felt they had enough information.[27]

As in the monetary policy experiment, we provided subjects with a clear loss function; it imposed a heavy penalty for guessing the wrong color (analogous to moving monetary policy in the wrong direction) and a cost for each draw (the decision lag). The sequence of play followed Table 1 closely, except that we did more repetitions because the simpler urn game went much faster than the more complex monetary policy game.

2.2 The Three Main Hypotheses

We were gratified to find that the urn experiment produced almost exactly the same answers to our three main questions as the monetary policy experiment, which gives us some confidence in the robustness of our results. Briefly:

H_1: Groups decisions are more inertial than individual decisions: In the urn experiment, the two mean lags were again not significantly different at conventional levels ($t = 1.1$).[28] But this time the average lag was actually longer for the groups: 6.60 draws versus 6.40 draws for individuals. Histograms for the variable L (the decision lag) in individual and group play look strikingly like Figure 1. We performed the same sorts of tests for learning effects, with results similar to those in the monetary policy experiment: learning occurred, but it was dwarfed by the difference in quality between individual and group decisions.

H_2: Groups make better decisions than individuals: The experimental data generated by the urn experiment strongly support the hypothesis that groups outperform individuals; efforts to account for learning effects do not upset this conclusion. The average score attained by the groups was 86.8 (on a 1–100 scale), versus only 83.7 for individuals. The difference is highly significant statistically ($t = 4.6$).[29] More

27. Actually, we imposed an upper limit on the number of draws. But it was binding in only five of 4200 cases.

28. The t-statistic for the HAC estimation is identical. Treating the session as the unit of observation and performing a rank-sum test yields a z-statistic of 0.4, which is also not significant at conventional levels.

29. The conventional and HAC estimators are identical. In the nonparametric Wilcoxon test which treats the *session* as the unit of observation, the z-statistic is 3.2—which is significant at any conventional level.

important, it seems large enough to be economically meaningful.[30] Strikingly, the 3.7% performance gap between groups and individuals almost exactly matches the 3.5% gap that we found in the monetary policy experiment. Even if we had tried to "rig the deck" to make the two performance gaps come out the same, we would have had no idea how to do so.

Obviously, since the mean lags are statistically indistinguishable, the groups must have acquired their overall edge through *accuracy* rather than through *speed*. Specifically, groups guessed the urn's composition correctly 89.3% of the time whereas individuals got the color right only 84.3% of the time. This gap of 5 percentage points is sizable, strikingly similar to what we found in the monetary policy experiment (6.2 percentage points), and statistically significant ($t = 4.5$ or $z = 1.9$).

H_3: Decisions by majority rule are less inertial than decisions under a unanimity requirement: The urn experiment repeated the surprising result we found in the monetary policy experiment: there were almost no differences between groups operating under majority rule and groups operating under the unanimity principle. In fact, contrary to our priors, decisions were actually made faster under unanimity.

3. CAN WE MODEL GROUP DECISIONMAKING?

It is possible to formulate and test several simple models of how groups aggregate individual views into group decisions. None of these are strictly "economic" models, however, because every *homo economicus* would make the *same* decision—after all, both the objective function and the information are identical for all participants. As will be clear shortly, none of these simple, intuitive models of group decisionmaking takes us very far.

3.1 Model 1: The Whole is Equal to the Sum of Its Parts

The simplest model posits that there are no group interactions at all: the group's decision is simply the *average* of the five individual decisions. This, of course, come closest to the pure economic model, which says that everyone agrees. However, this model has, essentially, already been tested and rejected in Sections 1 and 2. Let X denote any one of our three decision-related variables (L, S, or C), and let X_G be the average value attained by the group and X_A be the average values attained by the five people in the group *when they played as individuals*. As noted earlier, we consistently reject $X_G = X_A$ in favor of the alternative that groups do better.

So let us ask a slightly different question: looking across the 20 groups, does the average performance of the five people who comprise a particular group (X_A) take us very far in explaining—in a regression sense—how well the *group* does on that same criterion (X_G)? Since we have three different choices of X (L, S, and C) and data from two different experiments, we can pose six versions of this question. Rather than display the (rather unsuccessful) regression equations, Figure 3 shows

30. That difference is about 72% of the standard deviation across individual mean scores.

the corresponding scatter diagrams. Each is based on 20 observations, one for each session. What message do these six charts convey?

In general, they give the impression that a linear model of the form $X_G = a + bX_A + u$ does not fit the data at all well.[31] In one case, the correlation is even negative. Looking across the three variables, L_A does by far the best job of explaining L_G, although even here the simple correlations are just 0.58 in the urn experiment and 0.57 in the monetary policy experiment—corresponding to R^2's of about 0.33. (The regression coefficients are 0.84 and 0.90, respectively.) In the monetary policy experiment, the correlations for the other two variables, S (score) and C (percent correct), are nearly zero.

In a word, the average performance of the five individuals who comprise each group carries little explanatory power for how well the group performed. Most championship teams would be surprised—and would be spending too much on payroll—if this were true in professional sports.

3.2 Model 2: The Median Voter Theory

A different concept of "average" plays a time-honored role in one of the few instances of group decisionmaking that economists have modeled extensively: voting. Where preferences are single-peaked, as they must be in these applications, a highly-pedigreed tradition in public finance holds that the views of the median voter should prevail. It seems natural, then, to ask whether the performance of the *median player* can explain the performances of our five-person groups? Remember, we literally used either a majority vote or a unanimous vote to determine the group's decisions in our experiments.

Figure 4, which follows the same format as Figure 3, shows that the median voter model generally (but not always) is a better predictor of group outcomes than is simple averaging. In one case, the R^2 gets as high as 0.54. But, in general, these six scatters once again show that even the median-voter model has only modest success (and, in some cases, no success at all) in explaining the performance of the group. As before, the groups' L decisions are explained best; the R^2's of the two regressions are 0.54 for the urn data and 0.42 for the monetary policy data. In two cases (variables S and C in the monetary policy experiment), the correlation is actually negative.

3.3 Model 3: May the Best Man (or Woman) Win

In discussing our experiment with other economists, several confidently suggested that the group's decisions would be dominated by the *best player* in the group—as indicated, presumably, by his or her scores while playing alone.[32] This hypothesis struck us as plausible. So we tested models of the form $X_G = a + bX^* + u$, where X^* is the average outcome (on variable S, C, or L) of the individual who achieved the *highest* average score while playing alone.

31. It is apparent from the diagrams that linearity is not the issue. No obvious nonlinear model does much better.

32. The subject pool was very close to 50% male and 50% female.

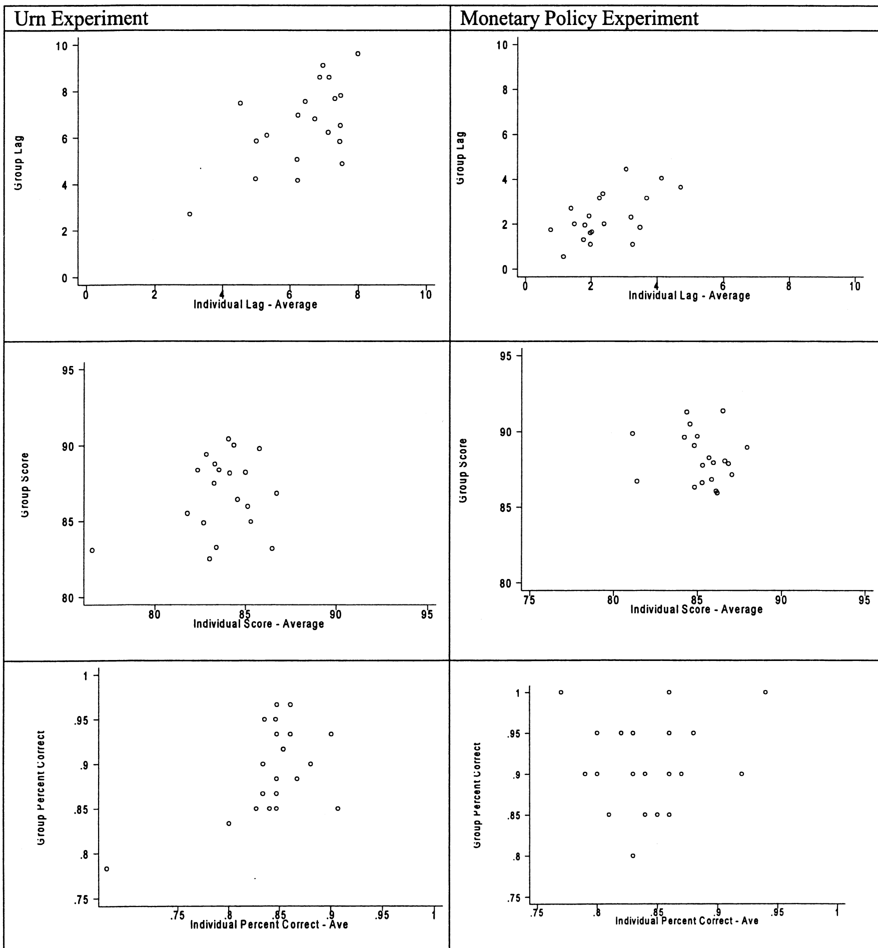

FIG. 3. Group Compared to Average Individual Play

There is, however, a logically prior question that we raised early in the paper: are there actually statistically significant individual fixed effects that can be used to identify "better" and "worse" players? To answer this question, we ran a series of regressions, one for each experimental session, explaining individual scores by five dummy variables, one for each player.[33] Perhaps surprisingly, this test of the idea that there is a "best player" turned up only very weak evidence that some players are better and others worse in the monetary policy experiment: 15 of the 100 individual dummies were significant at the 5% level. In the urn experiment, there

33. Thus, each regression was based on 150 observations in the urn experiment and 100 observations in the monetary policy experiment.

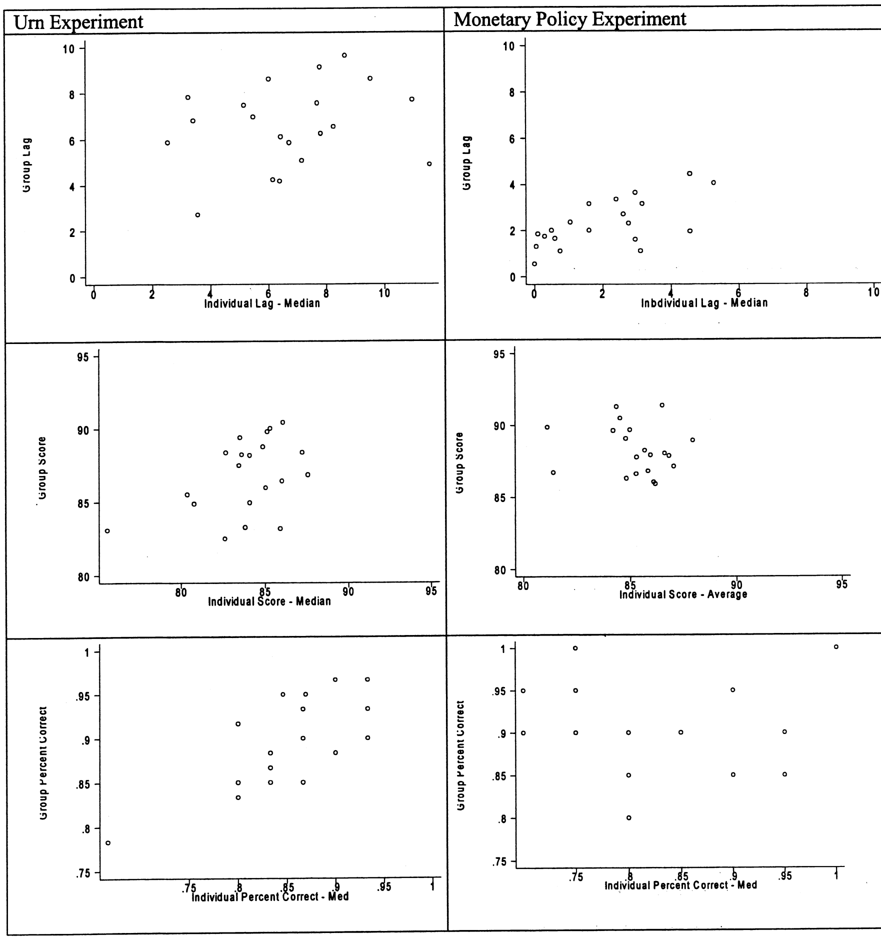

FIG. 4. Group Compared to Median Individual Play

was absolutely no evidence of individual fixed effects: only four of the 100 individual dummy variables were significant at the 5% level. A group can hardly be dominated by its best player, if there is no best player.

With this in mind, we can now look at Figure 5, which displays the six scatter diagrams. In general, the fits appears to be quite modest. (The highest R^2 among the six scatters is 0.28.) In only one of the six cases (explaining C_G in the monetary policy experiment), is this the best-fitting model; in three cases, it is the worst. Once again, the variable L is explained best.

Finally, we note that various multiple regressions using, say, both X_A and X^* do not appreciably improve the fit. In the end, we are left to conclude that neither the average player, nor the median player, nor the best player determine the decisions

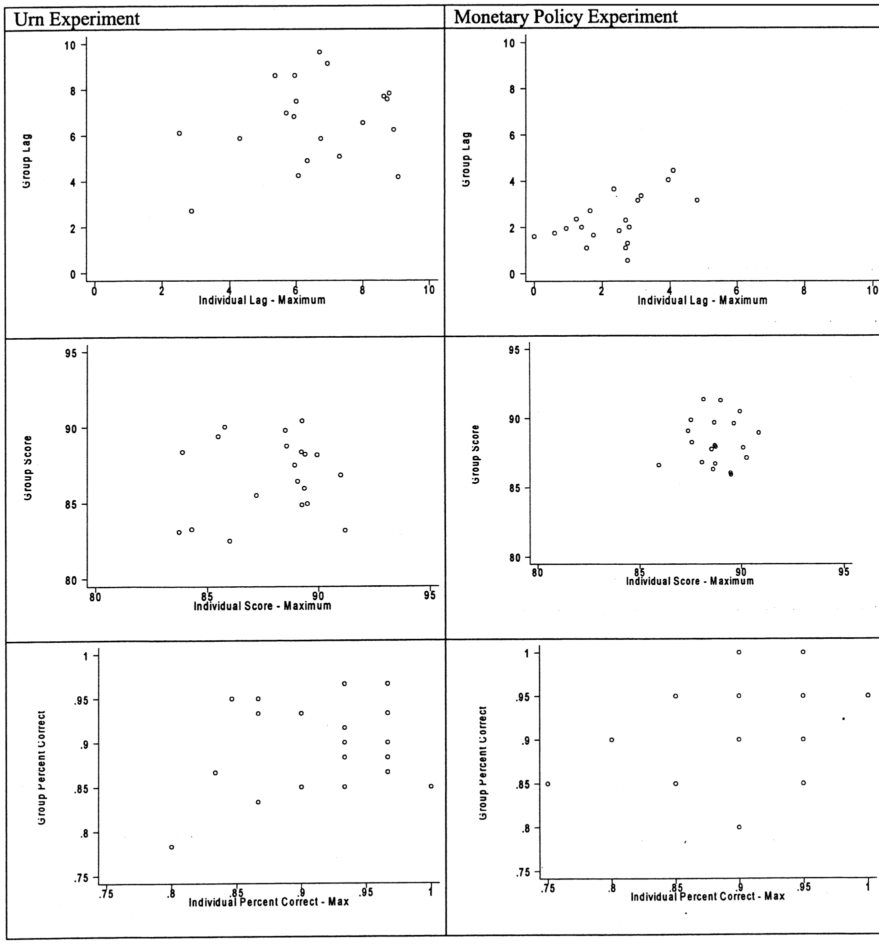

FIG. 5. Group Compared to Maximum Individual Play

of the group. The whole, we repeat, does indeed seem to be something different from—and generally better than—the sum of its parts.

4. CONCLUSIONS

Perhaps the best way to illustrate the similarity in findings from these two very different experiments is to rack them up, side by side, as we do in Table 2: while there are some minor differences (noted above) between the results of the urn experiment and those of the monetary policy experiment, the correspondence is little short of amazing.

TABLE 2
COMPARISON OF EXPERIMENTS

	Monetary policy experiment	Urn experiment
1.	Groups are not slower	Groups are not slower
2.	Groups score better by 3.5%	Groups score better by 3.7%
3.	Majority rule approximately the same as unanimity	Majority rule approximately the same as unanimity
4.	Early learning does not improve scores	Early learning improves scores
5.	Simple models of group behavior fit poorly	Simple models of group behavior fit poorly
6.	Some significant individual effects	No significant individual effects

From the start, our interest centered on the first two findings:

1. *Do groups reach decisions more slowly than individuals?* According to these experimental results, groups appear to be no slower in reaching decisions than individuals are.
2. *Do groups make better decisions than individuals?* The experimental answer seems to be yes. And the margin of superiority of group over individual decisions is astonishingly similar in the two experiments—about 3.5%.

If groups make better decisions and require no more information to do so, then two heads—or, in this case, five—are indeed better than one. Society is, in that case, wise to assign many important decisions, like monetary policy, to committees.

LITERATURE CITED

Arrow, Kenneth J. (1963). *Social Choice and Individual Values.* New York: Wiley.

Ball, Laurence (1997). "Efficient Rules for Monetary Policy." NBER Working Paper No. 5952 (March).

Blinder, Alan S. (1998). *Central Banking in Theory and Practice.* Cambridge, MA: MIT Press.

Blinder, Alan S. (2004). *The Quiet Revolution: Central Banking Goes Modern.* New Haven, CT: Yale University Press.

Bornstein, Gary, and Ilan Yaniv (1998). "Individual and Group Behavior in the Ultimatum Game: Are Groups more 'Rational' Players?" *Experimental Economics* 1, 101–108.

Cason, Timothy N., and Vai-Lam Mui (1997). "A Laboratory Study of Group Polarization in the Team Dictator Game." *Economic Journal* 107, 1465–1483.

Cox, James C., and Stephen C. Hayne (1998). *Group Versus Individual Decision Making in Strategic Market Games.* Mimeo, University of Arizona.

Gerlach-Kristen, Petra (2003). "Monetary Policy Committees and Interest-rate Setting." *European Economic Review*, forthcoming.

Gersbach, Hans, and Volker Hahn (2001). "Voting Transparency and Conflicting Interests in Central-bank Councils." Deutsche Bundesbank Economic Research Center Discussion Paper No. 03/01.

Goodfriend, Marvin (2000). "The Role of a Regional Bank in a System of Central Banks." *Economic Quarterly, Federal Reserve Bank of Richmond* (Winter 2000), 7–25.

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

Kocher, Martin G., and Matthias Sutter (2000). "When the 'Decision-maker' Matters: Individual versus Team Behavior in Experimental 'Beauty Contest' Games." University of Innsbruck Discussion Paper No. 2000/4.

J.P. Morgan & Co. (2000). *Guide to Central Bank Watching*, March 2000.

Newey, Whitney, and Kenneth West (1987). "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55, 703–708.

Rudebusch, Glenn D., and Lars E.O. Svensson (1999). "Policy Rules for Inflation Targeting." In *Monetary Policy Rules*, edited by J.B. Taylor, pp. 203–246. Chicago: University of Chicago Press.

Wallach, Michael A., Nathan Kogan, and Daryl J. Bem (1964). "Diffusion of Responsibility and Level of Risk Taking in Groups." *Journal of Abnormal Social Psychology* 68, 263–274.