Macroeconomics: A Survey of Laboratory Research*

John Duffy
Department of Economics
University of Pittsburgh
Pittsburgh, PA 15260 USA
Email: jduffy@pitt.edu

March 2008
Comments and Additional Citations Welcome

Abstract

This chapter surveys laboratory experiments addressing macroeconomic phenomena. The first part focuses on experimental tests of the microfoundations of macroeconomic models discussing laboratory studies of intertemporal consumption/savings decisions, time (in)consistency of preferences and rational expectations. Part two explores coordination problems of interest to macroeconomists and mechanisms for resolving these problems. Part three looks at experiments in specific macroeconomic sectors including monetary economics, labor economics, international economics as well-as large scale, multi-sector models that combine several sectors simultaneously. The final section addresses experimental tests of macroeconomic policy issues.

1 Introduction: Laboratory Macroeconomics

> "Economists can do very little experimentation to produce crucial data. This is particularly true of macroeconomics." Christopher A. Sims (1996, p. 107)

Macroeconomic theories have traditionally been tested using non-experimental field data, most often national income account data on GDP and its components. This practice follows from the widely-held belief that macroeconomics is a purely observational science: history comes around just once and there are no "do-overs". Controlled manipulation of the macroeconomy to gain insight regarding the effects of alternative institutions or policies is viewed as impossible, (not to mention unethical), and so, apart from the occasional *natural* experiment, many would argue that macroeconomic questions cannot be addressed using experimental methods.[1]

*Chapter prepared for the Handbook of Experimental Economics Vol. 2, J. Kagel and A.E. Roth, Eds. Frank Heinemann, Rosemarie Nagel, Charles Noussair and Andreas Ortmann provided helpful comments on an earlier draft.

[1] Indeed, the term "macroeconomic experiment" does not even typically refer to laboratory experiments involving human subjects but rather to *computational experiments* using calibrated dynamic stochastic general equilibrium models as pioneered in the work of Finn Kydland and Edward Prescott (1982). Even these experimental exercises have been ruled out as unacceptable by some. Sims again (1996 p. 113) writes: "What Kydland and Prescott call computational experiments are computations not experiments. In economics, unlike experimental sciences, we cannot create observations designed to resolve our uncertainties about theories; no amount of computation can resolve that."

Yet, as this survey documents, over the past twenty years, a wide variety of macroeconomic models and theories *have* been examined using controlled laboratory experiments with paid human subjects, and this literature continues to grow. The use of laboratory methods to address macroeconomic questions has come about in large part due to changes in macroeconomic modeling, though it has also been helped along by changes in the technology for doing laboratory experimentation, especially the use of large computer laboratories. The change in macroeconomic modeling is, of course, the now widespread use of explicit micro-founded models of constrained, intertemporal choice in general equilibrium, game-theoretic or search-theoretic frameworks. The focus of these models is often on how institutional changes or policies affect the choices of decision-makers such as household and firms, in addition to the more traditional concern with responses in the aggregate time series data (e.g. GDP) or to the steady states of the model. While macroeconomic models are often expressed at an aggregate level, for instance there is a "representative" consumer or firm or a market for the "capital good", an implicit, working assumption of many macroeconomists is that aggregate sectoral behavior is not different from that of the individual actors or components that comprise each sector.[2] Otherwise, macroeconomists would be obliged to be explicit about the mechanisms by which individual choices or sectors aggregate up to the macroeconomic representations they work with, and macroeconomists have been largely silent on this issue. Experimentalists testing non-strategic macroeconomic models have sometimes taken this representativeness assumption at face value, and conducted individual decision-making experiments with a macroeconomic flavor. But, as we shall see, experimentalists have also considered whether small groups of subjects interacting with one another via markets or by observing or communicating with one another might outperform individuals in tasks that macroeconomic models assign to representative agents.

To date, the main insights gained from macroeconomic experiments include 1) an assessment of the micro-assumptions underlying macroeconomic models, 2) a better understanding of the dynamics of expectations which play a critical role in macroeconomic models, 3) a means of resolving equilibrium selection (coordination) problems in environments with multiple equilibria, 4) validation of macroeconomic model predictions for which field data are not available and 5) the impact of various macroeconomic policy interventions on individual behavior. In addition, laboratory tests of macroeconomic theories have generated new or strengthened existing experimental methodologies including implementation of the representative agent assumption, overlapping generations and search-theoretic models, discounting and infinite horizons, methods for assessing equilibration, and the role of various market clearing mechanisms in characterizing Walrasian competitive equilibrium (for which the precise mechanism of exchange is left unmodeled).

The precise origins of "macroeconomic experiments" is unclear. Some might point to A.W. Phillips' (1950) experiments using a colored liquid-filled tubular flow model of the macroeconomy, though this did not involve human subjects! Others might cite Vernon Smith's (1962) double auction experiment demonstrating the importance of centralized information to equilibration to competitive equilibrium as the first macro-economic experiment. Yet another candidate might be John Carlson's early (1967) experiment examining price expectations in stable and unstable versions of the Cobweb model. However, I will place the origins more recently with Lucas's 1986

[2] Of course, this assumption is generally false. As Fisher (1987) points out in his *New Palgrave* entry on *aggregation problems*, "the analytic use of such aggregates as 'capital', 'output', 'labour' or 'investment' as though the production side of the economy can be treated as a single firm is without sound foundation." Fisher adds that "this has not discouraged macroeconomists from continuing to work in such terms." Indeed, one may think of macroeconomics as an impure language with bad grammar and borrowed words but a language nonetheless, and one with many users.

invitation to macroeconomists to conduct laboratory experiments to resolve macro-coordination problems that were unresolved by theory. Lucas's invitation was followed up on by Lim, Prescott and Sunder (1994) and Marimon and Sunder (1993, 1994, 1995), and perhaps as the result of their interesting and influential work, over the past decade there has been a great blossoming of research testing macroeconomic theories in the laboratory. This literature is now so large that I cannot hope to cover every paper in a single chapter, but I do hope to give the reader a good road-map as to the kinds of macroeconomic topics that have been studied experimentally as well as to suggest some further extensions.

How shall we define a macroeconomic experiment? One obvious dimension to consider is the number of subjects. Perhaps Sims's (and others') belief that macroeconomic models are difficult to test experimentally hinges on the problem of approximating the large scale of the macroeconomy with the necessarily small numbers of subjects available in laboratory experiments. However, as we have already noted, in micro-founded models, this issue of the number of subjects employed need not be the relevant focus. Further, as discussed in this survey, the evidence from many double auction experiments beginning with Smith (1962) reveals that equilibration to competitive equilibrium occurs reliably with just a few (3-5) buyers or sellers on each side of the market, so large numbers of subjects need not be necessary even in non-strategic market environments.

A second and more sensible approach is simply to define a macroeconomic experiment as one that tests the predictions of a macroeconomic model or its assumptions, or is framed in the language of macroeconomics, involving for example, intertemporal consumption and savings decisions, inflation and unemployment, economic growth, bank runs, monetary exchange, monetary or fiscal policy or any other macroeconomic phenomena. Unlike microeconomic models and games which often strive for generality, macroeconomic models are typically built with a specific macroeconomic story in mind that is not as easily generalized to other non-macroeconomic settings. For this reason, our definition of a macroeconomic experiment may be too restrictive. There are many microeconomic experiments - coordination games for instance - that can be given a macroeconomic interpretation. In discussing those studies as macroeconomic experiments, I will attempt to emphasize the macroeconomic interpretation.

The coverage of this chapter can be viewed as an update on topics covered in several chapters of the first volume of the *Handbook of Experimental Economics*, including discussions of intertemporal decision-making by Camerer (1995), coordination problems by Ochs (1995) and asset prices by Sunder (1995), though the coverage here will not be restricted to these topics alone. Most of the literature surveyed here was published since 1995, the date of the first *Handbook* volume. In addition, this chapter builds on, complements and extends earlier surveys of the macroeconomic experimental literature by myself, Duffy (1998, 2008), and by Ricciuti (2004).

2 Dynamic, Intertemporal Optimization

Perhaps the most widely-used model in modern macroeconomic theory is the one-sector, infinite horizon optimal growth model pioneered by Ramsey (1928) and further developed by Cass (1965) and Koopmans (1965). This model posits that individuals solve a dynamic, intertemporal optimization problem in deriving their consumption and savings plan over an infinite horizon. Both deterministic and stochastic versions of this model are workhorses of modern real business cycle theory and growth theory.

In the urge to provide microfoundations for macroeconomic behavior, modern macroeconomists assert that the behavior of consumers or firms can be reduced to that of a representative, fully rational individual actor; there is no room for any "fallacies of composition" in this framework. It is therefore of interest to assess the extent to which macroeconomic phenomena can be said to reflect the choices of individuals facing dynamic stochastic intertemporal optimization problems. Macroeconomists have generally ignored the plausibility of this choice-theoretic assumption preferring instead to examine the degree to which the time series data on GDP and its components move in accordance with the conditions that have been optimally derived from the fully rational representative agent model and especially whether these data react predictably to shocks or policy interventions.

2.1 Optimal Consumption/Savings Decisions

Whether individuals can in fact solve a dynamic stochastic intertemporal optimization problem of the type used in the one-sector optimal growth framework has been the subject of a number of laboratory studies, including Hey and Dardanoni (1988), Carbone and Hey (2004), Noussair and Matheny (2000), Lei and Noussair (2002), and Ballinger et al. (2003). These studies take the representative agent assumption of modern macroeconomics seriously and ask whether subjects can solve a discrete-time optimization problem of the form:

$$\max_{\{c_t\}} E_t \sum_{t=0}^{\infty} \beta^t u(c_t)$$

subject to:

$$c_t + x_t \leq \omega_t,$$

where c_t is time t consumption, $u(\cdot)$ is a concave utility function, β is the period discount factor, x_t represents time t savings (if positive) or borrowings (if negative) and ω_t is the household's time t wealth.

Hey and Dardanoni (1988) assume an endowment economy, where wealth evolves according to $\omega_t = R(\omega_{t-1} - c_{t-1}) + y_t$, with $\omega_0 > 0$ given. Here, R denotes the (constant) gross return on savings and y_t is the stochastic time t endowment of the single good; the mean and variance of the stochastic income process is made known to subjects. By contrast, Noussair and associates assume a non-stochastic production economy, where $\omega_t = f(k_t) + (1-\delta)k_t$, with $f(\cdot)$ representing the known, concave production function, k_t denoting capital per capita and δ denoting the depreciation rate. In this framework, it is public knowledge that all of an individual's savings s_t, are invested in capital and become the next period's capital stock, i.e., $x_t = k_{t+1}$. The dynamic law of motion for the production economy is expressed in terms of capital rather than wealth: $k_{t+1} = f(k_t) + (1-\delta)k_t - c_t$, with $k_0 > 0$ given. The gross return on savings is endogenously determined by $R = f'(k_t) + (1-\delta)$.

Solving the maximization problem given above, the first order conditions imply that the optimal consumption program must satisfy the Euler equation:

$$u'(c_t) = \beta R E_t u'(c_{t+1}),$$

where the expectation operator is with respect to the (known) stochastic process for income (or wealth). Notice that the Euler equation predicts a monotonic increasing, decreasing or constant consumption sequence, depending on whether βR is less than, greater than or equal to 1. Solving for a consumption or savings *function* involves application of dynamic programming techniques that break the optimization problem up into a sequence of two-period problems; the Euler equation above characterizes the dynamics of marginal utility in any two periods. For most specifications of preferences, analytic closed-form solutions for the optimal consumption or savings function are not possible, though via concavity assumptions, the optimal consumption/savings program can be shown to be unique.

In testing this framework, Hey and Dardanoni (1988) addressed several implementation issues. First, they had to rule out borrowing (negative saving) so as to prevent subjects from ending the session in debt. Second, they attempted to implement discounting and the stationarity associated with an infinite horizon by having a constant probability that the experimental session would continue with another period. Finally, rather than inducing a utility function, they supposed that all subjects had constant absolute risk aversion preferences and they estimated each individual subject's coefficient of absolute risk aversion using data they gathered from hypothetical and paid choice questions presented to the subjects. Given this estimated utility function, they then numerically computed optimal consumption for each subject and compared it with their actual consumption choice. To challenge the theory, they consider different values for R and β as well as for the parameters governing the stochastic income process, y.

They report mixed results. First, consumption is significantly different from optimal behavior; in particular, there appears to be great time-dependence in consumption behavior, i.e., consumption appears dependent on past income realizations, which is at odds with the time-independent nature of the optimal consumption program. Second, they find support for the comparative static implications of the theory. That is, changes in the discount factor β or in the return on savings, R have the same effect on consumption as under optimal consumption behavior. So they find mixed support for dynamic intertemporal optimization.

Carbone and Hey (2004) attempt to simplify the design of Hey and Dardanoni in several respects. First, they eliminate discounting and consider a finite horizon, 25 period model. They argue, based on the work of Hey and Dardanoni, that subjects "misunderstand the stationarity property" of having a constant probabilistic stopping rule. Second, they greatly simplify the stochastic income process, allowing there to be just two values for income – one "high" which they refer to as a state where the consumer is "employed" and the other "low" in which state the consumer is "unemployed." They use a two-state Markov process to model the state transition process: conditional on being employed (unemployed), the probability of remaining (becoming) employed was p(q), and these probabilities were made known to subjects. Third, rather than infer preferences they induce a constant absolute risk aversion utility function. Their treatment variables were p, q, R and the ratio of employed to unemployed income; they considered two values of each, one high and one low, and examined how consumption changed in response to changes in these treatment variables relative to the changes predicted by the optimal consumption function (again numerically computed). Table 1, shows a few of their comparative static findings.

An increase in the probability of remaining employed caused subjects to overreact in their choice of additional consumption relative to the optimal change regardless of their employment status (Unemployed or employed), whereas an increase in the probability of becoming employed – a decrease in the probability of remaining unemployed – led to an under-reaction in the amount of additional consumption chosen relative to the optimal prediction. On the other hand, the effect of a change in the ratio of high-to-low income on change in consumption was quite close to the optimum change. Carbone and Hey emphasize also that there was tremendous heterogeneity in terms of subjects' abilities to confront the life-cycle consumption savings problem, with most

Change (Δ) in treatment variable	Unemployed		Employed	
(from low value to high value)	Optimal	Actual	Optimal	Actual
Δp (Pr. remaining employed)	5.03	23.64	14.57	39.89
Δq (Pr. becoming employed)	14.73	-1.08	5.68	0.15
Δ ratio high-low income	0.25	0.24	0.43	0.76

Table 1: Average Change in Consumption in Response to Parameter Changes and Conditional on Employment Status, taken from Carbone and Hey (2004,Table 5).

applying too short, or too variable a planning horizon to conform to optimal behavior. They conclude that "subjects do not seem to be able to smooth their consumption stream sufficiently – with current consumption too closely tracking current income." Interestingly, the *excess sensitivity* of consumption to current income (in excess of that warranted by a revision in expectations of future income) is a well-documented empirical phenomenon in studies of consumption behavior using field data (see, e.g., Flavin (1981), Hayashi (1982), or Zeldes (1989)). This corroboration of evidence from the field should give us further confidence in the empirical relevance of the laboratory analyses of intertemporal consumption-savings decisions. Two explanations for the excess sensitivity of consumption to income that have appeared in the literature are 1) binding liquidity constraints and 2) the presence of a precautionary savings motive, (which is more likely in a finite horizon model). Future experimental research might explore the relative impacts of these two factors on consumption decisions.

Noussair and Matheny (2000) depart from the work of Hey and associates by adding a concave production technology, $f(k_t) = Ak_t^{\alpha}$, $\alpha < 1$, which serves to endogenize the return on savings in conformity with modern growth theory. They induce both the production function and a logarithmic utility function by giving subjects schedules of payoff values for various levels of k and c, and they implement an infinite horizon by having a constant probability that a sequence of rounds continues. Subjects made savings decisions (chose $x_t = k_{t+1}$) with the residual from their budget constraint representing their consumption. Noussair and Matheny varied two model parameters, the initial capital stock, k_0 and the production function parameter α. Variation in the first parameter changes the direction by which paths for consumption and capital converge to steady state values (from above or below) while variations in the second parameter affect the predicted *speed* of convergence; the lower is α, the greater is the speed of convergence of the capital stock and consumption to the steady state of the model. Among the main findings, Noussair and Matheny report that sequences for the capital stock are monotonically decreasing regardless of parameter conditions and theoretical predictions with regard to speed of convergence do not find much support. Consumption is, of course linked to investment decisions and is highly variable. They report that subjects occasionally resorted to consumption *binges*, allocating nearly nothing to the next period's capital stock, in contrast to the prediction of consumption smoothing, however, this behavior seemed to lessen with experience. A virtue of the Noussair-Matheny study is that it was conducted with both U.S. and Japanese subjects, with similar findings for both countries.

One explanation for the observed departure of behavior from the dynamically optimal path is that the representative agent assumption, while consistent with the reductionist view of modern macroeconomics, assumes too much individual rationality to be useful in practice.[3] Information on market variables (e.g., *prices*) as determined by *many different interacting agents*, may be a necessary aid to solving such complicated optimization decisions. An alternative explanation may be that the standard model of intertemporal consumption smoothing abstracts away from the importance of *social norms* of behavior with regard to consumption decisions. Akerlof (2007), for instance, suggests that people's consumption decisions may simply reflect their "station in life". College students (the subjects in most of these experiments) looking to their peers, choose to *live like college students* with expenditures closely tracking income. Both of these alternative explanations have been considered to some extent in further laboratory studies.

Ballinger et al. (2003) explore the role of social learning in a modified version of the noisy endowment economy studied by Hey and Dardanoni (1988). In particular, they eliminate discounting (presumably to get rid of time dependence) focusing on a finite 60-period horizon. Subjects are matched into three-person "families" and make decisions in a fixed sequence. The generation 1 (G1) subject makes consumption decisions alone for 20 periods; in the next 20 periods (21-40) his behavior is observed by the generation 2 (G2) subject, and in one treatment, the two are free to communicate with one another. In the next twenty periods (periods 41-60 for G1), (periods 1-20 for G2), both make consumption/savings decisions. The G1 subject then exits the experiment. The same procedure is then repeated with the generation 3 (G3) subject watching the G2 subject for the next twenty rounds etc. Unlike Hey and Dardanoni, Ballinger et al. induce a constant relative risk aversion utility function on subjects using a Roth-Malouf (1979) binary lottery procedure. This allows them to compute the path of optimal consumption/savings behavior. These preferences give rise to a precautionary savings motive wherein liquid wealth (saving) follows a hump-shaped pattern over the 60-period lifecycle.

Ballinger et al.'s (2003) main treatment variable concerns the variance of the stochastic income process (high or low) which affects the peak of the precautionary savings hump; in the high case they also explore the role of allowing communication/mentoring or not (while maintaining observability of actions by overlapping cohorts at all times). Among their findings, they report that subjects tend to consume more than the optimal level in the early periods of their lives leading to less savings and below optimal consumption in the later periods of life. However, savings is greater in the high as compared with the low variance case which is consistent with the comparative static prediction of the rational intertemporal choice framework. They also find evidence for time-dependence in that consumption behavior is excessively sensitive to near lagged changes in income. Most interestingly, they report that consumption behavior of cohort 3 is significantly closer to the optimal consumption program than in the consumption behavior of cohort 1 suggesting that social learning by observation plays an important role, and may be a more reasonable characterization of the representative agent.

Lei and Noussair (2002) pursue a complementary approach in the optimal growth model with capital. They contrast the "social planner" case, where a single subject is charged with maximizing the representative consumer-firms' present discounted sum of utility from consumption over an indefinite horizon (as in Noussair and Matheny (2000)), with a decentralized market approach wherein the same problem is solved by five subjects looking at price information. In this market treatment the production and utility functions faced by the social planner are disaggregated into five individual functions assigned to the five subjects; that aggregate up to the same functions faced by the social planner. For example, some subjects had production functions with marginal

[3] See Kirman (1992) for a discussion of the limitations of the representative agent assumption.

products for capital that were higher than for the economy-wide production function while others had marginal products for capital that are lower. At the beginning of a period, production took place, based on previous period's capital, using either the individual production functions in the market treatment or the economy-wide production function in the social planner treatment. Next, in the market treatment, a double auction market for output (or potential future capital) opened up. Agents with low marginal products of capital could trade some of their output to agents with high marginal products for capital in exchange for experimental currency units (subjects were given an endowment of such units each period, which they had to repay). The import of this design was that the market effectively communicated to the five subjects the market price of a unit of output (or future capital). As future capital could be substituted one-for-one with future consumption, the market price of capital revealed to subjects the marginal utility of consumption. After the market for output closed, subjects in the market treatment could individually allocate their adjusted output levels between future capital k_{t+1} or savings, and experimental currency units or consumption c_t. By contrast, in the social planner treatment, there was no market for output; the representative individual proceeded directly to the step of deciding how to allocate output between future capital (savings) and current consumption. At the end of the period, subjects' consumption amounts were converted into payoffs using the economy-wide or individual concave utility functions and loans of experimental currency units in the market treatment were repaid.

[Insert Figure 1 here.]

The difference in consumption behavior between the market and representative agent-social planner treatments is illustrated in Figure 1, which shows results from a representative session of one of Lei and Noussair's treatments. In the market treatment, there was a strong tendency for consumption (as well as capital and the price of output) to converge to their unique steady state values, while in the social planner treatment, consumption was typically below the steady state level and much more volatile.

In further analysis, Lei and Noussair (2002) make use of a linear, panel data regression model to assess the extent to which consumption and savings (or any other time series variable for that matter) can be said to be converging over time toward predicted (optimal) levels.[4] In this regression model, $y_{j,t}$ denotes the average (or economy-wide level) of the variable of interest by cohort/session j in period $t = 1, 2, ...$ and D_j is a dummy variable for each of the $j = 1, 2..., J$ cohorts. The regression model is written as:

$$y_{j,t} = \alpha_1 \frac{D_1}{t} + \alpha_2 \frac{D_2}{t} ... + \alpha_J \frac{D_J}{t} + \beta \frac{t-1}{t} + \epsilon_{j,t}, \quad (1)$$

where $\epsilon_{j,t}$ is a mean zero, random error term. The α_j coefficients capture the initial starting values for each cohort while the β coefficient captures the asymptotic value of the variable y to which all J cohorts of subjects are converging; notice that the α coefficients have a full weight of 1 in the initial period 1 and then have exponentially declining weights while the single β coefficient has an initial weight of zero that increases asymptotically to 1. For the dependent variable in (1), Lei and Noussair (2002) use: 1) the consumption and capital stocks (savings) of cohort j, $c_{j,t}$, and $k_{j,t+1}$, 2) the absolute deviation of consumption from its optimal steady state value, $|c_{j,t} - c^*|$, and 3) the

[4] This regression model was first proposed to study the convergence of experimental panel data in Noussair et al. (1995).

ratio of the realized utility of consumption to the optimum, $u(c_{j,t})/u(c^*)$. For the first type of dependent variable, the estimate $\hat{\beta}$ reveals the values to which the dependent variable, $c_{j,t}$ and $k_{j,t}$ are converging across cohorts; strong convergence is said to obtain if $\hat{\beta}$ is not significantly different from the optimal steady state levels, c^* and k^*. For the second and third types of dependent variable, one looks for whether $\hat{\beta}$ is significantly different from zero or one, respectively. Lei and Noussair (2002) also consider a weaker form of convergence that examines whether $\hat{\beta}$ is closer (in absolute value) to the optimal, predicted level than a majority of the $\hat{\alpha}_j$ estimates. Using all four dependent variables, they report evidence of both weak and strong convergence in the market treatment, but only evidence of weak (and not strong) convergence in the social planner treatment.[5]

Tests of convergence based on the regression model (1) are found in several experimental macroeconomic papers reviewed later in this chapter, and so I take the opportunity here to briefly comment on this model. First, the notion that strong convergence obtains if $\hat{\beta}$ is not significantly different from the predicted level, y^*, while weak convergence obtains if $|\hat{\beta} - y^*| < |\hat{\alpha}_j - y^*|$ for a majority of j's is somewhat problematic, as strong convergence need not imply weak convergence, as when the $\hat{\alpha}_j$ estimates are insignificantly different from $\hat{\beta}$. Second, if *convergence* is truly the focus, an alternative approach would be to use an explicitly *dynamic* adjustment model for each cohort j of the form:

$$y_{j,t} = \lambda_j y_{j,t-1} + \mu_j + \epsilon_{j,t}. \tag{2}$$

Using (2), *weak* convergence would obtain if the estimates, $\hat{\lambda}_j$, were significantly less than 1, while *strong* convergence would obtain if the estimate of the long-run expected value for y_j, $\frac{\hat{\mu}_j}{1-\hat{\lambda}_j}$, was not significantly different from the steady state prediction y^*; in this model, strong convergence implies weak convergence, and not the reverse.[6] Finally, analysis of joint convergence across the J cohorts to the predicted level y^* could be studied through tests of the hypothesis:

$$I_J \begin{pmatrix} \hat{\mu}_1 \\ \vdots \\ \hat{\mu}_J \end{pmatrix} + \begin{pmatrix} \hat{\lambda}_1 \\ \vdots \\ \hat{\lambda}_J \end{pmatrix} y^* = \begin{pmatrix} y^* \\ \vdots \\ y^* \end{pmatrix},$$

where I_J is a J-dimensional identity matrix.

2.2 Exponential discounting and infinite horizons

It is common in macroeconomic models to assume infinite horizons, as the representative household is typically viewed as a dynasty, with an operational bequest motive linking one generation with the next. Of course, *infinite* horizons are not operational but *indefinite* horizons are. As we have seen, in experimental studies, these have often been implemented by having a constant probability δ that a sequence of decision rounds continues with another round.[7] Theoretically this practice should

[5] Lei and Noussair (2002) also consider a planning agency treatment in which the social planner is replaced with a group of five subjects (as in the market treatment) who together attempt to solve the social planner's problem. Convergence results for this planning agency treatment are somewhat better than in the social planner treatment but still worse than in the market treatment, based on regression findings using the model (1)

[6] Starting in period 1 with $y_{j,1}$ and iterating on (2) we can write $E[y_{j,t}] = \lambda_j^{t-1} y_{j,1} + \sum_{i=0}^{t-2} \lambda^i \mu_j + \sum_{i=0}^{t-2} \lambda_j^i \epsilon_{j,t-i}$. Given $\lambda < 1$, and for t sufficiently large, we have $E[y_j] = \frac{\mu_j}{1-\lambda_j}$.

[7] The issue of whether the length of time taken up by a decision round matters is an unexplored issue. This issue is tied up with aggregation of decisions. Macroeconomic data are typically recorded at low frequencies, e.g., monthly

induce both exponential discounting of future payoffs at rate δ per round as well as the stationarity associated with an infinite horizon, in the sense that, for any round reached, the expected number of future rounds to be played is always $\delta + \delta^2 + \delta^3 + ...$, or $\frac{1}{1-\delta}$.

To better induce discounting at rate δ it is desirable to have subjects participate in *several* indefinitely repeated sequences of rounds within a given session - as opposed to a single indefinitely repeated sequence – as the former practice provides subjects with the experience that a sequence ends and thus a better sense of the intertemporal rate of discount they should apply to payoffs. A further good practice is to make transparent the randomization device for determining whether an indefinite sequence continues or not, e.g., by letting the subjects themselves roll a die at the end of each round using a rolling cup. A difficult issue is the possibility that an indefinite sequence continues beyond the scheduled time of a session. This may often be avoided by choosing a discount factor that is not too high. Of course, a discount factor should also not be so low that one-round (single-shot) sequences are a frequent occurrence. A good compromise might be to set δ to .80, implying an expected duration of 5 rounds from the start of each sequence (as well as from any round reached!).[8] Another good practice is to recruit subjects for a longer period of time than necessary, say several hours, and inform them that a number of indefinitely repeated sequences of rounds will be played for a set amount of time – say for one hour following the reading of instructions. Subjects would be further instructed at the outset of the session, that after that set amount of time had passed, the indefinite sequence of rounds currently in play would be the last indefinite sequence of the experimental session. In the unlikely event that this last indefinite sequence continued beyond the long period scheduled for the session, subjects would be instructed that they would have to return at a set date to complete that indefinite sequence.

In practice, as we have seen, some researchers feel more comfortable working with finite horizon models. However, replacing an infinite horizon with a finite horizon may not be innocuous; this change may greatly alter predicted behavior relative to the infinite horizon case. For instance, the finite horizon life-cycle model of the consumption savings decision greatly increases the extent of the precautionary savings motive relative to the infinite horizon case. Other researchers have chosen not to tell subjects when a sequence of decision rounds is to end as a way of implementing an indefinite horizon. A difficulty with that practice is that the experimenter loses control of subjects' expectations regarding the likely continuation of a sequence of decisions and appropriate discounting of payoffs. This can be a problem if, for instance, the existence of equilibria depend on the discount factor being sufficiently high. Yet another approach is to terminate a session after a finite number of rounds, but exponentially discount the payoffs subjects receive in each round leading up to the terminal round. The problem with this approach of course, is that does not implement the *stationarity* associated with an infinite horizon.

---

or quarterly "consumption," whereas in laboratory studies, the length of time between decisions is, out of necessity, much more compressed – a few seconds to a few minutes.

[8] This can be implemented using any fair randomization device, e.g., cards, balls in a bingo cage or dice. If using dice, it is best to use the five convex regular polyhedra (the platonic solids), i.e., 4-, 6-, 8-, 12-, or 20-sided dice which, when manufactured precisely, are regarded as "fair". Perhaps the simplest procedure is to roll a six-sided die at the end of each round and state that the game continues with another round so long as a six is not rolled, thus inducing $\delta = .833$.

2.3 Exponential or Hyperbolic Discounting?

Recently, there has been a revival of interest in time-inconsistent preferences with regard to consumption-savings decisions, where exponential discounting is replaced by a quasi-hyperbolic form so that the representative agent is viewed as maximizing

$$u(c_t) + \beta \sum_{i=1}^{T} \delta^i u(c_{t+i}),$$

where $\delta \in (0,1)$ is a discount factor and the parameter $\beta \leq 1$ characterizes the agent's *bias–for–the–present* (exponential discounting has $\beta = 1$). Agents who discount hyperbolically ($\beta < 1$) rather than exponentially may exhibit time-inconsistent behavior (self-control problems) in that they systematically prefer to reverse earlier decisions, e.g., regarding how much they have saved. Thus, a possible explanation for the departures from optimal consumption paths noted above in experimental studies of intertemporal decision-making may be that subjects have such present-biased preferences. Indeed, Laibson (1997), O'Donoghue and Rabin (1999) and several others have shown that consumers with such preferences save less than exponential consumers.

Although time-inconsistent preferences have been documented in numerous psychological studies (see, e.g. Frederick et al. (2002) for a survey) the methodology used has often consisted of showing inconsistencies in hypothetical (i.e. unpaid) money-time choices (e.g., Thaler (1981)). For example, subjects are asked whether they would prefer $\$D$ now or $\$D(1+r)$ t periods from now, where variations in both r and t are used to infer individual rates of time preference. Recently, non-hypothetical (i.e. paid) money-time choice experiments have been conducted that more carefully respect the time dimension of the trade-off (e.g. Coller et al. (2005) and Benhabib et al. (2006)). These studies cast doubt on the notion that discounting is consistent with either exponential or quasi-hyperbolic models of discounting. For instance, Benhabib et al. (2006) report that discount rates appear to vary with both the time delay from the present and the amount of future rewards in contrast to exponential discounting. However, Coller et al. (2005) show that in choices between money rewards to be received only in the future, e.g., 7 days from now versus 30 days from now, variations in the time delay between such future rewards do not appear to affect discount rates, which is consistent with both exponential and quasi-hyperbolic discounting, but inconsistent with continuous hyperbolic discounting. Consistent with quasi-hyperbolic discounting, both studies find that a small fixed premium attached to immediate versus delayed rewards, can reconcile much of the variation in discount rates between the present and the future and between different future rewards. However, this small fixed premium does not appear to vary with the amount of future rewards (Benhabib et al.) and may simply reflect transaction/credibility costs associated with receiving delayed rewards (Coller et al.) making it difficult to conclude definitively in favor of the quasi-hyperbolic model.

Even more recently, Anderson et al. (2007) make a strong case that time preferences cannot be elicited apart from risk preferences. Prior studies on time discounting all presume that subjects have risk neutral preferences. However, if subjects have risk averse preferences (concave utility functions) as is typically the case, the implied discount rates from the binary time preference choices will be *lower* than under the presumption of risk neutrality (linear utility functions). Indeed, Anderson et al. (2007) elicit joint time and risk preferences by having each subject complete sequences of binary lottery choices (of the Holt-Laury (2002) variety) that are designed to elicit risk preferences as well as sequences of binary time preference choices that are designed to elicit their discount rates (similar to those in the Coller et al. study). They find that once the risk aversion of individual subjects is taken into account, the implied discount rates are much lower than under the assumption of risk neutral preferences. This finding holds regardless of whether discounting is specified to be exponential or quasi-hyperbolic or some mixture.

While it is standard practice to implement exponential discounting in laboratory studies as most theories with indefinite horizons presume time-consistent preferences, it might be of interest to attempt to induce hyperbolic discounting (perhaps with an effort made to control for risk aversion, e.g., via the use of a Roth-Malouf (1979) binary lottery) by letting the probability of continuation (the discount factor) be time-dependent, i.e., $\delta_t = \frac{1}{1+rt}$, where r is the rate of time preference used in exponential discounting and t indexes time $t = 1, 2, ...$ One might then assess whether this different stochastic process for ending an indefinite sequence affected consumption-savings behavior relative to the case of exponential discounting.

2.4 Expectation Formation

In modern, self-referential macroeconomic models, expectations of future endogenous variables play a critical role in the determination of the current values of those endogenous variables, i.e. beliefs affect outcomes which in turn affect beliefs which affect outcomes, etc. Since Lucas (1972) it has become standard practice to assume that agents' expectations are *rational* in the sense of Muth (1961) and indeed most models are "closed" under the rational expectations assumption. The use of rational expectations to close self-referential models means that econometric tests of these models using field data are joint tests of the model and the rational expectations assumption, confounding the issue of whether the expectational assumption or other aspects of the model are at fault if the econometric evidence is at odds with theoretical predictions. While many tests of rational expectations have been conducted using survey data, (e.g. Frankel and Froot (1987)), these tests are beset by problems of interpretation, for example due to uncontrolled variations in underlying fundamental factors, or to the limited incentives of forecasters to provide accurate forecasts, etc. In the lab it is possible to exert more control over such confounding factors as well as to implement the self-referential aspect of macroeconomic models.

Early experimental tests of rational expectations involved analyses of subjects' forecasts of exogenously given stochastic processes for prices, severing the critical self-referential aspect of macroeconomic models, but controlling for potentially confounding changes in fundamental factors (see e.g., Schmalensee (1976) or Dwyer et al. (1993)). Later experimental tests involved elicitation of price forecasts from subjects who were simultaneously participants in experimental asset markets that were determining the prices being forecast (Williams (1987), Smith et al. (1988)). Marimon and Sunder (1993, 1994) were the first to elicit inflation forecasts which were then used to determine subjects' intertemporal consumption/savings decision and via market clearing the actual price level. Subjects were not aware of the overlapping generations model in which they were operating - instead they were engaged in a forecasting game, with optimal savings decisions being made for them by a computer program. As discussed in the prior handbook surveys by Camerer (1995) and Ochs (1995), many (though not all) of these papers found little support for rational expectations in that forecast errors tended to have non-zero means and were autocorrelated or were correlated with other observables. Further, the path of prices sometimes departed significantly from rational expectations equilibrium. However, most of these experimental studies involve analyses of price forecasts in environments where there is no explicit mechanism by which forecasts determine subsequent outcomes as is assumed in forward–looking macroeconomic models (Marimon and Sunder (1993, 1994) being an exception). Further, some of these experimental tests, e.g. Smith et al. (1988) involved analyses of price forecasts for relatively short periods of time or in empirically non-stationary environments where trading behavior resulted in price bubbles and crashes, providing a particularly challenging test for rational expectations hypothesis.

More recently some macroeconomists have come to believe that rational expectations presumes too much knowledge on the part of the agents who reside within these models. For instance, rational expectations presumes common knowledge of rationality. Further, rational expectations agents know with certainty the underlying model whereas econometricians are often uncertain of data generating processes and resort to specification tests. Given these strong assumptions, some researchers have chosen to replace rational expectations with some notion of bounded rationality and ask whether boundedly rational agents operating for some length of time in a known, stationary environment might eventually *learn* to possess rational expectations from observation of the relevant time series data (see e.g., Sargent (1993, 1999) and Evans and Honkapohja (2001) for surveys of the theoretical literature).

Two notions of boundedly rational expectation formation are found in the experimental literature relevant to macroeconomists. The first approach may be termed 'step-level' reasoning and was motivated by Keynes's (1936) famous comparison of financial market investors' expectations to newspaper beauty contests of that era in which participants had to select the six prettiest faces from 100 photographs. The winner of the contest was the person whose choices were closest to the average choices of all competitors. Keynes (1936, p. 156) noted that "each competitor has to pick, not those faces which he himself finds prettiest but those he thinks likeliest to catch the fancy of other competitors, all of whom are looking at the problem from the same point of view." Keynes went on to observe that individuals might form expectations not just of average opinion, but might also consider what average opinion expects average opinion will be, and he further speculated that there might be some who practiced still "higher degrees" of reasoning. These observations concerning expectation formation were tested experimentally by Nagel (1995) in a game developed by Moulin (1986) that has since come to be termed the "beauty contest" game in honor of Keynes's analogy.

In Nagel's design, a group of $N = 15-18$ subjects are each asked to 'guess' –simultaneously and without communication– a real number in the closed interval $[0, 100]$. They are further instructed that the person(s) whose guess is closest in absolute value to a known parameter p times the mean of all submitted numbers is the winner of a large cash prize, while all other participants receive nothing. Nagel's baseline experiment involves setting $p < 1$, e.g. $p = 2/3$. That game is straightforward to analyze: each player i wants to guess a number $x_i = p\bar{x}$, where $\bar{x}$ is the mean of all submitted numbers. Given this objective, in any rational expectations equilibrium we must have that $x_i = \bar{x}$ for all i. If $p < 1$, the only rational expectations solution is $x_i = \bar{x} = 0$, that is all N players guess *zero*.[9] To map this game into Keynes's (1936) example requires setting $p = 1$, in which case any number in $[0, 100]$ is a rational expectations equilibrium; the choice of $p < 1$ yields not only a unique equilibrium prediction but interesting insights regarding the extent of individual's higher degrees of reasoning.[10]

[9] Non-corner (interior) rational expectations solutions are possible via a simple change to the payoff objective, e.g., guess the number closest in absolute value to $100 - p\bar{x}$.

[10] The $p = 1$ case corresponds to a pure coordination game; see Ochs (1995) for the relevant experimental literature on such games.

[Figure 2 here.]

Nagel's experimental findings from three sessions of the $p = 1/2$-the mean game are shown in Figure 2 which reports the relative frequencies of number choices in the interval $[0, 100]$.[11] Notice first that the equilibrium prediction of 0 is never chosen. Second, there are large spikes in neighborhoods of the numbers 50, 25, and 12.5. A choice of 50 implies an expected mean of 100 in the $p = 1/2$ game and is thus *barely* rational - these players exhibit the lowest level of reasoning, which is often termed step or level 0. The somewhat more sophisticated level 1 types expect a mean of 50 and guess numbers that are 1/2 of their expectation around 25, while level 2 types are a step further ahead, anticipating a mean of 25 and thus guessing numbers around 12-13. A robust finding is that depths of reasoning in excess of level 2 are rarely observed; the winner of the beauty contest is typically a level-2 type. With repetition, subjects in these beauty contest games do eventually converge upon the unique rational expectations equilibrium prediction (0 in this case), but each individual's process of expectation revision over time typically follows the same level of reasoning they exhibited in the first round played, e.g. level $k = 1$ or 2 adjustment in each repetition. This experiment, which has now been replicated many times, (see, e.g. Duffy and Nagel (1997), Ho et al. (1998)), reveals that in multi-agent economies where all agents know the model, the common-knowledge-of-rationality assumption implicit in the rational expectations hypothesis may not hold. It further suggests that decision costs or cognitive constraints may lead individuals to adopt heuristic rules of thumb that result in predictable step-levels of belief revision, i.e., *systematic* forecast errors. That convergence to equilibrium does obtain in the limit is reassuring but suggests that rational expectations might be best viewed as a long-run phenomenon.

A second view of boundedly rational expectation formation imagines that agents don't know the model or data generation process and behave more as econometricians, using possibly misspecified forecasting rules which they update in real-time as new observations become available. This real-time, adaptive expectations approach been explored experimentally by Bernasconi et al. (2006), Hey (1994), Van Huyck et al. (1994), Kelley and Friedman (2002), Hommes et al. (2007) and Adam (2007). Here I focus on the latter two papers, which emphasize the feedback between agents' expectations and outcomes.

Hommes et al. (2007) consider expectation formation by groups of six subjects operating for a long time (in the laboratory sense)– 50 periods – in the simplest dynamic and self-referential model - the Cobweb model.[12] In each of the 50 periods, all six subjects are asked to supply a one-step-ahead forecast of the price that will prevail at time t, $p^e_{i,t}$ using all available past price data through time $t-1$; the forecast is restricted to lie in the interval $(0, 10)$. These price forecasts are automatically converted into supply of the single good, via a supply function $s(p^e_{i,t};\ \lambda)$ which is increasing in $p^e_{i,t}$ and has common parameter λ governing the nonlinearity of the supply function. Demand is exogenous and given by a linear function $D(p_t)$. The unique equilibrium price p^* is thus given by

$$p^*_t = D^{-1}\left(\sum_{i=1}^{6} s(p^e_{i,t})\right),$$

that is, it is completely determined by subjects' price forecasts. However, Hommes et al. add a small shock to exogenous demand which implies that prices should evolve according to $p_t = p^*_t + \epsilon$,

[11] Nagel (1995) also considers the case of $p = 2/3$ and $p = 4/3$, and repeated versions of all three games.

[12] Hommes et al. (2005) use a similar approach to study expectation formation in a simple asset-pricing model.

where $\epsilon \sim N(0, \sigma_\epsilon^2)$. Thus under rational expectations, all forecasters should forecast the same price, p^*. In the new learning view of rational expectations, it is sufficient that agents have access to the entire past history of prices for learning of the rational expectations solution to take place. Consistent with this view, Hommes et al. do not inform subjects of the market clearing process by which prices are determined. Instead, subjects are simply engaged in forming accurate price forecasts and individual payoffs are a linearly decreasing function of the quadratic loss $(p_t - p_{i,t}^e)^2$. The main treatment variable consists of variation in the supply function parameter λ which affects the stability of the cobweb model under the assumption of *naive* expectations (following the classic analysis of Ezekiel (1938)). The authors consider three values for λ for which the equilibrium is stable, unstable or strongly unstable under naive expectations.[13] Their assessment of the validity of the rational expectations assumption is based on whether market prices are biased (looking at the mean), whether price fluctuations exhibit excess volatility (looking at the variance) and whether realized prices are predictable (looking at the autocorrelations).

[Insert Figure 3 here].

Figure 3 shows a representative sample of prices and the autocorrelation of these prices from the three representative groups operating in the three different treatment conditions. This figure reveals the main finding of the study which is that in all three treatments, the mean price forecast is not significantly different from the RE value, though the variance is significantly greater (there is excess volatility) from the RE value $\sigma_\epsilon^2 = 0.25$ in the unstable and strongly unstable cases. Even more interesting is the finding that the autocorrelations are not significantly different from zero (5% bounds are shown in the figures) and there is no predictable structure to these autocorrelations. The latter finding suggests that subjects are not behaving in an irrational manner in the sense that there is no unexploited opportunities for improving price predictions. This finding is somewhat remarkable given the limited information subjects had regarding the model generating the data, though coordination on the rational expectations equilibrium was likely help by having a unique equilibrium and a limited price range $(0, 10)$

Adam (2007) uses a similar experimental methodology but in the context of the two-equation, multivariate New Keynesian "sticky price" model that is the current workhorse of monetary policy analysis (see, e.g., Woodford (2003)). In a linearized version of that model, inflation π_t, and output y_t, are determined by the system of expectational difference equations,

$$\begin{pmatrix} \pi_t \\ y_t \end{pmatrix} = a_0 + a_1 y_{t-1} + B \begin{pmatrix} \pi_t^e \\ \pi_{t+1}^e \end{pmatrix} + c v_t$$

where a_0, a_1, B and c are conformable vectors and matrices, π_t^e π_{t+1}^e are the one– and two–step ahead forecasts of future inflation using information available through time $t-1$, and v_t is a mean zero real monetary shock. Like Hommes et al. Adam provides information on all past realizations of π and y through period $t-1$ and asks a group of five subjects to provide one- and two-step ahead forecasts of inflation, π_t^e π_{t+1}^e repeatedly for 45–55 periods. The average forecasts each period are used in the model above to determine π_t and y_t. Subjects earn payoffs based on forecast accuracy alone and are uninformed regarding the underlying process generating data on π_t and y_t.

[13] Specifically denote the ratio of marginal supply to marginal demand at the equilibrium price by $\sigma(\lambda) = S'(p_t^*)/D'(p_t^*)$. Stability under naive expectations requires that $-1 < \sigma(\lambda) < 1$. Otherwise there is instability, and this can be determined by varying λ.

The rational expectation solution is of the form:

$$\begin{aligned} y_t &= y + v_t \\ \pi_t &= (\pi/y)y_{t-1} \end{aligned}$$

where y and π represent steady state values. Inflation lags output by one period due to predetermined (sticky) prices, and output deviates from its steady state only due to real monetary shocks. Thus a rational forecast model for π_t should condition on y_{t-1}, i.e. $\pi_t = \alpha_y + \beta_y y_{t-1}$. Of course, since subjects are given time series data on both y and π, Adam imagines that subjects might alternatively use a simple (but miss-specified) autoregressive forecast model of the form $\pi_t = \alpha_\pi + \beta_\pi \pi_{t-1}$. Thus, the issue being tested here is not simply one of whether agents can learn to form rational expectations of future inflation but more importantly whether subjects, like econometricians, can find the correct specification of the reduced form model they should use to form those rational expectations. Perhaps not surprisingly, the evidence on the latter question is somewhat mixed. Adam finds that in most of the experimental sessions, subjects forecast using the autoregressive inflation model and do not condition their forecasts on lagged output. However, he also shows that such behavior can result in a stationary, "restricted perceptions" equilibrium that is optimal in the sense that autoregressive inflation forecasts outperforms those that condition on lagged output. Adams further notes that this miss-specification in agents forecasts provides a further source of inflation and output persistence in addition to that implied by the model's assumption of sticky price adjustment.

Summing up, we have seen some ways in which three micro-level assumptions that are mainstays of macroeconomic modeling - intertemporal optimization, time-consistent preferences/exponential discounting and the rationality of expectations have been tested in the laboratory, primarily in individual decision-making experiments. The evidence to date suggests that human subject behavior is often at odds with the standard micro-assumptions of macroeconomic models. The behavior of subjects appears to be closest to micro-assumptions, e.g., intertemporal optimization, when subjects learn from one another or gather information on prices through participation in markets. Rational expectations appears to be most reasonable in simple, univariate models (e.g. the Cobweb model) as opposed to the more commonly used multivariate models. Hopefully these and other experimental findings will lead to a reconsideration of the manner in which macroeconomic modelers characterize the behavior of their "representative" agents, though so far, there is not much evidence that such a change is imminent.

3 Coordination Problems

In the previous section, we focused on individual behavior in dynamic intertemporal optimization problems where the optimal, rational expectations solution was unique. In many macroeconomic environments, this is not the case. Instead, multiple rational expectations equilibria exist and the question is which of these equilibria economic agents will choose to coordinate upon. Laboratory experiments can be quite useful in this regard. Indeed, Lucas (1986) argued that laboratory experiments were a reasonable means of resolving such coordination problems, because "economic theory does not resolve the situation [so] it is hard to see what can advance the discussion short of assembling a collection of people, putting them in the situation of interest, and seeing what they do."

Some coordination problems of interest to macroeconomists were previously addressed in Ochs (1995). In particular, that chapter surveyed experimental studies of overlapping generations models where money may or may not serve as a store of value (Lim et al. (1994)), or subjects can select between low or high inflation equilibria (Marimon and Sunder (1993, 1994, 1995)). Also included were experimental studies of stag-hunt and battle–of–the sexes games (surveyed also in Cooper 1999) and Bryant (1983)-type Keynesian coordination games (e.g., the minimum and median effort games of Van Huyck 1990, 1991, 1994).[14] The coordination games literature delivered a number of important findings on when coordination success was likely to be achieved and when coordination failure was likely. Importantly, the results have been replicated by many other experimenters leading to confidence in those findings. Rather than review those replications and extensions, in this section I report on more recent macro-coordination experiments. The environments tested in these experiments have a more direct resemblance to macroeconomic models than do the coordination games surveyed by Ochs, (with the exception of Marimon and Sunder's work on overlapping generations models). I also address some equilibrium selection mechanisms or refinements that have been proposed for resolving macro-coordination problems and the experimental studies of those mechanisms and refinements.

3.1 Poverty Traps

Lei and Noussair (2007) build on their (2002) experimental design for studying behavior in the one-sector optimal growth model by adding a non-convexity to the production technology, resulting in multiple, Pareto-rankable equilibria. Specifically, the production function used to determine output in Matheny and Noussair (2000) and Lei and Noussair (2002) is changed to:

$$f(k_t) = \begin{cases} \underline{A}k_t^{\alpha} & \text{if } k_t < k^* \\ \overline{A}k_t^{\alpha} & \text{if } k_t \geq k^*. \end{cases}$$

where $\underline{A} < \overline{A}$ and k^* is a threshold level of aggregate capital stock that is known to all 5 subjects. The threshold switch in productivity is a simple way of modeling positive externalities that may arise once an economy reaches a certain stock of capital (physical or human) (see, e.g. Azariadis and Drazen (1990)). An implication is that there are now two stationary levels for the capital stock (and output) $\overline{k}_l < k^* < \overline{k}_h$, with $\overline{k}_l$ representing the poverty trap and $\overline{k}_h$ representing the Pareto efficient equilibrium. The dynamics of the system (under perfect foresight) are such that for $k \in (0, k^*)$, $\overline{k}_l$ is an attractor whereas for $k \geq k^*$, $\overline{k}_h$ is the attractor. The main experimental question is which of these two equilibria subjects while learn to coordinate on.

One treatment variable was the initial aggregate level of the capital stock, either below or above the threshold level k^* and divided up equally among the 5 subjects. The other treatment condition was whether decisions were made in a decentralized fashion, with a market for the capital stock (subjects had different production technologies that aggregated up to the aggregate technology) or whether groups of subjects together made a collective consumption-savings decision, i.e. playing the role of a social planner. In both cases, the indefinite horizon of the model was implemented using a constant probability of continuation and subjects were paid on the basis of the utility value of the consumption they were able to achieve in each period. The main experimental finding is that in the decentralized treatment, the poverty-trap equilibrium is a powerful attractor; it is selected

[14] That material, while highly relevant to the literature on experimental macroeconomics will not be repeated here – the interested reader is referred to Ochs (1995). See also Camerer (2003 chp. 7) and Devetag and Ortmann (2007)

in all sessions where the initial aggregate capital stock is below k^* as well as in some sessions where the initial aggregate capital stock lies above k^*. There are some instances of convergence to the Pareto efficient stationary equilibrium $\overline{k}_h$, but only in the decentralized setting where the initial capital stock lies above k^*. In the social planner treatment, where 5-subject groups jointly decide on consumption-savings decisions, neither of the two stationary equilibria were ever achieved; instead there was either convergence to a capital stock close to the threshold level k^*, or to the golden-rule level that maximally equates consumption in every period. While the latter is close to the Pareto optimum it is inefficient, as it ignores the possibility that the economy may terminate (the rate of time preference is positive). Lei and Noussair (2007) conclude that additional institutional features may be necessary to both avoid and escape from the poverty trap outcome.

The possibility that various institutional mechanisms might enable economies to escape poverty traps is taken up in a follow-up experimental study by Capra et al. (2005). These authors begin by noting that laboratory studies of the role of institutions in economic growth may avoid endogeneity problems encountered in field data studies (where it is unclear whether institutions cause growth or vice versa), and more clearly explore environments with multiple institutions. The two institutions explored in this study are termed "freedom of expression," which involves free discussion among subjects prior to each round of decision-making and "democratic voting" in which subjects vote on two proposals for how to divide output up between consumption and savings (future capital) at the end of each period.

The baseline experimental design is essentially the same as the low initial capital stock treatment of Lei and Noussair (2007); there are five subjects who begin each indefinite sequence of rounds with capital stocks that sum up to an aggregate level that lies below the threshold level k^*.[15] This initial condition for the aggregate capital stock is the same in all treatments of this study, as the focus here is on whether subjects can escape from the poverty trap equilibrium. At the start of a period, output is produced based on last period's capital stock and then a market for capital (the output good opens). After the market for capital has closed, subjects independently and without communication decide on how to allocate their output between current consumption and savings (next period's capital stock). In the communication treatment, subjects are free to communicate with one another prior to the opening of the market for capital. In the voting treatment, after the capital market has closed, two subjects are randomly selected to propose consumption/savings plans for all five agents in the economy; these proposals specify how much each subject is to consume and how much to invest in next period's capital stock (if there is a next period). Then all five subjects vote on the proposal they prefer and the proposal winning a majority of votes is implemented. In a hybrid treatment, both communication and voting stages are included together.

[Insert Figure 4 here.]

The main findings examine the long-run values of two statistics for each session: 1) aggregate welfare (as measured by the sum of the period utility from consumption by all 5 agents $\sum_i u(c_t^i)$) and the aggregate capital stock ($\sum_i k_t^i$). Capra et al. use an equation similar to (1) to estimate the asymptotic values of these two measures for each 5-person economy.[16] These estimated values are shown as squares in Figure 4 and the line segment through each square represents the 95%

[15] One difference is that Capra et al. use a "call market" clearing mechanism for the capital market as opposed to the double auction mechanism used by Lei and Noussair (2007). The difference between these two mechanisms is discussed later in section 3.3.

[16] Specifically, for each session they estimate the equation $Y_{jt} = \alpha/t + \beta\frac{t-1}{t} + \epsilon_j + v_{j,t}$, where j indexes each indefinite

confidence region. The lower left intersection of the dashed lines shows the poverty trap level of aggregate welfare and capital, while the upper right intersection of the two dashed lines shows the Pareto efficient level of aggregate welfare and capital. This Figure reveals the main findings In the baseline treatment, consistent with Lei and Noussair, subjects are unable to escape from the poverty trap outcome. The addition of communication or voting helps some, though not all economies to escape from the poverty trap. In the hybrid model which allows both communication and voting, the experimental economies appear to always escape from the poverty trap (95% confidence bounds exclude poverty trap levels) and these economies are closest to the Pareto efficient equilibrium levels for welfare and the capital stock. Capra et al. argue that binding consumption/savings plans as in the voting treatment are important for achieving aggregate capital stock levels in excess of the threshold level, while communication makes it more likely that such consumption/savings plans are considered in the first place; not surprisingly then, the two institutions complement one another well and lead to the best outcomes.

While this experimental design involves a highly stylized view of the institutions labeled "freedom of expression" and "democratic voting" the same critique can be made of the neoclassical model of economic growth. The experimental findings suggest that there may be some causality from the existence of these institutions to the achievement of higher levels of capital and welfare, though the opposite direction of causality from growth to institutions remains an important possibility. More recently, macroeconomists have emphasized the role of human capital accumulation, so it would be of interest to consider whether subjects learn to exploit a positive externality from a highly educated workforce). And while several other studies have pointed to the usefulness of communication in overcoming coordination problems (see, e.g., Blume and Ortmann (2007), Cooper et al. (1992), these have been in the context of strategic form games. While the results of those studies are often cleaner, in the sense that the game is simple and communication is highly scripted, the Capra et al. study implements institutional features in a model that macroeconomists care about and this may serve to improve the nascent dialogue between experimentalists and macroeconomists.

3.2 Bank Runs

Another coordination problem that has been studied experimentally in the context of a model that macroeconomists care about is Diamond and Dybvig's (1983) coordination game model of bank runs. In this three period intertemporal model, depositors find it optimal to deposit their unit endowment in a bank in period 0, given the bank's exclusive access to a long-term investment opportunity and the deposit contract the bank offers. This deposit contract provides depositors with insurance against uncertain liquidity shocks; in period 1, some fraction learn they are have immediate liquidity needs (are impatient) and must withdraw their deposit early, while the remaining fraction learn they are patient and can wait to withdraw their deposit in the final period 2. The bank uses its knowledge of these fractions in optimally deriving the deposit contract, which stipulates that depositors may withdraw the whole of their unit endowment at date 1 while those who wait to withdraw until period 2 can earn $R > 1$. While there exists a separating, Pareto efficient equilibrium where impatient types withdraw early and patient types wait until the final period, there also exists an inefficient pooling equilibrium where uncertainty about the behavior of other

---

sequence or "horizon" within a session. The dependent variable $Y_{j,t}$ is either aggregate welfare, $U(c_t) = \sum_{i=1}^{5} u_i(c_t^i)$ or the aggregate capital stock $k_t^i = \sum_{i=1}^{5} k_t^i$. The two asymptotic estimates for each session –the estimates of β for each of the two dependent variables – are the squares shown in Figure 4.

Hypothetical No. of Withdrawl Requests	Amount Each Requester Would Receive	Projected Payment to Each Depositor
0	n/a	\$1.50
1	\$1	\$1.50
2	\$1	\$1.50
3	\$1	\$0
4	\$0.75	\$0
5	\$0.60	n/a

Table 2: Bank-Run Coordination Game Payoffs, Garratt and Keister (2005)

patient types causes all patient types to mimic the impatient types and withdraw their deposits in period 1 rather than waiting until period 2. In the latter case, the bank has to liquidate its long-term investment in period 1 and depending on the liquidation value of this investment, it may have insufficient funds to honor its deposit contract in period 1. The possibility of this bank-run equilibrium is the focus of experimental studies by Garratt and Keister (2005) and Schotter and Yorulmazer (2003).

Garratt and Keister's baseline experimental design dispenses with controlling for the two types and focuses on the pure coordination game aspect of the problem. Five subjects have \$1 deposited in a bank and must decide at one or more opportunities whether to withdraw their \$1 or leave it deposited in the bank potentially earning a higher return of \$1.50. Following each withdrawal opportunity, subjects learn the number of players in their group of 5 (if any) who have chosen to withdraw. As treatment variables, Garratt and Keister varied the number of withdrawal opportunities (1 or 3) and the number of early withdrawals a bank could sustain while continuing to offer those who avoided withdrawal a payoff of \$1.50 (i.e. variation in the liquidation value of the bank's long-term investment). Table 2 provides one parameterizations of Garratt and Keister's bank-run game.

Garratt and Keister report that for this baseline game, regardless of the liquidation value of the long-term investment, no group ever coordinated on the "panic equilibrium" (5 withdrawals) and a majority of groups coordinated on the payoff dominant equilibrium (0 withdrawals). In a second treatment that more closely implements the liquidity shock in the Diamond-Dybvig model, Garratt and Keister added "forced withdrawals" to the baseline game: at each withdrawal opportunity, there was a small known probability that one randomly selected player would be forced to withdraw; however whether a withdrawal was forced or not was unknown to subjects. The probabilities of forced withdrawals were chosen such that there continued to exist a payoff dominant equilibrium in which no player ever voluntarily withdrew at any withdrawal opportunity (if all adhered to this strategy they would earn an expected payoff greater than \$1) as well as a panic equilibrium where all withdraw. Garratt and Keister report that with forced withdrawals (liquidity shocks) the frequency of voluntary withdrawals and coordination on the panic equilibrium is significantly greater relative to the baseline treatment with unforced withdrawals. This increase in panic behavior was particularly pronounced in the forced withdrawal treatment where subjects had multiple withdrawal opportunities and could condition their decisions on the prior decisions of others. An implication of this finding is that panic behavior may require some conditioning on the decisions of others suggesting that the bank run phenomenon is perhaps best modeled as a dynamic game, as opposed to the simultaneous-move formulation of Diamond and Dybvig (1983).

Schotter and Yorulmazer (2003) arrive at a similar conclusion, using a rather different experimental design. Theirs involves a group of six subjects deciding in which of four periods to withdraw their deposit of $K in the face of uncertainty concerning both the withdrawal decisions of the other five subjects as well as the type of bank that all 6 have invested their deposits in. Subjects know that there are 5 possible bank types, that each type is equally likely to be drawn for the duration of each 4-period game, and that the mean return across types is r^*.[17] While the bank type is unobservable, the "promised" return is fixed at 12% per period, while the mean return r^* was varied across sessions, either .07, .08 or .14. Subjects were told that if they kept their $K deposit invested for ℓ periods, they could earn a return of $\$(1.12)^{\ell}K$ if the bank has sufficient funds left in period ℓ, but if not, the bank would pay all those withdrawing in that period an equal share of remaining funds on hand (if any). Subjects had to choose in which of the four periods to withdraw their money, with withdrawal being irreversible. The authors think of this as a model of a bank–run–in–progress, (the precipitating event is left unmodeled) and are interested in exploring three factors that may slow or hasten the period in which deposits are withdrawn. A first factor is whether the withdrawal decision across the four periods is implemented as a simultaneous-move normal form game, or as an extensive form game; in the former case subjects specify the period in which they want to withdraw their funds (1,2,3,or 4) while in the latter case subjects make withdrawal decisions period by period and may condition on the prior period withdrawal decisions (and in one treatment, the amounts earned) by others. The second and third factors are the use of deposit insurance to delay or slow down the run or the presence of insiders who know the mean return r^* of the banks, and may through their actions persuade other uformed subjects to run early or wait.

Schotter and Yorulmazer find that bank runs are less likely to be severe (withdrawal occurs later, e.g. in period 3 or 4) when r^* is known to be greater than the promised return of 12%. For fixed r^*, runs are also less severe in the extensive form version of their model, when agents can condition on the decisions of others and there is a high degree of information, in that subjects also know the amounts that others have received.[18] This finding is interesting in that theory does not predict that the game form should matter; the fact that it does again points to the value of thinking of bank runs as dynamic rather than static games. They further show that partial deposit insurance may work to diminish the severity of bank runs as can the presence of some depositor–insiders who know the type of bank with which funds have been invested.

The issue of the *contagious spread* of a bank run from one location to another is addressed experimentally by Corbae and Duffy (2008). They study a two stage, 4-player game. In the first stage, players simultaneously propose to form links with one another; mutually agreeable links are then implemented and comprise the set of each player's 'neighbors'. Corbae and Duffy interpret the players as 'banks' connected to one another via interbank reserve deposits that can serve to insure against risk. (a la Allen and Gale (2000)). In the second stage, each player plays τ rounds of an n-person, equal-weighted- payoff stag hunt game with his n=1,2 or 3 neighbors. As in Garratt

[17] The possible returns from the five banks are known to belong to the set: $\{(1/3)r^*, (2/3)r^*, r^*, (4/3)r^*, (5/3)r^*\}$.

[18] The latter finding may seem at odds with Garratt and Keister's findings but note that Schotter and Yorulmazer don't have forced shocks, so their set-up is closest to Garratt and Keister's setting without forced shocks, in which panics were rarely observed.

and Keister (2005), one of the 4-players is "shocked" i.e., randomly to play the inefficient 'hare' or run strategy in all rounds of the second-stage game. Corbae and Duffy define a contagion as a movement by all players away from the Pareto efficient 'stag' equilibrium to the inefficient hare equilibrium. While it is possible for subjects to implement a complete network of links (each of the 4 players has 3 links each) that provides insurance against the risk of being linked to player forced to panic as when all unshocked players play 'stag', Corbae and Duffy show that such a network configuration is not an equilibrium due to the free-rider problem. Instead, the network configurations that are predicted to emerge are *bilateral* networks (2-player networks where each player has a single link) which serves to limit the spread of the bank run outcome. Corbae and Duffy report experimental evidence that is broadly consistent with this prediction. Starting groups of 4 subjects out in different exogenous network configurations and then in subsequent games allowing them to choose the players they want to link to, they report that subjects consistently move in the direction of choosing to have a single link to one other player. Under this bilateral network, the bank-run equilibrium is isolated to just one of the 2-player networks; the other network achieves the efficient, payoff dominant equilibrium.

Summing up, we have discussed two kinds of macroeconomic-coordination experiments, poverty traps and bank runs. In the poverty trap model, the question of interest is how to get subjects to move *from* an inefficient equilibrium *to* an efficient one. We might think of this as a good contagion. In the bank run model the question of interest is precisely the opposite - how to keep funds deposited in a bank longer (earning higher returns) and avoiding a bad contagion to an inefficient panic equilibrium. Both types of movements are difficult to achieve in the laboratory. In the case of movement from an efficient to an inefficient equilibrium it seems necessary to force some players' hands in order to precipitate a transition to the inefficient outcome; that finding suggests that the precise mechanism precipitating a bad contagion has yet to be discovered.

We next explore experimental tests of two mechanisms that macroeconomists have used to resolve coordination problems.

3.3 Resolving Coordination Problems: Sunspots

In the bank-run coordination game, the question of equilibrium selection is left unmodeled. Diamond and Dybvig (1983) suggest that depositors might use realizations of some commonly observed, non-fundamental random variable, or "sunspot" in the language of Cass and Shell (1983) and Azariadis (1981) to resolve the question of which equilibrium to coordinate on.[19] Experimental tests of sunspot variables as coordination devices have been conducted by Marimon et al. (1993) and Duffy and Fisher (2005).[20]

Marimon and Sunder (1993) implemented a 2-period overlapping generations environment where, if agents have perfect foresight, there are multiple equilibrium: an interior steady state and a two-period cyclic equilibrium. Subjects in the role of young agents formed price expecta-

[19] John Maynard Keynes talked about "animal spirits" as a source of investment volatility. Charles McKay talked about the "madness of crowds" in documenting famous financial fiascos. These are references to the role played by non-fundamental factors or "sunspots" in economic activity. The term "sunspot" derives from the work of William Stanley Jevons, a nineteenth century economist and polymath who championed the notion that the solar cycle was responsible for the business cycle. Of course, business cycles are a lot more irregular than the 11-year sunspot cycle and there is considerable variation in the timing and duration of business cycles across countries. So today, we honor Jevon's folly by referring to non-fundamental or extraneous factors that may affect business activity as "sunspots".

[20] Here I emphasize the latter study, as the Marimon et al. (1993) study is thoroughly discussed in Ochs (1995).

tions which determined current prices, given the nonlinear model, $p_t = \phi(p^e_{t+1})$. Thus given price expectations, subjects optimal consumption and savings in the form of real money balances was determined (as in Marimon and Sunder (1993, 1994). Marimon and Sunder hoped that subjects would use realizations of a sunspot variable to coordinate their expectations on the cyclic equilibrium. Their sunspot variable consisted of a blinking cube on subjects' computer screens. The color of this cube alternated every period between red and yellow. Marimon and Sunder found that subjects essentially ignored the sunspot variable realizations and simply coordinated on the steady states. They later tried to add a correlation between the sunspot variable and a real endowment shock (alternating the size of the young generation between 3 and 4 subjects, i.e., 3-4-3-4) but this also did not lead to coordination on the sunspot variable when the endowment shock was shut off.

Duffy and Fisher (2005) consider a simpler, partial equilibrium framework that abstracts from a number of conceptual difficulties (e.g. implementing an infinite horizon). In this simple and static environment there are two equilibria that differ only in terms of the equilibrium price level; the equilibrium quantity is the same in both. The experimental design involves 5 buyers and 5 sellers, each with two units to buy or sell. Buyers seek to maximize consumer surplus, (valuation - price), while sellers seek to maximize producer's surplus (price - cost). Further each buyer (seller) had two possible valuations (costs) for each of his two units. If the state was "high" each buyer's (seller's) profits were calculated using his two high valuations (costs). If the state was "low" each buyer's (seller's) profits were calculated using his two low valuations (costs). The two sets of valuations/costs used in the experiment are shown in Figure 5.

[Figure 5 here.]

Two market clearing mechanisms were considered - the standard double auction where bids and asks can be observed in real-time and a sealed-bid variant known as a call-market, where bids and asks are submitted simultaneously, bids are sorted from highest to lowest, asks from lowest to highest and a single market clearing price is determined by the intersection of demand and supply (if there is one). All buyers (sellers) with bids (asks) above (below) the market price get to buy (sell) their units (provided there are enough units on both sides). The state of the world was determined by the median traded price in the double auction or by the market clearing price in the call-market. If either was greater than or equal to 150, then the high state was declared and subjects uses high valuations or costs in determining their surplus (payoff). Otherwise the low state was declared and low valuations and costs were used in the determination of payoffs. Thus the situation is akin to one in which there are multiple equilibria, each supported by different beliefs about the likely state of the world.

Duffy and Fisher's sunspot variable was one of two possible announcements made prior to each of 10 four-minute trading periods. The announcement chosen was determined by publicly flipping a coin. In one treatment, if the coin flip was heads, the public announcement was "the forecast is high" while if the coin flip was tails, the public announcement was "the forecast is low" and this scheme was public knowledge. Duffy and Fisher report that in sessions using a call market clearing mechanism, subjects perfectly coordinated on the high price equilibrium when the forecast was high and on the low price equilibrium when the forecast was low - that is the sunspot variable was shown to matter for economic volatility. On the other hand under the double auction market clearing mechanism, the sunspot announcements only sometimes served to coordinate subjects on the high or low equilibrium. Duffy and Fisher argue that the reason for this difference lies in the real-time information that was available in the double-auction; subjects could see bids and asks as they occurred and could use this fact to attempt to engineer an equilibrium outcome for prices (high or low) that was more favorable to them.[21] Thus the coordinating mechanism provided by the sunspot could be undone by the real-time information on bids, asks and trade prices. The same was not possible in the call-market where bids and asks had to be submitted simultaneously and hence the sunspot variable played an important coordinating role in the environment.

Duffy and Fisher further show that the semantics of the sunspot variable matter: replacing the forecast is "high" or "low" with the forecast is "sunshine" or "rain" eliminated the sunspot variable as a coordinating mechanism in the call market. Further research might seek to understand how the mapping from sunspot variable realizations to the action space matter in getting subjects to coordinate on sunspot equilibria. It would also be of interest to consider sunspot equilibria that are not simply randomizations over two certainty equilibria.

3.4 Resolving Coordination Problems: The Global Game Approach

Another view of multiple equilibria in macroeconomic modeling is that the equilibrium beliefs in support of these equilibria may not be as indeterminate as theory supposes. As Morris and Shin (2001) argue, these indeterminacies arise from assuming that economic fundamentals are *common knowledge* and that individuals are certain of the behavior of others in equilibrium. Relaxing these assumptions, e.g. by introducing some uncertainty about fundamentals, can remove the multiplicity, á la the Carlsson and van Damme's (1993) global game approach for 2×2 games.[22] The resulting game is one in which individuals adopt a unique threshold strategy -when fundamentals are weak, individuals are pessimistic about others' beliefs and the resulting outcome is poor, as in the bank run equilibrium. However if fundamentals are strong so will be beliefs about others' beliefs and the resulting outcome will be good, as in a payoff dominant equilibrium. This correlation between fundamentals and outcomes is missing from the sunspot approach.[23]

Heinemann et al. (2004) conducted the first experimental test of the global game approach to resolving equilibrium multiplicity in the context of a speculative currency attack model developed by Obstfeld (1996) and Morris and Shin (1998). Prior to the start of each $2 \times n$ player game, a payoff relevant random variable Y is drawn from a uniform distribution with known support. This variable represents the fundamentals of the economy with higher (lower) values of Y representing worse (better) fundamentals. In the complete information (CI) treatment, this variable is known to all 15 subjects while in the private information (PI) treatment, the value of Y is not known but each of the 15 subjects receives noisy signals of Y, X_i, that are uniform random draws from the known interval $[Y - \epsilon,\ Y + \epsilon]$ where ϵ is small. Subjects must then decide between two actions, A and B, where A is a safe choice resulting in a fixed payoff F (equivalent to not-running or not-attacking a currency). The other choice, B, is a risky choice (equivalent to attacking a currency, joining a rebellion, etc.) the payoff from which depends on the total number of players who choose B, as determined by a monotonically decreasing function $f(Y)$. If less than $f(Y)$ agents choose B all those choosing B earn 0 (the attack fails) while if at least $f(Y)$ agents choose B, then all those

[21] Notice from Figure 4 that 2 out of 5 buyers/sellers prefer the high equilibrium and 2 out of five prefer the low equilibrium and the remaining buyer and seller are indifferent. Thus, the equilibria are not Pareto rankable.

[22] In Carlsson and van Damme's approach, players facing a game of complete information with multiple equilibria behave as though it were a perturbed global game of incomplete information where the payoffs are determined by a random draw from a given class of games and players have a noisy signal of the chosen game.

[23] There is some debate about whether bank runs and financial crises are caused by fundamental or non-fundamental (sunspots).

choosing B earn Y points (the attack succeeds). Consistent with the theory, the distribution of Y values is chosen so that there exists values of $Y \leq F$, for which it is a dominant strategy to choose A and similarly there exist values of $Y \geq f^{-1}(1)$ for which a single individual can guarantee the success of an attack by choosing B, so that it is dominant for all to do so. For Y values in $(F, f^{-1}(1))$, under complete information, there are multiple equilibria: all choose A or all choose B, both of which can be supported by the belief that all others will choose A or B. However in the incomplete information game there exists a unique, threshold value of the noisy signal X for which all subjects should attack (choose B) if their signal is above the threshold and not attack otherwise. Taking the limit as $\epsilon \to 0$ it is possible to find a similar threshold Y^* in the complete information game.

The main question pursued by Heinemann et al. is whether the complete information game, with its multiplicity of equilibria is more unstable than the private information game, and whether subjects adopt threshold strategies consistent with the global game threshold prediction. They report that subjects do appear to adopt threshold strategies in both the private and complete information cases and these estimated thresholds generally lie below the global game predictions X^* or Y^* but are higher than the payoff-dominant prediction of choosing B whenever $Y > F$. The most interesting finding is in the complete information treatment, where Y is publicly known and there are in principle multiple equilibria. In that treatment Heinemann et al report less variance in entry decisions than in the incomplete information treatment and greater coordination on a common threshold in the former as compared with the latter. Heinemann et al conclude that "with public information the central bank has more control over trader's beliefs than when they get private information from other sources."

The global game refinement has been further tested in several other experimental studies. Cabrales et al. (2007) test the theory in two-person games with a more discrete state space. They find greater coordination on the global game prediction in the incomplete information case and on the payoff dominant equilibrium in the complete information case. Duffy and Ochs (2008) embedded Heinemann et al.'s design in a dynamic setting where subjects have multiple periods in which to decide whether to attack or not and may condition their decision on the prior decisions of others. They report little difference in the thresholds used in the dynamic game as compared with those used in the static game. Cornand (2006) adds two treatments, one with a private signal and a public signal and a second with two noisy public signals. She reports that subjects overreact to the public signal when they also receive a private one, but that predictability of an attack is higher in that case as compared to the case of two noisy public signals. This finding suggests that if officials are going to make public amusements they would do well to coordinate on a single message.

Summarizing, we have considered laboratory evidence on several mechanisms for selecting from among multiple equilibria in macroeconomic models including communication, voting, sunspots and threshold strategies based on the global game refinement. The laboratory is a natural testing ground for these mechanisms as other confounding factors can be minimized and attention can be focused on the hypothesized coordination device (perhaps too ideal a setting?) The experimental findings to date suggest mixed support for any single mechanism as the means by which individuals actually go about solving coordination problems. Still, many improvements on these studies remain to be conducted and it is likely that with further study we will have a better sense of which mechanisms work best in particular settings.

4 Sectoral Macroeconomics

In the following sections we review experimental studies that address issues in a particular sector of the macroeconomy. The macroeconomic sectors that have been studied in the laboratory include monetary economics (which has attracted the greatest attention to date), labor economics and international trade and finance. In focusing on specific sectoral phenomena, these laboratory studies follow the macroeconomic literature which often abstracts from certain sectors of the macroeconomy altogether, (e.g. the government sector) in order to better address a specific macroeconomic question (e.g. why money is used). A few studies have attempted to combine one or more of these sectors and these are reviewed in the last subsection on multi-sectoral macroeconomics.

4.1 Monetary Economics

What is the role of money in the macroeconomy? Traditionally, money has been assigned three roles: as a store of value, as a medium of exchange and as unit of account. As I have observed earlier (Duffy 1998), much of the theoretical and experimental literature on money can be divided up according to the primary role of money. Studies of money as store of value focus on the question of how an asset with no intrinsic value (i.e. fiat objects) may be used as storage devices even though they are subject to depreciation over time due to inflation. As a medium of exchange, money must serve as a store of value, but the opposite is not true; there are many stores of value that are not media of exchange. Thus researchers interested in money as a medium of exchange have sought to understand the frictions that give rise to use of certain stores of value as media of exchange. Finally, as the prices of goods and services are all stated in monetary terms, money's role as a unit of account is important for efficient decision-making. In addition to the primary roles of money, experimental studies can also be categorized according to the friction that enables money to be valued in equilibrium along with the mechanism by which exchange of money for goods takes place. Table 3 summarizes the approaches to studying money in the laboratory that are reviewed in this section.

Study	Primary Role of Money	Friction Enabling Money	Exchange Mechanism
McCabe (1989)	Store of Value	Cash-In-Advance	Clearinghouse with Rationing
Deck et al. (2006)	Store of Value	Cash-In-Advance	Double Auction
Hens et al. (2007)	Store of Value	Cash-In-Advance	Clearinghouse with Rationing
Marimon and Sunder (1993, 1994, 1995)	Store of Value	Overlapping Generations	Centralized Mkt. Clearing
Bernasconi and Kirchkamp (2000)	Store of Value	Overlapping Generations	Centralized Mkt. Clearing
Camera et al. (2006)	Medium of Exchange	Overlapping Generations	Double Auction
Brown (1996)	Medium of Exchange	Random Matching	Bilateral Exchange
Duffy and Ochs (1999, 2002)	Medium of Exchange	Random Matching	Bilateral Exchange
Duffy (2001)	Medium of Exchange	Random Matching	Bilateral Exchange

Table 3: Characteristics of Experimental Studies of Money

The store-of-value role of money is the focus of an early experimental study by McCabe (1989). That study focuses on whether fiat objects will be used as stores of value in an economy with a known finite end at which time the fiat object ceases to have any continuation value. McCabe's design involves three player types and six rounds of play. One of the three player types is initially endowed with a durable ticket (fiat money) that can be exchanged for one unit of any good and the other two types are endowed with nondurable goods. Exchanges of tickets for goods occurs via a centralized clearinghouse with known rationing rules; barter exchanges of goods for goods are not allowed, so effectively a cash-in-advance constraint operates. Holding a good at the end of a round yields different redemption values to different player types (either \$0.50, \$0.25 or \$0.0) and the endowments of these goods also varies across player types. If this game continued without end, the use of tickets would enable the efficient exchange of goods to types who most value those goods (direct barter is ruled out). However, since the game is known to have a finite end at which point tickets have zero value, via a backward induction argument, tickets should never be accepted in trade. McCabe however reports that tickets are indeed accepted, though with some fall-off near the end of each six round game. Despite repeating the six-round game 10-20 times with the same group of subjects, tickets continue to circulate in early rounds of the game. McCabe did eventually succeed in eliminating all trade in tickets but only after bringing back the same group of subjects for two further sessions, each a week apart. McCabe suggests that the inexperienced subjects' use of tickets may be sustained by strong home-grown prior beliefs that money-type objects such as tickets will be accepted in exchange as they are in everyday life.

[Figure 6 here].

Deck et al. (2006) follow up on the McCabe study by adding government agents who, unlike the other two player types in their study, are not budget constrained as to the quantity of tickets they can redeem for goods (i.e. they can "print money"). The two other player types, "A" and "B", are endowed each period with amounts of goods B and A respectively, but profit from acquiring certain amounts of goods A and B respectively; unlike the government player types, the A and B-type players are liquidity-constrained and must resort to trading the good they are endowed with in the two double-auction goods market for tickets in order to buy the good they desire to consume. Figure 6 provides an illustration. As barter is disallowed, the friction giving rise to a demand for money is a cash-in-advance constraint. As in McCabe's study, there is a finite horizon which is varied and in some treatments where money is "backed", tickets have a final cash redemption value. In treatments without government agents, subjects use money as a store of value (and hence as a medium of exchange) regardless of whether it is backed or not and despite the finite horizon as in McCabe (1989). The addition of the government agents who are not budget constrained and who desire additional units of both goods leads to a rapid escalation of the price level, which Deck et al. term a *hyperinflation*. This outcome arises in part because the government agents' ability to print tickets leads to a rapid increase in the supply of money but Deck et al. emphasize that the erratic means by which the government introduces newly printed money augments the corruption of the information revealed in market traded prices. The hyperinflation finding is consistent with the work of Sargent (1983) who attributes historical episodes of hyperinflations to excessive fiat money creation. Deck (2004) provides further experimental evidence that hyperinflations of the Deck et al. (2006) variety can be ended by either making the currency convertible or by limiting government spending to current tax receipts (a balanced budget). Such mechanisms are also consistent with the historical record on ending hyperinflations.

Similar to the Deck et al. study, Hens et al. (2007) address whether a fiat object can achieve a stable value, facilitating its use as a medium of exchange. However, their focus is on whether an optimal quantity of fiat money can be achieved. They present a model inspired by the Capitol Hill Baby Sitting Co-op, a natural experiment in the 1970s in which approximately 150 Capitol

Hill couples exchanged baby-sitting duties with one another for coupons (Sweeny and Sweeny (1977)). The co-op organizers found that too few coupons led to coupon hoarding (precautionary savings?) resulting in low demand for baby-sitting and a collapse of the system. An increase in coupons led to a thriving exchange of baby-sitting services, but eventually, over-issue of coupons resulted in excess demand for baby-sitting, and, given the fixed price of 1 coupon=1/2 hour of baby-sitting, led again to a collapse of the system. Hens et al. first develop a model wherein individuals face preference shocks for a perishable single good (they have either a high or low value for it) eliminating barter, and must choose whether to be buyers/sellers of the good in each period. Buy or sell decisions are made simultaneously via a centralized mechanism with a long-side of the market rationing rule. To buy a good, an individual must have money on hand, so a cash-in-advance constraint gives money value. Sales of goods augment an individual's money holdings; prices are fixed. The unique equilibrium prediction of their rational expectations, forward looking, infinite horizon model is that subjects who hold no money always offer to *sell* goods for money, regardless of their period valuation for the good. Whether subjects choose to *buy* goods using money depends on their period valuation for the good. In the high valuation state, exchanging money for goods is a dominant strategy. However, in the low valuation state, subjects should use money to buy goods only if their money holdings are sufficiently high; if below a critical level $\overline{m}$, subjects should sell goods to acquire more money. This critical level of money holdings is related to the supply of money, which is exogenously chosen. Hens et al. show that there is a unique optimal quantity of money that maximizes the number of trades possible (i.e. no trader is rationed), given that players are playing according to the optimal buy/sell strategy. The nicely designed experiment tests these predictions in a two stages. In the first stage subjects participate in individual decision-making experiments where they make buying and selling decisions and do or do not face exogenous rationing with regard to whether their buy or sell orders are satisfied; this gives subjects experience with the clearinghouse mechanism. In the second stage, subjects participate in a six-player market game where the probabilities of successfully buying or selling (rationing) using the centralized mechanism depend on the decisions of all agents. Hens et al report that subjects' strategies coincided well with the forward-looking optimal strategies of the theory. Furthermore, exogenous increases in the supply of money led to first to an increase in the volume of trade that was followed by a decrease in the volume of trade as the supply of money was further increased, with the peak corresponding to the predicted optimal quantity of money. The latter finding thus replicates the history of the Capitol Hill Baby sitting co-op and nicely illustrates the difficulty central banks face of determining an optimal quantity of money. Of course, the optimal quantity of money is complicated by the fact that the coupon price of baby-sitting is fixed, which is more typical of trade circles where fairness is a concern and less so of actual monetary systems.

A second friction giving rise to the use of money as a store of value is that of *overlapping generations (OG)* of trading agents; as is well known (Shell 1971) the double infinity of dated goods and traders in the OG model violates the standard assumptions of general equilibrium analysis and can give rise to competitive equilibria that are not Pareto optimal in violation of the second welfare theorem. This possibility provides a role for money (or other stores of value, e.g., social security promises) as Pareto improving devices (Samuelson (1958)). Lim, Prescott and Sunder (1994) were the first to implement an OG model of money in the laboratory with the aim of studying money as a store of value and the dynamics of price behavior. Further experimental studies involving monetary OG models that focused on questions of equilibrium selection were performed by Marimon and Sunder (1993, 1994) and are reviewed in the first volume of the *Handbook of Experimental*

$s_{i,t}$	β	σ_β	t-stat	Pr > $\|t\|$	95% conf. interval
$s^*_{i,t}$	1.015071	0.0012107	838.38	0.00	[1.012698, 1.017445]

Table 4: Regression of actual savings on recommended, optimal savings, Bernasconi and Kirchkamp (2000).

Economics by Ochs (1995). Here I want to review two OG money model experiments that have appeared more recently and which build on the design of Marimon and Sunder. Bernasconi and Kirchkamp (2000) re-examine Marimon and Sunder's experimental design regarding how young agents determine the fraction of youthful endowment they should save in the form of money for later purchase of old age consumption. Marimon and Sunder (1993) had subject cohorts alternate between youth and old age in their indefinitely repeated two-period OG model. Each subject i who was 'young' in period t forecast the gross inflation rate (π) of the price level between t and $t+1$, $E_{i,t-1}\pi_{t+1} = E_{i,t-1}(P_{t+1}/P_t)$, drawing on the past history of the aggregate price level, P, through period $t-1$. Based on this forecast, the computer program determined each subject i's optimal savings, $s_{i,t}$, given their lifetime utility function and budget constraint. As savings had to be held in the form of money, equilibrium market clearing required that the aggregate demand for real savings $\sum_i s_{i,t}$, equals the supply of real money balances M_t/P_t. Since the money supply M_t is exogenously determined, this market clearing condition determines the period t price level, P_t. Bernasconi and Kirchkamp were critical of the optimal derivation of individual savings based on inflation forecasts. The "learning how to forecast" design of Marimon et al. is only one dimension of forward-looking rational expectations models - the other being the ability of agents to solve intertemporal optimization problems given their forecasts. Bernasconi and Kirchkamp thus modified the design of Marimon and Sunder. Subjects still made forecasts of future inflation and the computer program continued to calculate optimal savings amounts for each subject, but subjects were now free to ignore the optimal savings suggestion when asked to state the fraction of their youthful endowment they actually wanted to save. In addition, they could consider information on the past savings decisions of other subjects. Another treatment variable concerned the money creation process–whether the supply of money followed a constant exogenous growth process or was endogenously determined by the need to finance a fixed real government deficit. Both money supply rules give rise to two monetary equilibria, one involving a high inflation rate and the other involving a low inflation rate; the latter steady state is precisely the same under the two money supply regimes. Under rational expectations the high inflation steady state is an attractor, but under first-order adaptive expectations, the low inflation steady state is an attractor. Similar to the findings of Marimon and Sunder (1993, 1994, 1995), Bernasconi and Kirchkamp find that actual inflation converges to a neighborhood of the low inflation monetary steady state under both monetary regimes, though inflation is systematically biased below the low inflation steady state. The later finding is consistent with the findings of Marimon and Sunder (1993, 1994, 1995). What differs is Bernasconi and Kirchkamp's finding that savings under both regimes is greater than the optimal level (which is not possible in Marimon and Sunder's design). Specifically, Bernasconi and Kirchkamp run a regression of actual individual savings s^i_t choices on optimal choices as recommended to subjects $s^*_{i,t}$. The results are reproduced in Table 4. As these results confirm, there is a significant difference between subjects' actual savings choice and the optimal savings amount given their forecast. Bernasconi and Kirchkamp argue that a precautionary saving motive arising from subjects' uncertainty regarding their inflation forecasts *can* rationalize the observed over-saving behavior. This finding would appear to invalidate the use of Marimon and Sunder's 'learning to forecast' experimental design; subjects do not make savings decisions as if they were certain of their forecasts of future inflation.[24] Given that agents in macroeconomic models must 1) form rational expectations of future variables and 2) choose current quantities optimally in response to those expectations, further experimental work on this important topic is required.[25]

Thus far, the experimental studies reviewed have considered environments where a single good, e.g., tickets, is long-lasting (durable); all other goods are perishable. If subjects perceive the durable good to be a store of value, (perhaps owing to its durability), then that good *necessarily* serves as a medium of exchange, as it is the only good that can serve in that capacity. By contrast, I regard experimental studies of the medium of exchange role of money as those which present subjects with multiple durable goods (candidates for money) and ask whether and which of these goods is adopted by subjects as money.

Camera et al. (2003) consider the overlapping generations model with fiat money that we have just discussed and add to it a second store of value, an interest-bearing consol.[26] The question addressed is whether fiat money continues to be used to transfer wealth from youth to old age when there is an interest-bearing alternative. Understanding why money is used as a medium of exchange when it is dominated in rate of return by other assets is a critically important issue in monetary theory. Camera et al. explore experimentally two complementary explanations for the rate of return dominance of fiat money. Their first, hoarding hypothesis – that assets bearing interest would be hoarded and not used as media of exchange when an alternative non-interest-bearing store of value exists – is tested by initializing the economy with stocks of both fiat money and consols, but requiring that consols be traded pre-dividend, i.e., the dividend accrues to the owner of the consol after trading is completed. Their second, hysteresis hypothesis - that the old habit of using zero-interest fiat money dies hard – is tested by initializing a sequence of two-period overlapping generations economies with a stock of fiat money that serves as the sole store of value and only later adding a stock of the interest-bearing consol, which trades either pre- or ex-dividend and seeing whether the fiat object continues to be used as a medium of exchange along with the consol. money object is used as a medium of exchange. Both of these hypotheses are purely behavioral; The stationary rational expectations equilibrium prediction in all treatments is that, in the presence of multiple stores of value, subjects will use the good offering the highest rate of return and a medium of exchange and eschew the other object. Consistent with the hysteresis hypothesis, the authors report that fiat money coexists with consols as a medium of exchange if there is a prior history of use of fiat objects alone as a medium of exchange. This coexistence is strongest when the consol dividend is paid after trade (consol is traded pre-dividend) consistent with the hoarding hypothesis. If the consol dividend is paid after trade, and consols and fiat objects are introduced simultaneously, then subjects cease to use the fiat object and exclusively use the consol as a medium of exchange.

The use of money as a medium of exchange even though it is dominated in rate of return by

---

[24] Surprisingly, Bernasconi and Kirchkamp do not consider the same parameterization of the OG money model as examined by Marimon and Sunder (1993, 1994, 1995), so a direct comparison is not possible.

[25] Experiments on learning in games similarly show that subjects' beliefs and action choices do not necessarily coincide, and that convergence may or may not obtain as subjects acquire experience. See, e.g. Ehrblatt et al. (2007).

[26] A consol is a bond with no terminal date paying a certain dividend per period forever.

other stores of value need not arise from irrational behavior. In the search theoretic approach to money as a medium of exchange, as pioneered by Kiyotaki and Wright (1989) and extended by many others, equilibria can be derived in which durable goods that are not the least costly to store (have lowest return) can nevertheless serve as media of exchange under the belief that these goods will be more readily accepted in exchange by others, thereby reducing the time it takes an individual to acquire goods he wants to consume. A second virtue of the search-theoretic approach over the models examined previously is that exchanges of goods and money is decentralized and occurs via the bilateral trading decisions of anonymous, randomly matched agents, which is an altogether different friction than cash-in-advance or overlapping generations. This third mechanism giving rise to the use of money seems closer to what actually occurs in monetary economies than does a centralized market clearing mechanism.

The predictions of the commodity money version of the Kiyotaki-Wright (1989) model are tested experimentally by Brown (1996) and Duffy and Ochs (1999). In this model, there are three goods (1,2,3) and three player types (1,2,3). Player type i desires to consume good i which yields a per period payoff of u but type i produces good $i+1$ modulo 3. Hence, there is an absence of a double-coincidence of wants and some players will have to trade for goods they do not desire to consume in order to obtain goods they do desire to consume; such goods may be regarded as commodity monies. Each player can store a single unit of a (perfectly durable) good in every period, but pays a per period storage cost c_i. In the parameterization studied by Brown and Duffy and Ochs, $c_1 < c_2 < c_3$. A trader starts out with a unit of his production good in storage. If he successfully trades for his consumption good he gets the period payoff for consumption and then produces a unit of his production good, so his payoff is reduced by the cost of storing the good. Under one parameterization of the model studied by Duffy and Ochs, all agents should play fundamental, storage-cost-minimizing strategies in equilibrium. For instance, type 2 players should trade their production good 2 with type 3 players in exchange for good 1, as this lowers type 2's storage cost and reduces the time it takes type 2s to acquire their consumption good 2, via trades with type 1. The predicted pattern of exchange in the unique equilibrium is as shown in Figure 7. Under a different parameterization, the unique equilibrium prediction -also illustrated in Figure 7– calls for some player types to adopt speculative strategies, wherein they trade lower storage cost goods for higher storage cost goods, e.g., type 1 players should agree to trade their production good 2 with type 2 players for the more costly to store good 3 as this reduces the time it takes type 1 to acquire its consumption good 1. This is a case where good 3 is used as a medium of exchange by type 1 even though it is dominated in rate of return (inverse of storage cost) by type 1's production good 2.

[Figure 7 here.]

Brown tested only the speculative pattern of exchange and made use of a strategy method, wherein each subject stated their trading decision for all possible player types storing all possible goods prior to being randomly matched with a player; trades were then executed in accordance with strategies. Duffy and Ochs tested both sets of equilibrium trading predictions. As in Browns' study, subjects were assigned a fixed player type, but unlike in Brown's study, following each random pairing with another player a subject had to decide whether to trade the good they had in storage for the good of the other player; mutually agreed upon exchanges were implemented. Despite these differences, the experimental findings of the two studies are quite similar as shown in Table 5, which reports the frequencies of exchange behavior in both the speculative and fundamental environments.

Speculative Parameterization	Type 1 trades 2 for 3	Type 2 trades 3 for 1	Type 3 trades 1 for 2
Brown (1996)	0.31	0.99	0.13
Duffy & Ochs (1999)	0.36	0.93	0.25
Spec. Eq. Prediction	1.00	1.00	0.00
Fundamental Parameterization			
Duffy & Ochs (1999)	0.30	0.97	0.13
Fund. Eq. Prediction	0.0	1.00	0.00

Table 5: Frequencies of trade offers by the three player types as reported by Brown (1996) and Duffy and Ochs (1999) in the speculative and fundamental equilibrium environments of the Kiyotaki and Wright (1989) model along with equilibrium predictions

The main finding of both studies is that, inconsistent with the theoretical predictions, subjects do not adopt the play of speculative strategies when such strategies constitute the unique equilibrium prediction. In particular, only around 1/3 of type 1 subjects storing good 2 agree to trade that good for the more costly-to-store good 3.[27] In the environment where the fundamental equilibrium is unique, Type 1s do not trade good 2 for good 3, behavior but Duffy an Ochs report that trading decisions by Type 1s in this environment are insignificantly different from decisions by the same type the speculative environment (see Table 5). Duffy and Ochs argue that subjects choose trading strategies based on immediate past payoff experiences as opposed to the more forward-looking marketability considerations that the theory emphasizes.

In an effort to make marketability considerations more transparent to Type 1 players, Duffy (2001) changed the distribution of the N subjects over the three types from 1/3 of each type to 1/3 of Type 1, 2/9 of Type 2 and 4/9 of Type 3. Thus, Type 1s were more likely to encounter a Type 3 player and might therefore appreciate the use of the more costly-to-store good 3 as a medium of exchange. Indeed, Duffy (2001) reports an increase in the acceptance of good 3 by Type 1 players from the 36% rate reported in Duffy and Ochs for the equal distribution of player across types to an acceptance frequency of 67% under the asymmetric distribution (still below the speculative frequency of 100%). Automating the decisions of Type 2 and 3 players with robot traders who played fundamental trading strategies also helped to boost speculative trades by Type 1 players to an average of 73%. These findings suggest that there exist certain parameterizations of the model in which a majority of subjects can learn to adopt speculative strategies where the money good is dominated in rate of return by other potential stores of value.

All of the goods in the search experiments described above had consumption value to one type of player. Duffy and Ochs (2002) add to this same environment an exogenous supply of a fourth good, 0 which is neither produced nor consumed by any player type. The question they pose is whether an intrinsically worthless fiat object that is not invested with value by legal restriction, would come to be used as a medium of exchange. Kiyotaki and Wright (1989) show that equilibria where this object is or is not traded coexist, so the issue is one of equilibrium selection. Duffy and Ochs' (2002) experimental finding is that an intrinsically worthless fiat object will circulate as a medium of exchange so long as it has the lowest storage cost; if it is not the least-costly to store

[27] Interestingly, this same lack of speculation finding is also obtained in agent-based model simulation conducted by Marimon et al. (1990), which was the inspiration for both the Brown and Duffy and Ochs studies.

good, i.e., if it is dominated in rate-of-return, then its circulation as a medium of exchange is more limited than predicted by the theory.

Money's role as a unit of account is uncontroversial; prices are typically quoted in terms of money units and not in terms of (say) artichokes. However, as money typically depreciates in value over time due to inflation, most macroeconomic models presume that agents evaluate all choice variables in real terms, taking into account changes in the purchasing power of money. Experimental studies of money as a unit of account assess the extent to which individuals evaluate magnitudes in real terms or, alternatively are subject to *money illusion* - the failure to adjust nominal values for changing prices. Motivated by survey evidence of money illusion (Shafir et al. 1997) and on the downward stickiness of nominal prices and wages (Bewley 1999), Fehr and Tyran (2001, 2007, 2008) have conducted several experimental studies documenting money illusion and its consequences for nominal inertia. In the first of these studies, Fehr and Tyran have subjects play an n-player "price–setting" game. In each of 2T periods, subject i chooses a price P_i and earns a real payoff that is a function of the time t average price chosen by other players, $P_{-i,t}$ and the time t nominal money supply M_t:

$$\pi_{i,t} = f(P_i, \overline{P}_{-i,t}, M_t)$$

The function f yields a unique, dominance-solvable equilibrium for every value of M, is homogeneous of degree 0 in all arguments, and $f_{P_{-i,t}} \geq 0$, so there is a weak strategic complementarity in price-setting. In addition to treatments where subjects are paid according to this *real* payoff function, there is also a *nominal* payoff treatment where subjects' earnings are reported to them in nominal terms, $\overline{P}_{-1}\pi_i$. Subjects are instructed on how they can deflate these payoffs into real terms by dividing by $\overline{P}_{-1}$. Fehr and Tyran characterize money illusion as a *framing effect*; behavior is predicted to differ depending on whether subjects are paid in real price adjusted terms, or in nominal terms. The difference comes in the adjustment to a nominal shock: the nominal money supply is known to be a constant level M for the first T periods and then to decline to a permanently lower level λM, $\lambda < 1$ for the last T periods. The issue addressed is whether subjects will adjust their prices downward at date T from P to λP, an adjustment that is more difficult in the nominal payoff function treatment where subjects have to correctly deflate their nominal payoff function. A second difficulty, arising from the strategic complementarity in price setting, is that the failure of some subjects to adjust to the nominal shock may make it a best response for others who are not subject to money illusion to only partially adjust to the shock themselves. To eliminate the latter possibility, Fehr and Tyran conduct individual-decision making experiments under both the real and nominal payoff functions where the other $n-1$ players are known to the human subjects to be robot players who are not subject to money illusion and who will adjust prices downward proportional to the shock and at the time of the shock.

[Figure 8 here.]

The experimental findings are nicely summarized in Figure 8, where we see that in three of the four treatments, the downward adjustment of prices to the new equilibrium occurs almost immediately following the fully anticipated reduction in M, whereas in the nominal payoff function with human opponents treatment, price adjustment is considerably more sluggish. Fehr and Tyran attribute behavior in the latter treatment to "the belief that there are subjects who take nominal payoffs as a proxy for real payoffs," which leads those who hold those beliefs to adjust their prices more slowly. When the payoff function is presented in real terms or there are computerized opponents, such beliefs are unwarranted, and so the extent of price sluggishness is greatly diminished, if not perfectly eliminated. Fehr and Tyran further show that prices adjust more rapidly in response to a positive shock than they do in response to a negative shock.

Fehr and Tryan (2007) consider a modified version of their price-setting game in which there are three Pareto-ranked equilibria. In real terms, the ranking of payoffs was $\pi_A > \pi_C > \pi_B$, but in nominal terms, the ranking was: $P_C\pi_C > P_A\pi_A > P_B\pi_B$. The treatments were as in their earlier study: whether payoffs were presented in real or nominal terms and whether subjects played against $n-1$ human or computer opponents. As before subjects are instructed in how to deflate nominal payoffs into real terms. Unlike the prior experiment, the focus here is not on adjustment to a shock but rather on equilibrium selection. In the computerized treatments, the $n-1$ robots play a best response to the past history of play of the human subject, effectively making the subject a Stackelberg leader. Fehr and Tyran's main finding is that in the nominal treatment with human opponents, subjects coordinate on the inefficient C equilibrium while in the real treatment with human opponents they coordinate on the efficient A equilibrium; they interpret this as evidence of money illusion. In the nominal or real treatments with computerized opponents, with experience subjects get close to the efficient equilibrium, though not as close as in the real payoff treatment with human opponents; they attribute the latter to imitation of the choices of other human actors, as reflected in prices observed each period.

In a third study, Fehr and Tyran (2008) consider not only the prior case where there is a strategic complementarity in price setting, but now also consider the case where there is strategic substitutability in price setting, i.e. $f_{P_{-i,t}} \leq 0$. They report that money illusion and the resulting nominal inertia in response to a fully anticipated monetary shock is greatly reduced in the case of strategic *substitutes* relative to the case of strategic *complements*. In the substitutes case, errors under adaptive learning are much greater following the money shock leading to much faster adjustment toward more rational behavior than in the complements case. Thus, it appears important to consider the strategic environment in assessing the extent to which money illusion may matter for nominal inertia.

Summing up, laboratory monetary experiments have examined whether individuals thinks in real or nominal terms, and have explored the circumstances under which a token object can serve as a store of value as well as the characteristics of stores of value that make them more readily acceptable as media of exchange. While the experimental literature on monetary questions is one of the largest in experimental macroeconomics, there remains much further work to be done. For instance, most of the experimental studies of money we have discussed have fixed rates of exchange between money and goods ignoring the important role of prices. Allowing for prices, one could then begin to think about exchange rate determination between multiple money objects.[28] While money illusion (together with the strategic environment) is an interesting explanation for nominal price stickiness, it is by no means the only explanation and indeed, most macroeconomists would point to other sources, including informational frictions, costly price or information adjustment or staggered contracting. Experimental studies of the behavioral relevance of these other mechanisms is an important an open question for future research.

4.2 Labor Economics

Empirical research in labor economics typically involves the use of large panel data sets as assembled by government agencies. However there is also a small and growing experimental literature

[28] For one early attempt, see the discussion of Arifovic (1996) below.

that exploits the greater control and identification of causal relationships that is afforded by the laboratory relative to the field (see, e.g., Falk and Gächter (2008) and Falk and Fehr (2003)). Here I focus on some of the labor economic experiments that should be of interest to macroeconomists.

An early experimental literature (previously reviewed by Camerer (1995)) examined individual behavior in intertemporal one-sided job search models that are commonly used to study unemployment and labor-market policies (e.g. as surveyed by Mortensen (1987)). Experimental studies testing many of the comparative static implications of job search models include Braunstein and Schotter (1981, 1982), Hey (1987), Cox and Oaxaca (1989, 1992), Harrison and Morgan (1990)). For instance, Braunstein and Schotter (1981) test a number of theoretical hypotheses involving the one-sided model of intertemporal optimal job search with or without perfect recall. In this model, an unemployed worker draws a wage offer each period and must decide whether to accept or reject each offer taking into account the known probability distribution of wage offers, search costs and the level of unemployment compensation (if any). The optimal search strategy involves calculation of a reservation wage level; wage offers at or above this level are accepted and those below it are rejected. Braunstein and Schotter (1981) report experimental evidence in support of the notion that individuals choose reservation wages that are nearly optimal and accept or reject offers relative to this wage level. Among the treatment variables they consider are different wage distribution functions, search costs and whether subjects could recall past wage offers or faced uncertainty about the wage distribution function they faced.

In addition to intertemporal labor force participation decisions, another labor market choice of interest to macroeconomists that has received some experimental attention, is the labor-leisure trade-off. An increase in wages may have both substitution and income effects on hours worked. The impact of wage changes on labor supply is an important empirical question, as most business cycle models require the (compensated) elasticity of labor supply to be positive and sufficiently large so that transitory shocks can generate the large volatility in hours worked that is observed in macroeconomic data.[29] Battalio, Green and Kagel (1981) report experimental evidence confirming positive compensated wage effects on time spent working, though their experiments involved pigeons rather than human subjects. In particular, Battalio et al. report that nearly all of their hungry Pigeons responded to a Slutsky-compensated wage *decrease* with a reduction in labor supply, which involved pecking a key.

Using human subjects, Dickinson (1999) has experimentally examined two extensions to the classical labor supply model. In the first, hours of work are no longer a choice variable, but are instead fixed - a situation that characterizes many (short-run) employment relationships; indeed some business cycle theorists have exploited this type of nonconvexity as a means of increasing volatility in hours worked. However, in contrast to the standard theory, which assumes that workers provide full effort when on the job, Dickinson allows subjects to choose the intensity of their work effort; essentially they can decide whether to take on-the-job leisure. Specifically, subjects must participate in a two hour experiment during which time they are asked to type an unlimited supply of paragraphs, earning a fixed wage for every paragraph they type with no more than a few errors. The intensity of their work effort is examined in response to compensated changes in the (piece-rate) wage. Compensation was achieved by varying the value of non-labor income. This kind of data on labor effort is typically unavailable to labor economists (who at most can observe labor hours) and serves to illustrate one of the advantages of study labor market theories in the

[29] Most estimates based on microeconomic data find the compensated elasticity to be small or even negative.

laboratory. In the second modification, subjects could choose both the hours worked - they did not have to stay for the duration of the 2-hour experiment - and the intensity of their work effort, and these are again examined in response to compensated wage changes. In both the intensity and the combined intensity and choice of hours treatments, Dickinson reports that a majority of subjects, worked harder (less hard) when given a compensated wage increase (decrease), i.e. the compensated elasticity of labor supply is, on average, positive. A notable feature of this experimental design as well as that of Battalio et al., is that subjects really much choose to exert a level of effort at a task (pecking or typing) as opposed to experimental designs (discussed below) involving costly-but-effortless effort.

More recently, experimental labor economics has moved in the direction of a more behavioral view of labor market dynamics arising out of the influential work of Akerlof (1982) on efficiency wage theory (see also the papers in Akerlof and Yellen (1986) and Akerlof (2002)). While standard neoclassical theory presumes that, in a perfectly competitive equilibrium all labor of a certain type is paid its marginal product, there is no involuntary unemployment and no problems of worker motivation, the efficiency wage theory disputes this view. In Akerlof's (1982) original model, firms sets wages above the competitive market level so as to better motivate employees and in exchange, employees' effort levels are in excess of minimum standards so that the labor contract involves "partial gift-exchange." A consequence of setting non-market "efficiency wages" and the reciprocity by workers it induces is that fewer workers are hired than in competitive equilibrium so some unemployment may be regarded as "involuntary".

The notion that labor market contracts are incomplete, e.g., on specification of effort levels, or on the monitoring of effort or both, so that reciprocity in the form of gift exchange may play a role has been tested experimentally in the form of the "gift-exchange game" first developed by Fehr, Kirchsteiger and Riedl (1993, 1998), with replications and variants subsequently studied by many others (see Gächter and Fehr (2002) for a survey of this now extensive literature). The gift exchange game is similar to a one-shot, sequential-move prisoner's dilemma game or the trust game. All versions share similar features. In the original formulation of Fehr et al. (1993), subjects are assigned roles as firms and workers and there are two stages to the game. In the first stage, firms post wage offers $w \in [\underline{w},\ \overline{w}]$, which may or may not be accepted by workers. Firms can only employ a single worker, workers can accept at most a single wage offer and there are more workers than firms so wage offers should be accepted immediately and should not exceed a worker's reservation value, that is, all rents should accrue to the firm. If a worker accepts a wage offer, then in the second stage she had to choose an effort level, $e \in [\underline{e}\ \overline{e}]$. Payoffs to workers are $w - c(e)$, where $c(e)$ is a convex cost of effort function, with the normalization that $c(\underline{e}) = \underline{w}$, which can be viewed as the workers' reservation value. (Effort here is of the costly-but-effortless variety). Payoffs to firms are $(v - w)e$ where v is the firm's redemption value. All payoff functions, wage and cost of effort schedules were public knowledge. In the baseline model, workers and firms are separated, interactions were anonymous so that each two-stage game can be viewed as one-shot; that is, reputational considerations cannot play a role. Thus, the subgame perfect equilibrium prediction is that workers will choose the lowest possible effort level $\underline{e}$ and recognizing this, firms will offer the lowest possible wage $\underline{w}$. The two-stage game is typically repeated 10-16 times. The main experimental finding, which has been replicated several times, is that workers reciprocate high wage offers with high effort. Figure 9 (from Fehr et al. (1993) illustrates this main finding.

[Figure 9 here.]

In this Figure, the competitive equilibrium (and lowest possible) wage $\underline{w} = 30$ which is associated with minimum effort level $\underline{e} = 0.1$. The maximum possible wage is 110 and the maximum effort level was 1. A wage of 30 was observed only once, and workers chose the minimum effort level only 16% of the time. The average wage was 72 and the average effort level was .4, both well above the competitive equilibrium predictions.

Fehr and Falk (1999) modify the first stage of the gift-exchange game so that both firms and workers can propose and accept wage offers via a double auction, following the standard improvement rules. In the second stage, worker effort was either exogenously fixed by the experimenter, so that the contract negotiated in the first stage was "complete," or workers were free to choose effort levels in the second stage, the case of "incomplete" wage contracts. As in the prior experiments, there were more workers than firms, so one would expect workers to underbid one another down to their minimum, reservation wage levels. Fehr and Falk report two main findings. First, when contracts are completely specified by a wage offer (effort predetermined) this wage tends to be close to the competitive equilibrium level, where all rents accrue to the firm due to the smaller number of firms relative to workers. Second, when workers are free to choose effort levels, wages are significantly above competitive equilibrium levels as in the earlier experiments where only firms could make wage offers. These higher wages are not because workers are refusing to undercut one another (a possibility suggested by Solow (1990)); Falk and Fehr report that there is, in fact, "massive underbidding" by workers seeking to secure wage offers. Interestingly most firms refuse to accept these low wage offers; while bid improvement rules force workers' wage offers to fall, firms are free to accept any wage offer and choose only to contract at wages well above workers' reservation levels. Subjects in the role of firms recognize that subjects in the role of workers will provide greater effort the greater is the wage offered, and this recognition results in sticky downward wage rigidity. This evidence is consistent with survey evidence, e.g. by Bewley (1999) indicating that managers recognize the impact of low wages on employee morale.

In a third set of experiments, Fehr et al. (1996, 1997) and Fehr and Gächter (2002) further modify the basic experimental design of Fehr et al. (1993) so that in the first stage, the wage contract specifies a wage, a desired effort level and a fine for effort below the desired level. A third stage is added in which the worker's effort level is probabilistically monitored by the experimenter; if below the desired level, the worker pays an fixed and publicly known fine to the firm. This design can be viewed as a version of Shapiro and Stiglitz's (1984) deterrence–of–shirking version of the efficiency wage model though in that model, a worker detected to be shirking is fired rather than fined. The issue explored in these experiments is whether the specification of desired effort levels, monitoring and fines i.e., *incentive contracting*, undermines the positive reciprocity observed in experiments where these feature of the wage contract are unspecified. The results are somewhat mixed. On the one hand, firms are able to obtain effort levels above the requested level by setting high, "efficiency wages" as in the earlier experiments. On the other hand, firms tended to request too much effort and set wages too low to enforce a no-shirking outcome given the fines workers faced. Consequently, there is a substantial amount of shirking, despite the no-shirking-in-equilibrium prediction of the Shapiro-Stiglitz model.

Summarizing, experimental research pertaining to the labor market finds some support for the comparative static implications of rational job search models and labor-leisure decisions. While that work focuses exclusively on labor supply decisions, work by Fehr and associates has considered both labor demand and supply decisions. Consistent with efficiency wage theories, Fehr and associates have provided evidence that incomplete labor contracts and reciprocity concerns can lead to above market clearing wages and involuntary unemployment. The collection of papers by Fehr and associates in particular is an excellent illustration of how a body of knowledge can be built up from a simple experimental game, to which additional features are incrementally added. The evidence provided in all of these studies, e.g. on the formation of reservation wages or the extent of involuntary unemployment would be difficult to observe or identify outside the controlled environment of the laboratory.

4.3 International Economics

A third sector of the macro-economy where experimental methods have been employed is the international sector. The justifications for an experimental approach to international economics are similar to those we have seen before: the available field data does not allow for precise tests of theoretical predictions nor is it possible to abstract away from complicating factors, for example, transport costs or multi-lateral as opposed to bilateral two-country trade (most theoretical models assume the latter).

Noussair et al. (1995) conducted the first experimental test of two key principles of international trade: comparative advantage and factor price equalization. They consider two experimental environments involving 8-16 subjects each. The first is a labor-only, Ricardian model and the second is one where both capital and labor are used as inputs into production.[30] In both environments there are two countries and within each country two player types: consumers and producers. Producers and consumers have induced desires to produce and consume quantities of the two goods Y and Z. In the Ricardian model, consumers inelastically supply labor L to producers for "francs" (money) which they use to buy quantities of the producers' goods Y and Z. Producers use labor as input into production of good Y and Z. There are equal numbers of consumers and producers in each country and all subjects have the same endowments of labor and money. The two countries differ only in their production technologies:

$$\begin{array}{lll} \text{Country 1} & Y_1 = 3L_1 & Z_1 = L_1 \\ \text{Country 2} & Y_2 = L_1, & Z_2 = 2L_2 \end{array}$$

Thus country 1 (2) has a comparative advantage in the production of good Y (Z). While labor supplies, L_1 and L_2, are not mobile across countries, trade in goods is possible, and there is no perceived difference in good Y (Z) produced by either country. Thus in the Ricardian model, there are six markets, two internal labor markets and four external goods markets for the two goods Y, Z, produced by each of the two countries. These were implemented using computerized double auctions and induced values for inputs (by producers) and for goods bought (by consumers) and sold (by producers). The main hypothesis tested in this design is the law of comparative advantage; in the competitive equilibrium, trade occurs in the sense that members of two countries buy and sell goods Y and Z to one another with county 1 completely specialized in the production (sales) of good Y and country 2 completely specialized in the production (sales) of good Z. This prediction may be contrasted with the inefficient autarkic outcome in which there is no trade between countries, and hence no specialization. The second environment which adds capital, differed in that the two countries had identical linear production technologies, i.e., $Y = L$ and $Z = K$ in both countries, but different aggregate endowments of labor and capital and there was now an internal market for both labor and capital (both immobile factors). Thus this economy had eight markets. The main

[30] The first experiment involving both input and output markets was conducted by Goodfellow and Plott (1990).

prediction of this environment is that both countries produced both goods and in the competitive equilibrium, country 1 would be a net exporter of good Y and country 2 a net exporter of good Z. Further, in the competitive equilibrium, prices of the two goods should be equalized across countries and this further implies factor price equalization. Such equalization does not occur under autarky.

The experimental results are somewhat mixed. On the one hand, there is strong support for the law of comparative advantage; in the Ricardian environment there is nearly complete specialization by producers in the two countries, and in the environment with capital, the two countries net exports are of the good for which they hold a comparative advantage. Further in the environment with capital, output prices are equalized across countries and given the identical linear production functions, so are factor prices. The latter finding is one that would be very difficult to observe outside of the controlled environment of the laboratory, as it only holds in special cases such as the one induced here. On the other hand, input and output prices are neither consistent with competitive equilibrium or with autarkic levels. Noussair et al. argue that production and consumption patterns appear to be converging toward competitive equilibrium levels especially under free trade (they also consider some environments with tariffs). As evidence for convergence, they make use of regression equations of the type (1) discussed earlier in section 2.1.

In a related paper, Noussair et al. (1997) focus on issues of *international finance*: exchange rate determination, the law of one price and purchasing power parity. They simplify the set-up from their prior experiment so that there is no longer any factor inputs or production processes; there is simply an endowment of two final goods X and Y in each of the two countries, A and B. A further difference is that each country now has its own money. Each country was populated by six subjects, three of whom were sellers of (endowed with) good X and buyers of good Y and the other three were sellers of (endowed with) good Y and buyers of good X. In addition subjects were endowed with amounts of their *home* currency only. As in the prior study, a demander of good X was indifferent between acquiring X from a supplier in his home country or in the foreign country. However, foreign country purchases required acquisition (cash) in advance of the foreign currency. A further restriction designed to force the use of currency markets was that residents of one country could not transport and sell goods abroad so as to obtain foreign currency for purchases abroad. (On the other hand, goods purchased abroad could be costlessly transported home). In each country, markets in the two goods and foreign currency were implemented using computerized double auctions. Subjects were induced to value quantities of goods X or Y and the home currency only; the end-of-session redemption value of any foreign currency holdings was zero.

The exchange rate e –the price of currency A in terms of currency B – is determined according to the *balance of payments* approach wherein e equates the demand and supply for currencies A and B arising out of the flow of international transactions, as predicted by comparative advantage: in the competitive equilibrium, country A (B) is an importer of good X (Y). Given this balance of payments view, and supposing that trade occurs, the main hypothesis tested concerns the law of one price:

$$eP_X^A = P_X^B \quad eP_Y^A = P_Y^B$$

or that adjusting for exchange rates, goods X and Y have a single world price. The alternative hypothesis is again, that the inefficient, autarkic, no-trade outcome is realized in which case the law of one price does not hold.

The experimental findings are somewhat mixed, though the authors conclude that their data are closer to the competitive equilibrium than to the autarkic predictions, again using regression equations of the type (1). On the one hand, they find somewhat remarkable (given the complexity of the environment) evidence of convergences to the competitive equilibrium exchange rate prediction $e = 47$ across four sessions as shown in Figure 10. On the other hand, the law of one price (and a variant, purchasing power parity, that is based on price level indices) fails to obtain. Noussair et al. (1997) conjecture that this failure arises because of different speeds of convergence of prices in the two domestic markets, which leads to a failure of the law of one price even though the exchange rate is at the competitive equilibrium level. Increasing the duration of the experiment beyond the ten 15-minute trading periods in a session might have allowed for such a convergence to take place.

[Figure 10 here.]

One observation regarding this pair of experiments is that the autarkic outcome, while soundly rejected, is something of a straw man; absent restrictions on trade, the no-trade outcome does not comprise an equilibrium and is rationalized as being plausible if subjects are so averse to foreign exchange market uncertainty that they refuse to engage in trade. Nevertheless, the important value of these experiments in illustrating how basic tenets of international trade and finance can be tested in the laboratory cannot be emphasized enough, and much further work could be done along these same, lines e.g., allow capital flows across countries.

Some theoretical work on exchange rate determination is in environments where there are no restrictions on portfolio holdings and the demands for currencies are endogenously derived, as opposed to the cash-in-advance induced demand for currency in the design of Noussair et al. 1997. In this more general environment, if two monies are perfect substitutes and there is no government intervention in currency markets or legal restrictions on currency holdings, the exchange rate may be *indeterminate*. Further, if agents have perfect foresight it is predicted that whatever the exchange rate turns out to be, it will be *invariant* over time, as in the overlapping generations model of Kareken and Wallace (1981). These two predictions are tested in an experiment by Arifovic (1996) that was designed for comparison with the predictions of an agent-based model (a genetic algorithm). In the experiment there was a single consumption good and equal, fixed supplies of two currencies, francs and lire. As the environment is an overlapping generations model, even/odd-numbered subjects alternated every even/odd period between being young and receiving endowment ω_y of the consumption good and being old and receiving endowment ω_o of the consumption good with $\omega_y > \omega_o$. They were then reborn as young agents, repeating the two-period cycle of life anew. Subjects were induced to hold log preferences over consumption in the two periods of life, so their optimal plan involves consumption smoothing, or selling some of their endowment for the two monies (the only stores of value) when young and redeeming these money holdings in the next period at prevailing prices for old-age consumption. Initial period "old" subjects were endowed with equal amounts of the 10 units of the two currencies. Each young subject was called on to make two decisions - how much of their youthful endowment to save (the remainder was consumed) and what fraction of their savings was to be held in domestic currency; the remainder was placed in foreign currency holdings. Old subjects inelastically supplied their money holdings for consumption. The exchange rate between the two currencies was that which equated youthful demands for, and old agent supplies of the two currencies. The main experimental finding (from just two experimental sessions!) was that the mean exchange rate was about 1, but counter to the stationary perfect foresight equilibrium prediction, there was persistent fluctuations in the exchange rate. Arifovic attributes this volatility to small changes in the portfolio decisions of young agents in response to immediate past differences in rates of return on the two currencies, which in turn generates volatility in the exchange rate in a continual feedback loop. Observed volatility in exchanges rates has been difficult to explain - many attribute it to "news" or "sunspots" – but Arifovic's experimental finding of adaptive learning dynamics with regard to portfolio decisions provides a new alternative.

Fisher (2001) revisits the issue of the law of one price and purchasing power parity that Noussair and associates failed to observe in their experiment by constructing a greatly simplified, version of the Noussair et al. (1997) environment. In Fisher's design, each country produces only a single good, the prices and supplies of which are perfectly controlled by the experimenter, so the main job of subjects (as in Arifovic (1996)), is to determine the nominal exchange rate. The two goods and currencies are "green" (domestic) and "red" (foreign), and green(red) currency is required in advance to buy green (red) goods (so, this is again the case of a cash-in-advance induced demand for currency). Each subject begins a session endowed only with a large supply of the green currency.[31] The price and end-of-session redemption value of a unit of the green good, p^g and v^g are fixed and known for the duration of a session as is the end-of session redemption value of a unit of the red good, v^r and the green currency. Red currency is in limited supply, has no end-of session redemption value and cannot be carried over from one period to the next; its main purpose is to purchase the red good. The red currency price of a unit of the red good in period t, p_t^r, –a treatment variable– is randomly determined from a set of values and announced at the beginning of each of the 10 periods that comprise a session. Supplies of the two goods are unlimited, but $v^r > v^g$, which motivates a demand for the red good and red currency. The limited supply of red currency each period, equal to just $k - 2$ units where k is the number of subjects, is held by the experimenter. After the unit price of the red good for the period (p_t^r) is announced, the supply of red currency is auctioned off in a second-price, sealed bid auction. Each subject could bid amounts of green currency for just one of $k - 2$ units of red currency during this first auction phase of a period. The market-clearing price of a unit of red (foreign) currency in terms of green (domestic) currency (equal to the second lowest bid submitted) is interpreted as the nominal exchange rate for period t, e_t. Once e_t is determined, subjects were free to buy units of green and red goods subject to cash-in-advance and budget constraints. Fisher's main hypothesis - a relative version of purchasing power parity -is that the *real* exchange rate in each period t, defined by $q_t = e_t p_t^r / p^g$, is invariant over time, i.e., that the market clearing, nominal exchange rate e_t immediately adjusts to the announced red good prices, p_t^r so as to keep the real rate, q_t, constant. A related hypothesis, absolute purchasing power parity, posits that the real exchange rate q_t equals v^r/v^g, the marginal rate of substitution between foreign and domestic goods, so the nominal exchange rate in each period is determined according to:

$$e_t = \frac{v^r p^g}{v^g p_t^r}.$$

Thus Fisher's exchange rate determination process arises out of purchasing power parity as opposed to the balance of payments approach to exchange rate determination followed by Noussair et al. which relies on trade flows between countries. With this stripped down experimental design involving perfectly controlled prices, Fisher finds convincing evidence for both the relative and absolute versions of purchasing power parity. This finding confirms a conjecture of Noussair et al. that the failure of purchasing power parity in their study was likely owing to the slow and differential convergence of prices in the goods markets; in Fisher's design there is no problem with non convergence of goods prices as these are pre-determined. Fisher also adds an interest rate to red currency

[31] As in Arifovic (1996), one can think of all the subjects in Fisher (2001) as residing in the domestic country only, but having access to foreign currency.

holdings over a subperiod of each period as well as uncertainty regarding the price of the red good in order to test hypotheses related to covered and uncovered interest parity. He finds support for these hypotheses as well.

Having studied a greatly simplified exchange rate environment, Fisher (2005), in a follow-up paper, seeks to understand two complicating factors that might account for the widespread lack of evidence in support of purchasing power parity and (un)covered interest parity in econometric analyses of historical field data.[32] He considers the role of 1) non-traded goods, which if sizeable, may lead to failures of purchasing power parity in analyses using aggregate price indices, and 2) non-stationary price level dynamics. Proxies for these two complicating factors are introduced into the design of Fisher (2001). Both non-traded goods and non-stationary goods prices are found to increase the deviation of exchange rates from theoretical predictions, with the largest deviations coming from the environment with non-stationary prices.

Summarizing, the laboratory has been used to test some basic principles of international economics including the law of comparative advantage, the law of one price and theories of exchange rate determination and the notion of purchasing power parity. These are phenomena that are either difficult to test (comparative advantage) explain (exchange rate volatility) or which have been refuted in econometric tests with available field data (purchasing power parity). We have seen how experimental methods can shed light on these topics and how building on prior experimental designs can help to clarify puzzling findings, such as Noussair et al.'s finding that purchasing power parity does not hold. Further work on this topic might consider adding dynamic, intertemporal linkages such as would occur by adding capital accumulation or considering intertemporal consumption savings/decisions. One shortcoming of the international experiments reported on here is that, with the exception of Fisher (2005), the number of experimental sessions of a treatment are too few. In implementing complex international economic environments, the temptation is to load up each session with many changes in treatment variables, a practice that is understandable, but one that should be avoided nonetheless.

4.4 Multi-sectoral Macroeconomics

A few courageous researchers (namely Charles Plott and associates), have sought to combine all three of the sectors we have explored in the last few sections by implementing large-scale laboratory macroeconomies. Such multi-sectoral systems involving simultaneous markets for factor inputs, goods, money as well as foreign goods and money, may be what many people have in mind when they hear the term "macroeconomic experiments." I hope it is clear by now that a macroeconomic experiment need not be *elephantine*; rather it suffices that the experiment addresses a topic of interest to macroeconomists. Nevertheless, it *is* of interest to understand the extent to which many interlinked experimental markets can operate simultaneously, so as to identify the source of inefficiencies.

A first effort at developing such a multi-sectoral laboratory macroeconomy is found in Lian and Plott (1998) who implement a static, Walrasian competitive general equilibrium model. There are two types of agents, consumers and producers, two goods, X and Y and a constant supply of fiat money. Consumers were induced to have a preference function $U(X, Y)$ over the two goods. Each period they were endowed with 0 units of X and a constant amount of Y; X can be interpreted as

[32] For instance, random walk models have consistently outperformed any economic theory of exchange rate dynamics.

a consumption good and Y as labor/leisure. Producers desired good Y only, and could consume it directly (e.g., labor services) or use it as input into production. Producers were endowed with a concave, labor only production technology yielding $f(Y)$ amount of good X for Y units of input. Producers were endowed with an amount of fiat money and good X in the first period only; these endowments were not refreshed in subsequent periods (e.g. a constant money supply). In simultaneously operating, multi-unit double auctions, consumers could trade good Y with producers for fiat money and consumers could purchase good X from producers in exchange for fiat money - i.e. a cash-in-advance constraint was binding. Units of X or Y that were consumed/used as input into production left the system (subjects received redemption values for these based on their induced utility/production functions). Somewhat strangely, remaining balances of X and Y were carried forward to the next period, investing consumption and labor with a durability -and asset value– that they would not ordinarily possess (and which would have obviated the need for fiat money, absent cash-in-advance constraints). Finally, all subjects (producers and consumers) had access to a financial market where they could borrow and lend to one another in fiat-money-denominated contracts through a one-period bond market. Default was discouraged through the use of large exogenous fines. Given the initial cash endowments, in the static competitive equilibrium resulting from consumer and producer optimization, there are no cash constraints and financial markets should not operate. However, there is a unique equilibrium volume of production and consumption of goods and ratio of the price of a unit of Y to a unit of X that is independent of the number of subjects. Each session consisted of a number of periods. The final period was not announced in advance; market prices during that final period were used to evaluate final inventory holdings which were redeemed into cash at a fixed rate.

The economy is illustrated in Figure 11, which also gives the induced utility and production functions used in the study.

[Insert Figure 11 here.]

Subjects for this study were nontraditional, consisting primarily of high school students participating in a summer school program at CalTech. In addition, one session involved science and engineering graduate students from the People's Republic of China. Aside from these different subject populations, the main treatment variables were variations in the exogenous money supply and the experience level of subjects (whether they participated in more than one session).

Among the main findings, Lian and Plott provide convincing evidence that there is considerable order to the observed economic activity. Using regression equations of the form (1), they show that convergence toward the competitive equilibrium outcome appears to be occurring, albeit slowly; indeed, they formally reject the hypothesis that the competitive equilibrium is actually achieved. Still, the ratio of the price of Y to the price of X, predicted to be 2, is found to be around this level in all sessions. Volume in both the input and output markets is only slightly less than predicted, and this is attributed to overconsumption of Y by consumers and underproduction of X by producers, who also overconsumed Y. Financial markets are rarely used as predicted. Experience is shown to matter greatly in reducing the volatility of prices and volume and improving efficiency. Changes in the money supply have proportionate effects on the price level but no real effects, and the velocity of circulation of money appears to hit a constant level, especially with experience. Perhaps the most intriguing findings are based on constructed measures of unemployment, inflation and real GNP. Using these, Lian and Plott 1) find no evidence for any inflation-output Phillips-curve type trade-off and 2) strong support for a negative trade-off between changes in the unemployment rate and changes in real GDP (a version of Okun's law[33] With a keen knowledge of how their macroeconomy operates, Lian and Plott interpret the latter phenomenon as "no surprise...A fall in unemployment translates to an increase in system efficiency and that becomes an increase in income and thus real GNP. (p. 62)."

Building on Lian and Plott (1998) as well as Noussair et al. (1995, 1997), Noussair et al. (2007) develop an experimental multi-sectoral macroeconomy which they claim [p. 50] is "far more complex than any laboratory economies created to date." This claim cannot be disputed. The economy has 3 output goods, x, y, and z, two factor inputs, labor l and capital k all of which are specific to one of three countries A, B and C, each of which has their own currency, a b c. Thus there are 21 double auction markets in simultaneous operation – 7 markets in each country - the three goods markets, the two input markets and two currency markets. Three experimental sessions were conducted each involving in excess of 50 subjects; two of the three experiments were conducted remotely via the internet. The subjects were divided up roughly equally into twelve types, with each type being characterized by a country of residence and typically assigned two of three possible roles: as a producer of output goods, consumer of two output goods or supplier of input goods. The precise roles of each subject type, their (continuous) induced production function $f(k, l)$, utility function over the two goods $U(\cdot, \cdot)$ and/or supply of input cost function $C(k, l)$ is given in Table 4.4

The actual functions were discretized and presented to subjects as tables. Using the induced functions in Table 4.4, aggregate demand and supply functions can be calculated. Using these, the competitive equilibrium can be found using 15 market clearing conditions for output and input markets, together with three law-of-one price (no arbitrage conditions) and three flow of funds equations determining exchange rates (country A,B,C has a comparative and absolute advantage in x, y, z, respectively. As in Noussair et al. (1997), the main comparison is between the efficient, full trade, competitive equilibrium prediction and the autarkic, no trade outcome. The main difference between Noussair et al. (2007) and Noussair et al. (1997) is the addition in the former of factor input markets for labor and capital. In essence, the Noussair et al. (2007) environment is a combination of Noussair et al (1995) and (1997) with a third country added, and a proportionate increase in the number of subjects.

What is the motivation for such an exercise? As in Lian and Plott (1998), it is to demonstrate that such an experiment is possible, and that competitive equilibrium remains an attractor despite the complexity of the environment. As the authors themselves say,

> The number of [excess demand] equations explodes as the number of commodities and resources increase, but theory itself suggests no effects of the increased complexity. On the surface, the thought that a decentralized system of competitively interacting humans might approximate the [competitive equilibrium] solution as the number of equations grows large is a staggering and contentious proposition that many cannot believe without demonstration. (Noussair et al. 2007, p. 50).

[33] More precisely, Okun's law predicts that a 1% increase in unemployment above the natural rate is associated with a reduction in real GDP of 2-3%. Lian and Plott find evidence for a negative and roughly proportional trade-off between changes in unemployment and real GNP.

The main finding of Noussair et al. (2007) is again that the most prices, wages, exchange rates, production, consumption and trade volumes are closer to the competitive equilibrium prediction than to the autarkic outcome again using regression equations of the form (1). In this study, however, the pattern is less obvious than in the simpler economies of Lian and Plott (1998) and Noussair et al. (1995, 1997), perhaps reflecting the additional complexity of this environment. Among other new findings, there appears to be much more pronounced "home bias", in the sense that imports are considerably lower than competitive equilibrium levels. Further, price volatility is greatest in exchange rates, intermediate for producer (input) prices and lowest for output prices. Interestingly, Noussair et al. attribute these findings to less than complete equilibration as opposed to the more traditional view that markets are in equilibrium and institutional factors, government policies or exogenous shocks are responsible for any observed inefficiencies or volatility.

The construction of complex large-scale macroeconomic laboratories to demonstrate that the static, competitive general equilibrium model works reasonably well is an important achievement. Further work along these same lines might seek to incorporate financial market frictions (e.g. binding liquidity constraint) and intertemporal, forward-looking behavior, in which expectations of future variables determine current quantities as in much of modern, dynamic macroeconomic modeling. Of course, a difficulty with this research agenda is that the systems studied are so complex to analyze, not to mention logistically difficult and costly to implement that other researchers may be discouraged from following up with the crucial replication and extension studies that are essential to scientific progress. Perhaps as computing, coordination and recruitment costs decline further with further innovations in social networking technology, multi-sectoral macroeconomic experiments of the scale studied by Plott and associates will become more commonplace.

An alternative possibility is that researchers interested in large-scale macroeconomic experiments will instead mine the synthetic worlds of massively multi-player online role-playing games (MMORPGs). For instance, Castranova (2006) provides some evidence of large-scale social coordination in response to "natural experiments" on two MMORPGs, Sony's "Everquest" and Mythic Entertainment's "Dark Age of Camelot." Generating any level of controlled intervention over outcomes in such environments seems difficult at best, unless of course, the experimenter was also the game designer.

5 Macroeconomic Policies

As we have seen, many researchers have felt confident that they could test the predictions of modern, micro-founded macroeconomic models in the small scale of the laboratory. It should not be surprising, then, to find that several researchers have also used the laboratory to examine the effects of macroeconomic *policies*. As such experimentation is not typically feasible (not to mention ethical) for macro policymakers, the laboratory provides an important and (to my mind) under-utilized environment in which to assess the likely impact of macroeconomic policies before such policies are actually implemented.

5.1 Ricardian equivalence

One important macroeconomic policy debate which experimentalists have contributed to concerns whether a temporary fiscal stimulus, financed by government borrowing is preferred to a tax-financed stimulus or as Barro (1974) put it, whether government debt is viewed as net wealth. In

Barro's reformulation of the Ricardian equivalence doctrine, given an operational intergenerational bequest motive, lump-sum taxes, perfect capital markets and no change in government purchases, the timing of tax levies (now or later) is irrelevant. An issue of government debt to finance temporary spending is readily absorbed by the public who anticipate using these bond holdings to pay for the future increase in taxes thus leaving all real variables, e.g. interest rates, unaffected. Thus, the consequences of a bond or tax financed stimulus are equivalent: there are no real effects. The empirical evidence using field data on whether Ricardian equivalence holds or not is mixed (for contrasting conclusions see Bernheim (1997) and Seater (1993)). However the environment in which the Ricardian doctrine holds, e.g., lump-sum taxes, strong intergenerational bequest motive, etc. is not one that is necessarily observed in nature. For this reason, the laboratory may be the more desirable place in which to explore the question of Ricardian equivalence and indeed, two experimental studies explore this question.

Cadsby and Frank (1991) design an experiment that closely mimics the overlapping generations model that Barro (1974) used to formalize the notion of Ricardian equivalence. In Cadsby and Frank's design, an experimental session involves 8-10 rounds with each round consisting of three periods, A, B and C. At the start each session, subjects were anonymously paired. Within each pair, one member played the role of generation 1 and while the other played the role of generation 2. Pairings and roles were fixed for all rounds of the session. Subjects were endowed with tokens in various periods, and these could be converted into certificates (consumption) at a price of 1 token =1 certificate, or tokens could be stored for future periods (savings). Members of generation 1 make consumption and savings decisions in period A, denoted by C_1^A, S_1^A and also in period B, denoted by C_1^B, S_1^B, and are inactive in period C. The savings of generation 1 in period B, S_1^B, which is constrained to be non-negative, is given as a *bequest* to their generation 2 partner and is available to that partner at the start of period C. A bequest motive for members of generation 1 was induced by the choice of preferences (as illustrated below). Members of generation 2 have no bequest motives - they can be viewed as the descendents of generation 1– and are inactive (unborn) in period A. Those in generation 2 also make consumption and savings decisions in period B, denoted C_2^B, S_2^B; they do this knowing the amount of any tax they will face in the final period C. In period C, the remaining savings of generation 2, including bequests received S_1^B from generation 1, are consumed (converted into certificates). After period C ends, the round was complete and if the last round has not been played, a new round began following the same sequence of choices, and refreshed endowments. The main treatment variables consisted of the token endowments generations 1 and 2 received in periods A and B, and the amount of deficit spending (the tax burden) generation 1 received in period B and generation 2 was required to repay in period C. There was also some variation in the induced preference functions, with a multiplicative utility function performing better than an additive one. The hypotheses concerned the amounts consumed, saved and bequeathed in response to temporary expansionary and contractionary government policies. To simplify the environment as much as possible, there was neither discounting nor interest payments on government debt. Here I will describe one experiment, #3, that seems representative of Cadsby and Frank's experimental design and findings.

In this experiment, generation 1 agents' induced utility function was of the multiplicative form: $U_1(C_1^A, C_1^B, U_2) = C_1^A C_1^B U_2$ and included as an argument the utility of generation 2, which was given by $U_2(C_2^B, C_2^C) = C_2^B C_2^C$; this was the manner in which a bequest motive was operationalized. Notice both agent types should seek to intertemporally smooth consumption and in that regard, the experiment can be viewed as another test of intertemporal optimization (as discussed at the beginning of this chapter), albeit now with a bequest motive added. In years 1-5, generation 1 received token endowment E_1^A in period A and 0 in period B, while generation 2 received token endowment of E_2^B in period B and 0 in period C. In years 6-10, generation 1 received endowment E_1^A in period A as before but now received an additional token endowment in period B of $E_1^B > 0$. The latter is viewed as temporary deficit spending. Generation 2 received endowments of E_2^B in period B as before but now had to pay a tax out of accumulated savings at the start of period C equivalent to E_2^B, an amount that was precisely equal to the amount of his parent's period B endowment. Under perfect foresight, the optimal consumption/savings plan is derived by solving generation 2's problem first:

$$\max_{C_2^B, C_2^C} U_2 = C_2^B C_2^C \text{ subject to: } C_2^B \leq E_2^A \text{ and } C_2^C \leq S_2^B + S_1^B - E_1^B$$

and using the maximized value of U_2^* to solve the first generation's problem:

$$\max_{C_1^A, C_1^B} U_1 = C_1^A C_1^B U_2^* \text{ subject to: } C_1^A \leq E_1^A \text{ and } C_1^B \leq S_1^A + E_1^B \text{ and } S_1^B \geq 0$$

The latter constraint requires that bequests from generation 1 to 2 cannot be negative. The endowments in experiment 3 were chosen in such a way that for the first 5 years, when there was no deficit spending, the optimal, perfect foresight bequest amount from generation 1 to 2, $S_1^{B*} = 7$. Beginning in year 6, when generation 1 started receiving an endowment (deficit spending) of $E_1^B = 42$ at the start of period B, that had to be repaid by generation 2 at the start of period C, the optimal bequest rose proportionately to $S_1^{B*} = 49$,i.e., the Ricardian prediction, $\Delta S_1^{B*} = \Delta E_1^B$, holds.

Cadsby and Frank show that in this experiment as well as several other treatments, the prediction of Ricardian equivalence is approximately correct, and the predictions of a purely myopic model in which no bequests are given $S_1^{B*} = 0$ can be soundly rejected. Figure 12 shows individual and average bequests in the treatment we have discussed. Following the change in endowment patterns in year 6, bequests jump from an average near 7 to a neighborhood of 49. As Cadsby and Frank acknowledge, however, the introduction of the deficit policy "produced slightly Keynesian results in every case" i.e., the Ricardian equivalence was not perfect. This can be seen in Figure 12 where the average bequest lies below 49 even in the final year 10. It may be that such small Keynesian effects account for the continued belief by many in the efficacy of deficit policies.

[Insert Figure 12 here.]

Two further experiments build on Cadsby and Frank's design. Slate et al. (1995) change the design so that subjects face uncertainty as to whether the full amount or a smaller fraction of the deficit spending must be repaid. The find that when the probability of full debt repayment is low, Ricardian equivalence fails to hold -generation 1 subjects overconsume and leave too little a bequest. As the probability of full debt repayment becomes larger, so do bequests, which more closely approximate the levels associated with Ricardian equivalence. Ricciuti and Di Laurea (2003) change the overlapping generations matching protocol so that players are not always in the same role or in fixed pairs. They consider the role of two additional complicating factors that may well prevent members of generation 1 from making neutral bequests - 1) liquidity constraints and 2) uncertainty about future (second period) income. They find that both of these complicating factors reduce the likelihood that subjects in the role of generation 1 make bequests that neutralize the debt burden on generation 2, relative to the baseline case.

Future work on the economic impact of deficit spending might consider environments where government bonds pays interest, and there also exist markets for private savings. In that case, the more mainstream, neoclassical view, that deficits crowd out private sector investment, could be explored as a rival to the Ricardian view that they have neutral effects.

5.2 Commitment versus discretion

Another important macroeconomic policy issue concerns the suboptimality of time-consistent, "discretionary" policies that do not commit the policymaker to a predetermined policy response but are instead optimal for the current situation only, taking current expectations as given and ignoring private sector expectations with regard to future policies. As Kydland and Prescott (1977) first showed in the context of a two-period, expected inflation-output (Phillips curve) model, following this time-consistent policy can result in the policymaker ratifying the inflation expectations of the public resulting in an excessive level of inflation and no change in unemployment relative to the social optimum, which involves a zero inflation rate. The social optimum could be implemented by a policymaker who was able to pre-commit once and for all to zero inflation, but such a "commitment technology" is not typically observed in nature. Kydland and Prescott thus argued in favor of policy rules, rather than discretionary policies. Barro and Gordon (1983) recast the inflation-unemployment trade-off as a non-cooperative game between the policymaker and the private sector, which is fully aware of the policymaker's objective function and forms expectations rationally. In an infinitely repeated version of this game, they show that if the policy maker and private sector care enough about the future (have high discount factors), the socially optimal policy (zero inflation, unemployment at the natural rate) may be sustainable as an equilibrium through the use of a grim trigger strategy (many other equilibria are possible as well, as the Folk theorem of repeated games applies). The recasting of the policymaker's problem as a game makes it amenable to testing in the laboratory, and indeed there are two experimental studies that take aim at this issue.

Van Huyck et al. (1995, 2001) use a "peasant-dictator" game to explore policymaking under 1) full pre-commitment of policy (not observed in nature and thus ripe for experimental testing) 2) discretionary, one-shot policymaking and 3) the repeated game case, where reputational concerns from repeated interactions with the private sector may induce the policymaker to embrace policies closer to the social optimum (commitment solution). Subjects in the two-player, two-period stage game are assigned roles as either 'dictators' or 'peasants'. In period 1, peasants are endowed with amount W of beans and must decide how much of these to consume $c_1 \geq 0$, or invest $k \geq 0$, earning a gross return of $(1+r)k$ in period 2; $r > 0$ is exogenous. The second period consumption $c_2 \geq 0$ depends on their investment and the fraction τ of the bean harvest taxed by the dictator. Formally, the peasant's problem is:

$$\max_k U = c_1 + c_2 \text{ subject to } c_1 = W - k \text{ and } c_2 = (1-\tau^e)(1+r)k$$

Here τ^e is the expected tax rate; in the commitment case only, there would be no uncertainty about τ as it is announced in advance of peasant's investment decisions. As utility is linear (no need to consumption-smooth) The dictator's best response function is:

$$k^* = \begin{cases} W & \text{if } (1-\tau^e)(1+r) > 1 \\ 0 & \text{if } (1-\tau^e)(1+r) < 1 \\ [0,1] & \text{if } (1-\tau^e)(1+r) = 1 \end{cases}$$

Under commitment, the dictator moves first and solves:

$$\max_{\tau \in [0,1]} R = \tau(1+r)k(\tau)$$

yielding the social optimum $\tau^* = r/1+r$. Given this, it is a weak best response for the peasant to set $k^* = W$, and this is the unique subgame perfect equilibrium. Under discretion, the dictator moves after the peasant has made an investment choice and so optimally chooses $\tau = 1$. Knowing this, peasants choose $k = 0$. A further solution they consider is the Nash bargaining solution which results in a split-the-surplus tax outcome: $\tau = 1/2\tau^*$. Finally, they note that in the infinitely repeated game, implemented with fixed pairings constant probability of continuation, if the discount factor is sufficiently high, trigger strategies can support the social optimum commitment solution, as well as other equilibria, e.g. equal division or the Nash bargaining solution. The experimental design involved three regimes: commitment (C) and discretion (D), implemented as a sequence of one-shot games (random matching) with there different timing of moves (dictator/peasant) or (peasant/dictator) and the reputational indefinitely repeated game, involving fixed pairings for each supergame and $\delta = 5/6$. The other main treatment variable was the peasant's endowment, $W$and the rate of return r, which were varied subject to the constraint that $W(1+r) = \$1$. Mean experimental earnings from at least 20 rounds of the stage game are shown in Figure 13 for various cohorts (C), (D) and (R) under various values of W. The shaded regions show feasible repeated game equilibrium payoffs.

[Figure 13 here.]

Generally speaking discretionary cohorts (D) are closer to the discretionary equilibrium, commitment cohorts (C) are closer to the commitment equilibrium, and reputational cohorts lie somewhere in between. In summary they find that reputation is indeed an imperfect substitute for commitment. It is also sensitive to r; as $r$decreases/Wincreases, reputational concerns are weakened with a corresponding efficiency loss.

Arifovic and Sargent (2003) pursue a similar question to that of Van Huyck et al. (2001) - whether the optimal, commitment solution can be implemented by policymakers lacking commitment. The Arifovic-Sargent experiment, however, is in the context of a repeated version of Kydland and Prescott's expectational Phillips curve model, where the policy maker controls the *inflation rate*. The motivation for this exercise is also different as it focuses the predictions of models where the private sector dues not have rational expectations (is unaware of the inflation output trade-off) but instead forms its expectations adaptively (the central bank is fully informed of the model). In one model of adaptive expectations due to Phelps (1967), with a sufficiently high discount factor, the government eventually chooses inflation rates consistent with the commitment level. In another model of adaptive expectations due to Sargent (1999) the discretionary "Nash" equilibrium is the only limiting equilibrium.

The experimental design involves $N + 1$ subjects with $N = 3 - 5$. N subjects play the role of the private sector, moving first by forming expectations of inflation. Unlike the peasants in the Van Huyck et al. experiments, the N private sector subjects in Arifovic and Sargent's design know nothing about the inflation unemployment trade-off, nor the central bank's objective, but do know the central bank controls inflation. Private sector subjects have access to the path of past inflation (and unemployment) and can use that information in forming expectations. Thus, the design induces them to form expectations of inflation adaptively, consistent with the theory being tested. The mean value of the N inflation expectations each period is regarded as the economy's expected inflation rate, π^e. The lone central banker, picked randomly, moves second. She also has access to the past history of unemployment, actual inflation and, in most treatments, past private sector expectations of inflation π^e. She *is* aware of how the economy works, and faces a problem of the form:

$$\min_{x_t} \sum_t \delta^t (U_t^2 + \pi_t^2) + v_{1t}$$

subject to

$$\begin{aligned} U_t &= U^* - (\pi_t - \pi_t^e) \text{ (Phillips curve tradeoff)}, \\ \pi_t &= x_t + v_{2t} \text{ (CB control of inflation)}. \end{aligned}$$

where U_t is the unemployment rate, U^* the natural rate (set equal to 5 in the experiment) π_t is inflation, x_t is the central bank's inflation choice variable (which was constrained only to be nonnegative) and v_{jt} are mean zero, random noise terms, with $Ev_{jt}^2 = \sigma_j^2$. The commitment solution has $x = \pi^e = 0$, while the discretionary equilibrium has $x = \pi^e = U^* = 5$. In the indefinitely repeated experiment, decision rounds continued with probability equal to the discount factor $\delta = .98$ [34] and central bank subjects were paid inversely to the session wide-average value of the policy loss function, $U_t^2 + \pi_t^2$. The N forecasters were paid based on average inflation forecast accuracy. The only treatment variable was the shock variance σ_j^2, either large, .3, or small .03 for both shocks.

The main finding is that in 9 of 12 sessions, inflation starts out close to the Nash equilibrium level, but over time, the subject in the role of the policymaker steers inflation rather smoothly to within a small neighborhood of the commitment equilibrium for the duration of the experimental session. Further, the private sector's expectations closely follow the same trajectory, and become much more homogeneous with experience. In the other three sessions, inflation fails to converge or remains close to the Ramsey equilibrium value. In four of the sessions where the commitment equilibrium is achieved, there is some 'backsliding' in the sense that inflation temporarily rises to near discretionary Nash equilibrium levels. Arifovic and Sargent conclude that Phelp's (1967) model of adaptive expectations appears to best characterize most sessions, as it predicts that the central bank exploits adaptive learning by the public to manipulate expectations in the direction of a zero inflation rate. However, they also not that this model predicts much faster convergence than is observed in the data, and does not predict instances of backsliding.[35]

5.3 Monetary policy decision-making

On the same subject of monetary policy, Blinder and Morgan (2005, 2007) also consider subjects in the role of central bankers. However, their main focus is on whether monetary policy as formulated

[34] A upper bound of 100 rounds was imposed, and sessions were conducted for two hours.

[35] One version of Sargent's (1999) adaptive learning dynamics, constant-gain learning, predicts long endogenous cycles ("escape dynamics") which can rationalize instances of backsliding from the commitment equilibrium to the discretionary equilibrium and back.

by *committees* (groups of policy makers) outperforms individuals (dictators) in stabilizing the economy and whether there is a difference in the speed of decision-making between groups and individuals. The motivation for this research is the observed switch in the 1990s among some developed nations to a more formal committee-based monetary policymaking, as opposed to the prior, informal single decision-maker policy regime.[36] By contrast with the studies discussed in the previous section, the private sector (peasantry) in the Blinder and Morgan studies is eliminated in favor of automated, stochastic, two-equation coupled system for unemployment U_t (an IS curve) and inflation π_t (a Phillips curve) that are used to generate data similar to that of the U.S. economy:

$$
\begin{aligned}
U_t - 5 &= 0.6(U_{t-1} - 5) + .3(i_{t-1} - \pi_{t-1} - 5) - G_t + e_t \\
\pi_t &= 0.4\pi_{t-1} + 0.3\pi_{t-2} + 0.2\pi_{t-3} + 0.1\pi_{t-4} - 0.5(U_{t-1} - 5) + w_t
\end{aligned}
$$

Here, the natural rate of unemployment is 5, e_t and w_t are mean zero random shocks with small known support, and G_t represents government fiscal activity, a treatment variable. In this environment, subjects playing the role of the central bank, must repeatedly choose the nominal interest in each period t, i_t. Notice that monetary policy impacts on unemployment with a one period lag, and via unemployment it impacts on inflation with a two–period lag. Subjects were informed of the data-generating process, equations (**??**–**??**) but *were* told that raising interest rates would lead to lower inflation and higher unemployment and that lowering interest rates resulted in the opposite outcome. Subjects were further told that G starts out at 0 and sometime during the first 10 periods, would permanently change to either 0.3. or -0.3, resulting in an equal and opposite change in U_t (via (**??**)).

The two-equation system was initialized at the equilibrium for $G = 0$, all lags of U at 5%, all lags of π at 2% and $i_{t-1} = 7\%$. The variables U_t and π_t were then drawn according to (**??**–**??**) and policy makers were instructed to choose i in each of the subsequent 20 periods so as to maximize a known, policy objective linear scoring function yielding $S = 100 - 10|U_t - 5| - 10|\pi_t - 2|$ points per period. Thus subjects were given the policy targets for U and π, 5%, and 2%, respectively. Changes made to the nominal interest rate i following the first period cost subjects 10 points per change. A within subjects design was followed: in the first 20 periods, 5 subjects made interest rate choices as individuals (no communication). Then, in the second 20 periods they made interest rate choices as a group under either a majority or unanimous voting rule. (The reverse order was not considered). Each member of the group received the group score S in points, so there was no difference in payoff opportunities between the two treatments.

Blinder and Morgan's main findings from 20, 5-player sessions are that 1) groups make decisions just as quickly as individuals and 2) groups make *better* decisions than individuals based on the scoring function, S, and 3) majority or unanimous voting rules in the group treatment yielded the same average scores. These same findings were replicated in a second purely statistical experiment (involving balls drawn from two urns) that was completely devoid of any monetary policy context. The main finding, that groups outperform individuals may rationalize the growing trend toward formal monetary policy committees.

Three other experimental studies examining monetary policy decision-making by individuals or groups have been conducted by Lombardelli et al. (2005), Blinder and Morgan (2007), and Engle-Warnick and Turdaliev (2006). Lombardelli et al. (2005) adopt a context-laden experimental design

[36] For instance, in 1997, the Monetary Policy Committee of the Bank of England replaced the Chancellor of the Exchequer as the primary decision-maker on short-term interest rates.

that is similar to one found in Blinder and Morgan (2005), though their exogenous, two-equation data-generating process for inflation and unemployment has fewer parameters and is calibrated to fit UK time series data. They divide their sessions into more than two phases, beginning with a pre-experiment survey of prior beliefs. The experiment begins with several periods of individual decision-making (choice of interest rates), followed by several periods of group decision-making with or without communication (in the latter case, the median interest rate chosen by group members is implemented), followed by several periods of individual decision-making and finally a repeat of the initial survey instrument. Subjects were given about the same amount of instruction about the economy as in Blinder and Morgan, but were asked challenging survey questions such as: "After how many quarters is the maximum impact of monetary policy on inflation felt?" Answers to such questions in the pre-experiment survey were (unsurprisingly) rather poor, but performance on most questions in the post-experiment survey showed some significant improvement. Consistent with Blinder and Morgan's findings, Lombardelli et al. (2005) also find that groups outperform individuals using the same kind of linear loss-function score. Interestingly, they report that the group learning experience is *not* sustained–when individuals return at the end of the experiment to making decisions individually, their scores significantly worsen –see Figure 14 - even comparing the median of individual scores to the score of groups. This provides even more powerful evidence on the efficacy of group over individual decisions regarding monetary policy.

[Insert Figure 14 here.]

Blinder and Morgan (2007) use their earlier experimental design to study two additional issues related to monetary policy decision-making: the role of group size and of leadership. They report results on four treatments: 1-2) four person groups with or without leaders and, 3-4) eight person groups with or without leaders. In treatments with a leader, the chosen leader was the subject with the highest score in part 1 (individual decision-making). However, the leader was endowed with rather weak *leadership powers*: the ability to communicate the group's decision, to cast a tie-breaking vote and to earn a payoff double that of other group members. While Blinder and Morgan are able to replicate their earlier finding that groups outperform individuals, they find that neither group size, nor leadership has any statistically significant effect. An implication of these findings are that, while monetary policy decision-making committees are a good idea, details of the composition of these committees - the size, or designating a leader - are of second-order importance. Future work on this topic might consider actual policymakers as subjects.

In all three of the prior studies of monetary policy decision-making, the focus is on whether subjects' interest rate choices enabled them to achieve *target levels* for inflation and unemployment given the stochastic data generating process for the economy. Engle-Warnick and Turdaliev (2006) ask whether the interest rate choices of subjects playing the role of central bankers can be characterized by an *instrument rule*–specifically, the *Taylor rule* (Taylor (1993))– which is *optimal* for the environment in which they place their subjects. The environment implemented is a purely backward-looking version of the New Keynesian model due to Svensson (1997). As in the prior studies of central bank decision-making, the data generating processes for inflation, π_t, and output, y_t, are exogenous and stochastic but are affected directly (in the case of output) or indirectly (in the case of inflation) by the nominal interest rate, i_t, chosen each period by the central bank. Subjects were not told the data generating processes for inflation or output, nor were the labels 'inflation' or 'output' used; instead reference was made to variables A and B. Subjects' payoff function induced an objective related to the problem of minimizing the expected loss function:

$$E\sum_{\tau=t}^{\infty}\delta^{\tau-1}1/2(\pi_t-\overline{\pi})^2,$$

where $\overline{\pi}$ is a target inflation rate, set to 5%. Discounting was not implemented; subjects were paid on the basis of their performance in a 50-round game. In one environment they study, the optimal policy rule (based on the quadratic objective and the linear laws of motion for inflation and output) is the Taylor rule:

$$i_t = \gamma_0 + \gamma_1\pi_t + \gamma_2 y_t,$$

while in a second model environment, subjects should additionally place some weight on π_{t-1}. Here the γ's represent coefficient weights for which their are precise (optimal) predictions. The optimal policy predictions involved varying the interest rate between 3.0 and 6.5. More generally, the Taylor *principle*, that stabilizing monetary policy requires a more-than-proportionate response of interest rates to changes in inflation, requires subjects to set $\gamma_1 > 1$. Among the main findings, Engle-Warnick and Turdaliev report that while most subjects did not precisely follow the predictions of the optimal Taylor rule, they did manage to keep inflation largely in check, in a neighborhood of the 5% target, and payoffs were not much lower relative to the optimal expected payoff. Further, a clear majority of subjects placed weight greater than 1 on inflation, in accordance with the Taylor principle, though this weight was typically less than the optimal level. Overall, the findings suggest that Taylor's rule and principle for monetary policy may occur rather naturally to subjects with no prior experience as central bankers, but who face a data generating process for which the Taylor rule is an optimal policy prescription.

Future work on monetary policy making in the workhorse New Keynesian model might consider *forward-looking* models of private sector behavior, where expectations about future inflation and output play a role, and enabling central bank policies to affect these expectations (see, e.g. Woodford, 2003). This would necessitate adding subjects to play the role of the private sector as in the study by Arifovic and Sargent (2003), and eliciting their forecasts. At the same time, it appears that modelling the central bank as a policy *committee* consisting of several subjects as opposed to a single decision-maker (as in Arifovic and Sargent and Engle-Warnick and Turdaliev) also has its advantages, and further experimental studies along the lines of Blinder and Morgan and Lombardelli et al. might be devoted to understanding characteristics of committee arrangements that improve monetary policy decision-making.

5.4 Fiscal and tax policies

Having considered monetary policy, we turn finally to experimental analyses of fiscal and tax policies.

Bernasconi et al. (2006) explore how subjects form expectations about fiscal variables, specifically about government expenditure levels and tax revenues. They present subjects with graphical displays showing the historical time path of government debt, B, the change in the government debt, ΔB, tax revenues, T and in one treatment, history of government expenditures, G. After viewing the historical time series, subjects have up to two minutes to form one-step-ahead forecasts of taxes and, in one treatment, government expenditures as well. The novelty of their design is that the data presented to subjects is the actual OECD historical time series data taken from one of 15 European states primarily between 1970-1998. Subjects were not informed of the country the data come from. In most treatments they were told the name of each historical series, e.g. "tax revenue." Subjects were not particularly knowledgeable about relationships between G, T, and B, a fact the experimenters view as a strength of their study, as it parallels the largely ad hoc, time-series-econometric approach that has been taken to understanding the sustainability of fiscal policies. Subjects are rewarded in a somewhat complicated fashion according to their forecast accuracy, which is assessed every two periods. Thus, this is a "learning-to-forecast" type of experiment. However, like the monetary policy experiments discussed in the last section, subjects are being presented with data that have a more realistic macroeconomic flavor, e.g., in terms of magnitudes, causal relationships etc. Unlike the monetary policy experiments however, there is no feedback from subjects' choices (expectations) to subsequent data realizations; subjects are truly atomistic in this environment. The main finding is that changes in subjects' expectations, ΔG^E and ΔT^E compare poorly with a time-series, vector autoregression model for ΔG and ΔT estimated using the same historical time series presented to subjects. The model that best fits the change in subjects' expectations appears to be one that is weighted-adaptive, with heaviest weight placed on recent forecast errors.

Riedl and van Winden (2001, 2007) design a one (closed) or two-country (international) experimental economy that is quite similar to the set-up of Noussair (1995, 1997) to explore government tax policies in the financing of unemployment benefits. This experimental work is particularly notable for being the first laboratory experiments ever commissioned by a government agency - the Dutch Ministry of Social Affairs and Unemployment - to inform on macroeconomic policymaking. Within each country there are two player types, consumers and producers, two production goods, K (capital) and L (labor) and two final goods X and Y. In the international economy, the goods K and X are tradeable between nations while L and Y are not. Producers are endowed with cash and a CES production function that uses both K and L as inputs. Consumers are endowed with preferences for the two goods and leisure and with amounts of $\overline{K}$, $\overline{L}$, and money. In the international setting, in the "large" country, consumer and producer endowments are seven times those for the other, "small" country; the number of subjects in each country is the same. For each unit of "unsold" labor, $\overline{L} - L$, consumers get an unemployment benefit, w^0 from the exogenous government entity (not a player); this becomes an additional source of money for consumers, in addition to money earned by selling L units of labor at wages w and K units of capital at rents r to producers who require these as inputs to produce X and Y. Consumers also earn money from consumption of these final goods according to their utility functions. Double auction markets for input goods open first, then production occurs, then double auction markets open for final goods. The main focus of these studies is on the unemployment benefits policy. Unemployment benefits in country k are financed (as in many European countries) by a tax rate τ^k, applied to units of labor income, $w^k L^k$. This tax is is paid by producers, who are induced in the design to want to maximize after-tax profits. In the first half (8 periods) of their experimental sessions, τ^k is is held constant at the general equilibrium level associated with a balanced budget. In the second part the benefits tax is adjusted dynamically up to some limit, so as to gradually close any deficits. Specifically, the tax rate is set according to the ratio of paid benefits to the tax base in the prior period,

$$\tau_{t+1}^k = \min\left[\frac{w^0\left(\overline{L} - L_t^k\right)}{w_t^k L_t^k},\ 0.9\right],$$

where w_0 is the constant benefit level.

Riedl and van Winden report that under the stable tax regime of the first half of sessions, wages are too low relative to the marginal revenue product and unemployment is too high, though both measures are moving slowly toward the induced equilibrium levels (as demonstrated in regressions models of the form (1)). This is attributed to producers' reluctance to employ sufficient labor and capital given uncertainties about prices and revenues earned on output. The result is a deficit in the employment benefits program.[37] Following the switchover to the dynamic tax policy, tax rates immediately rise in response to the employment benefits deficit and eventually plateau out rising from 38% to around 70%, and resulting in a more balanced budget. However, this steep tax increase in benefits taxes is associated with a rather large increase in unemployment and reduction in real GDP relative to the constant tax rate policy. It appears that the benefit tax increases on producers discourages them from hiring labor and this, together with an excess supply of labor by consumers, leads to much lower wages and higher unemployment, which leads to further demand for benefits, i.e. a "vicious cycle". Future work on this topic might consider alternative policies for maintenance of a balanced budget including variations in the amount and duration of the unemployment benefit, w^0.

Finally, two experimental studies address redistributive social policies associated with the welfare state.

van der Hiejeden et al. (1998) test a possible explanation for the widespread and sustained public support for pay-as-you-go social security systems in which old, retired agents are paid benefits from taxes on the income of the working young. Viewing such systems as a repeated game played between successive generations of young and old agents, they propose that the social norm of transfers from young to old may be sustained as a Nash equilibrium of the infinitely repeated game by a grim trigger strategy: if one young generation ever failed to make transfer payments to the old, subsequent young generations would revert to a perpetual punishment strategy of transferring zero to all future old generations, including the defecting generation. Their argument relies on the ability of generations to *monitor* transfers made by earlier generations, and thus in one treatment, this monitoring ability is present, while in another it is not. The experimental design involves implementation of a overlapping generations environment in which each of eight subjects takes a randomly ordered turn as a young agent making a *voluntary* transfer to an old agent. Subjects are young in one of the eight periods t, and old in period $t + 1$ and then no longer participate in the round (dead). Young agents have an endowment of 9 units of the consumption good, but only 7 of these units are transferable to the current 'old' subject, who has an endowment of 1 non-transferable unit. Payoffs are proportional to the product of consumption in the two periods of life. The payoff to the subject in the role of generation t is $C_{1t} \times C_{2t} = (9-T_t)(1+T_{t+1})$, where T_t is the transfer made by generation t. After 8 transfer decisions, the round is over and a new one begins involving the same subjects, who make transfer decisions in another random order. Unfortunately, an infinite horizon was not implemented: subjects knew that fifteen 8-round games would be played and consequently there are end-game effects.[38] Still, the results are interesting: while subjects did not achieve the efficient, payoff-maximizing transfer of $T = 4$ units from young to old, they did transfer on average about 2 units per period, with a slight drop-off over time. Further, the amount of transfers was independent of whether monitoring of past transfers was possible; this finding may be due to the (unnatural) repeated interactions among groups of 8 subjects. Indeed, these

[37] Budget balance requires that $\tau^k w^k L^k = w^0 \overline{L} - L_t^k$.

[38] Under these conditions, the repeated game equilibrium sustaining transfers does not exist, as it would unravel via backward induction.

results are reminiscent of experimental studies showing positive contributions in repeated, linear, voluntary contribution mechanisms (see Ledyard (1995)). However, in this case, the transfers are dynamic and intertemporal, hallmarks of macroeconomic systems. The willingness of subjects to sustain a social norm of (low) transfers from young to old, regardless of the ability to monitor, may nevertheless rationalize support of pay-as-you go systems as arising from hard-wired preferences for 'fairness'. Future laboratory work on social security ought to investigate the consequences of government imposed taxes on labor income for consumption, savings and capital formation under both pay-as-you go and fully-funded (private accounts) systems. Such studies could greatly inform current policy debates regarding the merits of these two different systems.

Cabrales et al. (2006) also study whether an efficient, redistributive social contract can emerge in the laboratory. In their case, the redistribution is not from young to old but from rich to poor and the extent of the redistribution implemented by the government is decided by voters under various voting procedures. The basic stage game involves 9 players and consists of two rounds. In the first round, subjects choose high or low effort with high effort costing c and low effort being costless. Those who choose high effort earn high income y^h with probability 2/3 and low income y^l otherwise. Those choosing low effort earn low income y^l with certainty. Once effort choices and incomes are determined and revealed to subjects, the next round of the game is played in which all subjects vote on whether to equalize ("redistribute") incomes so that each player $i = 1, 2, ..9$ receives $\frac{1}{9}\sum_{i=1}^{9} y_i$. The actual equalized income level is revealed to subjects in advance of the vote. Three voting procedures are considered: majority rule, unanimous consent, or majority rule voting only by those who chose high effort. In a fourth treatment, incomes are randomly assigned and subjects only vote in the second round under majority rule. If income equalization fails according to the voting procedure, then each subject gets the income they earned, y^l or y^h. A one-shot version of the two-round game under majority rule is like a stag hunt game with two Pareto-ranked equilibria: an inefficient "Hobbesian" equilibrium where all choose low effort and vote to equalize incomes and a Pareto superior equilibrium where all choose high effort and vote against equalization. However, in the finitely repeated game, which is the focus of this study, there exists an even better, social insurance equilibrium, which the authors label a Rousseau-type "social contract". In this sequential equilibrium, everyone chooses high effort but votes for equalization -i.e., they recognize that some (1/3 on average) choosing high effort earn low income due simply to back luck. This equilibrium is sustainable until a certain number of periods from the finite end (when there is a switchover to the outcome where all supply high effort but vote against equalization) via the threat to revert to the "Hobbesian" equilibrium of low effort and redistribution.

The main finding from several sessions involving 50 repetitions of the two-round, majority rule game is that the social contract equilibrium is not observed. With experience, most groups of subjects move closer to or achieve the inefficient Hobbesian equilibrium. When a majority of subjects is poor (which occurs 75% of the time) redistribution got a majority of votes 90% of the time, while when a majority of subjects was rich, redistribution succeeded only 15% of the time. Similar results are observed in the other three treatments - unanimous voting, voting restricted to those choosing high effort, and random exogenous effort with majority voting. These results suggest that social insurance contracts are unlikely to emerge on their own. However, the fact that redistributive welfare policies are observed in nature suggests that some critical element is missing from this experimental design. Some possibilities to consider are 1) whether longer-term, binding redistributive policies –in effect for multiple periods – might aid in the formation of social insurance policies or 2) whether political institutions, e.g., parliamentary or proportional representation systems might play some role in the implementation and sustenance of social insurance policies.

6 Conclusions

Certainly the most important development in macroeconomics over the past several decades has been the widespread adoption of fully rational, micro-founded, calibrated, dynamic stochastic general equilibrium *models as laboratories* for evaluation of macroeconomic theories and policies. In this chapter I have summarized the small but growing research on an alternative methodology, which can be characterized as the use of experimental *laboratories as laboratories* for evaluation of macroeconomic theories and policies.

As we have seen, contrary to the claim of Sims (1996), "crucial data" in support of macroeconomic models and theories –especially, (though not exclusively) those that are micro-founded– can be gathered in the laboratory. Such experimental tests can complement empirical analyses using field data, as in analysis of intertemporal consumption/savings decisions, rational expectations, efficiency wages or Ricardian equivalence. On the other hand, there are many macroeconomic theories, for instance on the origins of money, sunspots, speculative attacks and bank runs for which the data critical to an assessment of the theory are not available in the field. In the laboratory we can manufacture such data to meet the precise specifications of the theory being tested. In macroeconomic systems such data include not only individual choices over time, but also frequently involve individual expectations of future variables - data which are not readily available in the field.

Indeed, one innovation of macroeconomic experiments is the division of experimental designs into two basic types. In "learning-to-optimize" design, one observes whether individuals can learn over time to maximize some well-defined objective function as in most microeconomic laboratory experiments. However, many macroeconomic experiments make use of a less conventional "learning-to-forecast" design in which subjects' expectations of future variables are elicited and given these expectations, their optimization problem is solved for them by the experimenter (computer program) -they are then rewarded solely on the basis of expectations accuracy. Macroeconomic experiments have yielded other innovations, including the implementation of overlapping generations and search-theoretic environments in the laboratory, the used of indefinite repetition to implement discounting and the stationarity associated with infinite horizons and a methodology for assessing whether laboratory time series data are converging toward predicted equilibrium levels (as in equation (1).

Much further experimental research on macroeconomic topics remains to be done. Throughout this survey I have suggested a number of extensions to existing experimental studies that I believe would make for useful experiments. However, there are a number of macroeconomic topic areas for which there are no existing experimental studies and are therefore real targets of opportunity.[39] In this category I would place analysis of 1) sticky price mechanisms such as staggered wage and price setting, 2) habit formation, relative concerns and the durability of expenditures in intertemporal consumption decisions, 3) the search and matching approach to understanding unemployment, job creation and destruction (as developed by Mortensen and Pissarides (1994)), 4) Tobin's q-theory of investment determination and the observed lumpiness in aggregate investment dynamics, 5) various theories of the term structure of interest rates, 6), the irrelevance of financial structure (stock or

[39] If I had any sense, I would keep this list of topics under my own hat, though most seem (to me) to be fairly obvious candidates for experimental analysis.

bond financing) as in the Modigliani-Miller theorem, 7) the role of credit market imperfections in business cycle fluctuations, 8) policies that have been proposed to stabilize balance of payments crises in developing countries, 9) some of the explanations for cross-country differences in economic growth including legal-institutions and human capital accumulation, and 10) the existence of political business cycles.

The field of macroeconomics is among the final frontiers in the continuing transformation of economics into an experimental science. As this survey illustrates, that frontier is beginning to be populated, but only time will tell whether mainstream macroeconomists join their microeconomic brethren in accepting the relevance of laboratory methods. If past history is any guide, e.g., the rational expectations/microfoundations revolution of the 1970s and 1980s, another revolution in macroeconomic methodology may well be at hand.

References

Akerlof, G.A. (1982), "Labor Contracts as Partial Gift Exchange," *Quarterly Journal of Economics* 97, 543–69.

Akerlof, G.A. (2002), "Behavioral Macroeconomics and Macroeconomic Behavior," *American Economic Review* 92, 411-433.

Akerlof, G.A. (2007), "The Missing Motivation in Macroeconomics," *American Economic Review* 97, 5-36.

Akerlof, G.A., and J. Yellen (1986), *Efficiency Wage Models of the Labor Market*, (Cambridge: Cambridge University Press).

Adam, K. (2007), "Experimental Evidence on the Persistence of Output and Inflation," *Economic Journal* 117, 603–636.

Allen, F., and D. Gale (2000), "Financial Contagion," *Journal of Political Economy* 108, 1-33.

Anderson, S., G.W. Harrison, M.I. Lau and E.E. Rutström (2007), "Eliciting Risk and Time Preferences," working paper, University of Central Florida.

Arifovic, J. (1996), "The Behavior of the Exchange Rate in the Genetic Algorithm and Experimental Economies," *Journal of Political Economy* 104, 510-41.

Arifovic, J. and T.J. Sargent (2003), "Laboratory Experiments with an Expectational Phillips Curve," in D.E. Altig and B.D. Smith, eds., *Evolution and Procedures in Central Banking*, (Cambridge: Cambridge University Press), 23-55.

Azariadis, C. (1981), "Self-Fulfilling Prophecies," *Journal of Economic Theory* 25, 380-96.

Azariadis, C. and A. Drazen (1990), "Threshold Externalities in Economic Development," *The Quarterly Journal of Economics* 105, 501-26.

Ballinger, T.P., M.G. Palumbo and N.T. Wilcox (2003), "Precautionary Savings and Social Learning Across Generations: An Experiment," *Economic Journal* 113, 920-947.

Barro, R.J. (1974), "Are Government Bonds Net Wealth?," *Journal of Political Economy* 82, 1095-1117.

Barro, R.J., and D.B. Gordon (1983), "A Positive Theory of Monetary Policy in a Natural Rate Model," *Journal of Political Economy* 91, 589-610.

Battalio, R.C., L. Green and J.H. Kagel (1981), "Income-Leisure Tradeoffs of Animal Workers," *American Economic Review* 71, 621-32.

Benhabib, J., A. Bisin and A. Schotter (2006), "Present-Bias, Quasi-Hyperbolic Discounting and Fixed Costs," working paper, New York University.

Bernasconi, M. and O. Kirchkamp (2000), "Why Do Monetary Policies Matter? An Experimental Study of Saving and Inflation in an Overlapping Generations Model," *Journal of Monetary Economics* 46, 315-343.

Bernasconi, M., O. Kirchkamp and P. Paruolo (2006), "Do Fiscal Variables Affect Fiscal Expectations? Experiments with Real World and Lab Data," Universität Mannheim SPF 504 Discussion Paper No. 04-26.

Bernheim, B.D. (1997), "Ricardian Equivalence: An Evaluation of Theory and Evidence," in S. Fischer, ed., *1997 NBER Macroeconomics Annual* (Cambridge: MIT Press), 263-304.

Bewley, T.F. (1999), *Why Wages Don't Fall During a Recession*, (Harvard: Harvard University Press).

Blinder, A.S. and J. Morgan (2005), "Are Two Heads Better than One? Monetary Policy by Committee," *Journal of Money, Credit, and Banking* 37, 789-811.

Blinder, A.S. and J. Morgan (2007), "Leadership in Groups: A Monetary Policy Experiment," NBER Working Paper No. 13391.

Blume A. and A. Ortmann (2007), "The Effects of Costless Pre-play Communication: Experimental Evidence from Games with Pareto-ranked Equilibria," *Journal of Economic Theory* 132, 274-290.

Braunstein, Y.M. and A. Schotter (1981), "Economic Search: An Experimental Study," *Economic Inquiry* 19, 1-25.

Braunstein, Y.M. and A. Schotter (1982), "Labor Market Search: An Experimental Study," *Economic Inquiry* 20, 133-44.

Brown, P.M. (1996), "Experimental Evidence on Money as a Medium of Exchange, " *Journal of Economic Dynamics and Control* 20, 583-600.

Bryant, J. (1983), "A Simple Rational Expectations Keynes-Type Model," *Quarterly Journal of Economics* 98, 525-28.

Cabrales, A., R. Nagel and R. Armenter (2007), "Equilibrium Selection Through Incomplete Information in Coordination Games: An Experimental Study," *Experimental Economics* 10, 221-234.

Cabrales, A., R. Nagel and J.V. Rodriguez–Mora (2006), "It's Hobbes, Not Rousseau: An Experiment on Social Insurance," working paper, Universitat Pompeu Fabra.

Cadsby, C.B. and M. Frank (1991), "Experimental Tests of Ricardian Equivalence," *Economic Inquiry* 29, 645-664.

Camera, G., C.N. Noussair and S. Tucker (2003), "Rate-of-Return Dominance and Efficiency in an Experimental Economy," *Economic Theory* 22, 629-660.

Camerer, C.F. (1995), "Individual Decision Making," in J.H. Kagel and A.E. Roth, eds., *The Handbook of Experimental Economics*, (Princeton: Princeton University Press), 588–703.

Camerer, C.F. (2003), *Behavioral Game Theory*, (Princeton: Princeton University Press).

Capra, C.M., T. Tanaka, C.F. Camerer, L. Munyan, V. Sovero, L. Wang and C.N. Noussair (2005), "The Impact of Simple Institutions in Experimental Economies with Poverty Traps," working paper, CalTech.

Carbone, E. and J.D. Hey (2004), "The Effect of Unemployment on Consumption: An Experimental Analysis," *Economic Journal* 114, 660-683.

Carlson, J.A. (1967), "The Stability of an Experimental Market with a Supply-Response Lag," *Southern Economic Journal* 33, 305-321.

Carlsson, H. and E. van Damme (1993), "Global Games and Equilibrium Selection," *Econometrica* 61, 989-1018.

Cass, D. (1965), "Optimum Growth in an Aggregative Model of Capital Accumulation" *Review of Economic Studies* 32, 233-240.

Cass, D. and K. Shell (1983), "Do Sunspots Matter?," *Journal of Political Economy* 91, 193-227.

Castranova, E. (2006), "On the Research Value of Large Games: Natural Experiments in Norrath and Camelot," *Games and Culture* 1, 163-186.

Cooper, R. (1999), *Coordination Games*, (Cambridge: Cambridge University Press).

Cooper, R., D. De Jong, R. Forsythe and T. Ross (1992), "Communication in Coordination Games," *Quarterly Journal of Economics* 107, 739–771.

Coller, M., G.W. Harrison and E.E. Rutström (2005), "Are Discount Rates Constant? Reconciling Theory with Observation," working paper, Universities of South Carolina and Central Florida.

Corbae, D. and J. Duffy (2008), "Experiments with Network Formation," forthcoming in *Games and Economic Behavior*.

Cornand, C. (2006), "Speculative Attacks and Informational Structure: an Experimental Study," *Review of International Economics*, 14, 797-817.

Cox, J.C. and R.L. Oaxaca (1989), "Laboratory Experiments with a Finite-Horizon Job-Search Model," *Journal of Risk and Uncertainty* 2, 301-29.

Cox, J.C. and R.L. Oaxaca (1992), "Direct Tests of the Reservation Wage Property," *Economic Journal* 102, 1423-32.

Deck, C.A. (2004), "Avoiding Hyperinflation: Evidence from a Laboratory Economy," *Journal of Macroeconomics* 26, 147-170.

Deck, C.A., K.A. McCabe and D.P. Porter (2006), "Why Stable Fiat Money Hyperinflates: Results from an Experimental Economy," *Journal of Economic Behavior and Organization* 61, 471-486.

Devetag, G. and A. Ortmann (2007), "When and Why? A Critical Survey on Coordination Failure in the Laboratory," *Experimental Economics* 10, 331-344.

Diamond, D.W. and P. Dybvig (1983), "Bank Runs, Deposit Insurance and Liquidity," *Journal of Political Economy* 91, 401–419.

Dickinson, D.L. (1999), "An Experimental Examination of Labor Supply and Work Intensities." *Journal of Labor Economics* 17, 638-670.

Duffy, J. (1998), "Monetary Theory in the Laboratory," *Federal Reserve Bank of St. Louis Economic Review*, 80 (September/October), 9-26.

Duffy, J. (2001), "Learning to Speculate: Experiments with Artificial and Real Agents," *Journal of Economic Dynamics and Control* 25, 295-319.

Duffy, J. (2008), "Experimental Macroeconomics," in: L. Blume and S. Durlauf eds., *The New Palgrave Dictionary of Economics*, 2nd Ed., (London: Palgrave Macmillan).

Duffy, J. and E. O'N. Fisher (2005), "Sunspots in the Laboratory," *American Economic Review* 95, 510–529.

Duffy, J. and J. Ochs (1999), "Emergence of Money as a Medium of Exchange: An Experimental Study," *American Economic Review* 89, 847-77.

Duffy, J. and J. Ochs (2002), "Intrinsically Worthless Objects as Media of Exchange: Experimental Evidence," *International Economic Review* 43, 637-73.

Duffy, J. and J. Ochs (2008), "Equilibrium Selection in Entry Games: An Experimental Study," working paper, University of Pittsburgh.

Duffy, J. and R. Nagel (1997), "On the Robustness of Behavior in Experimental 'Beauty Contest' Games," *Economic Journal* 107, 1684-1700.

Dwyer, Jr., G.P., A.W. Williams, R.C. Battalio and T.I. Mason (1993), "Tests of Rational Expectations in a Stark Setting," *Economic Journal* 103, 586-601.

Ehrblatt, W.Z., K. Hyndman, E.Y. Özbay and A. Schotter (2007), "Convergence: An Experimental Study of Teaching and Learning in Repeated Games," working paper, New York University.

Engle-Warnick, J. and N. Turdaliev (2006), "An Experimental Test of Taylor-Type Rules with Inexperienced Central Bankers," working paper 2006s-05, Centre internuniversitaire de recherche en analyse des organizations, Montreal.

Evans, G.W. and S. Honkaphoja (2001), *Learning and Expectations in Macroeconomics*, (Princeton: Princeton University Press).

Ezekiel, M. (1938), "The Cobweb Theorem," *The Quarterly Journal of Economics* 52, 255-280.

Falk, A. and E. Fehr (2003), "Why Labor Market Experiments?," *Labor Economics* 10, 399-406.

Falk and Gächter (2008), "Experimental Labor Economics," in L. Blume and S. Durlauf, eds., *The New Palgrave Dictionary of Economics*, 2nd Ed., (London: Palgrave Macmillan).

Fehr, E. and A. Falk (1999), "Wage Rigidity in a Competitive Incomplete Contract Market," *Journal of Political Economy* 107, 106-134.

Fehr, E., S. Gächter (2002), "Do Incentive Contracts Undermine Voluntary Cooperation?," University of Zurich, Institute for Empirical Research in Economics Working Paper No. 34.

Fehr, E., S. Gächter, and G. Kirchsteiger (1997), "Reciprocity as a Contract Enforcement Device – Experimental Evidence," *Econometrica* 65, 833-860.

Fehr, E., E. Kirchler, A. Weichbold, and S. Gächter (1998), "When Social Norms Overpower Competition – Gift Exchange in Experimental Labor Markets," *Journal of Labor Economics* 16, 324-351.

Fehr, E., G. Kirchsteiger and A. Riedl (1993), "Does Fairness Prevent Market Clearing? An Experimental Investigation," *Quarterly Journal of Economics* 108, 437-460.

Fehr, E., G. Kirchsteiger and A. Riedl (1996), "Involuntary Unemployment and Noncompensating Wage Differentials in an Experimental Labour Market," *Economic Journal* 106, 106-21.

Fehr, E., G. Kirchsteiger and A. Riedl (1998), "Gift Exchange and Reciprocity in Competitive Experimental Markets," *European Economic Review* 42, 1-34.

Fehr, E. and J-F. Tyran (2001), "Does Money Illusion Matter?," *American Economic Review* 91, 1239-62.

Fehr, E. and J-F. Tyran (2007), "Money Illusion and Coordination Failure," *Games and Economic Behavior* 58, 246-268.

Fehr, E. and J-F. Tyran (2008), "Limited Rationality and Strategic Interaction: The Impact of the Strategic Environment on Nominal Inertia," *Econometrica* 76, 353-394.

Fisher, E.O'N. (2001), "Purchasing Power Parity and Interest Parity in the Laboratory," *Australian Economic Papers* 40, 586-602.

Fisher, E.O'N. (2005), "Exploring Elements of Exchange Rate Theory in a Controlled Environment," working paper, Ohio State University.

Fisher, F.M. (1987), "Aggregation Problems," in: Eatwell et al., eds., *The New Palgrave Dictionary of Economics*, (London: Macmillan), 53–55.

Flavin, M.A. (1981), "The Adjustment of Consumption to Changing Expectations about Future Income," *Journal of Political Economy* 89, 974-1009.

Frankel, J.A. and K.A. Froot (1987), "Using Survey Data to Test Some Standard Propositions Regarding Exchange Rate Expectations," *American Economic Review* 77, 133-153.

Frederick, S., G. Loewenstein and T. O'Donoghue (2002), "Time Discounting and Time Preference: A Critical Review," *Journal of Economic Literature* 40, 351-401.

Gächter, S. and E. Fehr (2002), "Fairness in the Labour Market – A Survey of Experimental Results," in: F. Bolle and M. Lehmann-Waffenschmidt, eds., *Surveys in Experimental Economics. Bargaining, Cooperation and Election Stock Markets*, (New York: Physica Verlag), 95-132.

Garratt, R. and T. Keister (2005), "Bank Runs: An Experimental Study," working Paper, University of California, Santa Barbara.

Goodfellow, J. and C.R. Plott (1990), "An Experimental Examination of the Simultaneous Determination of Input Prices and Output Prices," *Southern Economic Journal* 56, 969-83.

Harrison, G.W. and P. Morgan (1990), "Search Intensity in Experiments," *Economic Journal* 100, 478-486.

Hayashi, F. (1982), "The Permanent Income Hypothesis: Estimation and Testing by Instrumental Variables," *Journal of Political Economy* 90, 971-987.

Heinemann, F., R. Nagel and P. Ockenfels (2004), "The Theory of Global Games on Test: Experimental Analysis of Coordination Games with Public and Private Information," *Econometrica* 72, 1583–1599.

Hens, T., K.R. Schenk-Hoppe and B. Vogt (2007), "The Great Capitol Hill Baby Sitting Co-op: Anecdote or Evidence for the Optimum Quantity of Money?," *Journal of Money, Credit and Banking* 39, 1305-1333.

Hey, J.D. (1994), "Expectations Formation: Rational or Adaptive or ...?," *Journal of Economic Behavior and Organization* 25, 329-344.

Hey, J.D. (1987), "Still Searching," *Journal of Economic Behavior and Organization* 8, 137-44.

Hey, J.D. and V. Dardanoni (1988), "Optimal Consumption Under Uncertainty: An Experimental Investigation," *Economic Journal* 98, 105-116.

Ho, T., C. Camerer and K. Weigelt (1998), "Iterated Dominance and Iterated Best-Response in Experimental p-Beauty Contests," *American Economic Review* 88, 947-69.

Holt, C.A., and S.M Laury, (2002), "Risk Aversion and Incentive Effects," *American Economic Review* 92, 1644-1655.

Hommes, C.H., J. Sonnemans, J. Tuinstra and H. van de Velden (2007), "Learning in Cobweb Experiments," *Macroeconomic Dynamics* 11 (Supplement 1), 8–33.

Hommes, C.H., J. Sonnemans, J. Tuinstra and H. van de Velden (2005), "Coordination of Expectations in Asset Pricing Experiments" *Review of Financial Studies* 18, 955-980.

Kareken, J.H. and N. Wallace (1981) "On the Indeterminacy of Equilibrium Exchange Rates," *Quarterly Journal of Economics* 96, 207-222.

Kelley, H., and D. Friedman (2002), "Learning to Forecast Price," *Economic Inquiry* 40, 556–573.

Keynes, J.M. (1936), *The General Theory of Employment, Interest, and Money*, (New York: Harcourt, Brace and Co.).

Kirman, A.P. (1992), "Whom or What Does the Representative Individual Represent?," *Journal of Economic Perspectives* 6, 117-136.

Kiyotaki, N. and R. Wright (1989), "On Money as a Medium of Exchange," *Journal of Political Economy*, 97, 927-954.

Koopmans, T.C. (1965), "On the Concept of Optimal Economic Growth," in: *The Econometric Approach to Development Planning*, (Amsterdam: North-Holland), 225-287.

Kydland, F. and E. Prescott (1977), "Rules Rather than Discretion: The Inconsistency of Optimal Plans," *Journal of Political Economy* 85, 473-490.

Kydland, F. and E. Prescott (1982), "Time to Build and Aggregate Fluctuations," *Econometrica* 50, 1345-70.

Laibson, D.I. (1997), "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics* 62, 443-478.

Ledyard, J.O. (1995), "Public Goods: A Survey of Experimental Research," in: J.H. Kagel and A.E. Roth, eds., *The Handbook of Experimental Economics*, (Princeton: Princeton University Press), 111-194.

Lei, V. and C.N. Noussair (2002), "An Experimental Test of an Optimal Growth Model," *American Economic Review* 92, 549-70.

Lei, V. and C.N. Noussair (2007), "Equilibrium Selection in an Experimental Macroeconomy," *Southern Economic Journal* 74, 448-482.

Lian, P. and C.R. Plott (1998), "General Equilibrium, Markets, Macroeconomics and Money in a Laboratory Experimental Environment," *Economic Theory* 12, 21-75.

Lim, S. Prescott, E.C. and Sunder, S. (1994), "Stationary Solution to the Overlapping Generations Model of Fiat Money: Experimental Evidence," *Empirical Economics* 19, 255-77.

Lombardelli, C., J. Proudman, and J. Talbot (2005), "Committee Versus Individuals: An Experimental Analysis of Monetary Policy Decision Making," *International Journal of Central Banking* 1, 181-203.

Lucas, R.E. Jr. (1972), "Expectations and the Neutrality of Money," *Journal of Economic Theory* 4, 103-124.

Lucas, R.E. Jr. (1986), "Adaptive Behavior and Economic Theory," *Journal of Business* 59, S401-S426.

McCabe, K.A. (1989), "Fiat Money as a Store of Value in an Experimental Market," *Journal of Economic Behavior and Organization* 12, 215-231.

Marimon, R., E. McGrattan and T.J. Sargent (1990), "Money as a Medium of Exchange in an Economy with Artificially Intelligent Agents," *Journal of Economic Dynamics and Control* 14, 329-373.

Marimon, R. and S. Sunder (1993), "Indeterminacy of Equilibria in a Hyperinflationary World: Experimental Evidence," *Econometrica* 61, 1073-1107.

Marimon, R. and S. Sunder (1994), "Expectations and Learning under Alternative Monetary Regimes: An Experimental Approach," *Economic Theory* 4, 131-62.

Marimon, R., and S. Sunder (1995), "Does a Constant Money Growth Rule Help Stabilize Inflation?," *Carnegie-Rochester Conference Series on Public Policy* 43, 111-156.

Marimon, R., S.E. Spear and S. Sunder (1993), "Expectationally Driven Market Volatility: An Experimental Study," *Journal of Economic Theory* 61, 74-103.

Morris, S. and H-S. Shin (1998), "Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks," *American Economic Review* 88, 587–597.

Morris, S. and H-S. Shin (2001), "Rethinking Multiple Equilibria in Macroeconomic Modeling," *NBER Macroeconomics Annual* 15, 139-161.

Mortensen, D. (1987), "Job Search and Labor Market Analysis," in: O. Ashenfelter and R. Layard, eds., *Handbook of Labor Economics*, (Amsterdam: North-Holland), 849-919.

Mortensen, D.T. and C.A. Pissarides (1994), "The Cyclical Behavior of Job and Worker Flows," *Review of Economic Studies* 61, 397-415.

Moulin, Herve (1986), *Game Theory for the Social Sciences* 2nd ed., (New York: New York University Press).

Muth, J.F. (1961), "Rational Expectations and the Theory of Price Movements," *Econometrica* 29, 315–335.

Nagel, R. (1995), "Unraveling in Guessing Games: An Experimental Study," *American Economic Review* 85, 1313-1326.

Noussair, C.N. and K.J. Matheny (2000), "An Experimental Study of Decisions in Dynamic Optimization Problems," *Economic Theory* 15, 389-419.

Noussair, C.N., C.R. Plott and R.G. Riezman (1995), "An Experimental Investigation of the Patterns of International Trade," *American Economic Review* 85, 462-91.

Noussair, C.N., C.R. Plott and R.G. Riezman (1997), "The Principles of Exchange Rate Determination in an International Financial Experiment," *Journal of Political Economy* 105, 822-61.

Noussair, C.N., C. Plott and R.G. Riezman (2007), "Production, Trade, Prices, Exchange Rates and Equilibration in Large Experimental Economies," *European Economic Review* 51, 49-76.

O'Donoghue, T. and M. Rabin (1999), "Doing It Now or Later," *American Economic Review* 89, 103-124.

Obstfeld, M. (1996), "Models of Currency with Self-fulfilling Features," *European Economic Review* 40, 1037-47.

Ochs, J. (1995), "Coordination Problems," in: J.H. Kagel and A.E. Roth, eds., *The Handbook of Experimental Economics*, (Princeton: Princeton University Press), 195-251.

Phelps, E.S. (1967), "Phillips Curves, Expectations of Inflation and Optimal Unemployment Over Time," *Economica* 2, 22–44.

Phillips, A.W. (1950), "Mechanical Models in Economic Dynamics," *Economica* 17, 283-305.

Ramsey, F.P. (1928), "A Mathematical Theory of Saving," *Economic Journal* 38, 543-559.

Ricciuti, R. (2004), "Bringing Macroeconomics into the Lab," International Center for Economic Research, working paper no. 26.

Ricciuti, R. and D. Di Laurea (2004), "An Experimental Analysis of Two Departures from Ricardian Equivalence," *Economics Bulletin*, 8, 1-11.

Riedl, A. and F. van Winden (2001), "Does the Wage Tax System Cause Budget Deficits? A Macro-economic Experiment," *Public Choice* 109, 371-94.

Riedl, A. and F. van Winden (2007), "An Experimental Investigation of Wage Taxation and Unemployment in Closed and Open Economies," *European Economic Review* 51, 871-900.

Roth, A.E. and M.W.K. Malouf (1979), "Game-Theoretic Models and the Role of Information in Bargaining," *Psychological Review* 86, 574-594.

Samuelson, P.A. (1958), "An Exact Consumption-Loan Model of Interest With or Without the Social Contrivance of Money," *Journal of Political Economy* 66, 467-482.

Sargent, T.J. (1983), "The Ends of Four Big Inflations," in: R.E. Hall, ed., *Inflation: Causes and Effects*, (Chicago: University of Chicago Press), 41-97.

Sargent, T.J. (1993), *Bounded Rationality in Macroeconomics*, (Oxford: Oxford University Press).

Sargent, T.J. (1999), *The Conquest of American Inflation*, (Princeton: Princeton University Press).

Schmalensee, R. (1976), "An Experimental Study of Expectation Formation," *Econometrica* 44, 17–41.

Schotter, A. and T. Yorulmazer (2003), "On the Severity of Bank Runs: An Experimental Study," working paper, New York University.

Seater, J.J. (1993), "Ricardian Equivalence," *Journal of Economic Literature* 31, 142-90.

Shafir, E., P. Diamond, and A. Tversky (1997), "Money Illusion," *Quarterly Journal of Economics* 112, 341-74.

Shapiro, C. and J.E. Stiglitz (1984), "Equilibrium Unemployment as a Worker Discipline Device," *American Economic Review* 74, 433-44.

Shell, K. (1971), "Notes on the Economics of Infinity," *Journal of Political Economy* 79, 1002-11.

Sims, C.A. (1996), "Macroeconomics and Methodology," *Journal of Economic Perspectives* 10, 105-120.

Slate, S., M. McKee, W. Beck and J. Alm (1995), "Testing Ricardian Equivalence under Uncertainty," *Public Choice* 85, 11-29.

Smith, V.L. (1962), "An Experimental Study of Competitive Market Behavior," *Journal of Political Economy* 70, 111-137.

Smith, V.L., G.L. Suchanek and A.W. Williams (1988), "Bubbles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets," *Econometrica* 56, 1119-51.

Solow, R.M. (1990), *The Labour Market as a Social Institution*, (Oxford: Blackwell).

Sunder, S. (1995), "Experimental Asset Markets: A Survey," in: J.H. Kagel and A.E. Roth, eds., *The Handbook of Experimental Economics*, (Princeton: Princeton University Press), 445-500.

Svensson, L.E.O. (1997), "Optimal Inflation Targets, 'Conservative' Central Banks, and Linear Inflation Contracts," *American Economic Review* 87, 98–114.

Sweeney, J. and R.J. Sweeney (1977), "Monetary Theory and the Great Capitol Hill Baby Sitting Co-op Crisis," *Journal of Money, Credit and Banking* 9, 86-89.

Taylor, J. (1993), "Discretion vs. Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy* 39, 195-214.

Thaler, R. (1981), "Some Empirical Evidence on Dynamic Inconsistency," *Economics Letters* 8, 201–207.

van der Heijden, E.C.M., J.H.M. Nelissen, J.J.M. Potters and H.A.A. Verbon (1998), "Transfers and the Effect of Monitoring in an Overlapping-Generations Experiment," *European Economic Review* 42, 1363-1391.

Van Huyck, J.B., R.C. Battalio, and R.O. Beil (1990), "Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure," *American Economic Review* 80, 234-48.

Van Huyck, J.B., R.C. Battalio, and R.O. Beil (1991), "Strategic Uncertainty, Equilibrium Selection, and Coordination Failure in Average Opinion Games," *Quarterly Journal of Economics* 106, 885-910.

Van Huyck, J.B., R.C. Battalio, and M.F. Walters (2001), "Is Reputation a Substitute for Commitment in the Peasant-Dictator Game?," working paper, Texas A&M University.

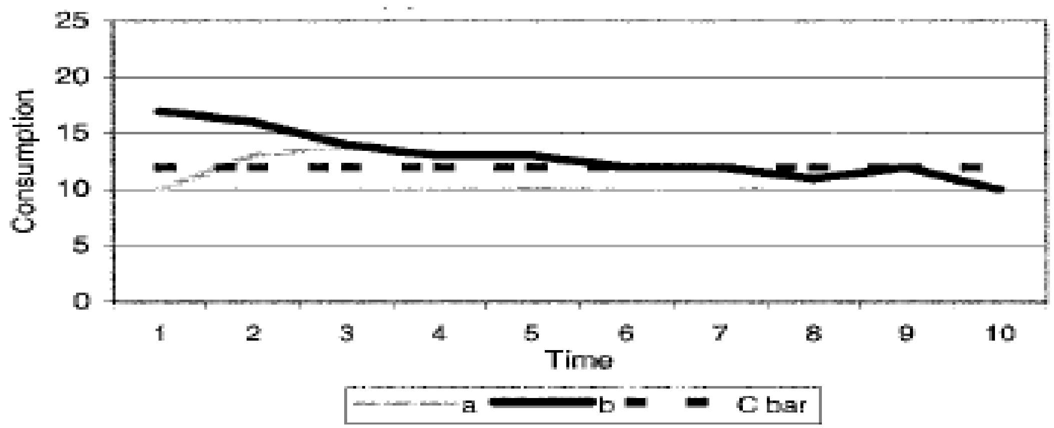
Van Huyck, J.B., R.C. Battalio, and M.F. Walters (1995), "Commitment versus Discretion in the Peasant-Dictator Game," *Games and Economic Behavior* 10, 143-71.

Van Huyck, J.B., J.P. Cook, and R.C. Battalio (1994), "Selection Dynamics, Asymptotic Stability, and Adaptive Behavior," *Journal of Political Economy* 102, 975-1005.

Williams, A.W. (1987), "The Formation of Price Forecasts in Experimental Markets," *Journal of Money, Credit and Banking* 19, 1-18.

Woodford, M. (2003), *Interest and Prices*, (Princeton: Princeton University Press).

Zeldes, S.P. (1989), "Consumption and Liquidity Constraints: An Empirical Investigation," *Journal of Political Economy* 97, 305-346.



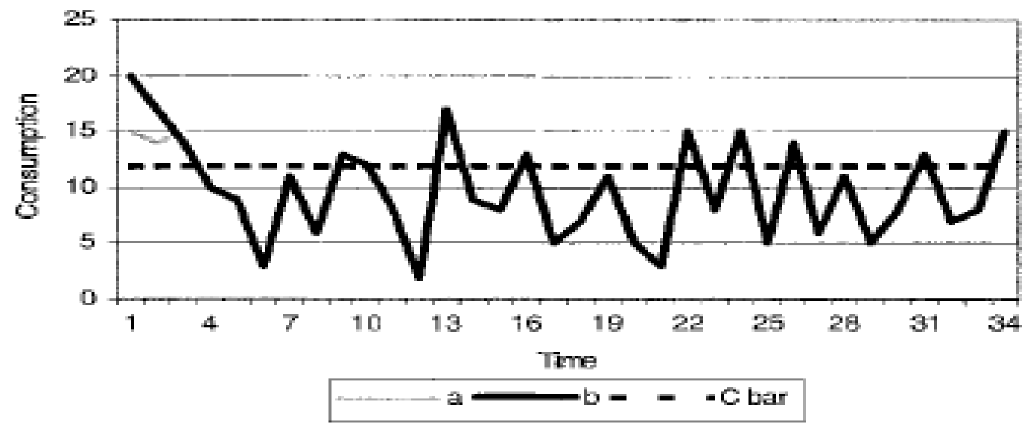


Figure 1 Consumption choices over 2 indefinite horizons (a,b) compared with optimal steady state consumption (C bar) Market treatment (top) versus Social Planner treatment (bottom) from Lei and Noussair (2002)

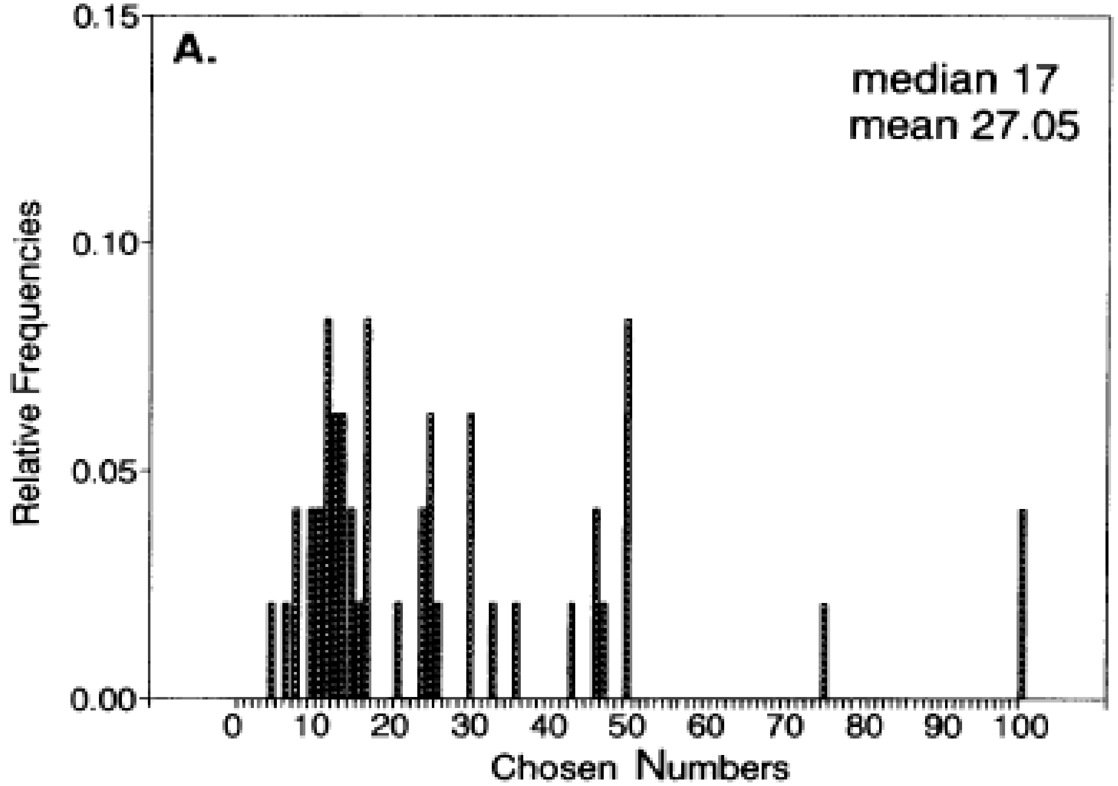


Figure 2 Relative frequencies of numbers in [0,100] chosen in Nagel's (1995) 1/2-mean game (beauty contest). Source: Nagel (1995).

Strongly Unstable
RE $p^*=5.91$, $\sigma^2=.25$

Unstable
RE $p^*=5.73$, $\sigma^2=.25$

Stable
RE $p^*=5.57$, $\sigma^2=.25$

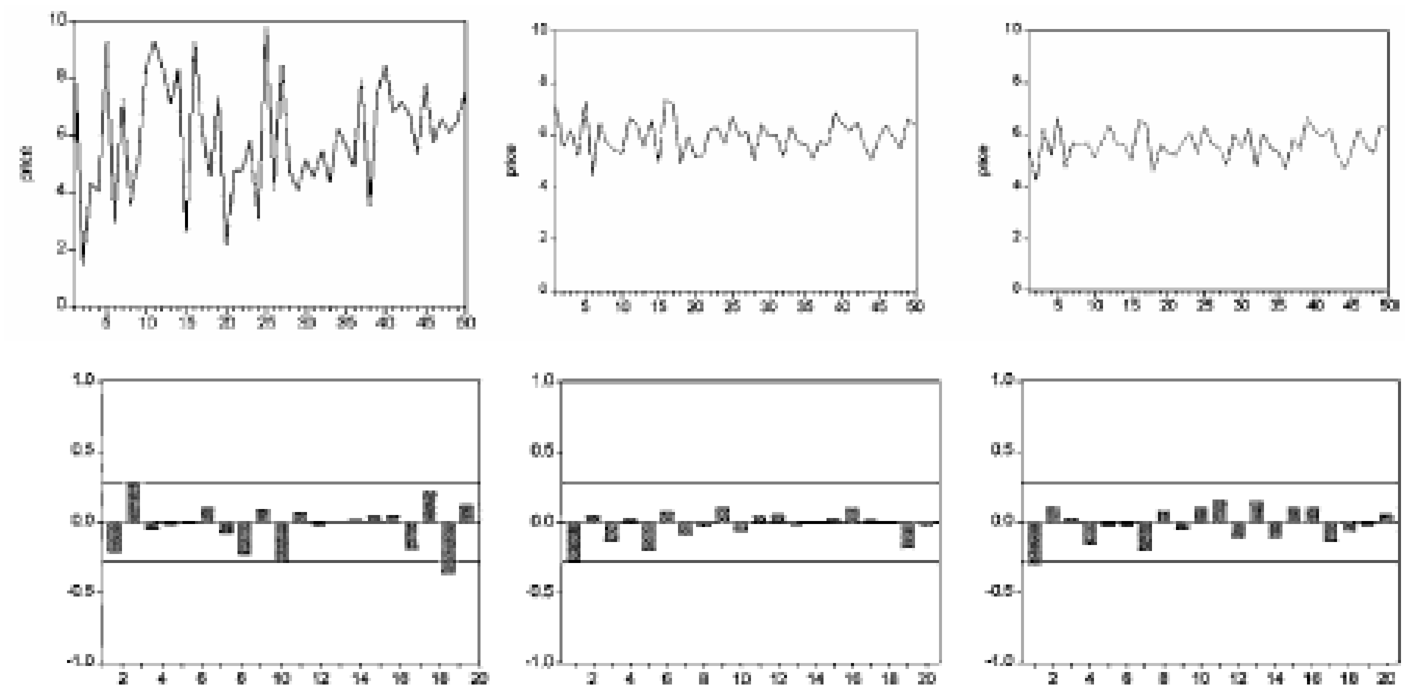


Figure 3 Actual prices (top) and autocorrelations (bottom) from three representative sessions of the three treatments of Hommes et al. (2007): strongly unstable, unstable and stable equilibrium under naïve expectations.

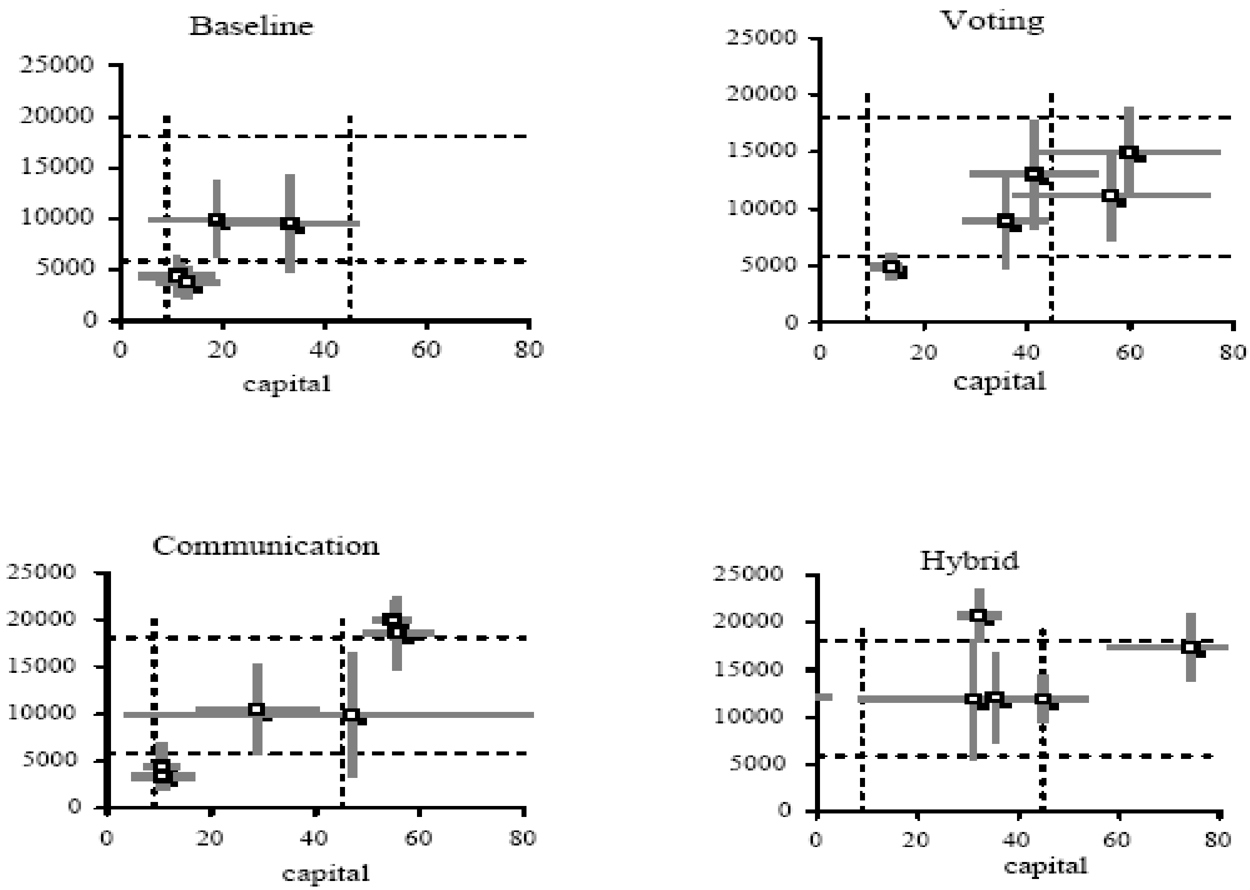


Figure 4 Asymptotic estimates of aggregate welfare (vertical axis) and capital (horizontal axis) for each session (square) of the four treatments of Capra et al. (2005). Line segments give 95% confidence regions. Poverty trap equilibrium is at the lower-left intersection of the two dashed lines, while the efficient equilibrium is at the upper-right intersection of the two dashed lines.

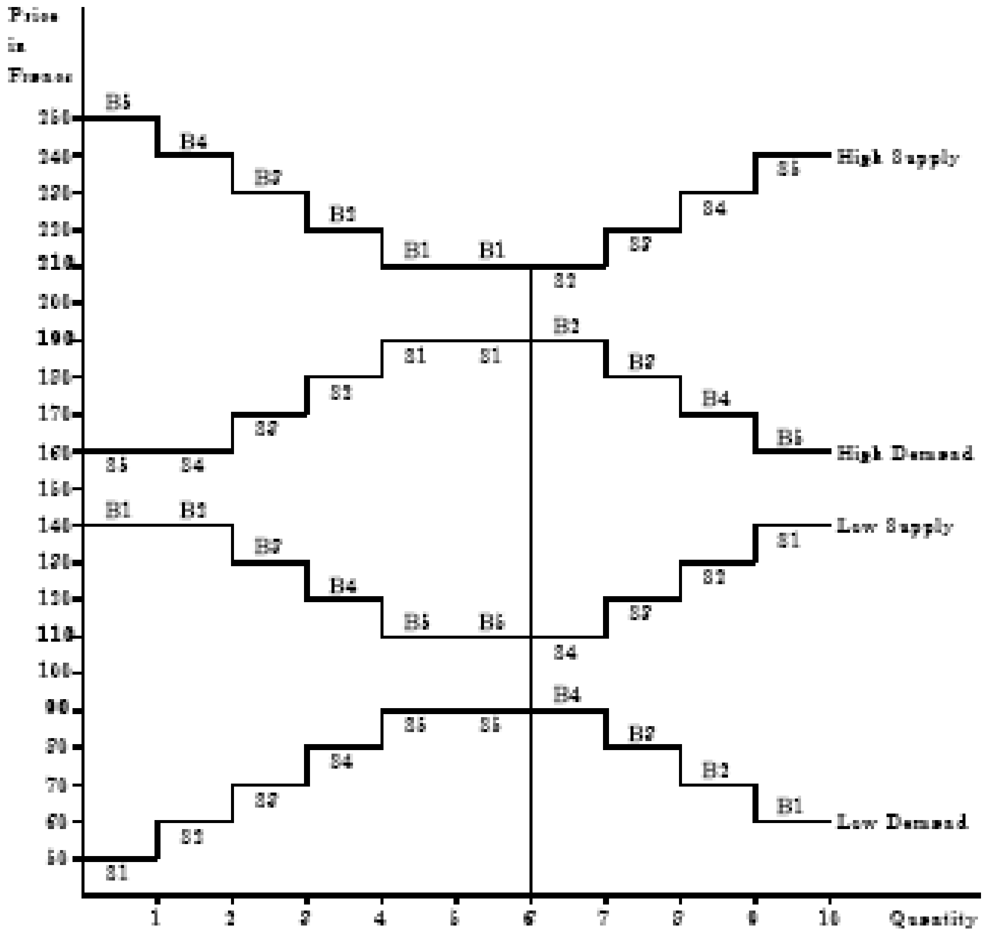


Figure 5: Induced High and Low Demand and Supply in Duffy and Fisher (2005). Buyers: B1--B5, Sellers: S1--S5. Market clearing prices with high demand and supply are [190,210] Market clearing prices with low demand and supply are [90,110]. The equilibrium quantity is always 6 units bought and sold.

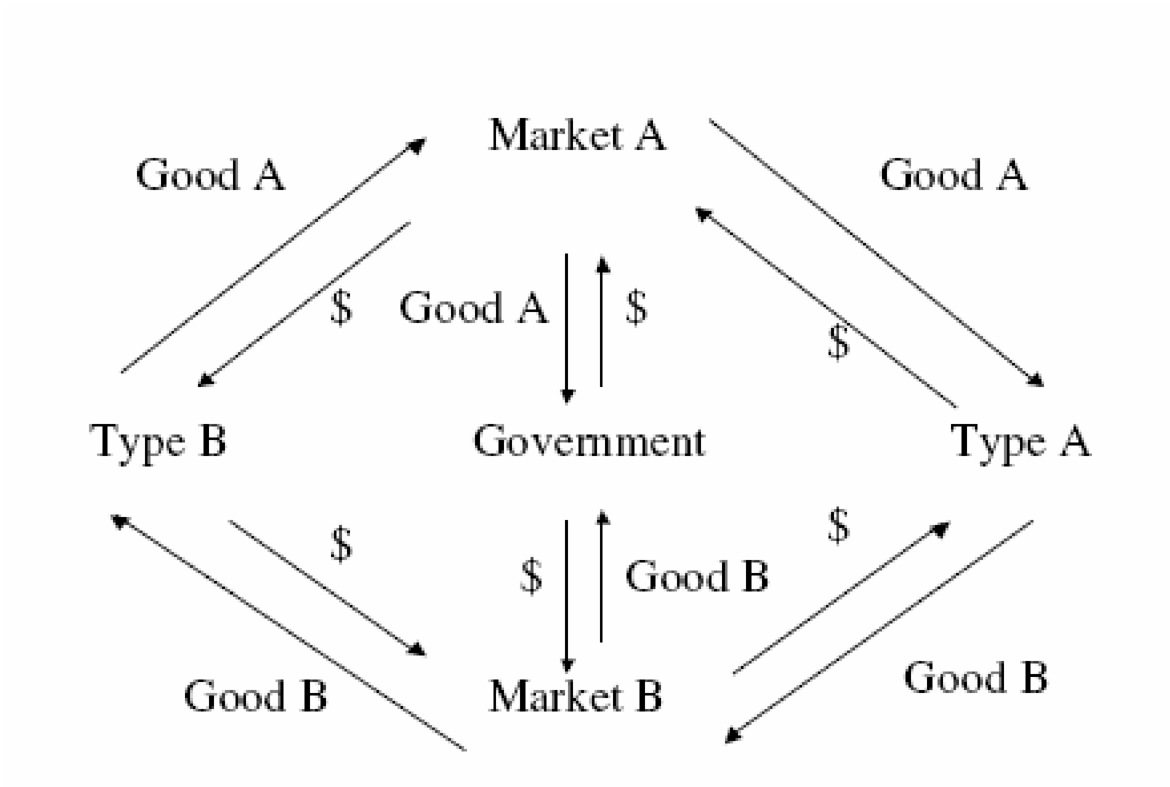


Figure 6: Experimental Design of McCabe et al. (2006)

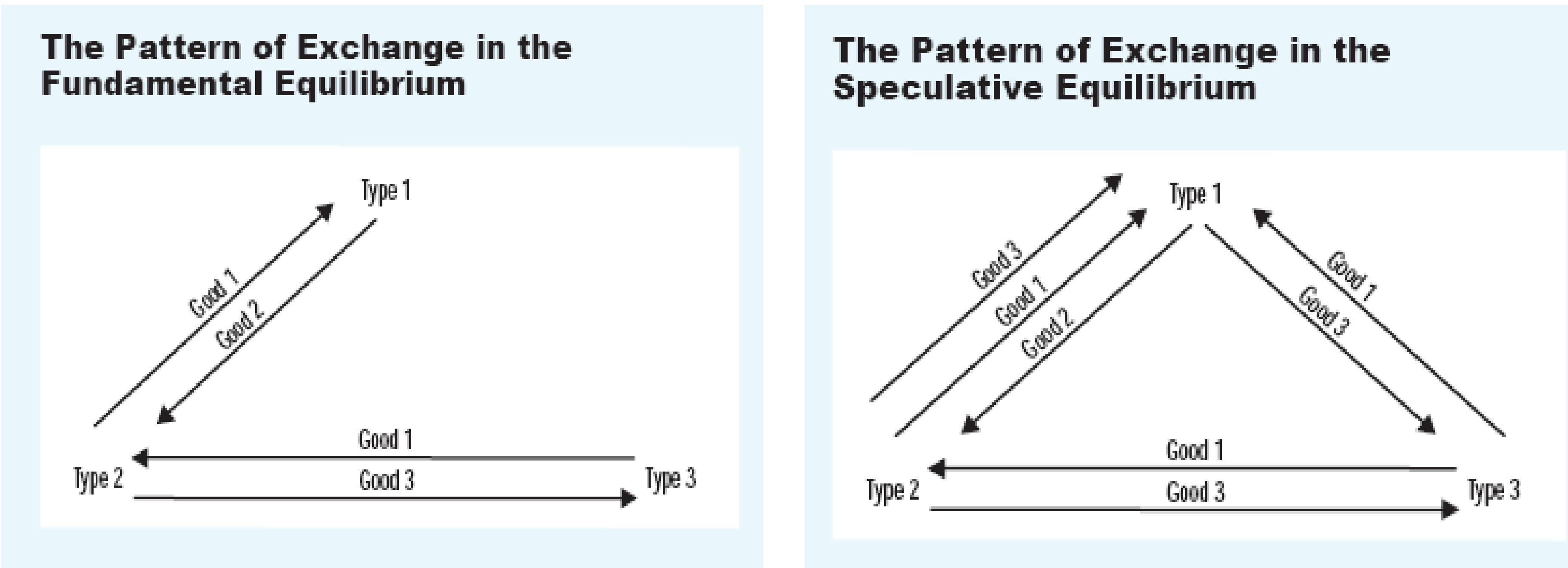


Figure 7: Predicted trading patterns in the fundamental (left) and speculative (right) equilibrium. In the fundamental equilibrium, Type 2 trades good 3 to Type 3 for the lowest storage cost good 1, and then trades good 1 to Type 1 for good 2.

In the speculative equilibrium, an additional trade is predicted: Type 1s agree to trade good 2 to Type 2 for the more costly to store good 3, and then trade good 3 to Type 3 for good 1. Goods 3 and 1 serve as media of exchange, though 3 is more costly.

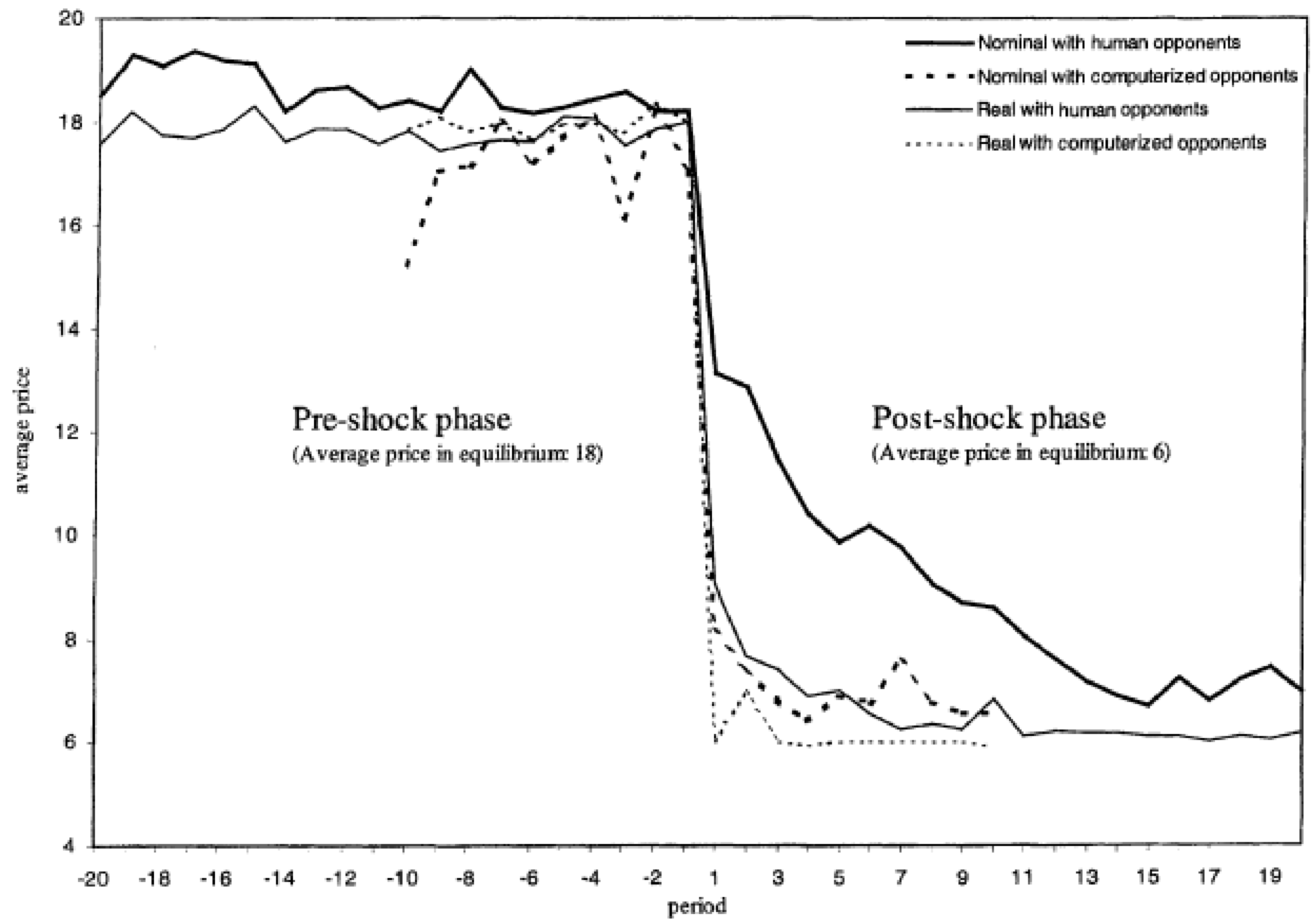


Figure 8: Path of Average Prices in the Four Treatments of Fehr and Tyran (2001). The nominal shock occurs in period 0.

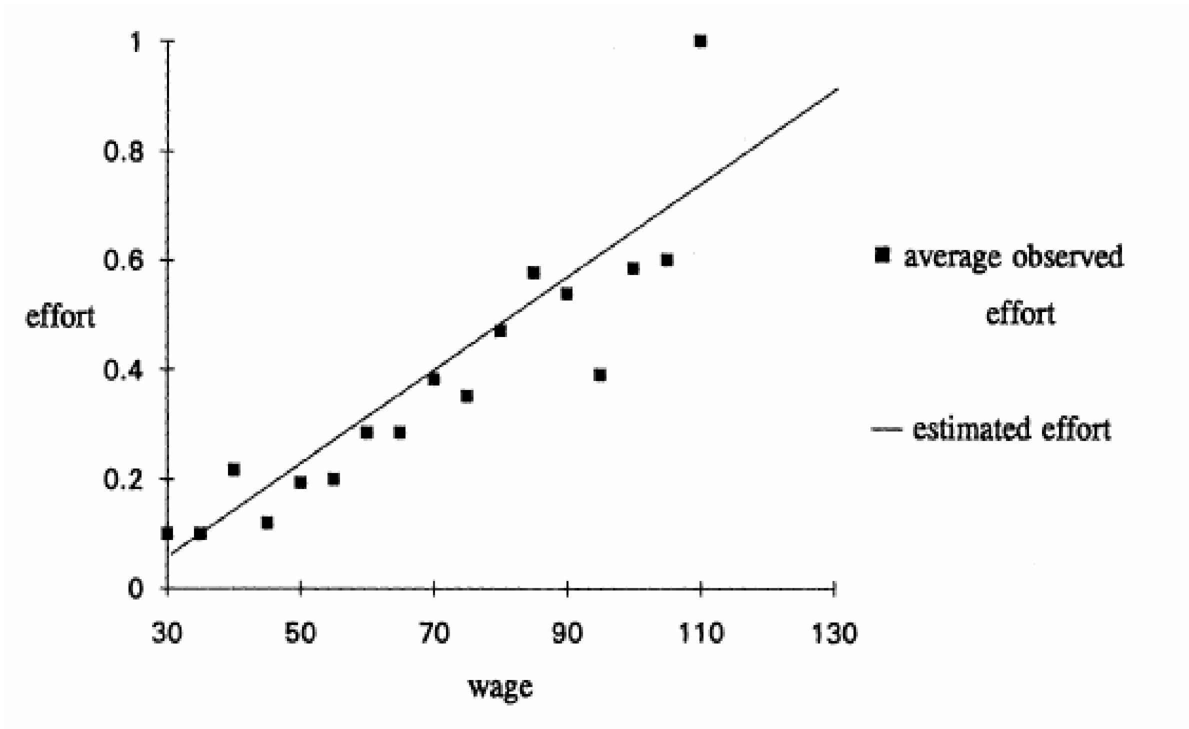


Figure 9: Average Observed Effort as a Function of Wages from Fehr et al. (1993)

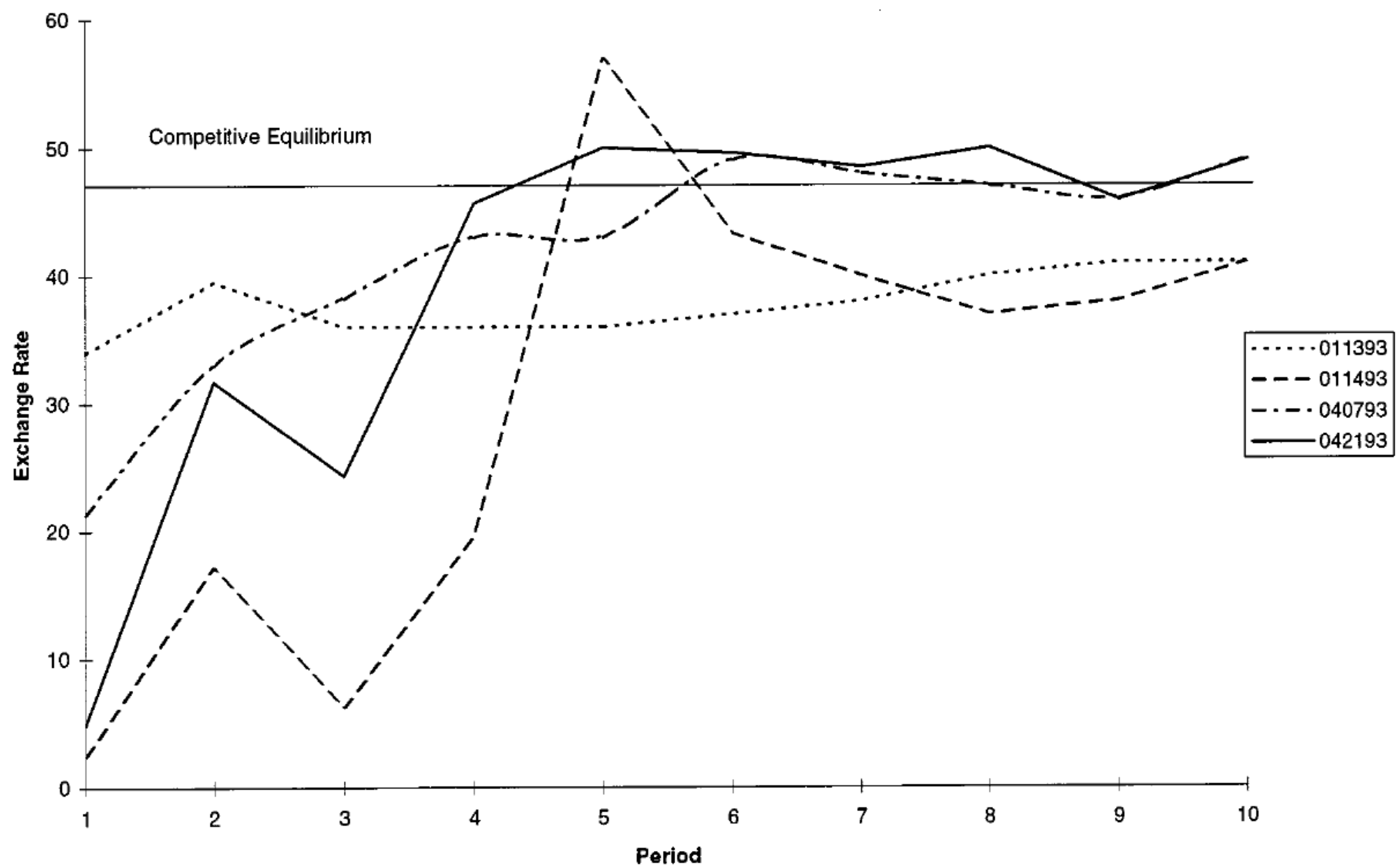


Figure 10: Mean Exchange Rate of Currency A for Currency B Over Ten Trading Periods of the Four Sessions of Noussair et al. (1997) (CE Prediction is 47).

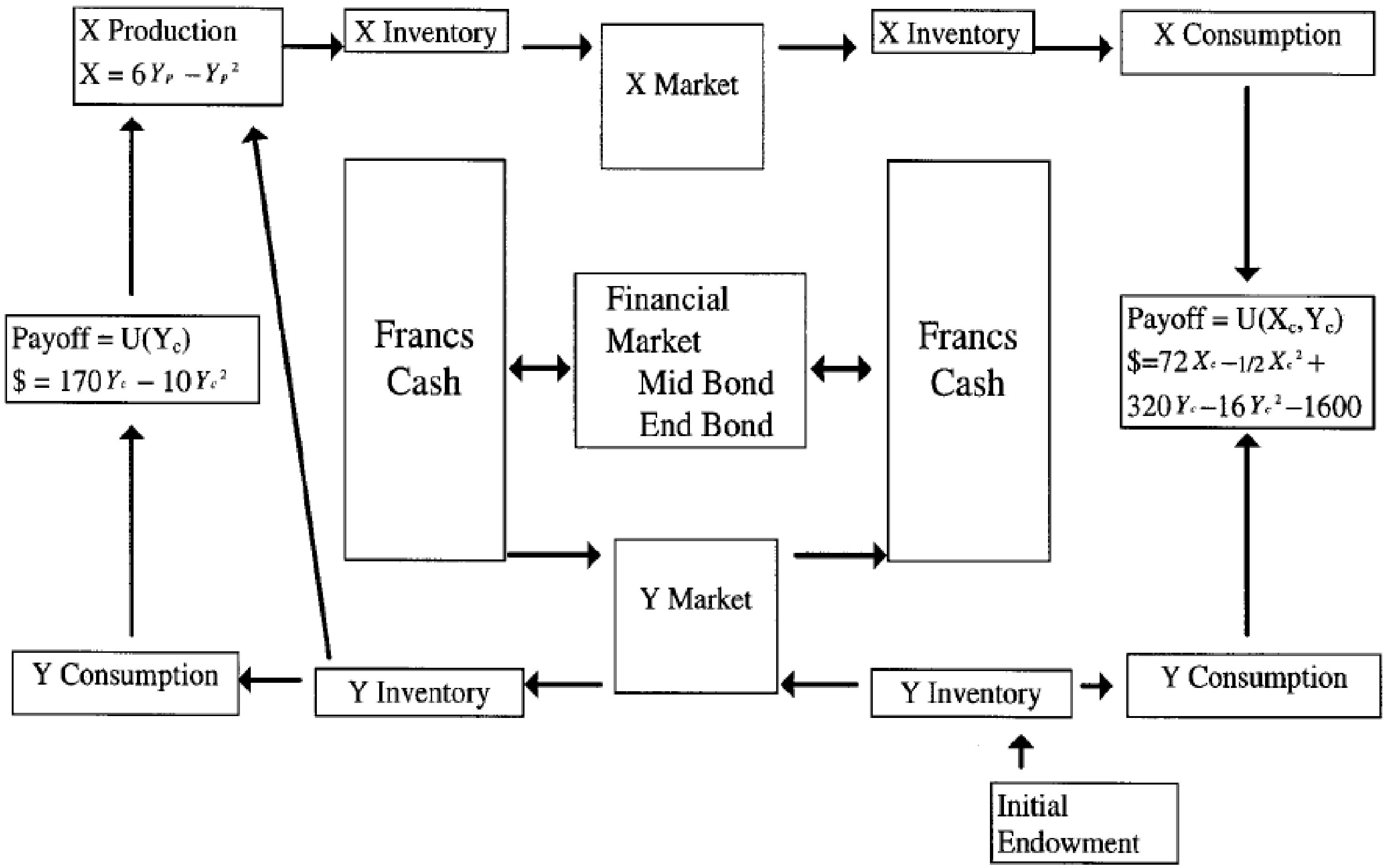


Figure 11: Circular Flow Model Illustrating the Experimental Environment of Lian and Plott (1998). (Source: Lian and Plott 1998-Figure 1)

Number of participants of each of the twelve types and their functions

Type	Country	Role	Parameter values	# In exp 1	# Iin exp 2	# In exp 3
1	A	Producer of X Consumer of Y	$f_x^A(k,l) = 4l^{0.25}k^{0.25}$ $U(y) = 1650y - 100y^2$	5	3	5
2	A	Producer of Y Consumer of X and Z	$f_y^A(k,l) = 2l^{0.25}k^{0.25}$ $U(x,z) = 700x - 100x^2 + 1900z - 100z^2$	5	3	5
3	A	Producer of Z Consumer of X and Y	$f_z^A(k,l) = 2l^{0.25}k^{0.25}$ $U(x,y) = 700x - 100x^2 + 1650y - 100y^2$	5	3	4
4	A	Supplier of L and K Consumer of Z	$C(l,k) = 26l + 2l^2 + 10k + 5k^2$ $U(z) = 1900z - 100z^2$	5	3	5
5	B	Producer of X Consumer of Y and Z	$f_x^B(k,l) = 2l^{0.25}k^{0.25}$ $U(y,z) = 3900y - 400y^2 + 5600z - 400z^2$	5	4	5
6	B	Producer of Y Consumer of X	$f_y^B(k,l) = 4l^{0.25}k^{0.25}$ $U(x) = 3800x - 400x^2$	5	4	5
7	B	Producer of Z Consumer of X and Y	$f_Z^B(k,l) = 2l^{0.25}k^{0.25}$ $U(x,y) = 3800x - 400x^2 + 3900y - 400y^2$	5	4	4
8	B	Supplier of L and K Consumer of Z	$C(l,k) = 48l + 15l^2 + 55k + 7.5k^2$ $U(z) = 5600z - 400z^2$	5	4	5
9	C	Producer of X Consumer of Y and Z	$f_x^C(k,l) = 2l^{0.25}k^{0.25}$ $U(y,z) = 13\,500y - 1000y^2 + 16\,000z - 1000z^2$	5	4	5
10	C	Producer of Y Consumer of X and Z	$f_y^C(k,l) = 2l^{0.25}k^{0.25}$ $U(x,z) = 12\,000x - 1000x^2 + 16\,000z - 1000z^2$	5	3	5
11	C	Producer of Z	$f_z^C(k,l) = 4l^{0.25}k^{0.25}$	5	2	4
12	C	Supplier of L and K Consumer of X and Y	$C(l,k) = 300l + 50l^2 + 220k + 20k^2$ $U(x,y) = 12\,000x - 1000x^2 + 13\,500y - 1000y^2$	5	3	5

Table 6: Twelve Subject Types, Preferences, Cost and Production Functions and Numbers of Each Type in Three Sessions of Noussair et al. 2007.

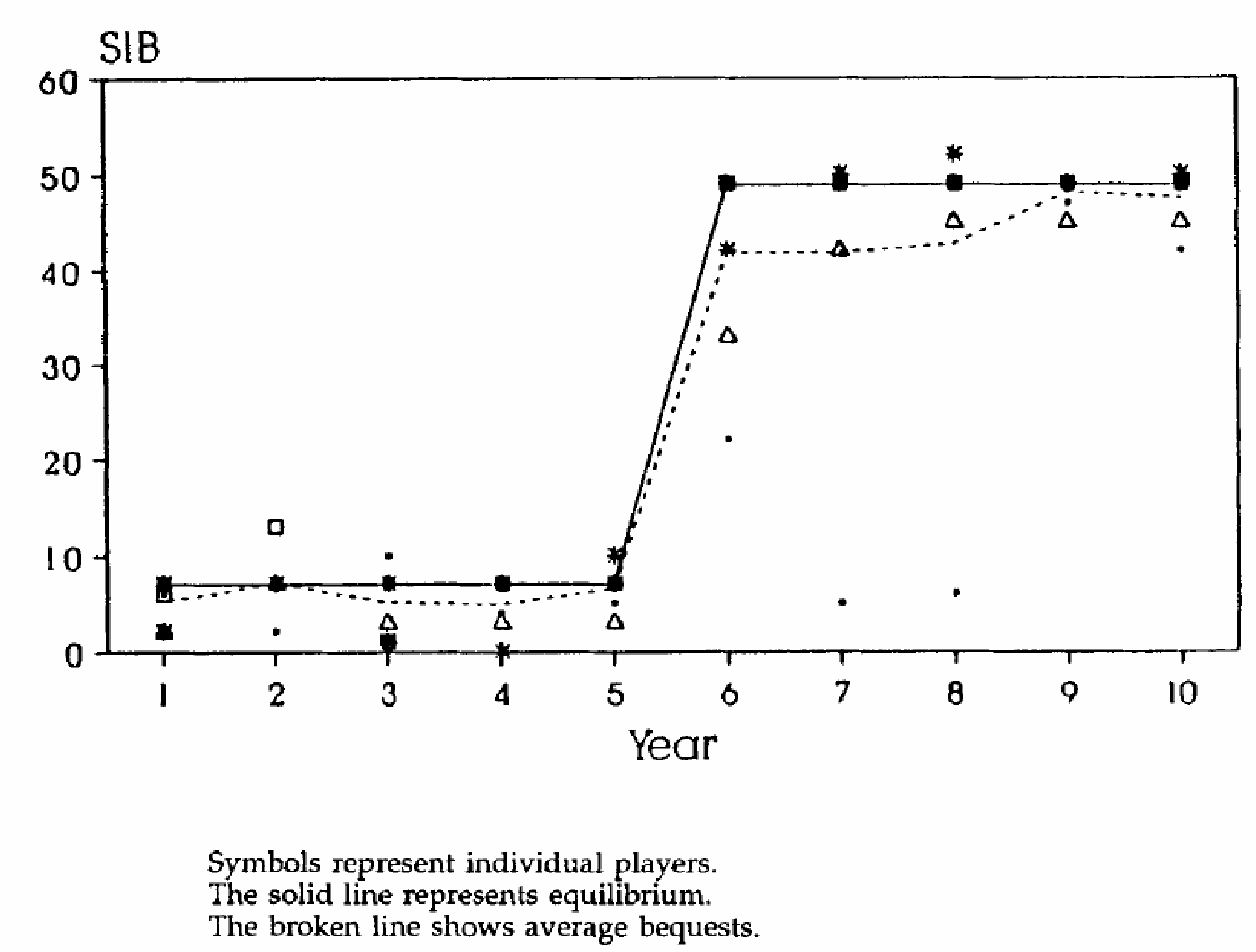


Figure 12: The temporal path of individual and average bequests S_1^B, in Cadsby and Frank's experiment #3. Source: Cadsby and Frank (1991, Figure 3).

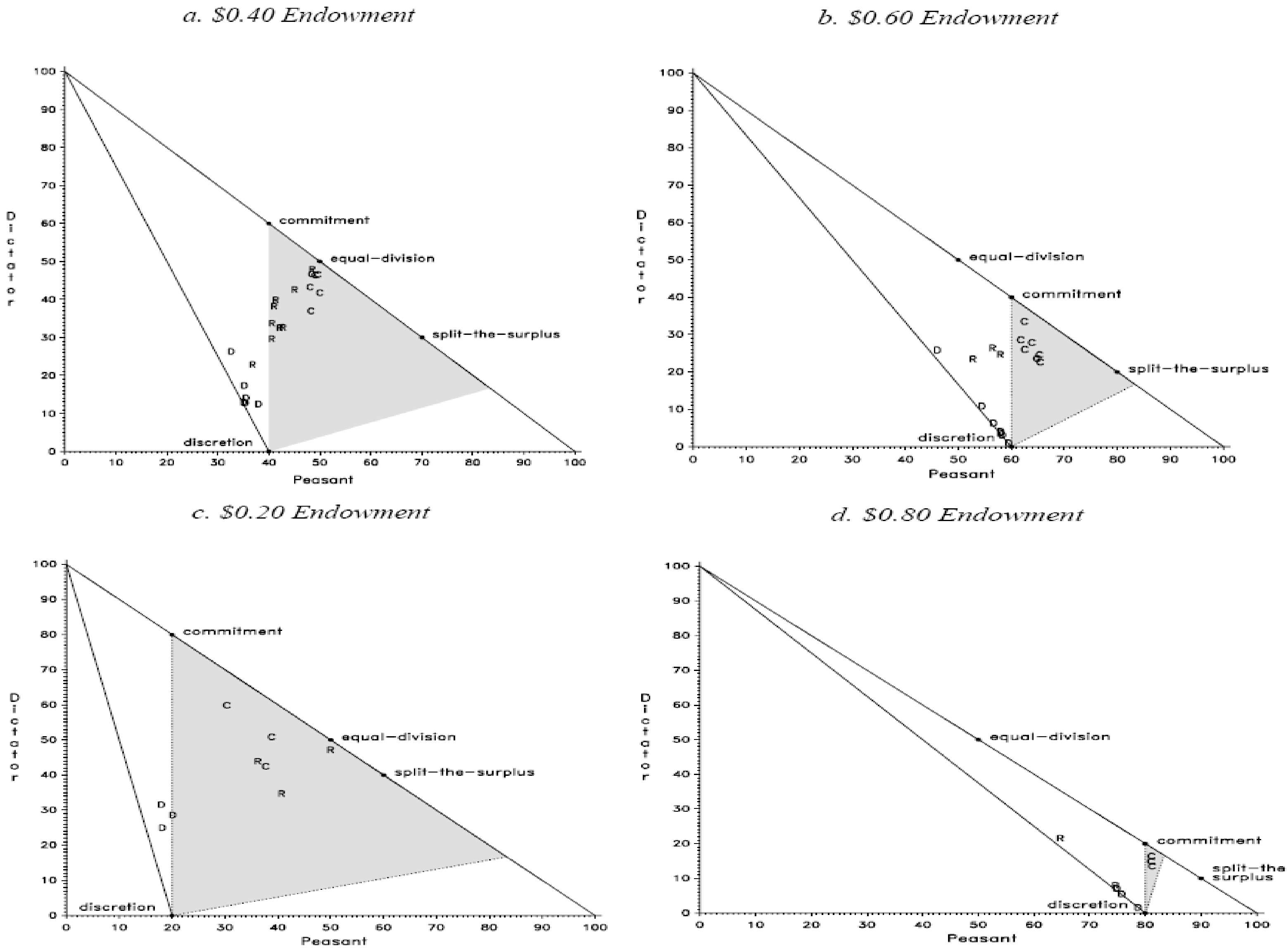


Figure 13: Mean payoffs by cohort: C=commitment, D=discretion, R=reputation in 4 (W,r) treatments of Van Huyck et al.'s (1995, 2001) peasant-dictator game.

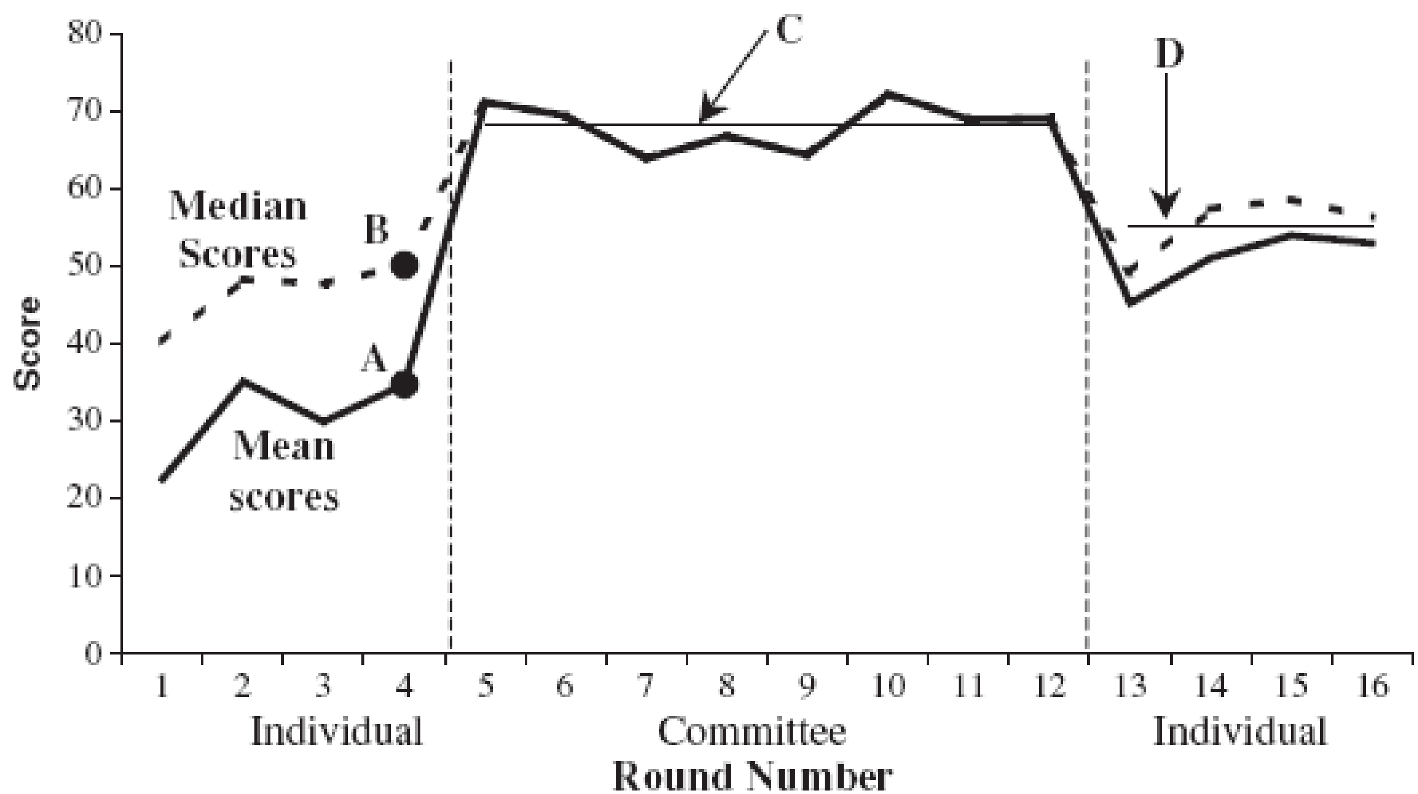


Figure 14: Mean/median scores for players over the various phases of the monetary policymaking experiment of Lombardelli et al. (2005): individual decision-making, group decision-making, and finally individual decision-making.