The Demand for Bad Policy when Voters Underappreciate Equilibrium Effects

Ernesto Dal Bó — UC Berkeley
Pedro Dal Bó — Brown University
Erik Eyster* — LSE

July 8, 2015

Abstract

The political economy literature provides several compelling explanations for inefficient policies and institutions that abstract from any responsibility on the part of voters. In this paper, we explore one way in which voters may also be to blame, by studying whether and why they may systematically err by demanding certain forms of bad policy. We show that a majority of subjects in an experiment vote against policies that would help them overcome social dilemmas. This is due to their failure to fully anticipate the equilibrium effects of new policies. More precisely, subjects systematically underappreciate the extent to which policy changes affect other people's behavior, and this results in increased demand for bad policy. In addition, we find that one-third of subjects do not appreciate how their own behavior will adapt to the new policy. The overall implication is that, in regimes where voter preferences affect policy selection, the underappreciation of equilibrium effects by voters could adversely affect the policies that get adopted.

JEL codes: C9, D7.

Keywords: voting, reform, political failure, endogenous policy, experiment.

*We thank Daniel Prinz and Santiago Truffa for excellent research assistance as well as Berkeley's XLab and Brown's BUSSEL for support. We are grateful to Anna Aizer, Ned Augenblick, Eric Dickson, Willie Fuchs, Alessandro Lizzeri, Matthew Rabin, Francesco Trebbi and Georg Weizsäcker for helpful discussions. We thank participants at various conferences and seminars for their comments and suggestions. Eyster thanks the ERC for financial support.

1 Introduction

Standard explanations for why bad policies are implemented when good policies are available typically point to the policy production process, including agency problems, incompetent policy-makers, and institutional failure to efficiently resolve competitive tensions.[1] In this paper, we shift the focus away from policy-makers and institutions, by providing experimental evidence that some of the blame for bad policies may lie with voters. Although it is standard in the literature to assume that voters correctly assess the merits of the options they face, we show with a simple experiment that people demand bad policies because they fail to correctly predict how new policies change equilibrium behavior and welfare. In our experimental setting, this prevents groups from resolving social dilemmas through democratic means.

Some economists have previously suggested that voters may not fully appreciate the effects of policies. Adam Smith emphasized the limited grasp of the implications of market equilibrium by the general public (see Smith 1776). North (1990) surmised that voters might misperceive the relative merits of different policies and institutions and, hence, demand suboptimal policy. More recently, Beilharz and Gersbach (2004) argued that unawareness of general equilibrium effects may lead citizens to support minimum wages above market ones. Caplan (2007) documents systematic divergences in the views of voters relative to those of experts on the benefits of, for instance, markets and foreign trade. In political science, the field of political behavior has long studied patterns of voter opinion with conflicting results

[1] Agency problems include discretion under limited electoral accountability (e.g., Barro 1973, Ferejohn 1986) and capture (e.g., Stigler 1971, Peltzman 1976, and Coate and Morris 1995). For accounts of why inept people may self-select into policymaking, see Caselli and Morelli (2004), Messner and Polborn (2004), Besley (2005), and Dal Bó, Dal Bó and Di Tella (2006) among others. Institutional failures to efficiently resolve collective disagreements may take the form of status quo bias (e.g., Romer and Rosenthal 1978), delay to reform (e.g., Alesina and Drazen 1991, and Fernandez and Rodrik 1991), and dynamic inefficiency due to the threat of losing political control (e.g., Alesina and Tabellini 1990, De Figueiredo 2002, and Besley and Coate 2007).

on the quality of those opinions (see Bartels 2012 for a review).

However, there are two challenges to establishing whether voters demand bad policies, and why. One is empirical, and the other conceptual. The empirical challenge is that policy in the real world is complex enough that it is difficult to demonstrate conclusively that voter opinion is wrong simply because it differs from expert opinion. Voters may value dimensions not valued by experts (and on occasion, experts may be wrong). We address this empirical challenge using experimental methods. In an experiment, we can cleanly identify the welfare effects of different policies and study whether voters choose policies that are bad (i.e., Pareto dominated). The conceptual challenge is to understand what aspects of policy pose problems for voters and how they can lead to the demand for bad policies. We address this challenge by presenting a simple conceptual framework, tightly related to the experimental design, which shows how the underappreciation of the equilibrium effects of policies can explain the demand for bad ones.

We present data from an experiment in which subjects choose whether to participate in a Prisoners' Dilemma game or an alternative game, which we call the Harmony Game. The Harmony Game derives from the Prisoners' Dilemma by taxing both cooperation and defection, the latter more than the former. These asymmetric taxes make cooperation a dominant strategy in the Harmony Game leading to mutual cooperation in equilibrium, which yields higher payoffs than the equilibrium of mutual defection in the Prisoners' Dilemma. In our experiment, subjects choose between the two games following several rounds of experience playing just one of the games. Although subjects on average earn significantly higher payoffs in the Harmony Game than in the Prisoners' Dilemma, a majority choose to play the Prisoners' Dilemma. This result does not depend upon the voting institution used nor upon which game subjects played before voting. The fact that most subjects choose to play the Prisoners' Dilemma game provides evidence that, under certain circumstances, voters demand bad

policies.

When and why do voters choose the wrong game? Since they lack experience with one of the games at the time of voting, subjects must forecast how other players will behave in that game. Voting for the Prisoners' Dilemma can only be justified if the voter underappreciates how differently the other players will behave in the two games. For example, a subject who believes that other players will cooperate as frequently in the Prisoners' Dilemma as in the Harmony Game will prefer to play the Prisoners' Dilemma. By eliciting subjects' beliefs about how their opponents will play the two games, we find that on average subjects underestimate the effect that the change in game will have on the behavior of other subjects. Moreover, those subjects who most underappreciate the difference in cooperation rates are most likely to vote for the Prisoners' Dilemma. This evidence is consistent with the demand for bad policy arising from a lack of understanding of the equilibrium effect of the game change.

In addition to the experimental evidence, we provide a simple conceptual framework to show how underappreciation of equilibrium effects can lead to the demand for bad policies. Our framework predicts voting failures when people tend to focus on the direct impact of policies – the impact on payoffs before changes in behavior are considered – over their indirect impact via equilibrium effects. This can bias demand towards policies with direct benefits over policies with direct costs, even when the latter may generate greater welfare for the voters. As an example, consider regulations proposed in the US to reduce CO_2 emissions, motivated by concerns about climate change. Experts recommend a Pigouvian tax on CO_2 emissions, but this policy is widely dismissed as politically infeasible due to a "reluctance of policymakers to adopt Pigouvian taxes that would affect petroleum consumption" (Knittel 2012, p. 111). Instead, Congress has enacted less efficient alternatives such as miles-per-gallon (MPG) regulations. But research has found that a Pigouvian tax could achieve the

same goals of MPG regulations at a fraction of the social cost (from 1/6 to 1/10 depending on the study, see Knittel 2012). MPG regulations are suboptimal compared to a carbon tax for at least two reasons: they impose significant development costs, and are partially undermined by "rebound effects", where the decreased fuel-cost of driving encourages people to drive more. Our framework and results on the underappreciation of equilibrium effects offer an explanation for the choice of inefficient policy. Policymakers may be reluctant to adopt Pigouvian taxes because they are unappealing to voters: their costs are direct (higher prices for gas) while their benefits are indirect (when people reduce their driving). In contrast, for MPG regulations the benefits are direct (increased efficiency given current driving) and the costs are indirect (e.g., rebound effects).

As with all empirical work, it is important to consider external validity. The fact that these experiments involve students from elite universities making choices in very simple environments suggests that, if anything, our results may understate the extent to which the average citizen underappreciates equilibrium effects in the more complex real-world policy realm. Of course, politicians or the media could mitigate voters' lack of understanding of equilibrium effects outside of the laboratory. But as Blinder and Krueger (2004, p.328) emphasize, even on matters admitting a technical answer "*the decisions of elected politicians are heavily influenced by public opinion.*" By focusing on the rigors of electoral discipline, most formal theories of electoral politics view politicians not as educators but instead as panderers to voters' policy positions, even wrong ones (Harrington 1993, Canes-Wrone, Herron and Shotts 2001, and Maskin and Tirole 2004). For-profit media may also pander rather than educate, since they have incentives to bias reporting to match consumers' priors (see Gentzkow and Shapiro 2006 for a theoretical argument, and Gentzkow and Shapiro 2010 for evidence). For these reasons, it is important to understand any potential problems in voters' demand for policies. Our experiments demonstrate the existence of problems and

help understand their causes.

The remainder of the paper is organized as follows. In Section 2, we situate our work in the literatures on behavioral political economy and failures in strategic reasoning. In Section 3, we introduce a framework for analyzing people who underappreciate equilibrium effects. We describe our experiments in Section 4, and our hypotheses in Section 5. In Section 6, we report data from the experiments, explaining how they demonstrate that people underappreciate the extent to which others react to policy changes. In addition, using a structural model, we estimate that one-third of subjects fail to appreciate how they themselves will react to a policy change. We conclude in Section 7.

2 Related Literature

This paper relates to the emerging political economy literature incorporating behavioral aspects. Examples include the study of the impact of cognitive dissonance on voting (Mullainathan and Washington 2009), the analysis of collective action with time-inconsistent voters (Bisin, Lizzeri and Yariv 2015, and Lizzeri and Yariv 2012), and the behavior of voters who fail to extract the right information from other voters' actions (Eyster and Rabin 2005, Esponda and Pouzo 2010, and Esponda and Vespa 2012).

This paper also relates to a growing experimental literature studying the choice of self-regulatory institutions (see Dal Bó 2014 for a survey). A few findings in that literature are worth highlighting here. Walker, Gardner, Herr and Ostrom (2000) study common-pool problems where players who would benefit from reduced extraction vote on extraction rules. They find that some voters propose inefficient extraction rules, and that the voting protocol affects efficiency. Magreiter, Sutter, and Dittrich (2005) show that subject heterogeneity exacerbates inefficiency. Dal Bó (2014) offers evidence consistent with the idea that a better

understanding of the strategic situation affects people's ability to select efficient institutions. Kallberkken, Kroll and Cherry (2011) use laboratory experiments to study people's attitudes towards Pigouvian taxes. Not only does their work focus on different issues than ours (the effects of using the term "tax" versus "fees", of education, and of the distribution of tax revenue), but, more importantly, their experiments cannot shed light on whether the non-imposition of taxes derives from a failure to understand equilibrium effects: all five types of voters predicted to benefit from imposing the tax *in equilibrium* would also (weakly) benefit from imposing the taxes *fixing others' behavior* (and strictly benefit in four of those five cases). Closer to this paper, Sausgruber and Tyran (2005, 2011) provide experimental evidence which suggests that people may not understand tax incidence: in a market where buyers bear all the burden of a tax, they prefer to impose a larger tax on sellers than a smaller tax on themselves. Because sellers' behavior is mechanized in the experiments, which buyers understand, it is unclear whether buyers choose the wrong tax because they misunderstand how sellers react to taxes or due to some simple aversion to paying tax of the form studied by Kallberkken, Kroll and Cherry (2011). Finally, Dal Bó, Foster and Putterman (2010), in their study of the direct effect of democracy, find that 46% of subjects prefer to play a prisoners' dilemma game over a coordination game derived from the prisoners' dilemma by taxing unilateral defections. This is not evidence that subjects do not understand equilibrium effects as the coordination game has two pure-strategy equilibria, only one of which results in higher equilibrium payoffs than under the prisoners' dilemma. In this paper we focus on demonstrating the connection between the underappreciation of equilibrium effects and the demand for bad policy, both by grounding the experiment in a conceptual framework and by choosing a design that can eliminate potential confounds challenging the identification of the effects of interest. One advantage of our design is it minimizes computational complications by facing subjects with the simplest possible equilibrium change, involving 2x2 games with

dominant strategies. In addition, as we will explain after presenting our main results, we rely on a variety of treatments to eliminate alternative explanations.

At an abstract level, our paper is related to the experimental literature documenting failures of backwards induction (e.g., McKelvey and Palfrey 1992, Bone, Hey and Suckling 2009, Palacios-Huerta and Volij 2009, Levitt, List and Sadoff 2011, and Moinas and Pouget 2013). We establish a systematic direction of departure from subgame perfection in a specific class of games: players underestimate how differently their opponents play across subgames with shared action spaces but different payoffs. This error resembles that embodied in Jehiel's (2005) Analogy-Based-Expectations Equilibrium (ABEE), where players think their opponents' play is constant across certain decision nodes. Despite the resemblance, Section 6.5 explains how subjects in our experiment exhibit a degree of partial sophistication about the relationship between the game and their opponents' actions inconsistent with ABEE.

3 Underappreciation of equilibrium effects and preferences over policies: conceptual framework

In this section, we present a simple conceptual framework to define the meaning and types of "underappreciation of equilibrium effects" and discuss how they affect voters' preferences over policies.

Consider an agent who will participate in a two-player Prisoners' Dilemma game – see left panel in Table 1. Participants in this game must choose between cooperate (C) or defect (D). Cooperation results in a cost to the agent of c and a benefit b to the other participant, with $b > c > 0$.[2] Given that $c > 0$, it is a dominant strategy to defect, and the Nash equilibrium of

[2]This specification of a Prisoners' Dilemma game has the property that a players' gain from defection is independent of the action of the other player. As discussed in Fudenberg, Rand and Dreber (2012), not every Prisoners' Dilemma game can be described this way. Regardless of the lack of generality, we follow

Table 1: The Games

Prisoners' Dilemma

	C	D
C	$b-c, b-c$	$-c, b$
D	$b, -c$	$0, 0$

Harmony Game

	C	D
C	$b-c-t_C, b-c-t_C$	$-c-t_C, b-t_D$
D	$b-t_D, -c-t_C$	$-t_D, -t_D$

this game is that both participants defect leading to (D, D) with payoffs $(0, 0)$; since (C, C) gives both players $b - c > 0$, the equilibrium outcome (D, D) is inefficient.

Imagine that the agent is now considering a policy proposal that would impose, on each player, taxes t_C on cooperation and t_D on defection. These taxes would modify the game to the one in the right panel in Table 1. Assume that these taxes are such that the following condition holds: $b > t_D > t_C + c$. Given the values of the taxes, the dominant strategy now is to cooperate, leading to an efficient equilibrium: (C, C) with payoffs $(b-c-t_C, b-c-t_C)$. Since there is no tension between personal incentives and group objectives in this game, we call it the "Harmony Game."

An agent who anticipates equilibrium behavior in both games will prefer to impose the taxes and play the Harmony Game, given that the equilibrium payoff in the Harmony Game is greater than the equilibrium payoff in the Prisoners' Dilemma game: $b - c - t_C > 0$. Therefore, if all agents think this way, they will vote for the imposition of taxes and the social dilemma will be resolved.

However, voters may not correctly appreciate how changes in policy will affect behavior and, as a result, may not demand the correct policies: they may demand the Prisoners' Dilemma, not the Harmony Game. In this section, we offer a taxonomy of the types of mistakes that voters may make when thinking about the effect of policies, and show how these mistakes can affect voters' demand for policy.

this description as it allows for a simpler analysis and includes the Prisoners' Dilemma game used in the experiment.

Assume that instead of holding equilibrium beliefs about behavior, voters may have any belief about their own behavior and the behavior of others in the two games. More precisely, assume that a voter believes that she will cooperate with probability α in the Prisoners' Dilemma and probability α' in the Harmony Game while believing that the other player will cooperate with probability β in the Prisoners' Dilemma and probability β' in the Harmony Game. Note that we do not attempt to explain here the origin of these beliefs. The goal is to understand how these beliefs affect preferences for the two games. Given her beliefs, the voter's preferences over the two games will depend on the difference in expected payoff between the two games. The expected gain from moving to the Harmony Game is $G(\alpha, \alpha', \beta, \beta') = EU(HG|\alpha', \beta') - EU(PD|\alpha, \beta)$, where $EU(HG|\alpha', \beta')$ is the expected payoff under the Harmony Game and $EU(PD|\alpha, \beta)$ is the expected payoff under the Prisoners' Dilemma game. The voter would prefer the Prisoners' Dilemma if her beliefs are such that the expected gain of imposing taxes is negative.

The expected gain can be decomposed into three terms. The first term is the direct effect of the game change: $DE = EU(HG|\alpha, \beta) - EU(PD|\alpha, \beta)$. The direct effect DE captures the change in expected payoff from going from the Prisoners' Dilemma to the Harmony Game assuming that the behavior of both players is not affected by the game change. The second term is the indirect effect due to the adjustment in behavior by self: $IS = EU(HG|\alpha', \beta) - EU(HG|\alpha, \beta)$. This indirect effect captures the change in expected payoffs due to the adjustment in one's own behavior while leaving the behavior of the other player unchanged. The third effect is the indirect effect due to the adjustment in behavior by the other player: $IO = EU(HG|\alpha', \beta') - EU(HG|\alpha', \beta)$. This indirect effect captures the change in expected payoffs due to the change in the behavior of others while leaving the behavior of the self constant at the new level (α'). Note that these three effects add up to

the total expected gain from a change in game: $G = DE + IS + IO$.[3]

For our two games, this decomposition can be simply expressed in terms of payoff parameters, taxes and beliefs. The direct effect is the expected payment of taxes under the belief that behavior will be as in the Prisoners' Dilemma: $DE = -(\alpha t_C + (1-\alpha)t_D)$. This effect is negative and its magnitude is decreasing in α. The indirect effect from the adjustment by self (IS) equals the change in the probability of cooperation by self times the amount saved by cooperating (the tax to defection minus the cost of cooperation and the tax to cooperation): $IS = (\alpha' - \alpha)(t_D - c - t_C)$. This effect is increasing in the believed change in cooperation by self $(\alpha' - \alpha)$. Finally, the indirect effect from the adjustment by other (IO) equals the change in the probability of cooperation by other $(\beta' - \beta)$ times the benefit from the other's cooperation (b): $IO = (\beta' - \beta)b$.

As an important benchmark, we can easily calculate the value of these effects if the player has Nash equilibrium expectations: $\alpha = \beta = 0$ and $\alpha' = \beta' = 1$: $DE^{NE} = -t_D$, $IS^{NE} = t_D - c - t_C$, and $IO^{NE} = b$. The total gain in equilibrium is $G^{NE} = b - c - t_C$ which is greater than zero given the assumptions on payoffs and taxes, so a voter holding equilibrium beliefs would prefer the Harmony Game.

We can compare these equilibrium effects on payoffs with those perceived by a voter who does not hold equilibrium beliefs. For a voter who does not hold equilibrium beliefs, the underappreciation of the indirect effect on payoffs due to the adjustment by self is proportional to the underappreciation of how much self will adjust behavior in equilibrium: $IS = (\alpha' - \alpha)IS^{NE} \leq IS^{NE}$. Similarly, a voter who does not hold equilibrium beliefs underappreciates the indirect effect on payoffs due to the adjustment by others in proportion to the underappreciation of how much the other player will adjust behavior in equilibrium:

[3]An alternative decomposition would consider first a change in the behavior of others and then a change in own behavior. As will be clear later, such a decomposition is equivalent to the one defined above for the class of games considered in this section.

$IO = (\beta' - \beta)IO^{NE} \leq IO^{NE}$.[4]

The following proposition establishes that there is a particular relationship between the underappreciation of equilibrium effects and a preference for the bad policy – the Prisoners' Dilemma game.

Proposition 1 *A voter has a preference for the Prisoners' Dilemma over the Harmony Game if and only if she sufficiently underappreciates the indirect effect due to the adjustment in behavior by the other player (IO) relative to equilibrium.*

Proof. A voter has a preference for the Prisoners' Dilemma over the Harmony Game if and only if her perceived gains from moving to the Harmony Game are negative:

$$G = D + IS + IO = -(\alpha t_C + (1-\alpha)t_D) + (\alpha' - \alpha)(t_D - c - t_C) + (\beta' - \beta)b < 0.$$

This condition holds if and only if $\beta' - \beta < \alpha\frac{t_C}{b} + (1-\alpha)\frac{t_D}{b} - (\alpha' - \alpha)\frac{t_D - c - t_C}{b}$. Given that $b > t_D > t_C + c$, the right hand side of the previous inequality attains a maximum of $\frac{t_D}{b} < 1$ with $\alpha = \alpha' = 0$ and a minimum of $\frac{t_C}{b} < 1$ with $\alpha = \alpha' = 1$. It follows that the voter with a preference for the Prisoners' Dilemma game must have $\beta' - \beta < \frac{t_D}{b}$ and that a voter with $\beta' - \beta < \frac{t_C}{b}$ must have a preference for the Prisoners' Dilemma. Therefore, a voter has a preference for the Prisoners' Dilemma if and only if $\beta' - \beta$ is sufficiently small (with how small depending on the direction of the implication). As $IO = (\beta' - \beta)b$ the proposition follows. ■

This proposition is developed for the case when the voter contemplates a move from the Prisoners' Dilemma to the Harmony Game, but it holds also for the reverse move from the Harmony Game to the Prisoners' Dilemma. In sum, in this section we study a class of

[4]To be clear, a person who mistakes the sign of equilibrium effects (e.g., $\alpha = \beta = 1$ and $\alpha' = \beta' = 0$) is also said to underappreciate equilibrium effects.

environments in which a voter can choose between two games. We showed that in this class of environments, the voter will have a preference for the Prisoners' Dilemma game if and only if she sufficiently underappreciates how the change in game will affect the behavior of the other player relative to equilibrium. Of course, for the equilibrium prediction to be a relevant benchmark, actual behavior must resemble equilibrium behavior. In the next section we explain the experimental design in which we bring the environment to the laboratory to study whether actual behavior matches equilibrium behavior, whether subjects underappreciate the indirect effect due to the adjustment of others and whether this affects subjects' preferences between games.

4 The Experiment

The experiment brings to the laboratory the environment studied in the previous section with a particular choice of parameters: $b = 6$, $c = 2$, $t_C = 1$ and $t_D = 4$ over a baseline payoff of 5.[5] This combination of parameters results in the payoff matrices in Table 2. The actions were labeled "1" and "2" in the experiment to maintain neutrality. The exchange rate was \$1 per 3 experimental points.

There were six experimental treatments. We begin by explaining the basic structure of the experimental sessions in all six treatments and then describe the differences across the treatments.

In Part 1 of the experiment, we divided subjects into groups of six. Each subject played against every other one in the group exactly once, resulting in five periods of (one-shot) play in this part of the experiment. The game played in Part 1 varied by treatment. In some

[5]There are two reasons why we considered taxes on both actions. One is realism: Pigouvian taxes under the socially optimal action are not necessarily zero. For example, that would be the case if CO_2 emissions were taxed and the socially optimal amount of emissions were positive. Second, we wanted to eliminate any unnecessary difference between the two actions that would arise from taxing only D.

Table 2: The Prisoners' Dilemma and Harmony Games

Prisoners' Dilemma (PD)

	C	D
C	9, 9	3, 11
D	11, 3	5, 5

Harmony Game (HG)

	C	D
C	8, 8	2, 7
D	7, 2	1, 1

treatments all groups played the Prisoners' Dilemma in Table 2 while in other treatments they played the Harmony Game in Table 2. All groups in a session were part of the same treatment.

After Part 1, new groups of six were formed randomly for Part 2, which included another five periods of play (6 to 10). At the beginning of Part 2, the game to be played in the next five periods was chosen. One of the main treatment variables is the way in which this choice was made, as described below. After the choice of game for Part 2, but before Period 6, subjects reported their beliefs about how a randomly selected opponent in a similar experiment would act in each of the two games.[6] As in periods 1 to 5 in Part 1, in Periods 6 to 10 every subject faced each other subject in the group exactly once. Subjects were paid for their earnings in all ten periods in parts 1 and 2.

Table 3: Experimental Design - Treatments

	Part 1	Part 2		
Treatment	Game	Game	Game Choice Institution	Game Choice Before
Control	PD	PD or HG	Random	Period 6
Reverse Control	HG	PD or HG	Random	Period 6
Random Dictator (RD)	PD	PD or HG	Random Dictator	Period 6
Reverse RD	HG	PD or HG	Random Dictator	Period 6
Majority Once	PD	PD or HG	Simple Majority	Period 6
Majority Repeated	PD	PD or HG	Simple Majority	Periods 6 to 10

The two treatment variables are the game that subjects played in Part 1 and the mech-

[6]The elicitation of beliefs was incentivized using a mechanism that gathers subjects' beliefs independently of their risk attitudes, as in Karni (2009). See also Grether (1992), Holt (2007, pages 384-385) and Möbius, Niederle, Niehaus and Rosenblat (2014).

anism used to choose the game for Part 2. The treatment arms labeled Control, Random Dictator, Majority and Majority Once had the subjects play the Prisoners' Dilemma game in Part 1, while Reverse Control and Reverse Random Dictator had the subjects play the Harmony Game in Part 1. We included treatments with different games in Part 1 to make sure that any demand for the Prisoners' Dilemma is not due to status quo bias. Table 3 summarizes the experimental design.

In the control treatments (Control and Reverse Control), the game for Part 2 of the experiment was chosen at random by the computer. This choice was made once at the beginning of Part 2, and applied for all players in a group and all periods (i.e., all subjects in a given group played the same game in all periods in Part 2). These treatments allow us to incentivize the belief elicitation in the other treatments. Together with the other treatments, they also allow us to compare behavior and payoffs between the two games and corroborate that subjects would be better off in the Harmony Game than in the Prisoners' Dilemma game as theory predicts.

The treatments Random Dictator and Reverse Random Dictator differed from the controls by asking all subjects to choose between the two games at the beginning of Part 2 and then implementing for the group the choice of a randomly selected subject. In the Majority Once treatment, the game chosen by the majority of the group before Period 6 was implemented for all periods in Part 2. Ties were randomly broken by the computer with equal probability. In the Majority Repeated treatment, subjects voted for a game before each period of Part 2. In this treatment, beliefs were not elicited so as not to affect voting behavior in future periods. In all the other treatments, the belief elicitation always occurred after voting, so as not to affect voting. Subjects were informed of the implemented game and not the voting distribution.

Both random dictator and majority institutions were considered to make sure that the

Table 4: Number of Subjects by Treatment and Place

	Berkeley	Brown	Total
Control	60	60	120
Reverse Control	60	60	120
Random Dictator	84	84	168
Reverse Random Dictator	60	60	120
Majority Once	60	60	120
Majority Repeated	60	60	120
Total	384	384	768

choice of subjects was robust to the voting institution. The majority repeated treatment was included to study the evolution of game choices by subjects.

During play, subjects were shown the payoff function corresponding to the game they were playing. This information was shown as a table with a row for each possible outcome of the game as shown in the slides in the appendix. Subjects knew the game was symmetric, so this representation carries the same information as the normal-form representation shown in Table 2.[7] At the time of voting, subjects were shown both tables side by side to facilitate comparison between the games. Moreover, since they did not face a time-limit to vote, participants had ample time to think about the two different games.

At the end of the experiment, subjects played a p-beauty contest (Nagel 1995) to assess their strategic sophistication in simultaneous-move games and filled out a questionnaire providing basic demographics (gender, political ideology, class, major and SAT scores).

We recruited 384 student subjects from UC Berkeley and 384 from Brown University to participate in the experiment. Table 4 shows the number of subjects from each university in each of the six treatments. Sessions lasted around half an hour and earnings ranged from \$16.75 to \$37 with an average of \$27.81 (earnings included a \$5 show-up fee). Appendix Table 10 displays summary statistics of demographics and beliefs.

[7]We find no evidence that this representation affected behavior as the levels and evolution of cooperation was consistent with those found in the literature.

5 Hypotheses

With the particular values of the parameters used in the experiment, we can now revisit the conceptual framework so as to provide the precise hypotheses that we test with the experimental data.

A voter who expects equilibrium behavior in both games will expect a gain of $G^{NE} = 8 - 5 = 3$ from moving to the Harmony Game. This gain can be decomposed into the three effects discussed in Section 3. The direct effect in equilibrium is just the reduction in payoff due to the move from PD to HG leaving the outcome (D, D) constant: $DE^{NE} = -4$. The indirect effect due to the adjustment by self is the increase in payoff due to the move from (D, D) to (C, D) in HG: $IS^{NE} = 2-1 = 1$. The indirect effect due to the adjustment by other is the increase in payoff due to the move from (C, D) to (C, C) in HG: $IO^{NE} = 8 - 2 = 6$.[8]

However, voters may have beliefs about their own behavior and the behavior of others in the two games that differ from equilibrium beliefs. As in Section 3, assume that a voter believes that she will cooperate with probability α in the Prisoners' Dilemma and probability α' in the Harmony Game while believing that the other player will cooperate with probability β in the Prisoners' Dilemma and probability β' in the Harmony Game. Given these beliefs, the advantage of moving from the Prisoners' Dilemma to the Harmony Game is:

$$G = EU(HG|\alpha', \beta') - EU(PD|\alpha, \beta) = -4 + \alpha - 2\alpha' + 6(\beta' - \beta).$$

Proposition 1 establishes that a voter has a preference for the Prisoners' Dilemma if and only if she sufficiently underappreciates the indirect effect associated with the adjustment of

[8]Similarly, a voter in a "reverse" treatment who expects equilibrium behavior in both games will expect a gain from moving from HG to PD equal to $G^{NE} = 5 - 8 = -3$. This total gain can again be decomposed in the three effects: $DE^{NE} = 9 - 8 = 1$, $IS^{NE} = 11 - 9 = 2$, and $IO^{NE} = 5 - 11 = -6$. Note that the absolute value of the indirect effect due to the adjustment by others is the same regardless of which game is played first.

others. This means that $\beta' - \beta$ must be sufficiently small, as $IO = 6(\beta' - \beta)$ in this case. We can calculate how small $\beta' - \beta$ must be by finding the condition on $\beta' - \beta$ such that $G < 0$. This condition is:

$$\beta' - \beta < \frac{4 - \alpha' - 2\alpha}{6}.$$

Note that the right hand side of this inequality defines a threshold that can reach values between $\frac{1}{6}$ and $\frac{2}{3}$ depending on the values of α and α'. Voters who estimate a difference in others' cooperation rates across games below the threshold (i.e., those with lower estimates of the indirect effect due to the adjustment of others) should prefer the Prisoners' Dilemma to the Harmony Game. Since the maximum value of this threshold is below one, a player with a preference for the Prisoners' Dilemma must have beliefs such that $\beta' - \beta < 1$. In other words, a voter with a preference for the Prisoners' Dilemma must underappreciate the adjustment by others. Consider for example a person who does not necessarily expect others to play the dominant strategy in each game despite knowing she will herself always play the dominant strategy. This person will have parameters $(\alpha = 0, \alpha' = 1, \beta, \beta')$ and prefer the Prisoners' Dilemma if $\beta' - \beta < \frac{1}{2}$. That is, she will prefer the Prisoners' Dilemma if she expects cooperation in the Harmony Game to be at most 50 percentage points higher than in the Prisoners' Dilemma.[9]

The experiment was designed to test whether subjects underappreciate the response of others to a change in game leading them to form the wrong preferences over games. For a preference for the Prisoners' Dilemma to be wrong, it must be that the Harmony Game

[9]This statement would not change much by introducing plausible risk aversion. For example, a subject with a quadratic utility function with no other income and who believes there is 10% cooperation in the Prisoners' Dilemma would prefer the Harmony Game if $\beta' - \beta$ is above approximately 0.56. If this subject instead has a baseline income of as little as \$10, then the critical value is below 0.51, and it becomes 0.5007 when baseline income is \$100. Other utility functions yield a similar picture even for subjects who are arbitrarily risk averse. Consider a subject with CRRA utility function $u(x) = \frac{x^\rho}{\rho}$ and who believes there is 10% cooperation in the Prisoners' Dilemma. The critical value of $\beta' - \beta$ converges to 0.61 when the subject becomes arbitrarily risk averse ($\rho \to 0$) and he has no income outside of the experiment. For baseline incomes of \$10 and \$100, the limit critical values become as low as 0.514 and 0.501 respectively.

actually results in higher average payoffs than the Prisoners' Dilemma. In other words, it is necessary for actual behavior to be close enough to equilibrium behavior, so that the observed ranking of games based on actual payoffs coincides with the theoretical one. We expect this to hold, leading to the following hypothesis.

Hypothesis 1 *The Harmony Game results in higher average payoffs than the Prisoners' Dilemma.*

We expect that even when the Harmony Game results in higher average payoffs than the Prisoners' Dilemma, a majority of subjects may underappreciate equilibrium effects leading to the following hypothesis.

Hypothesis 2 *A majority of subjects prefer the Prisoners' Dilemma to the Harmony Game.*

As discussed before, the preference for the Prisoners' Dilemma can only arise from an underappreciation of the indirect effect due to the adjustment by others. This leads to the next two hypotheses.

Hypothesis 3 *A majority of subjects underappreciates the indirect effects associated with the adjustment of behavior by others. The average belief differential about cooperation rates* $\beta' - \beta$ *is smaller than the equilibrium prediction* $\beta' - \beta = 1$*, and smaller than the empirical difference in cooperation rates between games.*

Hypothesis 4 *Subjects who appreciate less the indirect effect due to the adjustment of others are more likely to support the Prisoners' Dilemma over the Harmony Game.*

The core of our investigation concerns Hypotheses 1 to 4. We present next two secondary hypotheses. To motivate the first, note that subjects who vote for the Prisoners' Dilemma because they do not expect the behavior of others to change fail to make predictions based

on equilibrium considerations; even more, those predictions fail to recognize that others will follow dominant strategies. We conjecture that this failure may be related to a lack of strategic sophistication. We obtained one measure of strategic sophistication in simultaneous move games by having subjects play a p-beauty contest. We then hold the following:

Hypothesis 5 *Subjects who vote for the Prisoners' Dilemma are measured to be less sophisticated in the beauty contest.*

To motivate our other secondary hypothesis, note that some subjects may miss not only the fact that the behavior of others will change under a different game, but may also miss the fact that their own behavior will change. As discussed above, this is neither necessary nor sufficient to cause a preference for PD in our setting, and hence not central to our main argument. However, it highlights that subjects may display very basic departures from equilibrium thinking.

Hypothesis 6 *Some subjects underappreciate the indirect effect through the adjustment by self.*

6 Results

6.1 Benchmark: the Harmony Game leads to higher payoffs than the Prisoners' Dilemma

Do subjects play close enough to the Nash outcome in each game so that cooperation and payoffs are greater in the Harmony Game than in the Prisoners' Dilemma? The answer is yes, supporting Hypothesis 1. Considering all periods in Part 2 and all treatments, there is 95% of cooperation in the Harmony Game versus 23% under the Prisoners' Dilemma. While this

difference is lower than the 100 percentage points predicted by Nash equilibrium, it is well above the 50 percentage points needed for a rational player to prefer the Harmony Game to the Prisoners' Dilemma. Consistently, subjects playing the Harmony Game obtained significantly higher payoffs than those playing the Prisoners' Dilemma (7.66 versus 5.91 points, a 30% increase, which amounts to almost $3 in Part 2 of the experiment). The differences in cooperation rates and payoffs in Part 2 between the two games are statistically significant at the 1%-level and are robust to considering each treatment separately; see Appendix Table 11.[10] The changes in behavior are large even in the first interaction in Part 2 (Period 6): 94% of cooperation in the Harmony Game versus 33% under the Prisoners' Dilemma.

The evolution of cooperation and payoffs as a function of the game played in Part 2 can be seen, for all treatments, in Figure 1. While there is little difference in behavior in Part 1 as a function of the game played in Part 2, there is an immediate difference in behavior between the Harmony Game and the Prisoners Dilemma in Part 2. Similar comparisons hold for each treatment at a time (see Appendix Figures 5, 6, and 7).

Another way to see that behavior across games differs in the direction predicted by theory is to compare the cooperation rates across the two games in Period 5, once the players have already gained experience. Pooling across all treatments, we find that the cooperation rate in the Prisoners' Dilemma is 15.5% while that in the Harmony Game is 95% (p-value < 0.0001). The corresponding average payoffs are 5.62 and 7.65, respectively (p-value < 0.0001). Again, the Harmony Game leads to higher payoffs.

In conclusion, behavior and payoffs across the two games vary enough in the direction predicted by standard game theory that voting against the Harmony Game results in lower

[10]The p-values for all comparisons reported in this section are obtained from Wald tests. We adopt the most conservative clustering of standard errors, at the session level.

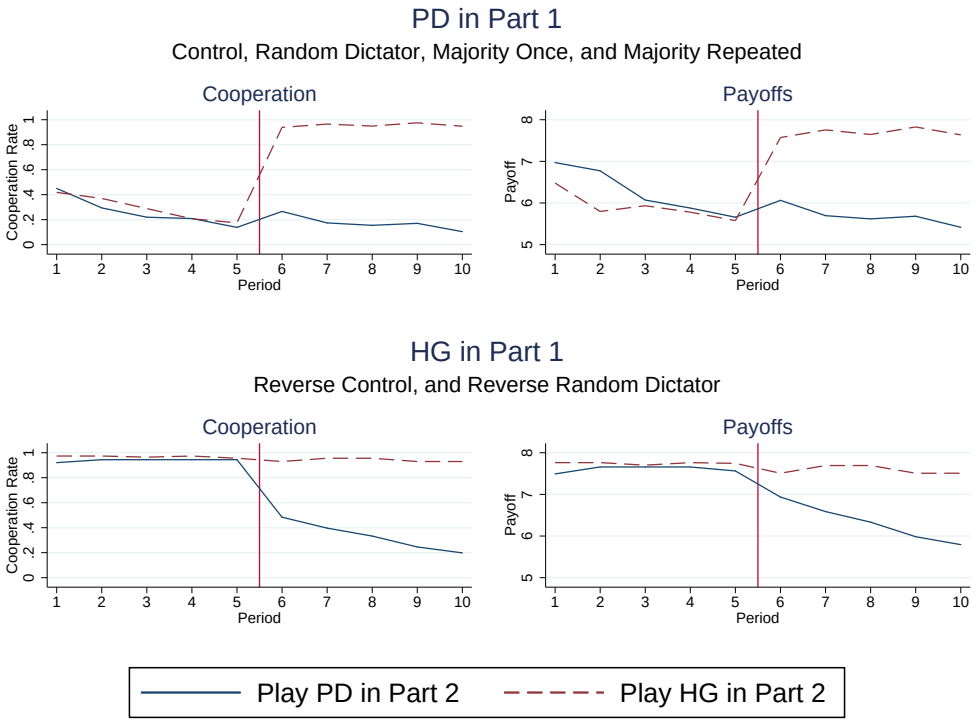

Figure 1: Comparing Prisoners' Dilemma and Harmony Game for all Treatments

payoffs in practice as well as in theory, as anticipated in Hypothesis 1.[11] Having established that the games' payoffs rank empirically as they do theoretically, we turn our attention to whether subjects choose games accordingly.

6.2 The demand for bad policy

Although choosing the Harmony Game leads to higher average payoffs for the subjects, a slight majority of subjects (53.60%) across all treatments voted for the Prisoners' Dilemma game at the beginning of Part 2, supporting Hypothesis 2 – see Table 5. The lowest share of subjects voting for the Prisoners' Dilemma game is 50.00% under Reverse Random Dictator,

[11] This ranking of games is unlikely to be affected by social preferences. For example, subjects with inequity aversion as in Fehr and Schmidt (1999) will have a stronger taste for the Harmony Game, which produces less inequality in practice than the Prisoners' Dilemma.

Table 5: Prisoner's Dilemma Vote Shares by Treatment at Beginning of Part 2

Treatment	Vote PD
Random Dictator	52.98%
Reverse Random Dictator	50.00%
Majority Once	60.83%
Majority Repeated	50.83%
Total	53.60%

while the largest is 60.83% under Majority Once. All of these shares differ significantly from the 0% that would be expected if subjects chose games according to theory.

This is the main result of the paper – a majority of subjects demanded the wrong game or policy. As a result of voting, a majority of subjects (54.55%) ended up in a game in Period 6 that led to lower payoffs than they would have achieved by voting for the Harmony Game. Subjects' tendency to support bad policy is remarkably stable across our various treatments varying the decision mechanism and timing; we will compare the voting shares across treatments later in the paper. We also study later the evolution of votes as subjects gain experience in Majority Repeated.

Note again that none of the usual explanations for the implementation of bad policies (bad institutions, incompetent or corrupt policymakers, etc.) apply to the simple environments of this experiment. Thus, responsibility for the implementation of bad policies can only be placed on the subjects, the citizens of this environment. But, what explains why a majority of subjects demanded bad policy?

6.3 Mechanism: failure to appreciate equilibrium effects

The conceptual framework presented in Section 3 showed that subjects in the environment we study in the experiment can only have a preference for the Prisoners' Dilemma if they underappreciate the indirect effect due to the adjustment in the behavior of others. This

led to the hypotheses that a majority of subjects will underappreciate the indirect effect due to the adjustment of others (Hypothesis 3) and that those who underappreciate this effect more will be more likely to vote for the Prisoners' Dilemma (Hypothesis 4). As we describe next, we find evidence that strongly supports both hypotheses.

First, we find that, on average, subjects grossly underestimate the effect of the game change on the behavior of others, consistent with Hypothesis 3. As Figure 2 shows, the distribution of the difference in the beliefs of cooperation between the Harmony and the Prisoners' Dilemma games ($\beta' - \beta$) is far from both the observed difference in behavior and the equilibrium one. The equilibrium difference is 100 percentage points and the observed difference is marked by a red line in Figure 2. The average difference in belief of cooperation between the Harmony and the Prisoners' Dilemma games is 35 percentage points in the pooled Random Dictator and Majority Once treatments while in reality cooperation is 76 percentage points higher in the Harmony Game. Similarly, the average belief difference is 30 percentage points in Reverse Random Dictator while the true difference in behavior is 63 percentage points. On average, subjects predict an effect of the game change half the size of its true effect.

Second, consistent with Hypothesis 4, subjects who more strongly underestimate the effect of the game on behavior are more likely to vote for the Prisoners' Dilemma. Figure 3 shows the average elicited belief of cooperation in each game broken down by vote of the subject. The solid line connects across games the beliefs of subjects who voted for the Prisoners' Dilemma while the broken line connects across games the beliefs of subjects who voted for the Harmony Game. The dots represent the observed behavior. In all three treatments in which beliefs were elicited, subjects who voted for the Prisoners' Dilemma expressed a lower belief that the behavior of others will differ across games. That is, subjects who voted for the Prisoners' Dilemma have a lower estimate of the effect of the game change

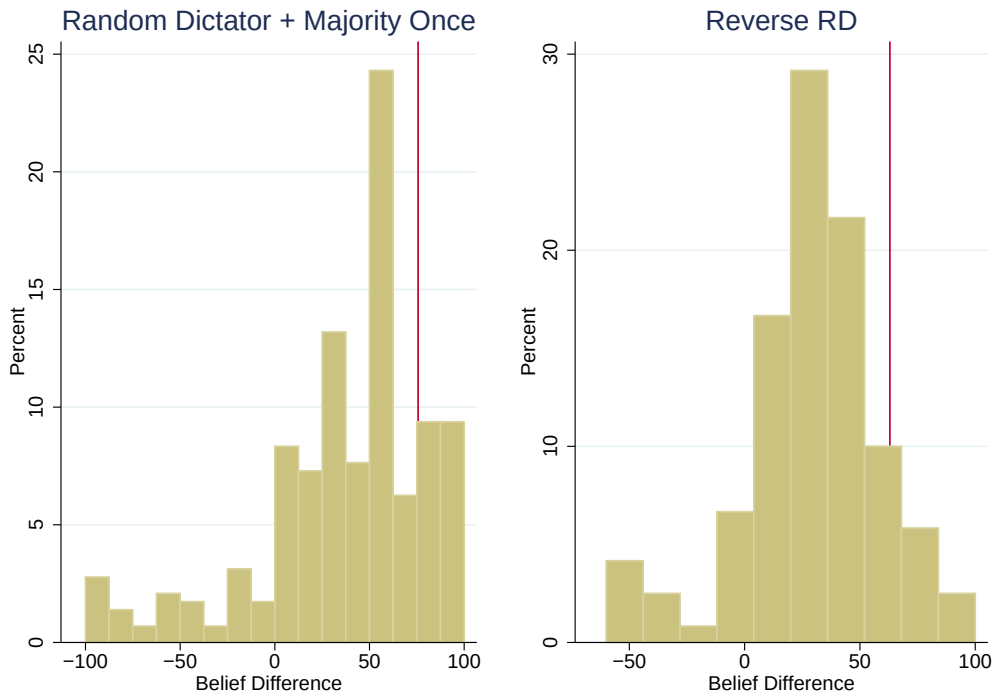

Figure 2: Distribution of Difference in Beliefs of Cooperation Between Games (Harmony Game – Prisoners' Dilemma)

on behavior.

The relationship between the difference in the beliefs of cooperation and voting is highly statistically significant across treatments – see Table 6. The "Belief Difference" variable denotes the difference in the belief of cooperation of other subjects under the Harmony Game relative to the Prisoners' Dilemma (i.e., $\beta' - \beta$). This OLS regression shows that an increase in the belief of cooperation of 100 percentage points (the theoretical prediction) would decrease the probability of voting for the Prisoners' Dilemma by around 50 percentage points.

Table 6 – columns 4 to 6 – also shows that the relationship between voting for the Prisoners' Dilemma and the belief difference is robust to controlling for personal characteristics.[12]

[12]We exclude self-reported SAT scores for two reasons: first, because not all subjects provided this infor-

Table 6: Beliefs and Voting for Prisoners' Dilemma (Dependent Variable: Vote for PD)

	(1) RD	(2) Reverse RD	(3) Majority Once	(4) RD	(5) Reverse RD	(6) Majority Once
Belief Difference	-0.005***	-0.004**	-0.005***	-0.005***	-0.004***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
Male				-0.126*	0.103	-0.207**
				(0.068)	(0.105)	(0.083)
Year				-0.007	-0.004	0.081**
				(0.032)	(0.036)	(0.039)
Ideology				0.035*	0.038**	0.045*
				(0.018)	(0.017)	(0.022)
Economics				0.051	0.169	-0.052
				(0.062)	(0.166)	(0.151)
Political Science				-0.194	0.232	0.093
				(0.222)	(0.169)	(0.119)
Brown University				0.070	0.075	0.149*
				(0.080)	(0.107)	(0.073)
Beauty Number				-0.001	0.001	-0.000
				(0.002)	(0.003)	(0.002)
Constant	0.698***	0.619***	0.774***	0.622***	0.343*	0.430**
	(0.037)	(0.078)	(0.044)	(0.122)	(0.165)	(0.185)
Observations	168	120	120	168	120	120
R-squared	0.202	0.058	0.160	0.244	0.124	0.244

Note: OLS specification. Belief Difference denotes the difference in beliefs of cooperation under HG and PD. Year denotes year in college. Ideology from 0 to 10 from most liberal to most conservative. Economics and Political Science denote subjects' major. Robust standard errors clustered by part 1 group: *** significant at 1%, ** at 5%, * at 10%

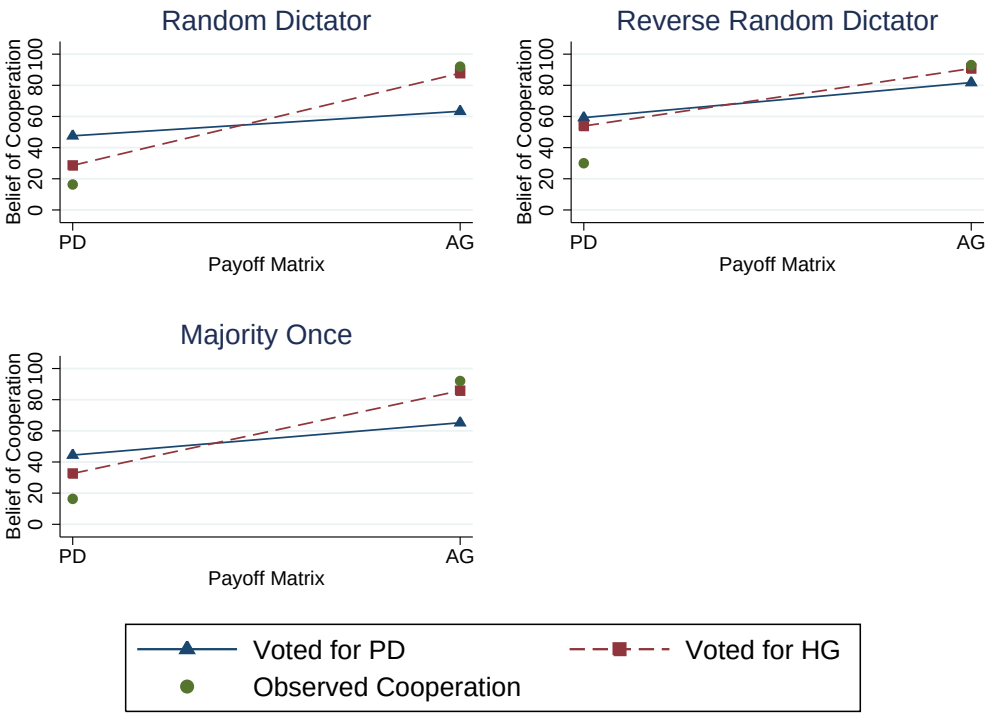

Figure 3: Belief of Cooperation and Voting

Most of these personal characteristics do not have a significant and consistent direct impact on voting across treatments, with the exception of ideology.[13]

The correlation of beliefs and voting documented in Table 6, while implied by our framework, does not necessarily imply itself that belief differences have a causal effect on voting. The reason is that people with different beliefs could also differ in dimensions that directly affect voting and are not observable.[14]

mation, including it would reduce the number of observations in the analysis; second, SAT scores do not significantly predict voting, and excluding them does not change our results.

[13]Appendix Table 12 shows the relationship between personal characteristics and voting for the Prisoners' Dilemma without controlling for the belief difference. This is important as personal characteristics could affect voting through their effect on beliefs. Appendix Table 12 shows that even when we do not control for beliefs there is no robust relationship between the personal characteristics we consider and voting for the Prisoners' Dilemma.

[14]Endogeneity may also occur if subjects examine the games further when being asked to report their beliefs, and report beliefs that justify their past voting choice. Costa-Gomes and Weizsäcker (2008) show evidence compatible with the idea that subjects re-examine strategic situations during the elicitation stage.

To show that beliefs have a causal effect on voting, we exploit exogenous variation in the beliefs held by a subject due to the behavior of the players encountered in periods 1 and 2. Because the identity and behavior of a subject's opponent in the first two periods is exogenous, it cannot correlate with any personal characteristic or past behavior of the subject (even the partner in period 2 cannot have played with someone who has played with the subject in question). A subject who observed more defection in the first two periods while playing a Prisoners' Dilemma should expect a larger increase in cooperation from the game change than a subject that observed more cooperation. Based on this idea, we use the observed behavior of the other players in periods 1 and 2 (measured as a cooperation rate of either 0, 0.5, or 1) as an instrument for beliefs. The approach is admittedly demanding, since by the time beliefs are elicited three more periods of play have occurred. We restrict attention to the three treatments where beliefs were elicited (Random Dictator, Reverse Random Dictator and Majority Once). Panel A in Table 7 shows that the cooperation rate observed in the first two periods has the expected effect on belief difference in all three treatments (positive for Random Dictator and Majority Once, and negative for Reverse Random Dictator). However, while the instrument is strong for Majority Once, it is very weak for the other two treatments.

Panel B in Table 7 shows the second-stage results. For the treatment for which we have an instrument, Majority Once, we can see that subjects who expect the change of games to have a greater impact on cooperation are less likely to vote for the Prisoners' Dilemma. Having a valid instrument in this treatment, we can now ask whether the instrumented estimate differs significantly from the OLS estimate. If not, then one cannot reject the hypothesis that beliefs are exogenous. The coefficients are very similar (-0.006 in the IV specification vs -0.005 in OLS) and a Hausman test cannot reject the null of belief exogeneity.

To sum up, the experimental evidence supports our core hypotheses that voters may

Table 7: Instrumenting for Beliefs

Panel A: First Stage (dependent variable: Belief Difference)

	(1) RD	(2) Reverse RD	(3) Majority Once	(4) RD	(5) Reverse RD	(6) Majority Once
Average other C	-7.122	15.008	-42.890***	-8.915	21.166	-47.546***
in periods 1 and 2	(10.489)	(21.063)	(7.552)	(10.032)	(19.966)	(6.701)
Personal Char.	N	N	N	Y	Y	Y
Constant	38.944***	15.630	52.572***	23.745	0.439	35.767**
	(5.609)	(20.935)	(4.740)	(21.358)	(19.130)	(16.855)
Observations	168	120	120	168	120	120
R-squared	0.003	0.008	0.137	0.074	0.063	0.168
F-test	0.461	0.508	32.25	0.790	1.124	50.35

Panel B: Second Stage (dependent variable: Vote for PD)

	(1) RD	(2) Reverse RD	(3) Majority Once	(4) RD	(5) Reverse RD	(6) Majority Once
Belief Difference	-0.031	-0.022	-0.007**	-0.026	-0.021	-0.006**
	(0.040)	(0.020)	(0.003)	(0.026)	(0.017)	(0.003)
Personal Char.	N	N	N	Y	Y	Y
Constant	1.656	1.154*	0.830***	1.112*	0.669**	0.449**
	(1.432)	(0.589)	(0.110)	(0.650)	(0.318)	(0.186)
Observations	168	120	120	168	120	120

Note: IV specification. Belief Difference denotes the difference in beliefs of cooperation under HG and PD. Year denotes year in college. Ideology from 0 to 10 from most liberal to most conservative. Economics and Political Science denote subjects' major. Robust standard errors clustered by part 1 group: *** significant at 1%, ** at 5%, * at 10%

demand bad policies, and that this demand is due to their inability to fully appreciate the equilibrium adjustments of others.

We now begin to investigate our secondary hypotheses. Going back to Table 6 and Appendix Table 12, we find that the strategic sophistication of subjects in simultaneous move games, as proxied by the number chosen in the p-beauty contest game, is not related to the voting decision in any of the treatments. Given that the p-beauty contest number may not be a perfect measure of strategic sophistication (for example, it is not the case that smaller numbers are necessarily a better choice given that others do not play Nash), we also studied whether there may be a non-linear relationship. We find that including a quadratic term does not change the lack of relationship. There is also no difference in voting between those with a number below and above the median, or below and above 66.66 (numbers above 66.66 are dominated). This lack of relationship between the the p-beauty contest number and voting refutes our secondary Hypothesis 5. This is not so surprising in the light of recent research showing that subjects' strategic sophistication is not persistent across games (see Georganas, Healy and Weber 2013). Moreover, this result could indicate that what is crucial in the voting decision is the capacity or inclination to think about the behavior of others in future stages and not in a simultaneous-move game as in the p-beauty contest game.

6.4 Do subjects understand how changing the game will affect their own behavior?

We have shown that subjects who vote for the bad policy greatly underestimate the effect that a policy change will have on the behavior of others. But it is also possible that subjects do not appreciate that their own behavior will change as well, as postulated in our secondary Hypothesis 6.

To study the share of subjects who fail to appreciate that their own behavior depends on the policy, we postulate a simple mixture model where individuals have one of two types $t \in (R, I)$ (for Rational, and Inertial, respectively) depending on the way they think about their actions in each game.[15] The Rational type is one who holds beliefs (β, β') about the cooperation rates by others in the Prisoners' Dilemma and the Harmony Game, respectively, but recognizes he will play his dominant strategy in each game. The Inertial type does not realize that his behavior will differ across games. This type considers that if he played action D(C) in the last round, he will continue to play it in the next, even if the game changes.

If we compute the expected payoff differential between the Prisoners' Dilemma and the Harmony Game for beliefs $[\alpha, \alpha', \beta, \beta']$, we obtain

$$\Delta u^t (\Delta\beta) = -6\Delta\beta + 4 - \alpha' - 2\alpha,$$

where $\Delta\beta = \beta' - \beta$. The key aspect differentiating the types is that the term $-\alpha' - 2\alpha$ is -1 for Rational types and is either 0 or -3 for Inertial types who defected or cooperated in period 5, respectively. Thus, $\Delta u^R (\Delta\beta) = -6\Delta\beta + 3$ and $\Delta u^I (\Delta\beta, c) = -6\Delta\beta + 4 - 3c$, where c is an indicator variable for whether the subject cooperated in period 5.

We postulate the existence of a share s of Rational types, and $1 - s$ of Inertial types. For the purposes of empirical identification of the share s, we assume that a Rational (Inertial) type votes for the Prisoners' Dilemma game with a probability that depends on the payoff differential $\Delta u^R (\Delta\beta)$ $(\Delta u^I (\Delta\beta, c))$. To account for empirical errors, we will assume that such probability is given by a logistic cdf with mean μ and standard deviation σ. Thus, a

[15] One way to test Hypothesis 6 would be to elicit beliefs about players' own actions. We did not do this in order not to disturb the elicitation of beliefs about others, which are key to our core hypotheses. An alternative was to add another condition, but given the large size of the experiment (768 subjects), we opted to investigate this secondary hypothesis via the structural approach presented in this section. While our Hypothesis 6 was formulated ex ante, the precise assumptions on types presented here were developed ex post.

player i with type t votes for Prisoners' Dilemma with a probability $F\left(\Delta u^t\left(\Delta\beta_i, c_i\right), \mu, \sigma\right)$, where F denotes the logistic distribution. It follows that the probability of a Prisoners' Dilemma vote by a player i, given a share s of Rational types is,

$$P\left(v_i = PD|\Delta\beta_i, c_i, s, \mu, \sigma\right) = sF\left(\Delta u^R\left(\Delta\beta_i\right), \mu, \sigma\right) + (1-s)\,F\left(\Delta u^I\left(\Delta\beta_i, c_i\right), \mu, \sigma\right).$$

We define the profile of period 5 actions as $\mathbf{c} = [c_1, ..., c_N]$ where $c_i = 1$ denotes cooperation by subject i in period 5 and $c_i = 0$ denotes defection. Similarly, we define the profile of votes as $\mathbf{v} = [v_1, ..., v_N]$ where $v_i = 1$ denotes a vote for Prisoners' Dilemma by subject i, and $v_i = 0$ a vote for Harmony Game. We have that the overall probability of such a profile of votes is,

$$\prod_{i=1}^{N} P\left(v_i = 1|\Delta\beta_i, c_i, s, \mu, \sigma\right)^{v_i}\left(1 - P\left(v_i = 1|\Delta\beta_i, c_i, s, \mu, \sigma\right)\right)^{1-v_i},$$

which yields the log-likelihood,

$$L\left(s, \mu, \sigma|\mathbf{v}, \mathbf{\Delta}\beta, \mathbf{c}\right) =$$
$$\sum_{i=1}^{N}\left\{\begin{array}{c} v_i \ln\left[sF\left(\Delta u^R\left(\Delta\beta_i\right), \mu, \sigma\right) + (1-s)\,F\left(\Delta u^I\left(\Delta\beta_i, c_i\right), \mu, \sigma\right)\right] \\ +(1-v_i)\ln\left[s\left(1 - F\left(\Delta u^R\left(\Delta\beta_i\right), \mu, \sigma\right)\right) + (1-s)\left(1 - F\left(\Delta u^I\left(\Delta\beta_i, c_i\right), \mu, \sigma\right)\right)\right] \end{array}\right\}.$$

We estimate the parameter s, by maximizing $L\left(s, \mu, \sigma|\mathbf{v}, \mathbf{\Delta}\beta, \mathbf{c}\right)$ given the voting data $\mathbf{v}$, the period 5 behavior $\mathbf{c}$ and the vector of elicited beliefs $\mathbf{\Delta}\beta$. Clearly, in this estimation we take beliefs to be exogenous – this is a maintained assumption with some support from the exogeneity test performed earlier in relation with the instrumental-variables findings. We pool the data for the three conditions where beliefs were elicited, namely Random Dictator, Reverse Random Dictator, and Majority Once, for a total of 408 observations.

The estimate of the share of Rational types s, presented in Table 8, is 67% when we consider all subjects. The Wald test for the share of Rational types being equal to 100% yields p-values of 0.042 and 0.057 depending on whether the standard errors are clustered respectively at the individual or group level, allowing us to reject the null of no Inertial types. These findings support the notion that a fraction of the players vote without appreciating how their own play will adjust following a policy change. The point estimate suggests that a full third of the players make this error. This is striking, given that all that is required is to forecast that one will play a different, dominant, strategy in a 2x2 game following the change in policy.

The existence of the logistic disturbance term implies that any voter could vote for or against the Prisoners' Dilemma; but the beliefs $\Delta\beta$ and the types (R,I) affect the likelihood of the vote going one way or another. For voters with $\Delta\beta$ in between $\frac{1}{6}$ and $\frac{2}{3}$, however, the type (R,I) alone can affect the vote, even with a zero realization of the disturbance term. One may consider the empirical variation offered by those voters as more central, and wonder about the robustness of the result to restricting the sample to those voters. We report the estimates of our model on this restricted sample in the second column of Table 8. Although we lose a large number of observations (from 408 to 232), the fraction of Rational types is again seen as significantly different from one (point estimate 0.4, p-value 0.084). The point estimate is not significantly different from the one obtained using the full sample.

6.5 Ruling out alternative mechanisms

The variation of treatments allows us to rule out some alternative mechanisms. One possibility is that the Prisoners' Dilemma obtains a majority in the Random Dictator and Majority treatments because a status quo bias causes some people to choose their initial game even if it is a suboptimal one. Instead, when the Harmony Game is played first, a reluctance to try

Table 8: Structural Estimates

	Subjects	
	All	$\frac{1}{6} \leq \Delta\beta \leq \frac{2}{3}$
s (Share of Rational Types)	0.670	0.399
	(0.173)	(0.348)
μ	0.801	1.097
	(0.149)	(0.549)
σ	1.336	0.978
	(0.350)	(0.493)
Observations	408	232
p-value of Wald test $s \neq 1$	0.057	0.084

Note: Pooled sample from Random Dictator, Reverse Random Dictator and Majority Once. Robust standard errors clustered by part 1 group in parentheses.

new things should reinforce a preference for the initial, and also optimal, game and secure virtually unanimous support for the Harmony Game. However, in the Reverse Random Dictator condition the Prisoners' Dilemma garnered 50% of the vote. This vote share is only 3 points smaller than (and statistically indistinguishable from) that under Random Dictator. This evidence rules out the status quo bias possibility.

The Random Dictator treatment allows us to rule out forms of pivotal thinking as a source of the demand for bad policy. Under majority rule, a vote matters only if pivotal. Two types of reasoning may dilute a subject's incentive to vote for the Harmony Game in that situation. First, the subject may have expectations of a large majority (in either direction), leaving her with negligible chances of being pivotal and therefore with no incentive to carefully consider how to vote. Second, even if the chance of pivotality is not negligible, a subject may interpret the event of being pivotal as a sign that a large share of the subjects are not expecting behavior to correspond to equilibrium (if they were, they would vote for the

Harmony Game). Subjects could take this as a sign that other subjects will themselves not respond to the change in game as theory predicts, and voting for the Harmony Game may be a bad idea. Under Random Dictator, pivotality has a clear, and non-negligible chance of 1/6. Moreover, the event of being pivotal does not depend on the votes of others and hence it cannot constitute a signal of how others may play in the Harmony Game. Therefore, the majoritarian 52.98% of votes for the Prisoners' Dilemma under Random Dictator cannot be explained by the previous pivotality concerns. The share of votes in favor of the Prisoners' Dilemma is greater under Majority Once (60.83%), but the difference is not statistically significant.

We derived our hypotheses from a framework where subjects may have difficulty appreciating how the behavior of others will adjust to a new game, and our main hypotheses are supported by the data. Nevertheless, our framework would warrant less attention if the majority vote for Prisoners' Dilemma could be explained by strictly rational motives. Yet, as shown above, a rational subject (one who knowingly plays the dominant action in each game) cannot prefer the Prisoners' Dilemma unless she expects cooperation in the Harmony Game to be at most 50 percentage point greater than in the Prisoners' Dilemma. Given the actual difference in cooperation rates is much higher (e.g., 76 and 63 percentage points in the Control and Reverse Control conditions, respectively), a rational subject who has fairly accurate beliefs about real behavior cannot prefer the Prisoners' Dilemma. Moreover, as we showed in footnote 8, even sizeable risk aversion makes rational subjects only slightly more willing to vote for the Prisoners' Dilemma. Thus, rationalizing the Prisoners' Dilemma majority in a model where the majority of players are both rational and correct about their environment appears difficult.

Several models of strategic naivety make predictions other than sub-game perfect ones in our setting, and one may wonder whether the underappreciation of equilibrium effects we

identify is a particular case of the phenomena explained by those theories. Here we discuss three such theories, namely the level-k model of strategic thinking (Stahl and Wilson 1994, Nagel 1995), Camerer, Ho and Chong's (2004) related "cognitive hierarchy model," and Jehiel's (2005) analogy-based-expectation equilibrium (ABEE). We then explain why they do not provide a fully satisfactory account of the patterns in our data.

The level-k model of strategic thinking summarizes players' strategic sophistication by the parameter k, where a level-k type of player best responds to beliefs that her opponent is a level-$k-1$ type of player (for $k \geq 1$), and a level-0 type randomizes uniformly over actions. Experimental work has estimated levels one and two to be the most frequent types across the universe of laboratory games where the model has been estimated (see, e.g., Crawford, Costa-Gomes and Iriberri 2013). In our setting, a level-0 type would cooperate and defect with equal probability in both games, leading a level-1 type to vote for the Prisoners' Dilemma and play the dominant strategy in both the Prisoners' Dilemma and the Harmony Game. Because level-1 types play dominant strategies, all higher levels vote for the Harmony Game. Hence, the level-1 type, better than any other type, fits the behavior of the majority of our subjects who vote for the Prisoners' Dilemma. However, a theory predicated on level-1 makes at least one prediction that is contradicted by the data, namely that subjects voting for the Prisoners' Dilemma predict that cooperation rates do not vary across the Prisoners' Dilemma and the Harmony Game. This does not match the fact that most individuals voting for the Prisoners' Dilemma in our experiments, even if they underestimate the difference in cooperation across the two games, still predict more cooperation in the Harmony Game than in the Prisoners' Dilemma. Those voting for the Prisoners' Dilemma estimate on average an increase in cooperation of about 20 percentage points (which is significantly different from zero with a p-value < 0.0001).

A similar prediction is made by Jehiel's (2005) ABEE model. In this approach each player

is modeled as bundling her opponents' decision nodes into partitional "analogy classes"; each player holds correct beliefs about her opponents' distribution of actions across each class, yet mistakenly believes that the frequency of each action played is constant across every node in an analogy class. There are two analogy classes of interest in our setting. Players with the finest analogy classes put the Prisoners' Dilemma and the Harmony Game in different analogy classes; each game has a different dominant strategy, and because players correctly predict actions in each class, the ABEE outcome coincides with subgame-perfect equilibrium. Alternatively, players who bundle the Prisoners' Dilemma and the Harmony Game into a single, coarser analogy class would predict that their opponents play the same way in both games. As with level-k models, this prediction is not supported by the data.

Since the level-k model constrains level-k players to believe their opponents are level-$k-1$, one may think it is not flexible enough to match our data, but that a more flexible framework in the same spirit could. Camerer, Ho and Chong's (2004) "cognitive hierarchy" model allows for more flexible beliefs: for instance, level-2 types believe that they face a distribution of level-0 and level-1 types that coincides with the population distribution in the experiment. This account faces the hurdle that few subjects play the dominated action in either the Harmony Game or the Prisoners' Dilemma, so level-2 types must assign a high probability to level-1 types in the cognitive-hierarchy model. But, if there are few level-0 players, higher level players have no reason to vote for the Prisoners' Dilemma in this model. Yet, we observe majoritarian support for the Prisoners' Dilemma in the data.

6.6 Learning Under Repeated Majority Voting

Our basic design presented subjects with a choice between a game with which they had experience and one with which they lacked experience. This sought to capture substantive situations of interest where some policy options are new to the population, as well as to

motivate subjects to make conjectures that ought to be informed by equilibrium predictions. In that context, we showed that beliefs about behavior are not tightly driven by equilibrium considerations and this led a majority of subjects to mistakenly prefer the Prisoners' Dilemma over the Harmony Game. Against such backdrop, one may conjecture that gaining experience with the less familiar game will lead subjects to rank the games correctly, and one may also wonder how quick and complete that learning will be.

The Majority Repeated treatment allows us to study the evolution of voting as subjects gain experience. The percentage of subjects voting for the Prisoners' Dilemma decreases from 50.83% in Period 6 to 28.33% in Period 10 – see Figure 4.

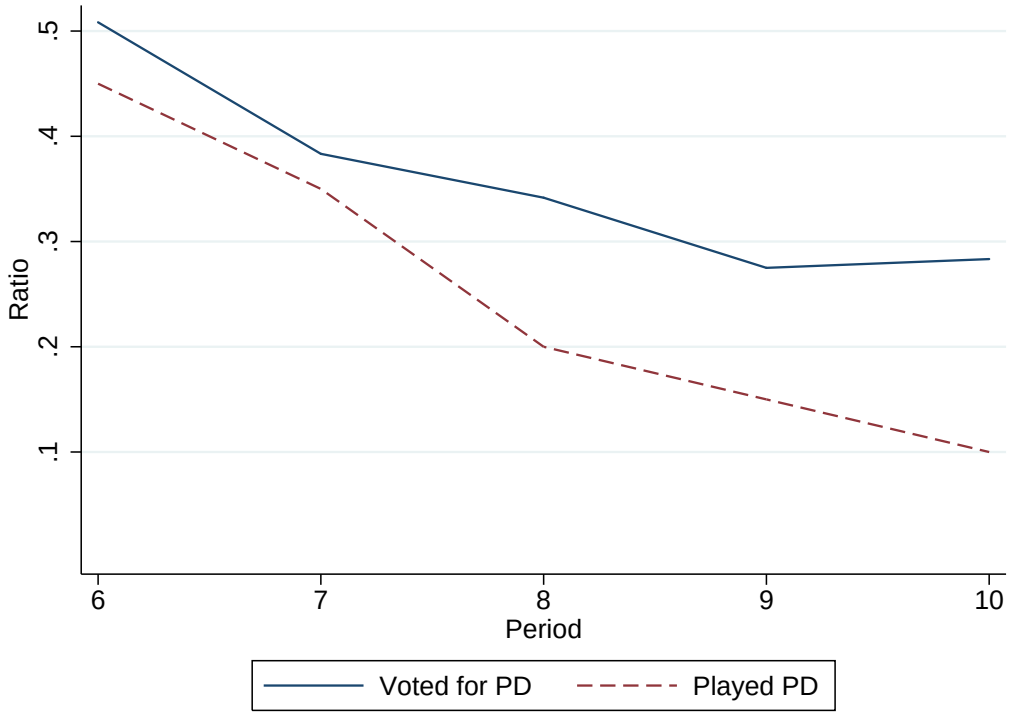

Figure 4: Evolution of Votes and Outcomes in Majority Repeated

The fact that vote shares for the Prisoners' Dilemma decrease with experience suggests that there is learning and that in our context experience can substitute for the ability to

make equilibrium predictions based on theory. However, it is noteworthy that even in the fifth period after the first vote, more than a quarter of subjects continued to choose the wrong game. One possibility is that many subjects still vote for the Prisoners' Dilemma simply because they have not been exposed to the Harmony Game. But this is not the case. The percentage of subjects playing the Prisoners' Dilemma decreases from 45% in Period 6 to 10% in Period 10.[16] In other words, a non-trivial share of votes for the Prisoners' Dilemma persists until Period 10 despite the fact that by then most voters have been exposed to the Harmony Game. More specifically, the second column in Table 9 shows, by period, the percentage of Prisoners' Dilemma voters who had played the Harmony Game before. Of course, none of them had played the Harmony Game before voting in period 6. However, by period 8 more than half of the Prisoners' Dilemma voters had played the Harmony Game before.

Why do some subjects that have experienced both games continue to vote for the wrong game? Is it the case that those subjects have met defectors in the Harmony Game? The third column in Table 9 shows the percentage of Prisoners' Dilemma voters who had played the Harmony Game before and had observed a difference in cooperation rates greater than 50 percentage points, the cutoff to prefer the Harmony Game over the Prisoners' Dilemma. Note that in every period, more than 70% of Prisoners' Dilemma voters who had played the Harmony Game before had observed cooperation rates in the Prisoners' Dilemma and the Harmony Game such that voting for the Harmony Game was warranted. So, a large fraction of Prisoners' Dilemma voters had information in favor of the Harmony Game but still preferred to vote for the Prisoners' Dilemma by the end of the experiment.

In conclusion, while many subjects who started voting for the Prisoners' Dilemma switch

[16]The reason is simple: given the simple majority rule, quickly after the vote share for the Prisoners' Dilemma becomes lower than 50%, the power of democracy kicks in: majorities for the Prisoners' Dilemma in each group become less frequent, and only a small share of subjects end up in the wrong game by Period 10.

Table 9: The information of Prisoners' Dilemma Voters

Period	Played HG Before	And Observed Cooperation Rate 50% over the one in PD
6	0%	
7	32.60%	80%
8	51.22%	85.71%
9	57.58%	78.95%
10	73.53%	72%

to vote for the Harmony Game as they gain experience, other voters do not learn even when they have observed behavior in both games that favors voting for the Harmony Game. This evidence yields a nuanced message: lack of familiarity with a policy option that requires making equilibrium predictions can generate substantial bias in policy preference; with more balanced experience, the bias in policy preference decreases, but it does not go away completely – some fraction of the biased preferences identified in our main experiments persists. While the particular levels of the demand for bad policy obtained in the lab are to be taken with caution, one interpretation of our experimental results is that in the presence of an unfamiliar policy option the prevalence of mistaken preference is such that groups may often select the wrong policy even when relying on majority voting. With more balanced experience, majoritarian mistakes are rare, but policy distortions could still occur when groups select policy through mechanisms that are less stark than majority voting, and which place positive weight on the opinion of all voters (e.g., collective bargaining).

7 Conclusion

We have experimentally identified a demand for bad policy driven by voters' underappreciation of how behavior changes when policies change. Voters in our experiment underestimate, on average, how much the behavior of others will change following a change in the game that

is played. In addition, a non-trivial share of voters appear to fail to appreciate that their own behavior will differ across games. Our evidence suggests that unfamiliar policy options can be a challenge for voters when these policies contain "hidden" costs or benefits that will accrue once behavior adjusts. An example of such a policy is a Pigouvian tax, which generates a direct monetary cost on taxpayers but it generates indirect benefits in equilibrium by inducing those same taxpayers to internalize externalities. More generally, our results help explain why voters may not always support the policy proposals of economists that are beneficial mainly through indirect equilibrium effects.

Of course, identifying a demand for bad policy in connection with a tendency to underestimate equilibrium effects in the laboratory does not necessarily mean that we have identified an important source of bad policy in real life. One could hope that public discourse and political competition would result in voters learning about the total effect of policies, thus bridging the gap between public opinion and reliable evidence. However, as discussed in the introduction, a vast literature in economics and political science – both theoretical and empirical – has considered politicians as reflecting, more than shaping, the positions of voters. To the extent that public opinion and voter preferences matter for the selection of policies, understanding how people think about policies appears relevant for our knowledge of how societies choose to regulate themselves. This paper makes a contribution to that understanding.

References

[1] Alesina, A. and G. Tabellini (1990). "Voting on the Budget Deficit," *American Economic Review* 80(1), 37-49.

[2] Alesina, A. and A. Drazen (1991). "Why are Stabilizations Delayed?" *American Economic Review* 81, 1170-1188.

[3] Barro, R.J. (1973). "The Control of Politicians: An Economic Model," *Public Choice* 14, 19-42.

[4] Bartels, L. (2012). "The Study of Electoral Behavior," in Jan Leighley (ed.) The Oxford Handbook of American Elections and Political Behavior. Oxford University Press.

[5] Beilharz, H.J., and H. Gersbach (2004). "General Equilibrium Effects and Voting into a Crisis," CEPR Discussion Paper no. 4454.

[6] Besley, T. (2005). "Political Selection," *Journal of Economic Perspectives* 19(3), 43-60.

[7] Besley, T. and S. Coate (1998). "Sources of Inefficiency in a Representative Democracy: A Dynamic Analysis," *American Economic Review* 88(1), 139-56.

[8] Bisin, A., A. Lizzeri and L. Yariv (2015). "Government Policy with Time Inconsistent Voters," *American Economic Review* 105(6), pp. 1711-37.

[9] Blinder, A. and A. Krueger (2004), "What Does the Public Know about Economic Policy, and How Does It Know It?" *Brookings Papers on Economic Activity* 2004(1), 327-97.

[10] Bone, J., Hey, J.D., and J. Suckling (2009), "Do People Plan?" Experimental Economics 12, 12-25.

[11] Camerer, Colin F., Ho, Teck-Hua, and Juin Kuan Chong "A Cognitive Hierarchy Model of Games," *Quarterly Journal of Economics* 119(3): 861-98.

[12] Canes-Wrone, B., M.C. Herron, and K.W. Shotts (2001). "Leadership and Pandering: A Theory of Executive Policymaking," *American Journal of Political Science* 45, 532-50.

[13] Caplan, B. (2007). *The Myth of the Rational Voter: Why Democracies Choose Bad Policies,* Princeton: Princeton University Press.

[14] Caselli, F., and M. Morelli (2004). "Bad Politicians," *Journal of Public Economics* 88(2), 759-82.

[15] Coate, S. and S. Morris (1995). "On the Form of Transfers to Special Interests," *Journal of Political Economy* 103(6), 1210-35.

[16] Costa-Gomes, Miguel and Georg Weizsäcker (2008). "Stated Beliefs and Play in Normal-Form Games," *Review of Economic Studies* 75: 729-62.

[17] Crawford, V., M.A. Costa-Gomes, and N. Iriberri (2013). "Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications," *Journal of Economic Literature* 51(1): 5-62.

[18] Dal Bó, E., Dal Bó, P., and R. Di Tella (2006). "Plata o Plomo?: Bribe and Punishment in a Theory of Political Influence," *American Political Science Review* 100(1), 41-53.

[19] Dal Bó, P. (2014). "Experimental Evidence on the Workings of Democratic Institutions," in Economic Institutions, Rights, Growth, and Sustainability: the Legacy of Douglass North, Cambridge University Press: Cambridge.

[20] Dal Bó, P., A. Foster and L. Putterman (2010). "Institutions and Behavior: Experimental Evidence on the Effects of Democracy," *American Economic Review* 100(5), 2205-29.

[21] De Figueiredo, R. (2002). "Electoral Competition, Political Uncertainty, and Policy Insulation," *American Political Science Review* 96(2), 321-33.

[22] Esponda, I. and D. Pouzo (2010). "Conditional Retrospective Voting in Large Elections," unpublished manuscript.

[23] Esponda, I. and E. Vespa (2012). "Hypothetical Thinking and Information Extraction: Strategic Voting in the Laboratory," Unpublished manuscript.

[24] Eyster, E. and M. Rabin (2005). "Cursed Equilibrium," *Econometrica* 73(5), 1623-72.

[25] Ferejohn J. (1986). "Incumbent Performance and Electoral Control," *Public Choice* 50(1/3) 5-25.

[26] Fernandez, R. and D. Rodrik (1991). "Resistance to Reform: Status Quo Bias in the Presence of Individual-Specific Uncertainty," *American Economic Review* 81(5), 1146-55.

[27] Fudenberg, D., D.G. Rand and A. Dreber. (2012). "Slow to Anger and Fast to Forgive: Cooperation in an Uncertain World" *American Economic Review* 102(2), 720-49.

[28] Fehr, E. and K.M. Schmidt. (1999). "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics* 114(3), 817-68.

[29] Gentzkow, M. and J. Shapiro (2006). "Media Bias and Reputation," *Journal of Political Economy* 114(2), 35-71.

[30] Gentzkow, M. and J. Shapiro (2010). "What Drives Media Slant? Evidence from U.S. newspapers," *Econometrica* 78(1), 35-71.

[31] Georganas, S., P.J. Healy and R.A. Weber (2013). "On the Persistence of Strategic Sophistication," unpublished manuscript.

[32] Grether, D.M. (1992). "Testing Bayes Rule and the Representativeness Heuristic: Some Experimental Evidence" *Journal of Economic Behavior & Organization* 17(1), 31-57.

[33] Harrington, J.E. (1993). "Policy, Economic Performance, and Elections" *American Economic Review* 83(1), 27-42.

[34] Holt, C.A. (2007). *Markets, Games, and Strategic Behavior*, Boston: Pearson/Addison Wesley.

[35] Jehiel, P. (2005). "Analogy-Based Expectation Equilibrium," *Journal of Economic Theory* 123 (2), 81-104.

[36] Kallbekken, S., Kroll, S. and T. Cherry (2011). "Do You Not Like Pigou, Or Do You Not Understand Him? Tax Aversion and Revenue Recycling in the Lab," *Journal of Environmental Economics and Management* 62, 53-63.

[37] Karni, E. (2009). "A Mechanism for Eliciting Probabilities," *Econometrica*, 77, 603-06.

[38] Knittel, C. (2012). "Reducing Petroleum Consumption from Transportation," *Journal of Economic Perspectives* 26(1), 93-118.

[39] Levitt, S., List, J. and S. Sadoff (2011). "Checkmate: Exploring Backward Induction among Chess Players," *American Economic Review* 101(2): 975-90.

[40] Lizzeri, A. and L. Yariv (2011). "Collective Self-Control," Unpublished manuscript.

[41] Margreiter, M., M. Sutter, and D. Dittrich (2005). "Individual and Collective Choice and Voting in Common Pool Resource Problem with Heterogeneous Actors," *Environmental & Resource Economics* 32, 241-71.

[42] Maskin, E. and J. Tirole (2004). "The Politician and the Judge: Accountability in Government," *American Economic Review* 94(4).

[43] McKelvey, R. and T. Palfrey (1992). "An Experimental Study of the Centipede Game", *Econometrica* 60: 803-36.

[44] Messner, M. and M.K. Polborn (2004). "Paying politicians," *Journal of Public Economics* 88 (12), 2423-45.

[45] Möbius, M.M., M. Niederle, P. Niehaus and T. Rosenblat (2014) "Managing Self-Confidence: Theory and Experimental Evidence," unpublished manuscript.

[46] Moinas, S. and S. Pouget (2013). "The Bubble Game: An Experimental Study of Speculation", *Econometrica* 81(4): 1507-40.

[47] Mullainathan, S. and E. Washington (2009). "Sticking with Your Vote: Cognitive Dissonance and Political Attitudes," *American Economic Journal: Applied Economics* 1(1), 86-111.

[48] Nagel, R. (1995). "Unraveling in Guessing Games: An Experimental Study," *American Economic Review* 85(5), 1313-26.

[49] North, D. C. (1990). *Institutions, Institutional Change and Economic Performance*, Cambridge: Cambridge University Press.

[50] Palacios-Huerta, I. and O. Volij (2009). "Field Centipedes," *American Economic Review* 99(4): 1619-35.

[51] Peltzman, S. (1976). "Toward a More General Theory of Regulation," *Journal of Law and Economics* 19, 211-40.

[52] Romer, T. and H. Rosenthal (1978). "Political Resource Allocation, Controlled Agendas and the Status Quo," *Public Choice* 33, 27-43.

[53] Sausgruber, R. and J-R Tyran (2005). "Testing the Mill Hypothesis of Fiscal Illusion", *Public Choice* 122: 39-68.

[54] Sausgruber, R. and J-R Tyran (2011). "Are We Taxing Ourselves? How Deliberation and Experience Shape Voting On Taxes," *Journal of Public Economics* 95, 124-76.

[55] Smith, Adam (1776). *An Enquiry into the Nature and Causes of the Wealth of Nations* W. Strahan and T. Cadell: London.

[56] Stigler, G. (1971). "The Regulation of Industry" *The Bell Journal of Economics and Management Science* 2, 3-21.

[57] Stahl, D.O. and Wilson P.W. (1994). "Experimental Evidence on Players' Models of Other Players ," *Journal of Economic Behavior & Organization* 25(3), 309-27.

[58] Walker, J., R. Gardner, A. Herr and E. Ostrom (2000). "Collective Choices in the Commons: Experimental Results on Proposed Allocation Rules and Votes," *The Economic Journal* 110(1), 212-34.

8 Appendix

Table 10: Summary Statistics

	Obs.	Mean	Std. Dev.	Min	Max
Male	768	0.43	0.50	0	1
Year	768	2.70	1.21	1	5
Ideology	768	3.54	2.14	0	10
Economics	768	0.15	0.36	0	1
Political Science	768	0.05	0.21	0	1
Brown U.	768	0.50	0.50	0	1
Beauty Contest Number	768	36.67	21.35	0	100
Math SAT	662	723.95	71.77	400	800
Verbal SAT	644	700.19	77.45	400	800
Belief of C in PD	408	44.26	25.79	0	100
Belief of C in HG	408	77.74	26.02	0	100
Belief Difference	408	33.47	-41.11	100	100
Earnings	768	27.81	3.27	16.75	37

Table 11: Comparison between Prisoners' Dilemma and Harmony Game by Treatment

Panel A: Cooperation

	Control			Reverse Control			Random Dictator		
	Part 1	Part 2		Part 1	Part 2		Part 1	Part 2	
Periods	All	6	All	All	6	All	All	6	All
HG in Part 2	27%	88%	92%	99%	92%	93%	28%	94%	96%
PD in Part 2	21%	30%	16%	95%	38%	30%	26%	21%	15%
Diff. p-value	0.288	0.001	0.000	0.319	0.001	0.000	0.786	0.000	0.000
	Reverse RD			Majority Once			Majority Repeated		
	Part 1	Part 2		Part 1	Part 2		Part 1	Part 2	
Periods	All	6	All	All	6	All	All	6	All
HG in Part 2	95%	94%	95%	37%	97%	94%	28%	97%	98%
PD in Part 2	93%	58%	36%	31%	33%	21%	24%	20%	17%
Diff. p-value	0.661	0.001	0.000	0.421	0.000	0.000	0.521	0.001	0.000

Panel B: Payoffs

	Control			Reverse Control			Random Dictator		
	Part 1	Part 2		Part 1	Part 2		Part 1	Part 2	
Periods	All	6	All	All	6	All	All	6	All
HG in Part 2	5.81	7.18	7.44	7.79	7.42	7.51	6.04	7.61	7.75
PD in Part 2	6.12	6.20	5.65	7.75	6.53	6.20	6.16	5.82	5.59
Diff. p-value	0.186	0.115	0.001	0.640	0.117	0.004	0.652	0.000	0.000
	Reverse RD			Majority Once			Majority Repeated		
	Part 1	Part 2		Part 1	Part 2		Part 1	Part 2	
Periods	All	6	All	All	6	All	All	6	All
HG in Part 2	7.70	7.61	7.66	5.95	7.77	7.58	5.82	7.79	7.83
PD in Part 2	7.48	7.30	6.44	6.42	6.33	5.85	6.35	5.81	5.67
Diff. p-value	0.258	0.392	0.001	0.019	0.008	0.005	0.227	0.001	0.000

Note: p-values calculated using Wald tests with s.e. clustered at session level.
For Majority Repeated, behavior in Part 1 as a function of game played in period 6.

Table 12: Personal Characteristics and Voting for Prisoners' Dilemma (Dependent Variable: Vote for PD)

	(1) RD	(2) Reverse RD	(3) Majority Once	(4) Majority Repeated
Male	-0.208**	0.061	-0.219**	-0.272**
	(0.083)	(0.116)	(0.097)	(0.100)
Year	0.006	-0.012	0.063	-0.023
	(0.031)	(0.040)	(0.041)	(0.043)
Ideology	0.027	0.040**	0.049*	-0.004
	(0.018)	(0.018)	(0.025)	(0.016)
Economics	-0.034	0.164	-0.081	-0.086
	(0.083)	(0.171)	(0.175)	(0.149)
Political Science	-0.147	0.263*	0.070	-0.102
	(0.215)	(0.141)	(0.180)	(0.164)
Brown University	0.066	0.044	0.136	0.043
	(0.087)	(0.121)	(0.084)	(0.116)
Beauty Number	-0.001	0.001	-0.000	0.003
	(0.002)	(0.003)	(0.002)	(0.002)
Constant	0.519***	0.260	0.321	0.589***
	(0.153)	(0.175)	(0.188)	(0.185)
Observations	168	120	120	120
R-squared	0.058	0.063	0.083	0.107

Note: OLS specification. Year denotes year in college. Ideology from 0 to 10 from most liberal to most conservative. Economics and Political Science denote subjects' major. Robust standard errors clustered by part 1 group:
*** significant at 1%, ** at 5%, * at 10%

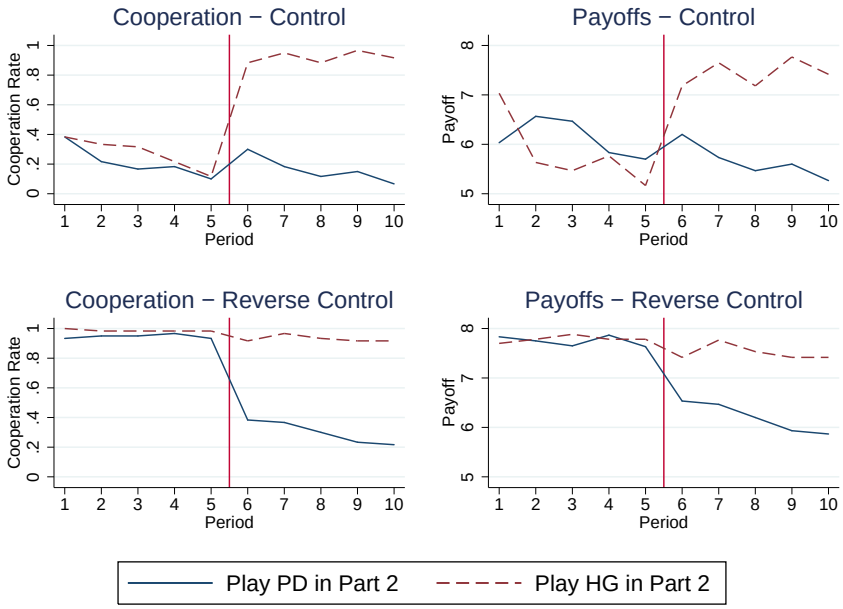

Figure 5: Comparing Prisoners' Dilemma and Harmony Game: Control and Reverse Control

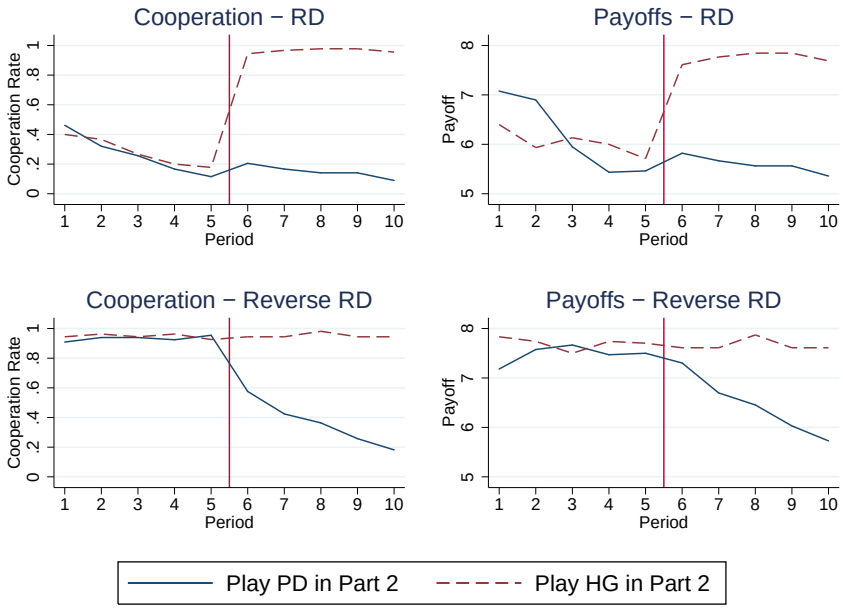

Figure 6: Comparing Prisoners' Dilemma and Harmony Game: Random Dictator and Reverse Random Dictator

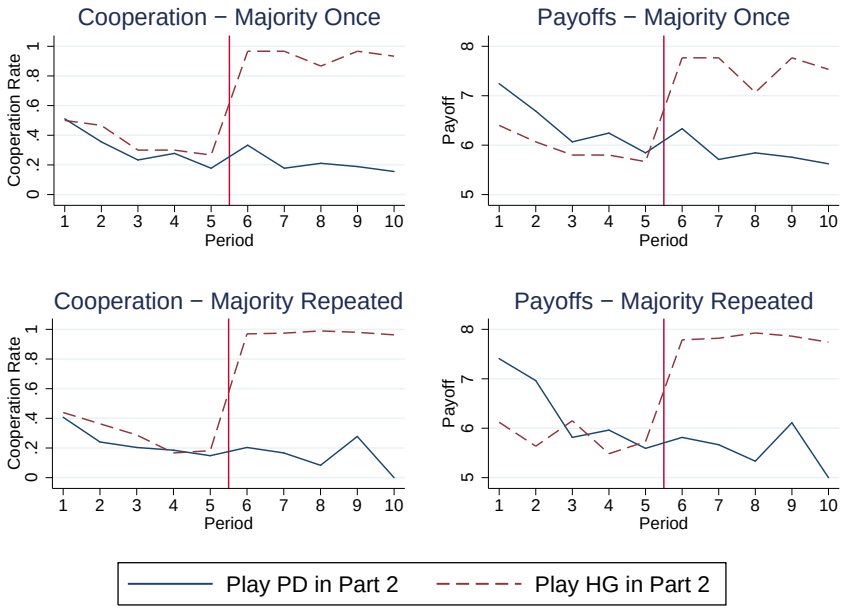

Figure 7: Comparing Prisoners' Dilemma and Harmony Game: Majority Once and Majority Repeated