Why macroeconomics needs experimental evidence

John Duffy[1]

Abstract

This paper discusses how macroeconomics can and already has begun to make use
of controlled experimental methods to address the assumptions and predictions of
macroeconomic models as well as to evaluate the impacts of macroeconomic policy
interventions. Specific issues addressed include rational expectations and alterna-
tives, intertemporal optimization with an application to household consumption and
savings decisions and the efficacy of various monetary policies.

Keywords Experimental economics · Macroeconomics · Rational expectations ·
Intertemporal optimization · Monetary policies

JEL Classification B40 · C90 · E00 · E30 · E50

1 Introduction

Economists were late to adopt experimental methods, perhaps due to a healthy
skepticism that they could use controlled laboratory methods on human subjects to
address economic questions. But following the pioneering contributions of Nobel
Laureates and experimenters Vernon Smith, Reinhard Selten, Elinor Ostrom, Alvin
Roth, Abhijit Banerjee, Esther Duflo and Michael Kremer, the use of controlled,
experimental methods has undoubtedly strengthened our understanding of micro-
economic theory, game theory and development, see, e.g., Falk & Heckman, (2009).
There is now a growing awareness that experimental evidence can be useful to *mac-
ro*economists as well. In this paper, I make the case for conducting "macroeconomic
experiments" by pointing to a number of recent experimental studies that have shed
light upon questions of relevance to macroeconomists and policymakers.

Research in Macroeconomics is currently dominated by the use of dynamic sto-
chastic general equilibrium (DSGE) models, the hallmarks of which are explicit

✉ John Duffy
duffy@uci.edu

1   Department of Economics, University of California Irvine, Irvine, CA 92697, USA

microfoundations involving optimizing agents, continuous market clearing and rational expectations. The use of DSGE models is often advocated on the grounds that experiments with humans subjects are not possible. For instance, Christiano et al., (2017) p. 2 in their defense of DSGE modeling write:

> "Macroeconomic policy questions involve trade-offs between competing forces in the economy. The problem is how to assess the strength of those forces for the particular policy question at hand. One strategy is to perform experiments on actual economies. This strategy is not available to social scientists. As Lucas (1980) pointed out roughly forty years ago, the only place that we can do experiments is in our models. No amount of a priori theorizing or regressions on micro data can be a substitute for those experiments."

Indeed, the dismissal of using any kind of experimental evidence for macroeconomic policy evaluation comes originally from Robert E. Lucas Jr. who was an early advocate for the DSGE approach. For instance, in Lucas, (1980), p. 710 he writes:

> "People interested in the way groups of monkeys solve problems of allocating scarce resources satisfy their curiosity by assembling groups of monkeys and tossing them scarce resources. I have taken it as given that we economists cannot proceed in this way, yet the allocation of scarce resources is something we are admired for being experts at."

Lucas advocates instead for "fully articulated, artificial economic systems that can serve as laboratories in which policies that would be prohibitively expensive to experiment with in actual economies can be tested out at much lower cost." (Lucas, 1980, p. 696).

In the 40 years that have followed, "experiments" using DSGE models have exploded in popularity. These experiments consist mainly of "counterfactual" estimation or simulation exercises, where the researcher departs from the baseline setup in some dimension in order to explore the impact of different policy interventions or some change in the underlying mechanism or modeling assumptions.[1] While the use of structural DSGE models avoids the "Lucas critique" of counterfactual policy evaluation that arises from using reduced form analyses, the counterfactual analysis of DSGE modelers is only valid to the extent that the micro-founded structural model and its assumptions are correctly specified in the first place! Indeed, there are good reasons to think that model misspecification remains an important issue, since DSGE models fail to satisfactorily explain a number of important macroeconomic phenomena, including the celebrated equity premium puzzle, the sources of

---

[1] There is also a small literature on *natural experiments* with macroeconomic implications. For a survey, see (Fuchs-Schündeln and Hassan, 2016). In such experiments, the control and treatment conditions are exogenously determined. The "subjects" are naturally embedded in the experiment by warrant of having lived through a "within-subject" design. However, the treatments, e.g., the reunification of Germany, while certainly interesting, may not be the most useful experimental interventions to macroeconomists for answering the most policy-relevant questions.

aggregate fluctuations, and the effects of money, credit and banking for real economic activity.

Even when DSGE models can account for movements in certain macroeconomic time series, there are often multiple modeling approaches that yield similarly good fits to the data. For instance, both real business cycle models with fully flexible prices and New Keynesian models with sticky prices, the two main variants of DSGE models, can explain observed co-movements in output, consumption and investment in the data rather well, but the two modeling approaches have very different implications for the effectiveness of monetary policy in smoothing aggregate fluctuations.

Toward the goal of building better models, I have advocated for and continue to promote the use of *experimental laboratories* as low-cost laboratories for macroeconomic research.[2] In encouraging this research agenda, I am not advocating against the DSGE modeling approach; that approach is, and remains, an important means by which macroeconomists communicate their ideas with one another. Instead, I see experimental evidence as an important *complementary resource*, along with other micro-level evidence to be used for evaluating and building better models, resolving indeterminacies and thinking about reactions to policy interventions. For instance, experimental evidence can be brought to bear on the reasonableness of DSGE modeling assumptions such as rational expectations and intertemporal optimization. More recent generations of DSGE models have exploited increased computing power to relax the assumptions of representative agents and rational expectations thereby allowing for a richer heterogeneity of outcomes. Experimental evidence can be useful in characterizing the nature and extent of this heterogeneity, whether it may lie in agents' preferences or in their degrees of bounded rationality, see, e.g. Arifovic & Duffy, (2018), and I will provide several examples here. Further, in models with multiple equilibria, e.g., those pertaining to the financial sector, experimental evidence can point to which of the various equilibria are more attractive to subjects and thus more empirically relevant for policy consideration. Finally, policy considerations, e.g., the effectiveness of different types of monetary policies, can be evaluated on a small scale using controlled experimental methods before being implemented in the field.

In addition to the changes that have occurred in computing power over past 40 years, so too has our knowledge of experimental methods, which has expanded to the point that experiments with human subjects that are relevant to macroeconomists can and have been conducted, and at relatively low cost as well. In this paper, I point to several macro assumptions and policy questions that experiments can and have helped to address. I conclude with some suggestions for the conduct of further macroeconomic experiments.

---

[2] See, e.g. Duffy, (2016).

2 Rational expectations

The rational expectations hypotheses proposed by Muth, (1961) and further popularized by Lucas, (1972) is a mainstay of modern DSGE modeling. If agents have rational expectations and there are no other frictions (e.g., price stickiness) then countercyclical monetary or fiscal policy interventions aimed at expanding output would be largely ineffective as agents can rationally foresee the consequences of these policy interventions and so respond by revising wage and price expectations accordingly so that there are no lasting real effects (Sargent & Wallace, 1975). More recent structural DSGE models that presume rational expectations have similar predictions; the dynamics of these estimated DSGE models imply faster adjustments toward steady states than what appears in the available field data. To address these differences in adjustment dynamics, various devices have been introduced to explain the sluggish adjustments of consumption or investment in the data, including habit formation, convex investment adjustment costs and indexation on past inflation.

Nevertheless, the experimental evidence for rational expectations is rather weak at least at the individual level. See Assenza et al., (2014) for a recent survey. As an illustrative example, consider the very simple first-order self-referential system for determination of the price level, p_t

$$p_t = \mu + \alpha E_t[p_{t+1}] + \epsilon_t, \tag{1}$$

where μ, $\alpha < 1$ are known constants and ϵ_t is a mean zero error term. This type of forward-looking expectational difference equation is the reduced form of the same "cobweb model" studied by Muth, (1961) and it lies at the heart of many macroeconomic models, though they may involve many more variables, a less direct relationship between expectations and outcomes and inflation rather than the price level. If, as under the rational expectations assumption, agents *know* the model, then they should be able to immediately solve for the rational expectations equilibrium (REE) value:

$$p_t = \frac{\mu}{1 - \alpha} \quad \text{for all } t.$$

In Bao & Duffy, (2016), we compared individual forecasts for p_t which we interpreted as those chosen by monopoly firms versus the forecasts of p_t given by groups of N individuals (an oligopoly firm treatment), where $E_t[p_{t+1}] = N^{-1} \sum_{i=1}^{N} E_{i,t}[p_{t+1}]$. Consistent with the rational expectations assumption, we endowed agents with complete knowledge of the data generating process, (1), including properties of the error term, ϵ_t. Subjects in both treatments were incentivized monetarily to forecast the future price level, p_{t+1} as accurately as possible; their payoff was a decreasing function of their forecast error, $E_{i,t}[p_{t+1}] - p_{t+1}$, in what is now known as a "learning to forecast" design (Duffy, 2010).

We compared the time it took for subjects' forecasts to converge to the REE within a maximum of 50 rounds in the monopoly and oligopoly
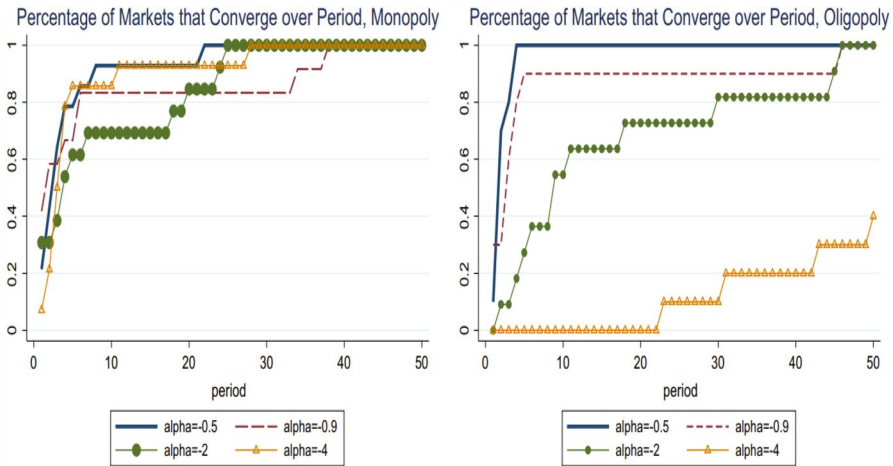
Fig. 1 CDFs of time to convergence to the REE in monopoly and oligopoly treatments. Source: Bao & Duffy, (2016)

treatments. We also varied the feedback parameter, α, which took on values in $\{-0.5, -0.9, -2, -4\}$. A summary of the experimental results is shown in Fig. 1

The main finding is that it *takes time* for individuals and groups to learn the REE, even if they perfectly know the data generating process! How much time depends on the parameters of the model and whether there is one (representative agent) as in the monopoly treatment or many heterogeneous agents as in the oligopoly treatment. We concluded that bounded rationality and/or strategic uncertainty can play an important, but largely ignored role in the equilibration process of models that presume rational expectations on the part of either a representative agent or groups of heterogeneous agents.

A natural reaction to such findings is: why not simply relax the rational expectations hypothesis? The immediate response by most macroeconomists would be that if we were to relax the REE assumption, then we would enter into what Sims, (1980) has termed the "wilderness" of bounded rationality and irrational expectations with no clear path forward. Some theorists, e.g., Sargent, (1993) and Evans & Honkapohja, (2001) have nevertheless ventured forward by proposing models of boundedly rational agents in macroeconomic models and how the presence of such agents would alter standard predictions. Evans & Honkapohja, (2009) propose the useful "cognitive consistency principle", which requires that agents in macroeconomic models should not be more knowledgeable than (good) economists, who generally lack knowledge about the specification of the actual data generating process. Still, and somewhat surprisingly, theorizing about how individuals form expectations has not been informed very much by or subjected to much experimental testing until only recently.

Even the cognitive consistency principle leaves a lot of wiggle room. For instance, Evans & Honkapohja, (2001), advocate that agents might be modeled as econometricians who run regressions of the variables they are seeking to forecast on

past data from model relevant factors. Under certain conditions, such econometric learning can lead to convergence to a REE, particularly if that REE is a fixed point of perceived law of motion that agents use as their regression model. But if the REE is not a fixed point of that perceived law of motion, i.e., if the regression model is *misspecified*, then the learning dynamics can lead to what Sargent, (1999) has termed self-confirming equilibria.

But let us consider the notion that the agents in our models *can* be modeled as econometricians or more generally as adaptive agents, each of whom averages past data in some way so to form expectations of the variables of interest. What is the experimental evidence on this dimension?

Here, the experimental evidence suggests that agents may not be as sophisticated learners as econometricians. Rather, subjects appear to use even simpler heuristics and there is heterogeneity in the heuristics that populations of subject use so that no one single model of adaptive learning behavior may be appropriate.

For example, Anufriev & Hommes, (2012) posit that agents forming expectations in models similar to equation (1) consider several different heuristic rules-of-thumb for forecasting $E_t[p_{t+1}]$ and switch among these based on their relative performance. Specifically, they find substantial support for three main types of rules in the experimental data:

Adaptive rules:	$E_t[p_{t+1}] = \lambda p_{t-1} + (1 - \lambda)E_{t-1}[p_t],$
Trend-following rules:	$E_t[p_{t+1}] = p_{t-1} + \gamma(p_{t-1} - p_{t-2}),$
Anchor and adjust rules:	$E_t[p_{t+1}] = \phi(p_{t-1}^{\text{avg}} + p_{t-1}) + (p_{t-1} - p_{t-2}).$

The first of these, the adaptive learning rule, was a mainstay of macroeconomic modeling *before* the rational expectations revolution (see, e.g., Nerlove, 1958). The other two rules involve, in part, the exploitation of recent trends in the data as reflected in the $(p_{t-1} - p_{t-2})$ terms that can arise in self-referential forward-looking expectations models (such as (1)) where future expectations of variables such as prices matter for their current realizations. That such trend extrapolation can matter is definitely a consequence of putting experimental subjects in the modern forward-looking models that are of interest to macroeconomists. But notice that none of these models is very sophisticated econometrically. With the exception of the Anchor and Adjust model, there is not much use of lagged data or past averages.

Consider, by contrast, the notion that agents are *least squares learners*, which is a common assumption in Evans & Honkapohja, (2001) and the literature on learning in macroeconomics. In particular, suppose that agents form expectations of future prices by running regressions on past data just as an econometrician would. How does such learning compare with what people actually do? In Bao et al., (2021) we use experimental methods to make a detailed comparison. We consider the Cobweb model which has a reduced form equilibrium price expression of the form:

$$p_t = \mu + \alpha p_t^e + \delta w_t + \epsilon_t, \tag{2}$$

where p_t is the time t price level, p_t^e is the expected price level, w_t is an exogenous random, "weather" variable whose time t value is known at time t, μ, α and δ are
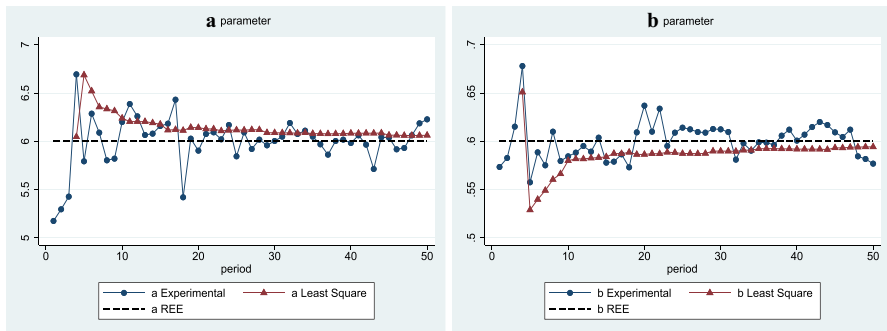
Fig. 2 Point estimates for parameters **a** and **b**, experimental data versus least-squares estimates. From Bao et al., (2021)

coefficients unknown to market participants and ϵ_t is a mean 0, i.i.d. noise term. As in the macro-learning literature, we endow agents with knowledge of the correctly specified "perceived law of motion":

$$p_t^e = a + bw_t + \epsilon_t, \tag{3}$$

with which to form expectations of p_t^e, using available data through time $t-1$. One aim, as in the macro-learning literature is to again assess whether agents can learn the REE, i.e., whether $a \to (1-\alpha)^{-1}\mu$ and $b \to (1-\alpha)^{-1}\delta$. But here we go a step further and consider the *process* by which agents adjust their estimates of a and b over time. If subjects were least-squares learners, then $p_t^e = \hat{a}_t + \hat{b}_t w_t$, where

$$\begin{pmatrix} \hat{a}_t \\ \hat{b}_t \end{pmatrix} = \left(\sum_{i=1}^{t-1} z_i z_i' \right)^{-1} \left(\sum_{i=1}^{t-1} z_i p_i \right),$$

and $z_i' = (1, w_i')$. In the experiment, we ask subjects to choose the values to use for a and b in each of 50 iterations with the understanding that these will be used in (3) to generate a forecast for p_t^e for which they will be paid based on their forecast accuracy as in standard learning-to-forecast experiments. Here, the subjects form expectations in groups of 6 and the average expectation of the group determines the actual realization of the price level via equation (2). Figure 2 shows the average choice of the a and b numbers from one of the treatments (which differ in terms of the persistence of the w variable) in relation to the recursive least-squares predictions for the a and b numbers. As the figure reveals, while there is some evidence for convergence to the REE values for a and b by the experimental subjects with the passage of time, the average values for a and b fluctuate much more than the least squares estimates, which are smoother and converging only very slowly to the REE estimates. Indeed, the data suggest that some other type of learning algorithm, perhaps a more vigilant, constant gain algorithm, would provide a better fit than least-squares learning. This illustration provides a very micro-level but still useful rationale for the use of

experimental evidence in thinking about how to model departures of behavior from rational expectations, as expectations data is hard to come by.

The ultimate aim of the learning and expectation formation literature in macroeconomics should be to develop a robust *behavioral model* of expectations that could serve as a replacement for rational expectations and which would also be useful for policy evaluation. We are only now in the initial stages of such a transformation.[3] Experimental evidence should be a complementary tool in developing this new approach.

3 Savings and intertemporal optimization

In addition to rational expectations, the other main pillar of macroeconomic modeling is intertemporal optimization. Given expectations of future wages, prices, profits and rates of return, it is a standard assumption that agents (households and firms) can solve complex, intertemporal optimization problems that may have no closed form solutions and are often solved numerically. As in the case of rational expectations, if we are to improve our models, it would be useful to understand the extent to which such an optimization assumption is valid. Intertemporal optimization is the legacy of Frank Ramsey (1928) and was further refined by Cass, (1965) and Koopmans (1965) in an optimal growth model setting where infinitely lived households with perfect access to credit markets are assumed to make optimal consumption and savings plans given perfectly competitively determined rates of return. An earlier literature, e.g., the Solow, (1956) growth model, is agnostic about how households choose to save. An even earlier literature was quite *behavioral* arguing that consumption/savings decisions were influenced by social factors, e.g., consumption or income relative to others, e.g., Veblen, (1899) and Duesenberry, (1949), and not in pursuit of some utility maximization objective. For illustration purposes, I will focus on households' savings decisions using the modern optimization approach as this has been the subject of the greatest amount of experimental research. While there is a lot of field data on household savings behavior, data on the information and choices available to households when making those savings decisions can be hard to come by, and aside from the occasional natural experiment, there is no good means of evaluating household's responses to changes in policies that may foster or inhibit savings behavior.

Consider the problem of a household seeking to maximize consumption over a T-period horizon.[4] The problem can be written as:

$$\max_{\{c_t\}} E_0 \sum_{t=0}^{T} \beta^t u(c_{t+1}) \tag{4}$$

subject to

---

[3] For a good first step see (Hommes et al., 2019).

[4] The infinite horizon case has also been studied, see, e.g., Anderhub et al., (2000), Duffy & Li (2021).
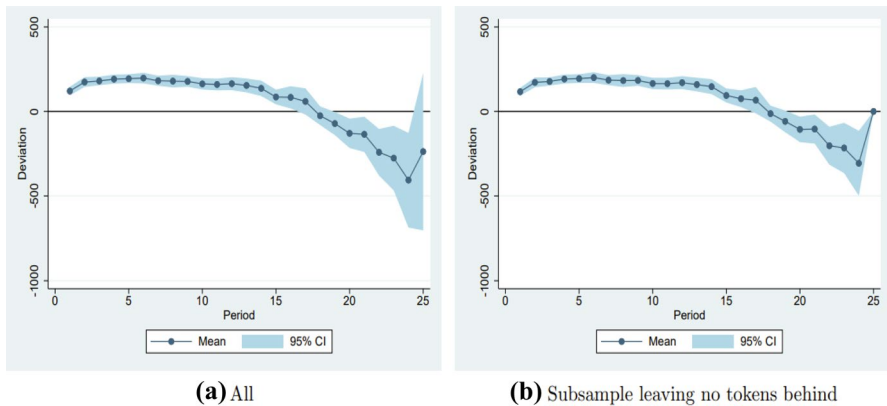
(a) All   **(b)** Subsample leaving no tokens behind

Fig. 3 Mean consumption deviations from the conditionally optimal path. Source: Duffy & Li, (2019)

$$a_{t+1} = (1 + r)a_t + y_t - c_t = (1 + r)a_t + s_t, \quad a_0 \quad \text{given} , \tag{5}$$

where u is a concave utility function, c_t denotes period t consumption, a_t wealth, s_t savings, and y_t income and β is the period discount factor while r is the exogenous interest earned on savings (we shall consider a partial equilibrium solution only). The solution can be found via backward induction. It involves solving $T - 1$ Euler equations of the form,

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

together with the budget constraint (5) and the fact that $a_{T+1} = 0$. Depending on the specification chosen for u and the number of periods, T, this can become a very complicated task. Indeed, many macroeconomists would solve this and more complicated versions of the problem numerically.

Various versions of this intertemporal optimization task have been studied in laboratory experiments. See, e.g., Hey & Dardanoni (1988), Anderhub et al., (2000), Carbone & Hey (2004), Ballinger et al., (2003), Ballinger et al., (2011), Carbone, (2006), Brown et al., (2009), Carbone & Duffy (2014), Duffy & Li, (2019) and Duffy & Li (2021). These versions have involved either stochastic or known deterministic processes for income, $\{y_t\}$, and have typically involved an induced concave utility function, u that serves as a mapping from consumption choice levels to monetary payoffs made to subject participants. Interest rates are typically exogenously fixed or set to zero, and discounting is not usually considered (i.e., $\beta = 1$) due to the short time horizons of most experiments.[5] Generally T has been set to many periods, e.g., 20–30 so as to simulate a lifecycle setting, with each period representing some length of time, e.g., 1–2 years.

---

[5] In the infinite horizon case, where $T \to \infty$, the discount factor represents the constant probability that a sequence of consumption decisions continues from one round to the next.

Consider as an illustration, Duffy & Li, (2019)'s experiment where agents face a known, deterministic income profile over $T = 25$ periods that drops following period 18, representing the retirement phase of life. In this experiment, there is a fixed interest rate of $r = 0.10$ and no discounting. A main finding to come out of this and many similar experiments is that individuals *over-consume* in the early periods of their T-period lifetime relative to the optimal path. As a result, they acquire less wealth, and so they end up under-consuming in the later periods of life. Figure 3 provides an illustration of the mean deviations of consumption from the conditionally optimal path in one of Duffy and Li's treatments.[6] The left panel includes all data including some subjects who do not fully consume all of their assets and income in the final, known period T, (i.e., $a_T + 1 > 0$) while the right panel excludes such subjects. The pattern of over- and under- consumption over the lifecycle remains the same regardless, and is also similar if we consider deviations from the unconditionally optimal path.

How might we explain this behavior? As in the case of relaxing the rational expectations hypothesis, one solution is to introduce type heterogeneity. Following (Campbell & Mankiw 1989) some agents are classified as "hand to mouth" consumers, who consume *all* of their income in each period ($s_t = 0$) for all t, while others act in a more rational fashion. In Duffy & Li, (2019) we find support for this view. Specifically, we find a mix of hand-to-mouth and conditionally optimal agents. Further, we show that such a mixture of types arises naturally from a rational inattention model as first proposed by Sims, (2003), Matějka & McKay (2015) and Gabaix, (2016), that is beginning to make some inroads in macroeconomic modeling. Following this approach, subjects are assumed to differ in their abilities to solve the optimization problems and also face some information processing costs to solving the intertemporal problem. Their incentives to solve the problem depend on their ability levels and the earnings they could get from adopting some default rule, e.g., the simpler strategy of just consuming their endowments in each period, or being a "hand-to-mouth" consumer. If their ability is high, the costs of solving the optimization problem are more than offset by the utility gains from following the conditionally optimal path over the heuristic of consuming endowments. On the other hand, if their ability is low the costs of solving the optimization problem might outweigh the utility gains relative to the strategy of consuming endowments, and so those agents rationally act as hand-to-mouth consumers. More work is needed to identify differences in cognitive abilities among subjects in order to further rationalize this result, but it seems a promising explanation.

Another, related approach is to model all agents as being forward-looking, but having bounded planning horizons as in Caliendo & Aadland (2007). For example, a planning horizon of τ periods so that in period t, the subject acts as if period $t + \tau$ is the final period. In this approach, both hand-to-mouth and optimal agents emerge as special limiting cases, where the length of the planning horizon τ is set to zero or

---

[6] The conditionally optimal consumption path considers the wealth of the agent at each date in time, as opposed to the unconditionally optimal path which is the optimal path for consumption starting from date $t = 1$.

the entire lifetime, respectively. A further advantage of this approach is that it can generate "hump-shaped" lifecycle profiles for consumption that are consistent with empirical data on household consumption expenditures over the lifecycle. Some corroborating experiment evidence for bounded planning horizons comes from Carbone, (2006), who reports that a sizeable fraction of subjects use shorter than optimal planning horizons of just a few (i.e., 1–5) periods ahead when making their consumption/savings decisions in lifecycle planning experiments.

A third approach is to assume some departure from the discounted utility model such as present bias e.g., Laibson, (1997), O'Donoghue & Rabin (1999). While in most laboratory experiments, discounting is not imposed, it remains possible that subjects nevertheless do discount or attach more weight to the utility from current consumption relative to future consumption, even in the short-time frame of the experiment. Present bias is typically captured using the $\beta - \delta$ formulation, where instead of maximizing (4), it is imagined that agents maximize:

$$u(c_1) + \beta E_1 \sum_{t=1}^{T} \delta^t u(c_t),$$

where β is now the present bias parameter, and δ is the traditional period discount factor. Experimental evidence has been brought to bear on this question, but with mixed results. Some studies find some evidence for present bias, while other find that exponential discounting characterizes behavior rather well. See for example (Andersen et al., 2008; Benhabib et al., 2010) and Andreoni & Sprenger, (2012) among others. Proxying the utility value of consumption via monetary payments as is done in most of these experiments may, however, be problematic; for instance, Augenblick et al., (2015) finds greater present bias with respect to *effort* than with respect to money. Further, in thinking about time preferences, one may want to relax the representative agent assumption. For instance, Jackson & Yariv, (2014) show that present bias should be expected and is in fact quite common in collective action experiments where a subject chooses a consumption stream for other subjects who differ in their discount factors, as might be done by a social planner or policymaker. Finally, there is also some relevant field evidence from Brown & Previtero, (2014) revealing that the same individuals who procrastinate in signing up for health insurance coverage also have less in accumulated savings.

A fourth and final explanation for under-saving is what has been termed "exponential growth bias" (EGB). This is the failure to properly account for the compounding of interest earned on assets over time. Levy & Tasoff (2016) model the EGB phenomenon by supposing that agents mistakenly perceive, in whole or in part, that they earn *simple* rather than compound interest on savings, so that their asset positions grow linearly rather than exponentially. Under certain utility conditions, if agents have EGB, the price of later consumption relative to present consumption will be perceived to be high, so earlier period consumption will be greater than that of an agent without exponential growth bias. Further, as the future value of the agent's asset position is reduced by EGB, the incentives to delay income into the future by saving more are also reduced. Levy & Tasoff (2016) report experimental evidence suggesting that one-third of subjects can be classified as making lifecycle
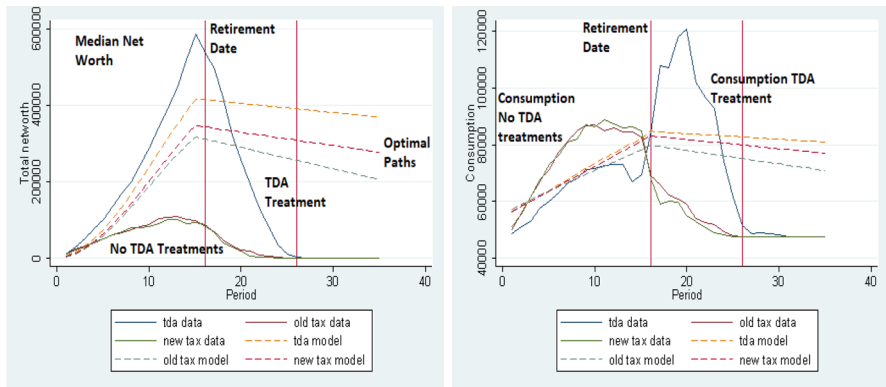
Fig. 4 Median net worth (left panel) and consumption (right panel) over the lifecycle, TDA treatments versus no TDA treatments. Dashed lines are optimal policies while solid lines represent data. Source: Duffy & Li (2021)

consumption decisions according to the simple interest rate specification of their model, while only 4% behave as if they fully understand compound interest rate calculations; the remaining 65% lie somewhere in between. Given these difficulties in saving over the lifecycle, an obvious policy question to explore is what kinds of treatment interventions, e.g., education, nudges, framing, etc., might help to improve lifecycle savings behavior? Levy and Tasoff explore the use of graphical illustrations of asset growth in an effort to *de-bias* subjects from EGB but report that this has little impact. Ballinger et al., (2003) have subjects participate in a kind of intergenerational learning experiment, where one generation (cohort) of subjects observes the lifecycle decisions made by their parents, before moving on to make their own lifecycle savings decisions. This design does yield some improvement in behavior by the experienced observer cohorts. Thaler & Benartzi, (2004)'s natural experiment offering an automatic pre-commitment to save pay raises ("save more tomorrow") is shown to have tripled savings rates over the first four years. Further as Choi et al., (2004) showed, enrolling employees automatically in retirement savings plans and forcing them to opt-*out* if they do not want to participate, harnesses people's natural inertia to do nothing and raises both enrollment and savings rates relative to the conventional opt-in approach. However the horizon of such field studies is limited; both the Thaler & Benartzi, (2004) and Choi et al., (2004) studies considered impacts/ participation in just the first four years.

Duffy & Li (2021) consider the role played by tax deferred retirement accounts for improving *lifecycle* savings using laboratory experiments. In one treatment, called the TDA treatment, subjects have access to a tax deferred account (TDA) that is calibrated to match the current US system, while in two other treatments they do not. The latter two differ in tax policies; one of them keeps the same tax policy as in the TDA treatment while the other one generates the same expected government revenue as in the TDA treatment. The main finding is that in the TDA treatment, the median individuals' net worth at the retirement date is considerably greater than what obtains in the no TDA treatments—see left panel of Fig. 4 and is even greater

than the predicted rational choice model level. Further, while in the no TDA treatments there is over-consumption in the pre-retirement periods—see right panel of Fig. 4 this over-consumption disappears if agents have access to a TDA. The main take-away here is that both the deferred tax benefits of a TDA and the commitment that such accounts provide (savings cannot be accessed until retirement) has a significant effect on net asset positions at retirement. This is an example of the kind of counterfactual analysis (turning TDAs on or off) that would be difficult to do in the field and it illuminates once again the value of the experimental method.

4 Monetary policies

By contrast with savings decisions, monetary policies do not have much to do with long-run growth. Rather, monetary policies are thought to be effective in the short-run for addressing business cycle fluctuations. As noted earlier, we have a variety of DSGE models that can match co-movements of macroeconomic time series data but which often lead to different conclusions regarding the effectiveness of monetary policies. Central bankers themselves seem puzzled about which modeling approaches to use.[7]

Toward improving our understanding of which monetary policies are effective and which are not, there may be no better approach than trying out different policies in laboratory environments and observing how incentivized subjects react to them. The costs of doing so are low and insights gained can be substantial.

4.1 The Friedman rule

The Friedman rule (Friedman, 1969) is perhaps the most celebrated monetary policy rule of all time. This rule for monetary policy stipulates that the central bank should implement monetary policy so as to make the nominal interest rate zero. The logic is that, while money is useful for carrying out transactions it is also costly to hold because of the interest that could be earned on alternative assets. To maximize the demand for money and transactions associated with that demand, the optimal monetary policy should therefore be to set nominal interest rates to zero. While the Friedman rule is the optimal monetary policy in a wide variety of macroeconomic models, perhaps surprisingly, it has not been implemented by any central bank. Reasons include difficulties with the control of monetary aggregates, the absence of lump-sum taxes and transfers and other technical difficulties. But these considerations are largely abstracted from in the theoretical literature and can also be set aside in the laboratory.

Indeed, in Duffy & Puzzello, (2021) we implement the Friedman rule in a laboratory economy using the micro-founded model of Lagos & Wright (2005),

---

[7] For instance, there is currently some debate as to whether the Phillips curve trade-off between inflation and unemployment that lies at the heart of New Keynesian models remains empirically relevant (e.g., Hooper et al., 2020).
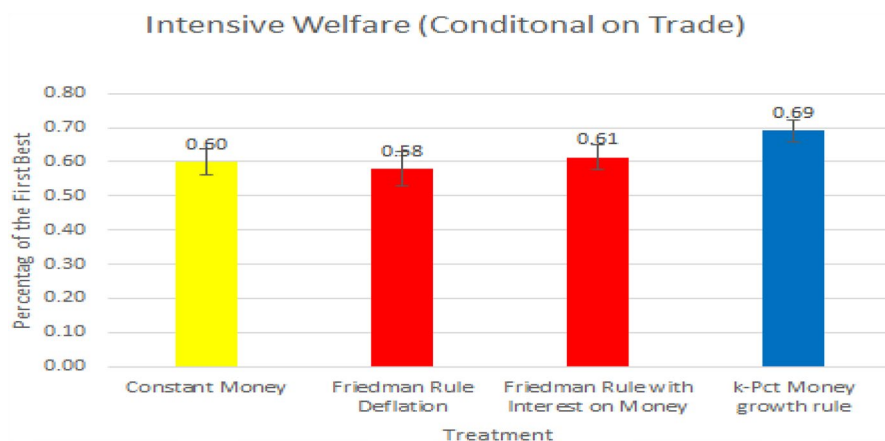
Intensive Welfare (Conditonal on Trade)

Fig. 5 Welfare relative to the first best across each of the four treatments in Duffy & Puzzello, (2021)

which enables welfare comparisons. We consider two implementations of the Friedman rule. In both, the gross rate of expansion of the money supply, μ, is set so that $\mu = \beta(1 + i)$, where β is the period discount factor and i is the nominal interest rate. In the deflation treatment, FR-DFL, i is set equal to 0 and so $\mu = \beta$; that is, the money supply contracts over time at rate $\beta - 1$. The reduction in the supply of money is implemented as a lump-sum tax on individual money holdings. In a different implementation of the Friedman rule, interest is paid on money holdings each period so as to encourage money demand and the interest payments are financed via lump-sum taxes. In this case, the money supply is held constant with $\mu = 1$, so the interest rate was set to $i = \frac{1-\beta}{\beta}$. Two other treatments considered a constant money supply regime where $\mu = 1$ with no interest payments on money as a no Friedman rule baseline treatment as well as a k-percent (k-PCT) inflationary monetary policy regime (which Friedman also advocated), where $k = 1 - \beta$ ($\mu = 2 - \beta$) which is the opposite of the deflation rate $\beta - 1$ of the FR-DFL treatment. The experimental findings are summarized in Fig. 5, which reveals the welfare achieved in the four treatments relative to the first best outcome, which is theoretically only attainable under the two Friedman rule policy regimes.

The two Friedman rule policies do no better, in welfare terms, than a constant money supply regime. By contrast, a growing money supply regime, the k-PCT treatment (which more closely approximates actual monetary policy) provides significantly higher welfare than in either of the two Friedman rule treatments or the Constant Money regime treatment. Duffy & Puzzello, (2021) show that this failure is due to a combination of liquidity constraints and precautionary motives; as subjects face taxes paid in tokens (fiat money) they sub-optimally accumulate/ hoard tokens (money). With a growing money supply and no lump-sum taxes, they are less likely to engage in such behavior. These findings suggest that despite providing a simple, micro-founded experimental economy where the Friedman

rule can improve welfare, it does not do so. Such findings provide a possible hint as to why the Friedman rule has not been widely adopted in practice, even though it continues to be featured prominently in the prescriptions of monetary theorists.

4.2 The Taylor rule

A monetary policy rule that *does* capture the behavior of many central banks (CB) is the Taylor rule (Taylor, 1993). The Taylor rule holds that the central bank should adjust nominal interest rates in response to deviations of inflation, π_t, and possibly the output gap, y_t, from certain target values, π^* and y^*, respectively, subject to interest rates being non-negative:

$$i_t = \max[\pi^* + r + \alpha_\pi(\pi_t - \pi^*) + \alpha_y(y_t - y^*), 0].$$

Here, r is the real, natural interest rate, i_t is the nominal interest rate under the control of the central bank and α_π and α_y, are the weights that the CB assigns to deviations of inflation and the output gap from target levels. While the Taylor rule is frequently used as both a description and a prescription for monetary policy, the efficacy of this rule in managing private sector expectations, and stabilizing inflation and output is not so clear as there can be many confounding factors (e.g., macroeconomic shocks) impacting on the economy, and thus there is a role for experimental evaluation.

Experimental tests of this model have typically employed New Keynesian DSGE models where prices are sticky and expectations of future values of inflation π_t and the output gap y_t matter for the realizations of these same variables in a learning-to-forecast design. See for example Assenza et al., (2021); Arifovic & Petersen, (2017); Cornand & M'baye, (2018); Pfajfar & Žakelj (2018)) and Mauersberger, (2021) among others. In the typical experiment, subjects forecast inflation and/or the output gap and their forecast accuracy matters for their payoff. Forecasts are then fed into the NK model to produce realizations for actual inflation and output. A key policy question is the size of the weights, α_π, and α_y that central bankers should assign to the Taylor rule and the efficacy of such a rule for stabilizing prices and output.

Assenza et al., (2021) studied the case where $\alpha_y = 0$, which can be viewed as an inflation targeting regime. Their four treatments involved variations in α_π: 1.00 (T1), 1.005 (T2) 1.015 (T3) and 1.5 (T4). Only the last 3 treatments, T2–T4, are consistent with the "Taylor principle", i.e., that $\alpha_\pi > 1$; the last treatment, T4, uses Taylor's preferred coefficient choice that α_π=1.5 (Fig. 6).

The experimental results suggest that the Taylor principle is *not* as sharp a consideration for whether policy stabilizes inflation and the output gap or not. When α_π is low, either 1.0 or 1.005, most economies do not converge to target values for inflation and the output gap. It is only if α_π is sufficiently high (not just greater than 1) that convergence is achieved.[8] Taylor's preferred coefficient choice of $\alpha_\pi = 1.5$

---

[8] Assenza et al., (2021) show that this result obtains if agents are modeled as using the heuristic switching model discussed earlier.
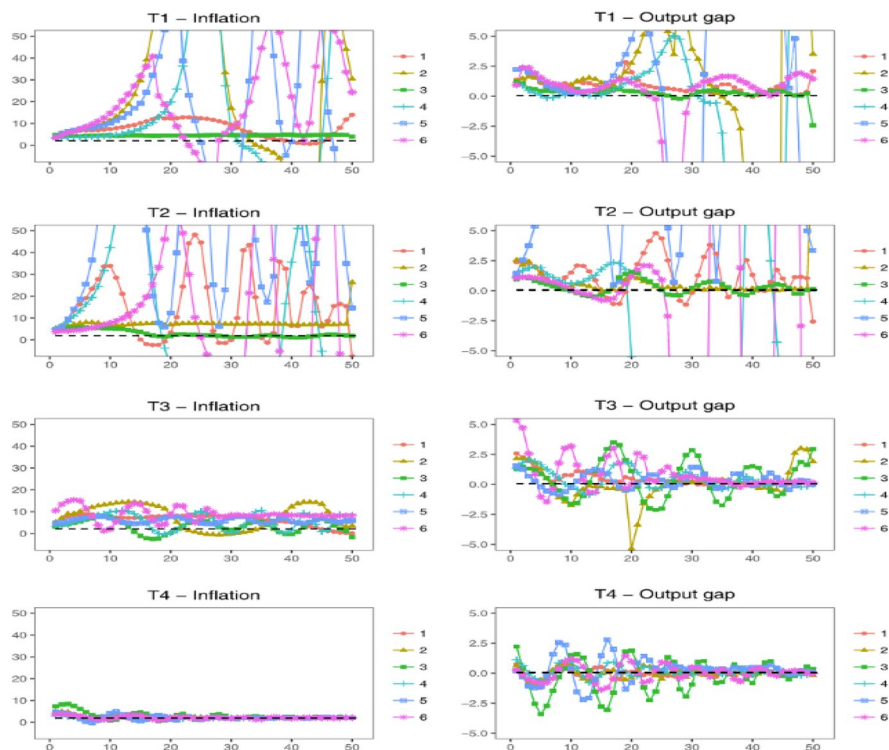
Fig. 6 Inflation and output under four different weights on deviations of inflation from a target value of 2%. Source: Assenza et al., (2021)

fares the best in terms of stabilizing expectations and achieving convergence to target values.

Some central banks, e.g., the US and New Zealand, have a *dual* mandate to maintain both price stability (low inflation) and full employment (a zero output gap). Relying on a single instrument, the nominal interest rate to achieve both objectives can be difficult, as shown in an experiment by Duffy & Jenkins, (2019) where subjects play both the role of the private sector forming expectations about inflation and the role of the central bank. Subjects in the private sector role were incentivized to form correct inflation expectations, which entered into the formula determining the actual inflation rate in a New Keynesian type model. The human subjects assigned to the central bank role were incentivized to choose interest rates so as to minimize deviations of inflation and output from targets values, in this case $\pi^* = 2.5$, $y^* = 0$ in a manner similar to a Taylor rule, but without an explicit rule dictating the interest rate policy that the human subject CB player actually chose. Specifically the CB's payoff function was

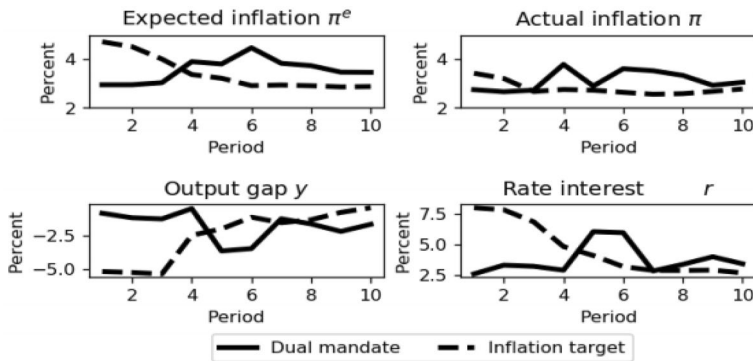$$\pi^{CB} = \text{ Constant } - (\pi_t - \pi_t^*)^2 - \lambda y_t^2.$$

Fig. 7 Mean realized values for π^e, π, y and r under the inflation targeting ($\lambda = 0$) and dual mandate regimes $\lambda = 0.10$. REE predictions are $\pi^e = \pi = 2.5$, $y = 0$ and $r = 2$. Source: Duffy & Jenkins, (2019)

In the inflation targeting regime $\lambda = 0$ (as in Assenza et al., 2021) while in the dual mandate regime $\lambda = 0.10$. The main finding from this study is that the inflation targeting regime yields better management of inflationary expectations, actual inflation, the output gap, and interest rates than does the dual mandate regime as shown in Fig. 7

4.3 Central bank communication

Another means of managing inflation and inflationary expectations is direct communication with the public (so-called "open mouth operations"). The aim of such operations is to better anchor inflationary expectations while also clearly communicating changes in the interest rate the central bank controls. Following the Fisher equation:

$$i_t = \pi_t^e + r_t,$$

if inflationary expectations π_t^e, are well-anchored, then central bank changes in the nominal interest rate, i_t, can affect real rates of return, r_t and thus real activity, at least in the short run, before inflationary expectations adjust. Nevertheless, the evidence suggests that inflationary expectations are far off-the-mark from actual inflation levels and the private sector does not always immediately react to changes in the central bank's policy rate changes. For instance, Coibion et al., (2020) report that households' and firms' expectations about inflation (across many low inflation countries) are much greater than actual inflation. Similarly, Diamond et al., (2020) report that 2/3 of Japanese households expected inflation to be no less than 2% in 2014 even though the official rate at the time was 1.5% and has since fallen.

Several experiments have been conducted examining the role of central bank communication for managing inflationary expectations and for various interventions that might improve the public's comprehension of policy changes or their reaction to such policy changes.
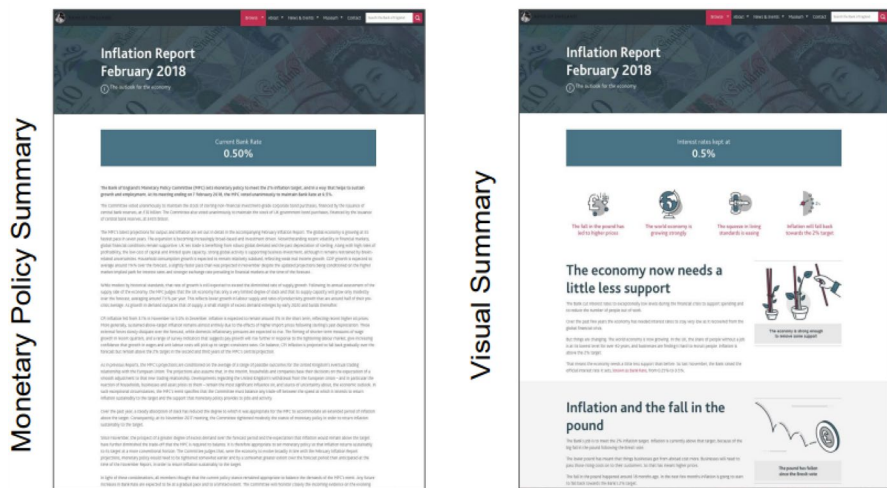
Fig. 8 Control and treatment versions of central bank communication summaries. Source: Bholat et al., (2019)

Bholat et al., (2019) varied the presentation of the Bank of England's real published communication summary on inflation and monetary policy. The control environment was the *actual* published Bank of England summary released to the public and the various treatments modified this summary adding (1) visual features, (2) reduced word-counts, (3) icons or (4) used "more relatable" language. Figure 8 provides an illustration of the control versus one of these treatment conditions.

They find that summaries using relatable language were the best, improving comprehension scores by 25% as well as participants' trust in the information they had read.

Kryvtsov & Petersen (2021) consider how subjects in a learning-to-forecast experiment respond to monetary policy changes as determined by a Taylor rule as part of a New Keynesian model that admits heterogeneous expectations. In their control treatment there is no CB communication, but in three other treatments they vary whether the communication from the CB is about interest changes that have taken place in the immediate past period, or changes that are planned for the immediate future period, or whether the CB provides *forward guidance* that interest rates will not change over some duration of time in the future. They show that communication always reduces forecast errors relative to its absence but that backward-looking CB communication is most effective in reducing subjects' forecast errors, while future-oriented communication is less useful. Interestingly, it is the communication of immediate past actions that helps subject learn the CB's reaction function in a way that leads to better future forecasts.

Cornand & M'baye, (2018) show that communication also interacts well with Taylor rule objectives. In a learning-to-forecast experiment in a New Keynesian model, they show that, if the central bank cares only about inflation stabilization as
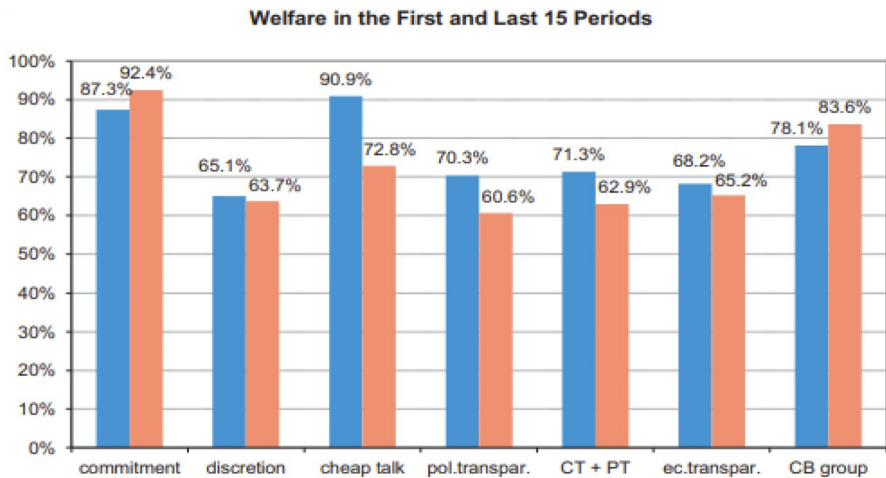
Fig. 5. Welfare in the First and Last 15 Periods.
Note: In CT+PT we compared the averages from the first and last 12 periods.

Fig. 9 Welfare comparisons of different monetary policy regimes using experimental data. Source Duffy & Heinemann, (2021)

in an inflation targeting regime and follows the Taylor principle, then communication of its inflation target does not make a difference in terms of macroeconomic performance. However, if the Taylor rule also conditions on output stabilization as in a dual mandate regime, then communicating the inflation target helps to reduce the volatility of inflation, interest rates and the output gap.

Duffy & Heinemann, (2021) consider whether reputational considerations, cheap talk, policy transparency and economic transparency serve effectively as mechanisms for overcoming CB commitment problems in a repeated monetary policy setting game modeled after (Barro & Gordon, 1983). They compare these various mechanisms with a commitment regime where the CB pre-commits to a monetary policy (which is possible in the laboratory, but may not be so possible in the field!) They find that only the cheap talk regime, without any policy transparency, where the CB promises to follow a certain policy (but can renege on its promise) achieves welfare levels that are close to that achieved under commitment. See Fig. 9

However, as the same figure shows, the welfare gains from this cheap talk regime are not persistent. The private sector eventually learns to discount the CB's promises of low inflation in light of their own actual experience with inflation and as a consequence, welfare declines in the second half of the Cheap Talk treatment. In essence, the finding is that insincere central bank communication can work for awhile, but in the long run it is simply not credible.

4.4 Quantitative easing

Recent worldwide experience with the zero lower bound for nominal interest rates has led most of the world's leading central banks to experiment with large scale open market purchases of riskier, longer-term assets including corporate bonds and asset-backed securities. These purchases have been paid for by crediting the reserve accounts of banks with the intent of lowering interest rates and stimulating lending. This approach stands in contrast with traditional open market operations which swapped risk- free short-term government bonds with bank reserves. However, many central banks now offer interest on bank reserves, a new lever for monetary policy. Further, the interest rates earned on reserves and short-term government bonds are presently so low that they are essentially substitutes for one another and so traditional open market operations may accomplish little. Purchasing riskier long term assets with higher interest rates might thus have some effect on lowering yields at higher maturities. Still, the rationale for the central bank policy of quantitative easing (QE) is not so clear and experimental evidence, even in the small scale of the laboratory, would be useful in understanding this policy choice. In frictionless markets with rational actors, QE should merely restructure the maturity of the government debt held in private hands toward shorter-term assets.

Penalver et al., (2020) propose a behavioral explanation for why QE might work. Their baseline treatment involves traders buying and selling coupon bonds with a fixed maturity date of 11 periods. The interest rate on cash and the dividend from the coupon bonds are set so that, if agents are risk neutral, the equilibrium fundamental value of the bond is constant over time. To this baseline environment, they consider a "buy-and-hold" treatment where the central bank (the experimenter) buys 1/3 of the outstanding bonds via a discriminatory auction prior to period 4 and holds them to maturity. They also consider a buy-and-sell treatment where the central bank repeats the purchase of 1/3 prior to period 4 but then sells the bonds following period 8. They demonstrate that both the buy-and-hold and buy-and-sell purchases by the central bank can have the short-term impact of pushing bond prices above their fundamental value and thereby lowering bond yields. The effect is shown to persist even among highly experienced subjects and provides a nice behavioral rationale for the current practice of quantitative easing by central banks.[9]

4.5 Discussion

In all of these policy experiments, caution is always warranted in extrapolating from the outcomes of experimental studies which often (though not always) use the convenience sample of student subjects to real macroeconomic settings. The real

---

[9] A related policy, of "helicopter money" drops has yet to be considered by central banks (or experimenters) though U.S. Fed Chair Ben Bernanke suggested it as a possibility. This policy differs from quantitative easing in that the CB does not get anything of value in exchange for its printing money (and dropping of money) which would go to private citizens and not to banks and so this policy cannot be unwound.

economy can involve a number of complexities that are not well-approximated by the models studied in the laboratory. However, this critique often applies as well to the theories or models being evaluated. Further, there is increasing evidence that student subjects' behavior is usually a very good approximation to that of the more general public operating in more naturalistic settings. For instance, Cornand & Hubert, (2020) report that the inflation forecast errors made by participants in laboratory learning-to-forecast experiments are similar to those made by households, industry and professional forecasters in survey data, as well as with the implicit inflation expectations from financial market (swap) data. The only inflation forecasts that were found to be less error-prone (more accurate and less systematically biased) and consistent with the full information rational expectation benchmark were those made by central banks! Carbone & Hey (2004) report that laboratory student subjects over-react in terms of consumption expenditures to a change in working status from being unemployed to being employed. They relate this behavior to the phenomenon of excess sensitivity of consumption to changes in income, a violation of the rational expectations/permanent income hypothesis, which is also found in macroeconomic data. Alm et al., (2015) look at tax compliance using a sample of US household tax returns that were subject to government audit and compare tax compliance by student subjects in a laboratory setting where they faced an income reporting task and audit risk that was similar to that of the U.S. households. Both the tax compliance rates and the distribution of those rates for U.S. households and the student laboratory subjects were found to be remarkably similar, suggesting that laboratory studies can be quite useful for informing policymakers about tax policies. More generally, Snowberg & Yariv (2021) find that student subjects behave similarly to members of the general population in a wide variety of individual and strategic tasks but that student subject's choices are often less noisy. Thus, the evidence suggests that we should not discount findings from laboratory studies in thinking about reactions to policy changes. Instead, the laboratory can and should continue to serve as an effective testbed for monetary and other macroeconomic policies before such policies are implemented in the field. The cost of such experiments is cheap, and the benefits can be substantial.

5 Conclusions and suggestions for further research

One of the great achievements of modern macroeconomics is the use of explicit structural and micro-founded models of the behavior of firms, households, governments and other agents that make up the macro-economy. Such microfoundations enable one to quickly assess the value added of various frictions or policy interventions on behavior. But these models are only as good as the behavioral assumptions that underlie them. If agents do not possess rational expectations or cannot solve dynamic optimization problems, then the conclusions derived from such models may not be valid. In this paper, I have suggested how experimenters have tested these modeling assumptions and provided new and alternative approaches that could be incorporated into macro models to make them more behavioral and thus relevant

for policy analysis. I have also shown how experiments are already being used to evaluate or understand the efficacy of various different types of monetary policies.

What lies ahead is always difficult to forecast, but as of this writing, there are several interesting macroeconomic questions that could be addressed using well-designed experiments, both in the lab and/or in the field.

For instance, there has been considerable interest in universal basic income policies and how these might operate relative to the negative income tax and transfer policies currently used by many developed nations. It would be of interest to consider the tax-financing consequences of those different types of income/tax policies as well as their effects on labor market participation using experimental methods.

Another topical question concerns the relevance of "modern monetary theory" (MMT). Is it really the case that, so long as a government's debt is denominated in it's own currency, there is no limit to government borrowing as debt can always be paid for by printing money? Indeed, Japan is often cited as the "poster child" for MMT (NY Times, 2019). But what is the evidence that such government borrowing has little macroeconomic consequences. A well-designed experiment would be useful in addressing this issue.

Also on the horizon is the implementation and acceptance of central bank digital currency (CBDC) as an alternative to cash. In a CBDC system, payments are recorded on a *ledger*, which means that transactions are no longer private. What policies would be necessary to get agents to switch to CBDC in a world where they can always flee to cash (fiat money) or crypto currencies? What are individuals willing to pay for privacy in monetary exchanges?

Finally, there is the question of the public's reaction to negative interest rates. While central banks have acted as though the zero lower bound is a hard constraint, interest rates have been allowed to go negative, for example on excess bank reserves held a the European Central Bank and the Bank of Japan. What would be the impact of negative interest rates more generally on household or firm behavior? How negative would interest rates have to go to stimulate demand, or possibly a flight to competing means of payments?

These are just a few of the many interesting policy questions on the current macroeconomic agenda that experimental evidence could help to answer.

References

Alm, J., Bloomquist, K. M., & McKee, M. (2015). On the external validity of laboratory tax compliance experiments. *Economic Inquiry*, *53*, 1170–1186.

Anderhub, V., Gäuth, W., Mäuller, W., & Strobel, M. (2000). An experimental analysis of intertemporal allocation behavior. *Experimental Economics*, *3*, 137–152.

Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2008). Eliciting risk and time preferences. *Econometrica*, *76*, 583–618.

Andreoni, J., & Sprenger, C. (2012). Estimating time preferences from convex budgets. *American Economic Review*, *102*, 3333–56.

Anufriev, M., & Hommes, C. (2012). Evolutionary selection of individual expectations and aggregate outcomes in asset pricing experiments. *American Economic Journal: Microeconomics*, *4*, 35–64.

Arifovic, J., & Duffy, J. (2018). Heterogeneous agent modeling: experimental evidence. In C. Hommes & B. LeBaron (Eds.), *Handbook of Computational Economics Volume 4,* (pp. 491–540). North-Holland.

Arifovic, J., & Petersen, L. (2017). Stabilizing expectations at the zero lower bound: Experimental evidence. *Journal of Economic Dynamics and Control*, *82*, 21–43.

Assenza, T., Bao, T., Hommes, C., & Massaro, D. (2014). Experiments on expectations in macroeconomics and finance. In J. Duffy (Ed.), *Experiments in macroeconomics* (pp.11–70). Emerald.

Assenza, T., Heemeijer, P., Hommes, C. H., & Massaro, D. (2021). Managing self-organization of expectations through monetary policy: A macro experiment. *Journal of Monetary Economics*, *117*, 170–186.

Augenblick, N., Niederle, M., & Sprenger, C. (2015). Working over time: Dynamic inconsistency in real effort tasks. *The Quarterly Journal of Economics*, *130*, 1067–1115.

Ballinger, T. P., Hudson, E., Karkoviata, L., & Wilcox, N. T. (2011). Saving behavior and cognitive abilities. *Experimental Economics*, *14*, 349–374.

Ballinger, T. P., Palumbo, M. G., & Wilcox, N. T. (2003). Precautionary saving and social learning across generations: An experiment. *The Economic Journal*, *113*, 920–947.

Bao, T., Dai, Y., & Duffy, J. (2021). "Least Squares Learning? Evidence from the Laboratory," Working paper.

Bao, T., & Duffy, J. (2016). Adaptive versus eductive learning: Theory and evidence. *European Economic Review*, *83*, 64–89.

Barro, R. J., & Gordon, D. B. (1983). A positive theory of monetary policy in a natural rate model. *Journal of Political Economy*, *91*, 589–610.

Benhabib, J., Bisin, A., & Schotter, A. (2010). Present-bias, quasi-hyperbolic discounting, and fixed costs. *Games and Economic Behavior*, *69*, 205–223.

Bholat, D., Broughton, N., Ter Meer, J., & Walczak, E. (2019). Enhancing central bank communications using simple and relatable information. *Journal of Monetary Economics*, *108*, 1–15.

Brown, A. L., Chua, Z. E., & Camerer, C. F. (2009). Learning and visceral temptation in dynamic saving experiments. *The Quarterly Journal of Economics*, *124*, 197–231.

Brown, J. R., & Previtero, A. (2014). "Procrastination, present-biased preferences, and financial behaviors", *Unpublished Manuscript, University of Illinois at Urbana-Champaign and University of Western Ontario*.

Caliendo, F., & Aadland, D. (2007). Short-term planning and the life-cycle consumption puzzle. *Journal of Economic Dynamics and Control*, *31*, 1392–1415.

Campbell, J. Y., & Mankiw, N. G. (1989). Consumption, income, and interest rates: Reinterpreting the time series evidence. *NBER Macroeconomics Annual*, *4*, 185–216.

Carbone, E. (2006). Understanding intertemporal choices. *Applied Economics*, *38*, 889–898.

Carbone, E., & Duffy, J. (2014). Lifecycle consumption plans, social learning and external habits: Experimental evidence. *Journal of Economic Behavior & Organization*, *106*, 413–427.

Carbone, E., & Hey, J. D. (2004). The effect of unemployment on consumption: An experimental analysis. *The Economic Journal*, *114*, 660–683.

Cass, D. (1965). Optimum growth in an aggregative model of capital accumulation. *The Review of Economic Studies*, *32*, 233–240.

Choi, J. J., Laibson, D., Madrian, B. C., & Wise, D. A. (2004). For better or for worse: Default effects and 401 (k) savings behavior. In D. A. Wise (Ed.), *Perspectives on the Economics of Aging* (pp. 81–121). University of Chicago Press.

Christiano, L. J., Eichenbaum, M. S., & Trabandt, M. (2017). "On DSGE Models," November 27, 2017 working paper version downloaded from https://faculty.wcas.northwestern.edu/yona/research/DSGE.pdf.

Coibion, O., Gorodnichenko, Y., Kumar, S., & Pedemonte, M. (2020). Inflation expectations as a policy tool? *Journal of International Economics*, *124*, 103297.

Cornand, C., & Hubert, P. (2020). On the external validity of experimental inflation forecasts: A comparison with five categories of field expectations. *Journal of Economic Dynamics and Control*, *110*, 103746.

Cornand, C., & M'baye, C. K. (2018). Does inflation targeting matter? An experimental investigation. *Macroeconomic Dynamics*, *22*, 362–401.

Diamond, J., Watanabe, K., & Watanabe, T. (2020). The formation of consumer inflation expectations: New evidence from Japan's deflation experience. *International Economic Review*, *61*, 241–281.

Duesenberry, J. S. (1949). *Income, saving, and the theory of consumer behavior*. Harvard University Press.

Duffy, J. (2010). "Experimental macroeconomics". *New Palgrave Dictionary of Economics, 2nd Edition.* Reprinted in S.N. Durlauf and L.E. Blume (Eds.) *Behavioural and Experimental Economics* (pp. 113–119). Macmillan

Duffy, J. (2016). Macroeconomics: A survey of laboratory research. In J.H. Kagel & A.E. Roth (Eds.), *Handbook of experimental economics Volume 2*, (pp. 1–90). Princeton University Press.

Duffy, J., & Heinemann, F. (2021). Central bank reputation, cheap talk and transparency as substitutes for commitment: Experimental evidence. *Journal of Monetary Economics*, *117*, 887–903.

Duffy, J., & Jenkins, B. (2019). A classroom experiment in monetary policy. *The Journal of Economic Education*, *50*, 89–107.

Duffy, J., & Li, Y. (2019). Lifecycle consumption under different income profiles: Evidence and theory. *Journal of Economic Dynamics and Control*, *104*, 74–94.

Duffy, J., & Li, Y. (2021). "Do Tax Deferred Accounts Improve Lifecycle Savings? Experimental Evidence," Working Paper, University of Albany.

Duffy, J., & Puzzello, D. (2021). "The Friedman rule: Experimental evidence," Working paper, Indiana University.

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

Evans, G. W., & Honkapohja, S. (2009). Learning and macroeconomics. *Annual Review of Economics*, *1*, 421–449.

Falk, A., & Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, *326*, 535–538.

Friedman, M. (1969). *The Optimum Quantity of Money*, Adline Publishing.

Fuchs-Schündeln, N., & Hassan, T. A. (2016). Natural experiments in macroeconomics. In J. B. Taylor & H. Uhlig (Eds.), Handbook of Macroeconomics Volume 2 (pp. 923–1012). North-Holland

Gabaix, X. (2016). "Behavioral macroeconomics via sparse dynamic programming," National Bureau of Economic Research Working Paper No. 21848.

Hey, J. D., & Dardanoni, V. (1988). Optimal consumption under uncertainty: An experimental investigation. *The Economic Journal*, *98*, 105–116.

Hommes, C., Massaro, D., & Weber, M. (2019). Monetary policy under behavioral expectations: Theory and experiment. *European Economic Review*, *118*, 193–212.

Hooper, P., Mishkin, F. S., & Sufi, A. (2020). Prospects for inflation in a high pressure economy: Is the Phillips curve dead or is it just hibernating? *Research in Economics*, *74*, 26–62.

Jackson, M. O., & Yariv, L. (2014). Present bias and collective dynamic choice in the lab. *American Economic Review*, *104*, 4184–4204.

Koopmans, T. C. (1965). On the concept of optimal economic growth. *The Econometric Approach to Development Planning*, North-Holland, 225–287.

Kryvtsov, O., & Petersen, L. (2021). Central bank communication that works: Lessons from lab experiments. *Journal of Monetary Economics*, *117*, 760–780.

Lagos, R., & Wright, R. (2005). A unified framework for monetary theory and policy analysis. *Journal of Political Economy*, *113*, 463–484.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, *112*, 443–478.

Levy, M., & Tasoff, J. (2016). Exponential-growth bias and lifecycle consumption. *Journal of the European Economic Association*, *14*, 545–583.

Lucas, R. E. J. (1980). Methods and problems in business cycle theory. *Journal of Money, Credit and Banking*, *12*, 696–715.

Lucas, R. E, Jr. (1972). Expectations and the neutrality of money. *Journal of Economic Theory*, *4*, 103–124.

Matějka, F., & McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, *105*, 272–98.

Mauersberger, F. (2021). Monetary policy rules in a non-rational world: A macroeconomic experiment. *Journal of Economic Theory*, in press, 105203.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica, 29,* 315–335.

Nerlove, M. (1958). Adaptive expectations and cobweb phenomena. *Quarterly Journal of Economics*, *72*, 227–240.

O'Donoghue, T., & Rabin, M. (1999). Doing it now or later. *American Economic Review*, *89*, 103–124.

Penalver, A., Hanaki, N., Akiyama, E., Funaki, Y., & Ishikawa, R. (2020). A quantitative easing experiment. *Journal of Economic Dynamics and Control*, *119*, 103978.

Pfajfar, D., & Žakelj, B. (2018). Inflation expectations and monetary policy design: Evidence from the laboratory. *Macroeconomic Dynamics*, *22*, 1035–1075.

Ramsey, F. P. (1928). A mathematical theory of saving. *The Economic Journal*, *38*, 543–559.

Sargent, T. J. (1993). *Bounded rationality in macroeconomics: The Arne Ryde memorial lectures*. Oxford University Press.

Sargent, T. J. (1999). *The conquest of American inflation*. Princeton University Press.

Sargent, T. J., & Wallace, N. (1975). "Rational" expectations, the optimal monetary instrument, and the optimal money supply rule. *Journal of Political Economy*, *83*, 241–254.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica, 48,* 1–48.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, *50*, 665–690.

Snowberg, E., & Yariv, L. (2021). Testing the waters: Behavior across participant pools. *American Economic Review*, *111*, 687–719.

Solow, R. M. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics*, *70*, 65–94.

Taylor, J. B. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester conference series on public policy, 39*, 195–214.

Thaler, R. H., & Benartzi, S. (2004). Save more tomorrow™: Using behavioral economics to increase employee saving. *Journal of Political Economy*, *112*, S164–S187.

Veblen, T. (1899). *The theory of the leisure class: An economic study of institutions*. Macmillan.