Contingent Reasoning and Dynamic Public Goods Provision*

Evan M. Calford† and Timothy N. Cason‡

April 11, 2022

Abstract

Contributions toward funding a public good often reveal information that is useful to others who are considering their own contributions. This experiment compares static and dynamic contribution decisions to determine how hypothetical contingent reasoning differs in dynamic decisions where equilibrium requires understanding future information can inform about prior events. We fit a structural model that identifies three main types of individuals: A Nash type who is effective at contingent thinking; a partially cursed type who extracts only partial information from contingent events; and a type who is better at extracting information from past, rather than future or concurrent, events.

Keywords: Cursed equilibrium; Voluntary contributions; Club goods; Laboratory experiment

JEL Codes: C91, D71, D91, H41

*We thank William Brown for excellent research assistance, Gary Charness, Anujit Chakraborty, Caleb Cox, Simon Siegenthaler and Manu Vespa for valuable comments, as well as audiences at the ESA and M-BEES conferences and ANU, Innsbruck, UC-Davis, Melbourne, Oklahoma and New Zealand eSeminars. Funding for experimental subject payments was provided by Purdue University.

†John Mitchell Fellow, Research School of Economics, Australian National University, Canberra, Australia. Evan.Calford@anu.edu.au

‡Department of Economics, Krannert School of Management, Purdue University, Indiana, USA. cason@purdue.edu

1 Introduction

Collective action requires coordination and often involves uncertainty. Besides strategic uncertainty about other agents' behavior, in many realistic situations the value of taking collective action is unknown until it takes place. Moreover, in many, or perhaps even most, of these problems the uncertain value has a strong common value component that correlates individuals' benefits from public good provision. Examples range from global challenges such as climate change, to specific goals addressed in thousands of crowdfunding campaigns that support civic objectives, create art or develop new products. Optimal decision making in such environments therefore requires agents to recognize that others' support for a common goal encodes a positive signal about the value of collective action.

The temporal structure of collective decision making affects the nature, and difficulty, of this signal extraction problem. When decision making is simultaneous, agents' expectations of the value of collective action must be contingent on concurrent, unobservable, decisions of others. If decision making is dynamic, then expectations in early stages must condition on both concurrent and future decisions of others, while in later stages expectations need only be conditioned on others' realized, observable, behavior.

In this paper we seek to understand better these types of contingent reasoning (failures) by studying public good provision across static and dynamic environments. We report a laboratory experiment examining equilibrium predictions arising for fully rational agents who correctly condition on private and public information, as well as boundedly rational agents who have difficulty with contingent reasoning. We then estimate a novel structural model, decomposing failures of contingent reasoning into complexity and awareness components, that suggests three key types of subjects are present in our data: A Nash type, who fully extracts information from both the concurrent and prior behavior of others; a partially cursed Eyster-Rabin (2005) type who extracts partial information from the behavior of others; and an Esponda-Vespa (2014) type who extracts information only from others' prior, and not concurrent or future, behavior.

The implications of limited statistical reasoning by humans, and particularly the failure to understand how others' actions provide valuable information about their private information, began with early evidence of the "winners curse" in common value auctions (Capen et al., 1971; Kagel and Levin, 1986, 2002). Eyster and Rabin (2005) formalized this intuition, introducing the notion of "cursed equilibrium" and applying it to more general environments. In a cursed equilibrium for common value auctions, bidders best respond to incorrect beliefs that fail to account for how rival bidders' bids depend on their signals (for a survey see (Eyster, 2019)).[1]

Although there is evidence of contingent reasoning failure with respect to realized events (see Araujo et al. (2021), for example), reasoning about hypothetical contingencies is especially difficult. Esponda and Vespa (2014) found that subjects in their voting experiment were much better at drawing inferences from actual previous decisions of others, available only in a sequential setting, than hypothetical events needed to guide choices in a static setting.

In this paper we compare dynamic and static public goods provision, with the overarching goal to provide more insight into the source of difficulties people have with contingent reasoning. In contrast to previous studies, however, in which the sequential ordering is enforced, the agents choose the timing of their decisions in our dynamic environment, as illustrated in Table 1. That is, in our public goods setting, any agent in the Dynamic treatment can choose to contribute to the public good in the first round or may elect to delay the decision until later.

Here, we posit two possible sources for failures of contingent reasoning. One possibility is that individuals simply fail to recognize that there is information in others' decisions that they could find useful for their own judgments and

[1]Robust evidence of this type of limited rationality arises in a range of environments, from simplified nonstrategic settings such as the "Acquire a Company" problem (Bazerman and Samuelson, 1983; Charness and Levin, 2009) to voting (Esponda and Vespa, 2014) and non-auction market environments (Ngangoue and Weizsacker, 2021; Bochet and Siegenthaler, 2021; Carrillo and Palfrey, 2011; Magnani and Oprea, 2017). Few previous studies have explored how limitations for contingent reasoning affect choices in a common value public good setting (Cox, 2015).

	Timing		
	$T = 1$	$T = 2$	$T = 3$
Static (simultaneous)	P1, P2, P3		
Dynamic	P1, P2, P3	P1, P2, P3	P1, P2, P3

Table 1: Order of decision making under various timing structures, where $P1$ denotes player 1, $P2$ denotes player 2 and $P3$ denotes player 3.

belief updating; we refer to this type of naivete as *unawareness*. Alternatively, individuals may be aware that information exists that could be extracted and useful, but they have difficulty doing so when reasoning must be hypothetical. We call this *complexity*.

For unawareness to be a candidate for the *differential* of contingent reasoning between hypothetical and realized events, observation of an event must trigger awareness.[2] Thus, we allow for the arrival of information about others' behavior to trigger awareness of the information extraction problem. Unawareness is therefore characterized, here, as a failure to consider the correlation between the future actions of others and the private information that others hold.[3] Our structural model assumes that an unaware individual ignores the possibility that future decisions of others might encode information and, therefore, the individual behaves myopically in the Dynamic treatment.

On the other hand, complexity is characterized by the difficulty of extracting information from the behavior of others. The extraction problem is more complex when considering hypothetical events, and less complex when considering realized events. Our structural model introduces two dimensions of complexity: one that captures the complexity of reasoning about concurrent or future (i.e. hypothetical) events and one that captures the residual complexity of reasoning about past (i.e. realized) events. The distinction between

[2]For example, an individual who observes the prior behavior of others might ask themselves the question "why did a majority of others support this public good project? Do they know something I do not?" The same individual might simply consider only private information if choosing concurrently with others.

[3]We do not explicitly model the cause of this failure. Individuals are assumed to have full knowledge of the structure, timing, and payoffs of the game, however.

hypothetical and realized contingent thinking can be explained by complexity only when there is a divergence between these two measures.[4]

Several studies have shown how decision-makers are better able to exhibit contingent reasoning when information is not hypothetical, for example by providing information or making choices sequential so that payoff consequences of each contingency are more transparent (Esponda and Vespa, 2014, 2021; Levin et al., 2016; Ngangoue and Weizsacker, 2021) or by reducing the underlying uncertainty in the environment (Martinez-Marquina et al., 2019; Brocas and Carrillo, 2020). These previous papers were not designed to distinguish unawareness and complexity and so their identification is somewhat confounded: the introduction of information both simplifies the choice environment and brings attention to the information extraction problem.

Ali et al. (2021) provides some suggestive evidence that failures of contingent reasoning are caused by complexity, rather than unawareness. They found that over 75% of subjects who fail to correctly apply contingent reasoning to an adverse (or advantageous) selection environment were able to correctly answer factual questions that isolated each step of the contingent reasoning process. Correct answers to the factual questions implies that these subjects had an awareness of the logic of contingent reasoning, yet they were still unable to apply this logic to a more complicated, multi-step, environment.

The dynamic structure of our experiment allows for a separation of the unawareness and complexity explanations to better understand failures of contingent reasoning. First, note that a comparison of the second stage ($T = 2$) of the dynamic treatment with the static treatment is analogous to the comparison in the previous literature: information arrives before the second stage begins, and this information both simplifies the choice problem and highlights the existence of the information. Importantly, we can also compare behavior in the first stage of the dynamic treatment with behavior in the static treatment. If *complexity* is the source of contingent reasoning failures, then we should observe that subjects often prefer to delay decision making to future

[4] In Section 5 we discuss other, distinct, notions of complexity from the recent literature (Li, 2017; Martinez-Marquina et al., 2019; Oprea, 2020).

stages where the arrival of information will make the decision less complex.[5] However, if *unawareness* is the source of contingent reasoning failures then subjects will not recognize the value of waiting for information to arrive and should therefore behave identically across the static treatment and the first stage of the dynamic treatment.[6]

Our experiment also addresses new issues in information extraction and contingent reasoning. In previous studies comparing simultaneous and sequential choices to explore contingent reasoning, the environments are isomorphic in the sense that optimal choices and equilibrium outcomes do not vary when introducing the sequential game form.[7] This is not the case for our public goods provision problem, which features a more complex signal space, greater payoff uncertainty, and a richer dynamic structure.[8] This confronts decision makers with a novel information extraction problem not addressed in previous research. Here, agents have an option value from deferring their decision about whether to contribute to the public good whereas previous experiments with endogenous timing have mainly considered strictly informational interactions, without payoff consequences (Ivanov et al., 2009). The option value of waiting in our environment leads some to initially delay contributions, which they must trade off against the opportunity to signal to others their favorable information about the good's common value. Effective contingent reasoning

[5]Alternative explanations for delaying are also possible, of course, such as *herding*–i.e., waiting to copy what others do. Our results show an under-reaction of subjects' responses to earlier choices by others on average, however, providing evidence against this explanation.

[6]As is standard in lab experiments, our subjects play the same game multiple times. Thus, unawareness could diminish over time as subjects learn about the strategic environment. Our data exhibits only weak learning effects, however. This suggests that either unawareness never existed among the population, or that providing an opportunity to overcome unawareness is insufficient and so it could persist through multiple repetitions of the game. Esponda and Vespa (2014) also document suprisingly weak learning across sequential and simultaneous versions of their voting game. We return to this point in Section 5.

[7]Esponda and Vespa (2021), for example, effectively change the framing of the decision tasks on five classic problems, helping subjects focus on the set of states where their choice matters.

[8]Multiple equilibria exist in all of our treatments, which also raises interesting new questions about behavioral equilibrium selection with contingent reasoning. For our empirical analysis we focus on symmetric equilibria in which the public good is provided with positive probability.

leads to information sharing across multiple decision rounds in the dynamic treatment that more efficiently reveals information through the "richer" message space. The public good is, in equilibrium, provided less frequently in the dynamic than the static treatment, with a pronounced drop in provision when it has a low common value and should not be provided.

Our results indicate that a large fraction of subjects appreciate the benefits of deferring choice to learn about the contribution decisions of others when their signals about the public good value are near the margin. They also react to the information conveyed by others choices, and how others choice to select the public good signals a higher common value. The bias away from Nash equilibrium choices is in the direction of Cursed equilibrium on average, particularly in the static treatment. Overall, however, public good provision rates and errors in overprovision do not differ between the static and dynamic treatments, contrary to the equilibrium prediction. That is, while there is substantially less support for the public good in the first stage of the dynamic treatment than the static treatment, the aggregate provision rate in the dynamic treatment increases to static levels via additional contribution opportunities in the later stages.

In order to parsimoniously summarize our complete data set, we propose and estimate a simple structural model that decomposes a subject's potential failures of contingent reasoning into three components: cursedness when considering hypothetical events, cursedness when considering realized events, and an awareness component. The reduction in cursedness when moving from hypothetical to realized contingent thinking provides a measure of the effects of complexity on contingent thinking.

Using a clustering algorithm to classify subjects into groups we find that 71% of subjects exhibit awareness and perform equally well across hypothetical and realized contingent reasoning, including a Nash cluster who perform well in both cases and an Eyster-Rabin cluster who perform moderately in both cases. A third cluster, comprising of 24% of subjects, exhibits unawareness and the fingerprint of the results of Esponda and Vespa (2014): performing poorly in the case of hypothetical contingent reasoning but moderately well in the case of realized contingent reasoning.

2 Experimental design and hypotheses

2.1 Environment

In order to focus on individuals' possible difficulties with contingent reasoning our design simplifies the public good provision problem, as in Cox (2015), by making consumption of the public good (PG) excludable. Only individuals who support provision of the PG may receive PG benefits, which eliminates the usual free-rider problem and associated social preference concerns arising from relative payoff comparisons. Previous experiments have shown that exclusion of the lowest contributors (Croson et al., 2015), exclusion of individuals who fail to meet minimum contribution thresholds (Swope, 2002), or excluding those who do not pay a small "membership fee" (Ali Bchir and Willinger, 2013) usually raises total contributions.[9] Although it may be more accurate to refer to the good here as a club good, given its excludability, we follow convention in the experimental literature and use the term public good.

Several previous experiments have considered uncertainty in social dilemmas, including uncertain returns to contribution.[10] Our design features two key stylistic departures from the standard paradigm of public goods contribution games. First, we frame the decision as a binary choice between a PG and a private good to ameliorate any status quo bias that might arise when a subject is deciding whether to contribute, or not, to a PG from a private

[9]Unlike our experiment, these previous studies of excludable PGs employed a private value, complete information environment. Gailmard and Palfrey (2005) have incomplete information (but still independent private values) and compare the performance of the serial cost sharing mechanism, which achieves incentive compatibility through exclusion, to two alternative mechanisms that do not employ exclusion and are not incentive compatible.

[10]Some of these studies find that contributions are lower with uncertain public returns (Dickinson, 1998; Gangadharan and Nemes, 2009; Levati et al., 2009), while others do not indicate contribution impacts of uncertainty, such as Stoddard (2017). Few studies have considered uncertain returns to common-value public goods, other than Cox (2015). See Cox and Stoddard (2021) for further discussion, and a static public goods provision experiment with information sharing about public returns through (binary) cheap talk messages.

endowment.[11] Second, we introduce uncertainty in the value of the private good in order to equate the ex-ante risk profile of the two options and thereby avoid risk aversion biasing choice towards the private good.

More precisely, let us denote the agents by $i \in \{1, 2, 3\}$. The common value of the PG is given by $P = s_1 + s_2 + s_3$ where each s_i is an independent draw from a uniform discrete distribution over the interval 0 to 100. Agent i observes only signal s_i. The value of the private good, V_i, differs for each agent, and is given by $V_i = D_0 + D_{1,i} + D_{2,i}$, where D_0 is exogenous, common, and common knowledge across all three agents. The six other signals, $D_{j,i}$ for $j \in \{1, 2\}$ and $i \in \{1, 2, 3\}$, are unobserved and are each independent draws – also from a uniform discrete distribution over the interval 0 to 100. Therefore, after observing their own signals, each agent knows that the value of the PG is a known amount plus two iid draws from a uniform distribution, and that the value of the private good is also a known amount plus two iid draws from the same uniform distribution, providing the same level of uncertainty for both the private and public goods. The values of each good are summarized in Table 2. Note that contingent reasoning is not needed when forming expectations of the private good value, since these expectations are independent of any behavior.

	Player 1	Player 2	Player 3
Value of Public Good	$s_1 + s_2 + s_3$	$s_1 + s_2 + s_3$	$s_1 + s_2 + s_3$
Value of Private Good	$D_0 + D_{1,1} + D_{2,1}$	$D_0 + D_{1,2} + D_{2,2}$	$D_0 + D_{1,3} + D_{2,3}$

Table 2: A summary of the value of the Public and Private goods for each player. Values in red are observable, and values in black are unobservable.

A subject receives the PG if they and at least one other subject select the PG, and otherwise receives the private good.[12] In the static treatment, all three subjects make decisions simultaneously. In the dynamic treatment

[11] Cox (2015) also frames the decision choice as a binary one.

[12] We did not include a simplified treatment in which subjects interacted with computerized, robot players, as in the voting experiment of Esponda and Vespa (2014). Although such robots could be programmed to implement the equilibrium cutoff strategies described in the next subsection, such strategies are not best responses to suboptimal human play; moreover, describing such robot strategies to subjects could bias their behavior.

decision making occurs in three stages, with simultaneous decisions within each stage. In the first stage, each agent has the option to select either the PG or private good. If an agent selects the PG, then the decision is final and is revealed to others in the group, and that agent does not participate in stages two or three.[13] Signals always remain private information. If an agent selects the private good in stage one, then in stage two they observe how many other group members selected the PG in stage one. In this second stage they may switch to select the PG or continue to choose the private good. Once again, if the agent selects the PG then the decision is final and the agent does not participate in stage three. Agents who selected the private good in both stages one and two observe the number of PG decisions made by others in stages one and two and then, for the third and final time, they select either the PG or private good. The environment is deliberately simplified, as agents have only a single binary choice (whether or not to select the PG) each round, and only the signal draw varies from round to round. This simplicity limits potential subject confusion.

2.2 Equilibrium and hypotheses

Multiple equilibria exist in all of our treatments. For example, given the requirement that at least two agents must select the PG for it to be provided, it is always an equilibrium for no agent to select the PG.[14] We focus on (symmetric) equilibria in which the PG is provided with positive probability.

Appendix A presents details for the static and dynamic treatment equilibria for the experimental environment. Here we provide a short intuitive summary. Given the symmetric distribution of signals, a subject who chooses the private good will earn in expectation $\mathbb{E}[V_i] = \mathbb{E}[D_0 + D_{1,i} + D_{2,i}] = D_0 + 100$. A subject who ignores selection effects and chooses the PG would expect to earn $\mathbb{E}[P] = \mathbb{E}[s_1 + s_2 + s_3] = s_1 + 100$, because they ignore the fact that other

[13]Revealing prior commitments to the PG is analogous to the continuously updated cumulative prior contributions made by others on crowdfunding sites such as Kickstarter.

[14]Previous research has identified conditions in which this type of inefficient equilibrium is not trembling hand perfect in private value environments (Bagnoli and Lipman, 1989).

agents choice of the PG is good news indicating the PG has higher value. This comparison between the private good and PG naive expected value suggests the simple but incorrect decision rule of selecting the PG if and only if $s_1 \geq D_0$. This is exactly the decision rule implied by fully cursed equilibrium (Eyster and Rabin, 2005), which is characterized by every agent making an optimal decision under the erroneous assumption that other players make decisions that are not conditioned on their private information.

Given that the expected value of the PG is strictly increasing in player i's signal s_i, while the value of the private good is independent of s_i, it is always optimal for a subject to use a cutoff rule, selecting the PG only for signals above some threshold. Whenever this cutoff point is positive, the PG choice of other agents is informative of their private signal, and changes the expected value of the PG. If the equilibrium cutoff is X, for example, then an outside observer expects that the average signal for agents who select the PG is $(X + 100)/2$. This exceeds the unconditional expected value of 50 for any $X > 0$. Consequently, when an agent's PG choice is pivotal (because at least one other agent also chose the PG) it has an expected value that exceeds the unconditional average. Agents should therefore choose the PG more frequently when they account for this selection. In other words, the selection effect lowers the threshold cutoff value for choosing the PG.

We show in Section A.1, for the static treatment, how much lower these Nash equilibrium cutoffs are than the cursed equilibrium cutoffs for any $D_0 > 0$. As shown in Table 3 for the three values of D_0 used in the experiment, the PG is chosen weakly more often when agents correctly condition on the "good news" that they are more likely to be pivotal when other agents have high signals and also opt for the PG.

Calculations are more complex for the dynamic treatment, because knowledge that other agents did or did not choose the PG in previous stages affects the estimates of the PG value. Section A.2 provides derivations for the equilibrium using backward induction, where agents with sufficiently high signals select the PG in early stages rather than delaying.[15] Two forces determine

[15] Although equilibria exist with delay, they lead to lower expected payoffs and our ex-

the first stage cutoff value. First, there is an option value from deferring a decision to select the PG: the longer I wait, the more I can infer about the private signals of others. The option value of waiting pushes the first stage cutoff value higher in the dynamic treatment, relative to the static treatment.

Second, there is a signaling effect: if I have a good signal I wish to communicate this to others, and induce their entry, by entering as soon as possible. The signaling effect increases the value of selecting the PG immediately for high private signals (as this will encourage entry by others and increase the chances that the PG is provisioned), but it decreases the value of selecting the PG immediately for low private signals (as encouraging entry by others in this case can lead to inefficient provisioning of the PG).

Understanding the option value of waiting requires a subject to recognize that there is information that can be extracted, in the future, from current decisions of other players. In contrast to equilibrium reasoning in the simultaneous treatment, the extraction of this information in the second stage does not require hypothetical thinking. On the other hand, the signaling effect requires hypothetical thinking about the future behavior of other players, but does not require an ability to extract information from a signal.

In general, as illustrated in Table 3, cutoffs decline as more group members choose the PG in earlier stages. Due to the greater information dissemination from the sequential PG decisions, in the no-delay equilibrium players choose the PG more often in the rounds where it is efficient to do so in the dynamic treatment relative to the static treatment. Of course, these predictions are based on common knowledge of full rationality.

In addition to the equilibrium cutoffs for the static and dynamic treatments, Table 3 also summarizes the likelihood of the PG being provisioned in the static Nash, dynamic Nash and cursed equilibrium, and expected frequency of inefficient PG choices (due to lower earnings than the private good) based on the uniform distribution of signal draws. These treatment differences

perimental data provide no evidence consistent with them. Further, although the game is formally a Bayesian game, it is easily established that every optimal strategy is a simple cutoff strategy and, given this, that the game can be solved via backwards induction.

Private good base value (D_0):	0	30	70
Cursed equilibrium cutoff	0	30	70
Static equilibrium cutoff	0	25.0	52.1
Dynamic equilibrium cutoffs:			
Stage 1	47.7	58.9	73.0
Stage 2 (One prior PG choice)	17.5	33.7	55.7
Stage 3 (One prior PG choice in each stage)	0	4.3	19.2
Stage 2 (Two prior PG choices)	0	0	0
Public Good frequency:			
Cursed equilibrium	1.00	0.784	0.216
Static equilibrium	1.00	0.844	0.468
Dynamic equilibrium	0.844	0.656	0.360
Loss frequency (PG value < private good value):			
Cursed equilibrium	0.189	0.181	0.034
Static equilibrium	0.189	0.225	0.154
Dynamic equilibrium	0.073	0.088	0.080

Table 3: Top panel: Equilibrium cutoffs. Middle panel: Frequency of public good provision. Bottom panel: Probability that PG is provisioned and total utility is lower than if PG was not provisioned.

lead to the following hypotheses that are based on proper contingent reasoning and equilibrium choices, as well as naive (unconditional) cursed beliefs for Hypotheses 1(a) and 3.

Hypothesis 1: (Outcomes) (a) The frequency of selecting the PG decreases as the private good base value (D_0) increases; and (b) the conditional probability of (inefficiently) receiving the PG when the private good has a higher value is larger for $D_0 = 30$ than either $D_0 = 0$ or $D_0 = 70$.

Hypothesis 2: (Outcomes) (a) The PG is chosen more frequently in the static than the dynamic treatment; and (b) the PG is chosen when it has a lower value than the private good more frequently in the static than the dynamic treatment.

Hypothesis 3: (cursed equilibrium) Estimated PG choice signal cutoffs correspond to the private good base value (D_0) for both the static and dynamic treatments.

The remaining hypotheses are based on Nash rather than cursed equilibrium.

Hypothesis 4: (a) Subjects choose the PG with lower frequency in stage 1 of the dynamic treatment than in the static treatment; and (b) estimated signal cutoffs for choosing the PG are higher in stage 1 of the dynamic treatment than in the static treatment.

Hypothesis 5: (a) Subjects choose the PG at higher rates in later stages of the dynamic treatment if more other agents have previously selected the PG; and (b) estimated signal cutoffs in the dynamic treatment decrease for later stages when more other agents previously selected the PG.

Note that in the last two hypotheses each part (a) is closely related to part (b) in the sense that, assuming subjects are using cutoff strategies, part (a) holds if and only if part (b) holds in the limit as the number of observations per subject increases. We test both parts, however, as a robustness check on our results.

2.3 Laboratory procedures

The experimental design varied the common, baseline value of the private good at three levels, $D_0 \in \{0, 30, 70\}$, and whether the binary PG choice was static or dynamic. The three D_0 values allow for a wide range of equilibrium cutoffs and PG choice frequency to identify types of reasoning failures and their implications for efficiency (cf Table 3). The D_0 value varied between subjects, as it was kept constant throughout each experimental session. The static and dynamic treatments were varied within subject: each session included 20 consecutive rounds of the static treatment and 40 consecutive rounds of the dynamic treatment, in two blocks; the ordering was varied so exactly one half of the sessions in each D_0 treatment began with the dynamic treatment and one half began with the static treatment. Independent signals s_i and $D_{j,i}$ were drawn each round. We conducted twice as many rounds for the dynamic treatment in order to obtain a greater number of observations for stage two and three decisions in different subgames (0, 1 or 2 earlier PG choices by others).[16] At the end of each round, subjects learned the signals received by all 3 members of their group, as well as all 3 components of their private project value. They also learned the number of other subjects in their group who chose the PG, but not the specific signals received by those who did or did not choose the PG.

We collected data from a total of 144 subjects, with 48 subjects in each D_0 treatment. Subjects were randomly reassigned to new groups of 3 each round, out of matching groups of size 12, so each treatment included 4 independent observations. The subjects were all undergraduate students at Purdue University, recruited from a database of approximately 3,000 volunteers drawn across a wide range of academic disciplines and randomly allocated to treatment conditions using ORSEE (Greiner, 2015). The experiment was implemented using oTree (Chen et al., 2016). We used neutral framing, referring to choices between the "Group Project" or the "Private Project." Details are provided in the instructions given to subjects (see Online Appendix).

These written instructions were read aloud at the start of the session by an experimenter, after distributing a hardcopy to subjects. New complete instructions were distributed at the treatment switch (from simultanous to dynamic or vice versa), but only the changes were highlighted and read aloud. Each session concluded with two short "acquiring-a-company" game choices (both paid) for a separate measure of subjects' contingent reasoning. Sessions lasted about 1 hour each, including instructions and payment time. At the conclusion of each session earnings were paid out privately in cash for one randomly-drawn round for the main PG provision task. Subjects earned $26.69 on average per person, with an interquartile range of [$21.68, $28.21].

[16]Conducting the dynamic treatment using the strategy method was not an option given our research objective to study contingent reasoning.

3 Results

3.1 Public Good Choices and Provision

Hypothesis 4(a) states that agents should choose the PG with lower frequency in stage 1 of the dynamic treatment than in the static treatment. Figure 1 summarizes these individual choices for different PG value signals, providing support for this prediction for all three D_0 treatments. The figure also documents that subjects choose the PG more frequently for the treatments with a lower base value (D_0), consistent with Hypothesis 1(a), and they choose the PG with low s_i signals at substantial rates only for the lowest $D_0 = 0$. Aggregate PG choices do not, however, exhibit the sharp shift at equilibrium threshold signal levels (indicated on the figure as vertical lines). We consider individual threshold strategies in Section 3.3.

For signal values below the static Nash equilibrium threshold the predictions coincide for both the static and stage 1 dynamic treatments. Similarly, for signal values above the stage 1 dynamic Nash equilibrium threshold, the predictions also coincide for both treatments. Therefore, we should expect to see treatment differences only between these signal ranges. Figure 1 demonstrates that this is indeed the case, as the treatment differences are substantial for signals that fall between the two equilibrium cutoffs, denoted by vertical red lines, for all three values of D_0. Differences in PG choice frequencies for the static and dynamic treatments in these key signal ranges are highly statistically significant, based on linear probability models with standard errors clustered on individual subjects, controlling for time trends and treatment ordering. (Estimated $p -$ values < 0.01 for all comparisons.)

Result 1. *Subjects choose the PG more frequently when the private good has a lower value (support for Hypothesis 1(a)) and more frequently in the static treatment than in stage 1 of the dynamic treatment (support for Hypothesis 4(a)).*

Hypothesis 5(a) concerns later stage choices in the dynamic treatment, in particular that agents will choose the PG at higher rates in later stages of the

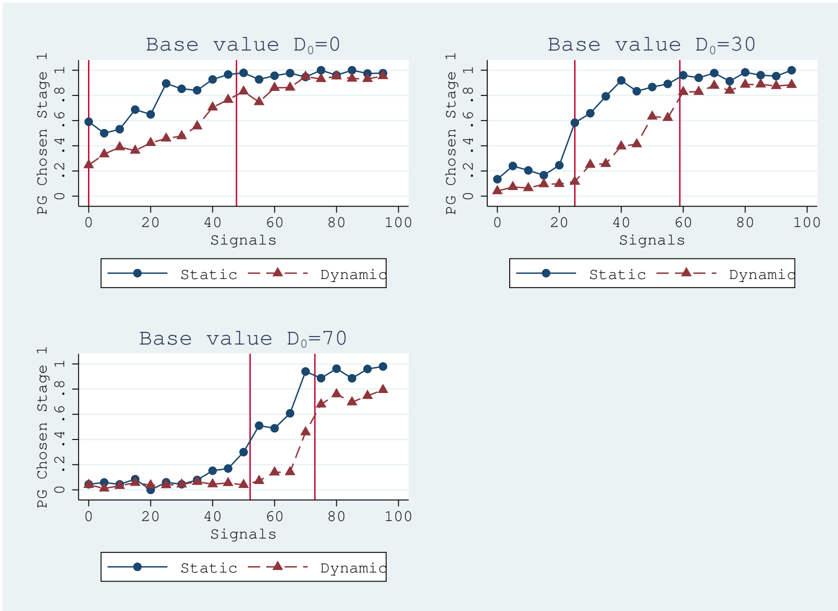

Figure 1: PG Choice Frequency by Signal, Static and Stage 1 Dynamic Treatments. Vertical lines denote Nash equilibrium predicted cutoffs, with the left (right) line corresponding to the static (dynamic) equilibrium.

dynamic treatment when more agents in their group previously selected the PG. Table 4 reports a series of linear probability models of subjects' choice of the PG in the second stage, conditional on the number of others in the group who selected the PG in the first stage. The omitted case is for zero other group members selecting the PG in the first stage. The models include the subject's own received signal (s_i) to control for the nonrandom selection (lower signal draws) of subjects who reach the later stages without having previously committed to the PG, and they also control for a time trend and treatment ordering.

The odd numbered columns report estimates without additional controls, while the even numbered columns add demographic characteristics as well as responses on the "acquiring-a-company" questions asked of subjects at the end of their session.[17] Results are similar with and without these controls.

[17]We employ multiple elicitations of this separate measure of individuals' comprehension of contingent reasoning and apply the *obviously related instrumental variables* method of Gillen et al. (2019) to attenuate measurement error.

The regression results show that having two rather than just one other subject choosing the PG previously has a particularly strong impact on the stage 2 PG decisions. For all six models the coefficient on two previous entries is significantly greater than for one previous entry (all p-values< 0.001). Subjects with higher signals are also significantly more likely to choose the PG in stage 2.[18] To summarize:

Result 2. *Subjects choose the PG at higher rates in later stages of the dynamic treatment if more members of their group have previously chosen the PG, particularly for two previous PG choices (support for Hypothesis 5(a))*

Hypothesis 2 concerns the overall provision rate for the PG, an outcome that depends on the decisions of multiple agents given that at least two people must select the PG for it to be provided. Table 5 displays the rate of PG provision by both treatment type (dynamic or static) and the value of D_0.[19] Hypothesis 2(a), that the PG is provided at a higher rate in the static than the dynamic treatment, is rejected by the data. In opposition to this hypothesis, the PG provision rate is actually higher in the dynamic than the static treatment for the $D_0 = 30$ and $D_0 = 70$ treatments.[20] In terms of point predictions, the average PG provision rate is remarkably close to the Nash Equilibrium (cf Table 3) for the dynamic treatment. Surprisingly, the PG provision rate for the static treatment is also similar to the dynamic equilibrium prediction, except for $D_0 = 70$ where the rate is even lower. The downward bias in the static treatment for the PG provision rate is in the direction of the cursed equilibrium for $D_0 > 0$, particularly for $D_0 = 70$.

Hypothesis 2(b) concerns errors in PG provision, in particular the provision

[18]Similar results obtain for stage 3 decisions, although we do not include them in Table 4 because the number of observations is lower and so the statistical significance is weaker, and the selection effect of nonrandom, low signal choices in the third stage is much stronger.

[19]No discernible time trend exists for the static treatment, but the dynamic treatment exhibits a modest decreasing time trend in the disaggregated early (periods 1-20) and later (periods 21-40). See Appendix C for details.

[20]We establish this using a linear probability model with clustered standard errors, controlling for time trends and treatment ordering. The differences between the static and dynamic treatment are significant at the 1-percent level for both the $D_0 = 30$ and 70 treatments, considering only the comparable first 20 periods of each treatment or all periods.

	Stage 2 PG for $D_0 = 0$		Stage 2 PG for $D_0 = 30$		Stage 2 PG for $D_0 = 70$	
	(1)	(2)	(3)	(4)	(5)	(6)
One other previous PG choice	0.083	0.054	0.063	0.064	0.135	0.134
	(0.044)	(0.114)	(0.034)	(0.032)	(0.021)	(0.020)
Two other previous PG choices	0.614	0.603	0.549	0.559	0.609	0.602
	(0.057)	(0.096)	(0.045)	(0.044)	(0.054)	(0.053)
Own signal (s_i)	0.0045	0.0049	0.0076	0.0075	0.0062	0.00611
	(0.0011)	(0.0018)	(0.0016)	(0.0014)	(0.0007)	(0.0007)
Round number t in treatment	0.0022	0.0020	0.0006	0.0009	-0.0001	-0.0002
	(0.0016)	(0.0023)	(0.0011)	(0.0011)	(0.0007)	(0.0007)
Treatment order in session	0.0025	0.0607	-0.0359	-0.0692*	-0.0158	-0.0111
	(0.0653)	(0.1980)	(0.0426)	(0.0329)	(0.0238)	(0.0243)
Demographic and ATC game controls	No	Yes	No	Yes	No	Yes
N	604	604	968	968	1439	1439
adj. R^2	0.347		0.343		0.334	

Standard errors (clustered on individual subjects) in parentheses.

Table 4: Stage 2 Public Good Choices in Dynamic Treatment.

of the PG to agents in rounds where the private good would have generated a higher payoff. One of the key theoretical advantages of the dynamic mechanism is that, in equilibrium, such over-provision of the PG is reduced compared to the static mechanism. The lower half of Table 5 shows that over-provision rates in the static treatment are lower than predicted, and over-provision rates in the dynamic treatment are higher than predicted. Differences between static and dynamic treatment error rates are not statistically significant in the $D_0 = 0$ and 30 treatments, but for the $D_0 = 70$ treatment the over-provision errors are significantly *greater* in the dynamic than static treatment.[21] Thus, the

[21]This conclusion is based on the same type of regression summarized in the previous

Private good base value (D_0):	0	30	70
Public Good frequency:			
Static	0.835	0.645	0.283
(standard error of mean)	(0.012)	(0.015)	(0.015)
Dynamic	0.836	0.672	0.359
(standard error of mean)	(0.008)	(0.011)	(0.011)
Loss frequency (PG value < private good value):			
Static	0.152	0.165	0.094
(standard error of mean)	(0.012)	(0.012)	(0.009)
Dynamic	0.158	0.178	0.123
(standard error of mean)	(0.008)	(0.009)	(0.008)

Table 5: Realized PG provision and overprovision for all treatments.

symmetric Nash equilibrium does not accurately predict the rates of provision and over-provision of the PG in aggregate.

Hypothesis 1(b) predicts that errors in PG provision are greater in the $D_0 = 30$ treatment than either the $D_0 = 0$ or $D_0 = 70$ treatments. Table 5 indicates that mean error rates are consistent with this hypothesized ordering, and regression-based statistical tests indicate that these differences are statistically significant for the $D_0 = 30$ to $D_0 = 70$ comparison in both the static and dynamic treatments (p-values< 0.001 in both cases). For the $D_0 = 30$ to $D_0 = 0$ comparison the $D_0 = 0$ treatment has significantly more errors in the dynamic treatment (one-tailed p-value$= 0.05$) but no significant difference in the static treatment. Thus, this hypothesis receives support in 3 out of the 4 relevant comparisons.

Result 3. *The PG provision rate in the dynamic treatment is greater than or equal to the rate in the static treatment, and the over-provision (error rate) is not lower in the dynamic treatment (Hypothesis 2 is not supported). Consistent with Hypothesis 1(b), over-provision of the PG is greatest in the intermediate*

footnote. Differences are significant at the two-percent level for $D_0 = 70$ regardless of whether all periods or only the first 20 periods are compared.

($D_0 = 30$) treatment.

Overall, the Nash equilibrium provides a useful approximation for aggregate PG provision rates in the dynamic treatment but not for the static environment. Nevertheless, aggregated outcomes may mask the choice of strategies at the individual level. In particular, the strategy choices of subjects may help explain the bias away from Nash outcomes, and towards outcomes predicted by Cursed equilibrium, in the static treatment. Thus, we now turn to a study of the strategies used by subjects in both our static and dynamic treatments.

3.2 Estimating strategies: Cutoff Intervals

Recall that a rational agent will use a cutoff rule in all scenarios: if the agent observes a signal s_i for the PG that is above some threshold the agent will prefer the PG and otherwise prefers the private good. Therefore, we summarize each subject's strategy with four points x_S, x_D, x_1 and x_2. x_S denotes the cutoff in the static treatment, x_D denotes the cutoff in the first round of the dynamic treatment, x_1 denotes the cutoff in the dynamic treatment when observing that one other player has already committed to the PG, and x_2 denotes the cutoff in the dynamic treatment when observing that both other players have committed to the PG.[22]

The subjects' binary choice data do not reveal their cutoff points directly.[23] We therefore infer their cutoffs using a deterministic process that provides interval identification of each cutoff point. Our procedure is maximally efficient: assuming a subject is using a cutoff rule we identify, conditional on the observed data, the smallest possible interval that contains the cutoff point. The

[22] Theory suggests that subjects might use a different cutoff in the cases where both other players are observed to commit to the PG in the first stage, and where one player commits in the first stage and another player commits in the second stage (cf Table 3). We do not have enough observations per subject to observe stage 3 cutoffs reliably in the data, so we instead focus only stage 2 decisions. This does not affect Hypothesis 5.

[23] Footnote 16 outlines why we did not use the strategy method in the dynamic treatment. We do not employ the strategy method in the static treatment either as doing so will make the use of signal contingent strategies more salient for subjects and, therefore, might also affect contingent thinking.

procedure is as follows. Order the k signals observed by the subject from smallest to largest, labeled s^1 through s^k, and denote the ordered set by S. The data are summarized by the mapping $d : S \to \{0,1\}^k$ where $d_j(S) = 1$ indicates that the subject selected the PG when signal s^j was observed, and $d_j(S) = 0$ indicates that the subject selected the private good.[24] Next, identify the set of $k+1$ possible intervals $I = \{I^0 = [0, s^1],\ I^1 = [s^1, s^2], \ldots, I^k = [s^k, 100]\}$. Then, for each interval, calculate an error index $E(I^j)$ for $j \in [0,k]$ as follows:

$$E(I^j) = \begin{cases} \sum_{i=j+1}^{k} 1 - d_i(S) & \text{if } j = 0 \\ \sum_{i=1}^{j} d_i(S) + \sum_{i=j+1}^{k} 1 - d_i(S) & \text{if } 1 \leq j < k \\ \sum_{i=1}^{k} d_i(S) & \text{if } j = k \end{cases}$$

If $\arg\min_{j\in[0,k]} E(I^j)$ is unique, then we conclude that the cutoff point must lie in I^j. If $\arg\min_{j\in[0,k]} E(I^j)$ is not unique, then we conclude that the cutoff point must lie in the interval $[s_{\underline{j}}, s_{\overline{j}+1}]$ where $\underline{j}$ and $\overline{j}$ are the smallest and largest minimizers, and we adopt the convention that $s_0 = 0$ and $s_{k+1} = 100$.

Finally, we adjust the observed intervals in the dynamic treatment to be consistent with an intuitive monotonicity condition: the cutoff at which a subject selects the PG should be non-increasing across stages of the dynamic treatment.[25] In some cases, subject behavior is not compatible with monotonicity. This usually occurs for subjects who have a high error index, suggesting that these subjects are not playing a cutoff strategy in the first place. On aggregate, however, subjects do appear to be implementing cutoff rules.[26]

In the remainder of this section we exclude subjects who violate the monotonicity conditions, and subjects who have a total error index of 6 or more.[27]

[24] Where $d_j(S)$ denotes the j-th element of the k-dimensional vector $d(S)$.

[25] For example, a subject who uses the same cutoff point in all stages of the dynamic treatment is never observed to select the PG in the second or third stages (implying that the upper bound of the interval is 100). We account for this be restricting the upper bound of the interval for x_1 and x_2 to be no greater than the upper bound for x_D.

[26] For x_S, with 20 observations per subject, the median error index is 0.5. For x_D, with 40 observations per subject, the median error index is 2. For x_1, with an average of 9.1 observations per subject, and x_2, with an average of 5.0 observations per subject, the median error indices are both 0. These low error indices provide evidence for consistency and against confusion among subjects.

[27] We calculate the total error index by aggregating errors made by a subject across all

Out of 144 subjects, 23 are excluded for monotonicity violations and a further 14 are excluded for an elevated error rate.

3.3 Analysis of strategies

We use interval regression techniques to estimate average cutoff strategies for the various stages and scenarios. The true, unobserved, cutoffs for each subject in each scenario are modeled as a normally distributed latent variable.[28] The interval regression estimates maximize the likelihood that the unobserved cutoffs lie within the intervals calculated above.

Once again we estimate our model both with and without demographic controls, including performance in the acquire a company game. Table 6 presents the average predicted cutoff value for each scenario. We estimate the model separately for $D_0 = 0, 30$ and 70. The estimated coefficients are remarkably robust to the demographic controls as are the standard errors, which are clustered at the subject level.

It is clear from the table that the average cutoff in the first stage of the Dynamic treatment is greater than the average cutoff in the static treatment, and also greater than the cutoff in the Dynamic treatment after others are observed to select the PG. These differences are significant at $p < 0.001$, except for the Dynamic treatment with one other selecting the PG where $p < 0.05$.

Result 4. *Estimated signal cutoffs for choosing the PG are higher in stage 1 of the dynamic treatment than in the static treatment (support for Hypothesis 4(b)) and estimated signal cutoffs decrease for later stages in the dynamic treatment when more other agents in the group choose the PG (support for Hypothesis 5(b)).*

The largest deviations from the Nash equilibrium point predictions occur in the Dynamic treatment when both opponents have already chosen the PG.

decision stages.

[28]The scenarios are the Static treatment, the first stage of the Dynamic treatment, the Dynamic treatment after observing one other subject select the PG, and the Dynamic treatment after observing both other subjects select the PG.

	$D_0 = 0$		$D_0 = 30$		$D_0 = 70$	
	(1)	(2)	(3)	(4)	(5)	(6)
Static	13.09	13.07	28.46	28.53	59.81	59.82
	(2.667)	(2.526)	(1.837)	(1.686)	(1.486)	(1.312)
Dynamic	28.58	28.55	55.35	55.35	76.97	77.05
(Stage 1)	(3.161)	(3.174)	(2.694)	(2.645)	(1.930)	(2.036)
Dynamic	24.30	24.23	42.35	42.50	59.58	59.54
(One previous entrant)	(3.016)	(2.941)	(2.095)	(2.074)	(1.462)	(1.479)
Dynamic	8.722	8.746	15.75	15.49	33.11	33.13
(Two previous entrants)	(1.806)	(1.596)	(2.020)	(1.854)	(4.321)	(4.094)
Treatment order controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographic and ATC game controls	No	Yes	No	Yes	No	Yes
N	128	128	148	148	152	152

Standard errors, clustered at the subject level in parentheses.

Table 6: Average predicted cutoff value by treatment and number of observed entrants, calculated via interval regression. Restricted to subjects who do not violate the monotonicity constraints and exhibit 5 or fewer choice errors.

The NE predicts that the PG should always be selected in this case, i.e., a cutoff of 0, but Table 6 shows that the estimated cutoffs are substantially larger than 0 in all treatments ($p < 0.001$ for all cases.) Subjects under-react to the information encoded in observing both opponents selecting the PG. Given the observed first stage behavior, the best response cutoff is 1, 0 and 0 in the $D_0 = 0, 30$ and 70 treatments respectively.[29]

Subjects are much closer to best responding to the information encoded in observing only one opponent choose the PG, although there is still some

[29] $p < 0.001$ for all treatments. Best responses are calculated using Equation 2 in Appendix A, using $\chi_R = 0$ and replacing p_0^* with the empirical distribution of first stage behavior.

Figure 2: Estimated cutoff values with 95% confidence intervals, controlling for demographic characteristics. Cursed equilibrium prediction denoted by dashed red lines, and Nash equilibrium predictions denoted by solid green lines.

under-reaction in the $D_0 = 30$ and 70 treatments. We calculate the best response cutoff after observing one selection of the PG to be 24, 35 and 54 in the $D_0 = 0, 30$ and 70 treatments, respectively. From Table 6 the estimated values of these cutoffs are 24, 42 and 60, respectively.[30]

Figure 2 plots the estimated cutoff values along with equilibrium predictions. Nash equilibrium predicts the direction of treatment effects across variation in both D_0 and the timing of the game. As a point prediction Cursed equilibrium does not perform well in our data, and Hypothesis 3 is rejected in 11 out of 12 cases. However, Cursed equilibrium does help to organize our data in the static treatment, where deviations from Nash equilibrium are in the direction of Cursed equilibrium (whenever the Nash and Cursed equilibrium differ). In the case of $D_0 = 30$, the 95% confidence interval of the average cutoff value covers the Cursed equilibrium but not the Nash equilibrium. In

[30] $p = 0.937$, $p < 0.001$ and $p < 0.001$, respectively, for tests comparing the estimated cutoffs to these best responses. Best responses are calculated using Equation 3, assuming $\chi_R = \chi_H = 0$ and replacing p_0^* with the empirical distribution of first stage behavior.

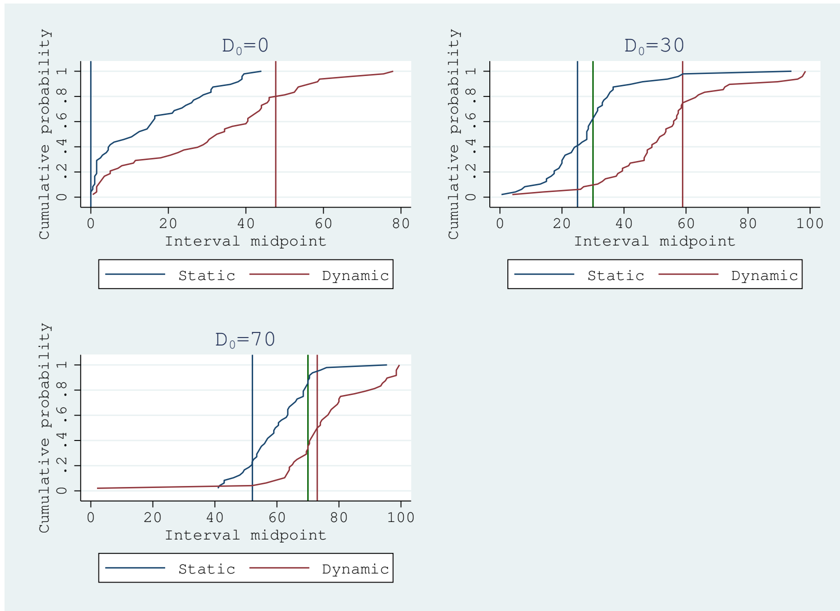

Figure 3: Cumulative Density Functions of the midpoint of the cutoff intervals (see Section 3.2). Vertical lines denote equilibrium predictions for Static Nash equilibrium, Dynamic Nash equilibrium and Cursed equilibrium in navy, maroon and dark green, respectively.

the Dynamic treatment, the Nash equilibrium point predictions perform well in the first stage, across the $D_0 = 30$ and $D_0 = 70$ treatments, and after observing one other player select the PG but not otherwise. Average cutoff values tend to lie between the Cursed and Nash equilibrium point predictions when D_0 is non-zero.

Figure 3 plots the CDF of the midpoint of the cutoff intervals for individual subjects in the Static treatment and the first stage of the Dynamic treatment. There is a clear rightward shift in the distribution of cutoffs in the Dynamic treatment, relative to the Static treatment, along with a substantial amount of within treatment heterogeneity across subjects.

In order to better understand the source of this heterogeneity, we next turn to estimation of a structural model that clusters subjects based on their belief updating biases.

4 Structural model overview and estimation

The structural model extends the model of cursed thinking that underlies the Cursed Equilibrium of Eyster and Rabin (2005). In the Eyster and Rabin model, an agent is parsimoniously modeled by the degree of cursedness, $\chi \in [0, 1]$. An agent with $\chi = 0$ is not cursed at all, and updates beliefs in a rational, Bayesian, fashion. An agent with $\chi = 1$ is fully cursed and thereby ignores any correlation between the actions of other agents and the private information that those agents hold. Agents with $0 < \chi < 1$ are then modeled as holding beliefs that are a linear combination of the beliefs of the rational and fully cursed types, with the relative weight being determined by χ.

To illustrate, in the context of our public goods game, suppose that an agent believes that each opponent will select the public good with probability p and, for ease of exposition, we approximate the discrete signal space of our experiment with a continuous signal space on the interval $[0, 1]$. The opponents use a cutoff strategy, such that they select the public good for signals $s \geq (1 - p)$ and select the private good for signals $s < (1 - p)$. If the agent is Bayesian ($\chi = 0$) then the agent believes the expected value of an opponents signal, conditional on the opponent selecting the public and private goods, respectively, to be $\mathbb{E}[s_j|PG] = 1 - \frac{p}{2}$ and $\mathbb{E}[s_j|RG] = \frac{1-p}{2}$.[31] In contrast, a fully cursed agent ($\chi = 1$) holds beliefs such that $\mathbb{E}[s_j|PG] = \mathbb{E}[s_j|RG] = \frac{1}{2}$. The general case, for $0 \leq \chi \leq 1$ is then given by $\mathbb{E}[s_j|PG] = 1 - \frac{p}{2} - \chi\frac{1-p}{2}$ and $\mathbb{E}[s_j|RG] = \frac{1-p}{2} + \chi\frac{p}{2}$.

Appendix A provides technical details of the model, which includes three parameters that capture different aspects of potential belief updating failures. Each parameter of the model is easily interpreted. There are two cursedness parameters, χ_H and χ_R. χ_H is the subject's cursedness when considering hypothetical events or, equivalently, when extracting information from the concurrent or future decisions of others. χ_R is the subject's cursedness when considering realized events or, equivalently, when extracting information from

[31] We use the notation $\mathbb{E}[s_j|X]$ to denote the expectation of agent j's signal, conditional on agent j selecting X, where $X = PG$ denotes selecting the Public Good and $X = RG$ denotes selecting the pRivate Good.

the past decisions of others. We impose a natural restriction that $\chi_H \geq \chi_R$ to reflect the intuition that extracting information from hypothetical events is more difficult than extracting information from realized events (Esponda and Vespa, 2014). Both χ_H and χ_R allow for a standard interpretation, in the style of Eyster and Rabin (2005), as a measure of a subject's misperception of the correlation between an opponent's behavior and an opponent's private information.

The third parameter is a binary variable, ψ, that captures the (un)awareness of a subject. To operationalize the somewhat abstract notion of awareness, we specify that a subject who is unaware of the information contained in others' decisions reveals this unawareness by ignoring future periods where this information could be utilized. We define $\psi = 1$ to represent an unaware subject, who ignores the future, and $\psi = 0$ to represent an aware subject who contemplates the future. Thus, our structural model of dynamic Cursedness may also be interpreted as a model of attention (Gabaix, 2014). In this interpretation, the structural parameters χ_H and χ_R dictate how *much* attention a subject focuses on a particular event. On the other hand, the parameter ψ indicates *which* events the subject focuses attention on.

Figure 4 displays the parameter space of the model, and demonstrates how it nests key models of conditional thinking from the literature. The right panel represents unaware agents with $\psi = 1$, and the left panel agents with $\psi = 0$. In each panel, χ_H is displayed on the horizontal axis and χ_R on the vertical axis. A Nash subject, who always correctly applies Bayesian updating, is represented in the bottom left corner of the left hand panel. Subjects who conform to the Eyster and Rabin (2005) model lie along the $\chi_R = \chi_H$ diagonal in the left panel, with fully cursed subjects at the top right of the panel at $\chi_R = \chi_H = 1$.[32] Subjects in the bottom right corner of each panel behave consistent with Esponda and Vespa (2014): fully cursed in the Static treatment, but not

[32]When $\chi_R = \chi_H = 1$, the ψ parameter has no effect on behavior. Intuitively, if a subject ignores all potential correlation of others actions and their possible information, and never updates initial beliefs, then it does not matter whether the subject is aware of the future arrival of information as the subject will not use the information.

cursed in the later stages of the Dynamic treatment.[33]

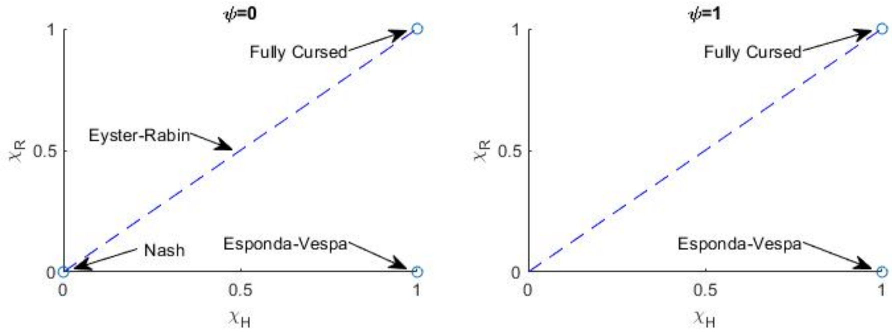

Figure 4: The parameter space of the structural model. The right panel represents unaware agents with $\psi = 1$, and the left panel agents with $\psi = 0$. The model resides in the triangles below the 45-degree line in each panel.

The intuition underlying the structural model is readily apparent in the figure. The complexity cost of hypothetical reasoning can be represented by the relative weights placed on cursed beliefs for hypothetical and realized events, measured by the difference $\chi_H - \chi_R$. In the diagram, this is captured by the vertical distance between any given point and the dashed 45-degree line. A subject who exhibits unawareness (right panel) ignores the future when making any decision, while one who is aware (left panel) allows the shadow of the future to affect current decisions. Thus, an unaware subject will behave identically in the first stage of the Dynamic treatment as in the Static treatment, while a subject with awareness will behave differently across the two environments.

The model is estimated in three steps. First, the theoretical model generates an equilibrium mapping from the three preference parameters χ_H, χ_R and ψ to a five dimensional strategy profile $(y_S, y_0, y_1, y_2, y_{1,1})$. y_S denotes the cutoff for the Static treatment, y_0 the cutoff for the first stage of the Dynamic treatment (i.e. after observing 0 other players select the PG), y_1 and y_2 the second stage cutoffs after observing 1 or 2 other players select the PG

[33]A subject at $\chi_H = 1, \chi_R = 0$ and $\psi = 1$ will be distinguished from a subject at $\chi_H = 1, \chi_R = 0$ and $\psi = 0$ in the first stage of the Dynamic treatment, but the framework of Esponda and Vespa (2014) does not provide a behavioral hypothesis in this case.

in the first stage, and $y_{1,1}$ the third stage cutoff after observing one other player select the PG in each of the first two stages. We restrict attention to symmetric equilibria to ensure a unique mapping from preference parameters to strategies. Formally, our symmetry assumption is that each agent believes that others will select the PG with the same *probability* as the agent.[34] One possible interpretation, but not the only one, is that the preference parameters χ_H, χ_R and ψ are identical across players. The Static treatment is solved as a simultaneous game. The Dynamic treatment is solved using backwards induction when $\psi = 0$, but solved as a sequence of static, simultaneous, games when $\psi = 1$.

Second, given the equilibrium cutoffs, y_t, as a function of χ_H, χ_R and ψ, we estimate the parameters at the individual subject level using maximum likelihood (Section 4.1). We assume that agent i selects action $a_{i,r,t} = 1$ in round r, when faced with decision state t, whenever

$$\lambda_i(s_{i,r,t} - y_{i,t}) + \varepsilon_{i,r,t} > 0 \tag{1}$$

where λ_i is a positive scale parameter and $\varepsilon_{i,r,t}$ is a random error term with a logistic distribution, and selects $a_{i,r,t} = 0$ otherwise.[35] This functional form imposes larger likelihood penalties for "mistakes" that are made when a subject's signal is further from the equilibrium cutoff, and the scale parameter λ can be interpreted as a logit goodness-of-fit parameter. Large values of λ indicate a good model fit, and $\lambda = 0$ indicates that the model fits subject behavior no better than a uniform random distribution over actions. The maximum likelihood estimates for a subject are the set of structural parameters that generate a set of cutoff points that best match the data for that subject.

Third, we employ a maximum likelihood clustering model to identify statistically similar groups of subjects.[36] This algorithm endogenously determines the number of clusters, and the preference parameters associated with each

[34] Note that, because agents may be cursed, this does not necessarily imply that the agent believes that others are using the same cutoff strategy as the agent.

[35] The decision state, t, takes a value in the set $\{S, 0, 1, 2, \{1, 1\}\}$.

[36] A distinct class of machine learning clustering algorithms reaches similar conclusions.

cluster.

4.1 Estimation results

Figure 5 displays the subject level structural parameter estimates.[37] We only estimate the model for subjects in the $D_0 = 30$ and $D_0 = 70$ treatments; because the Nash and Cursed equilibrium coincide in the Simultaneous treatment when $D_0 = 0$, this treatment does not provide enough variation in cutoffs to reliably estimate the structural parameters.

Figure 5 displays substantial heterogeneity across subjects, a result that is to be expected given the heterogeneity already documented in Section 3.3. The clustering algorithm endogenously determines both the number and the locations of the clusters within the parameter space. That is, we do not specify the clusters ex-ante. Remarkably, three out of the four clusters identified by the algorithm are easily recognizable as having antecedents in the prior literature. This is illustrated in Figure 6, where the size of each circle is proportional to the number of subjects contained in that cluster.

The largest cluster, consisting of 39% of subjects, exhibits Eyster and Rabin (2005) partially cursed beliefs. This cluster exhibits awareness ($\psi = 0$) and has cursedness parameters of $\chi_R = \chi_H = 0.54$. This first cluster exhibits a constant and partial degree of cursedness across all decision environments, and considers future stages when making decisions in the Dynamic treatment.

The second cluster, consisting of 32% of subjects, is a Nash cluster with $\chi_R = \chi_H = 0$ and awareness ($\psi = 0$). That is, about one third of the subjects are estimated to exhibit no cursedness, and they consider future stages when making decisions in the Dynamic treatment.

The third cluster, consisting of 24% of subjects, exhibits behavior that is broadly consistent with the behavioral hypothesis of Esponda and Vespa (2014). This type is almost fully cursed when considering hypothetical decisions ($\chi_H = 0.94$), and is myopic ($\psi = 1$). When extracting information

[37] A table showing the subject level estimates, including bootstrapped confidence intervals, is available in Appendix C. The table documents substantial width in the confidence intervals for many subjects.

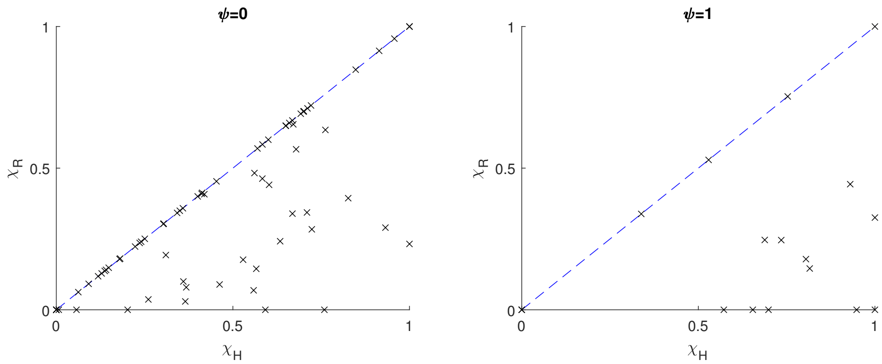

Figure 5: Individual level structural parameter estimates.

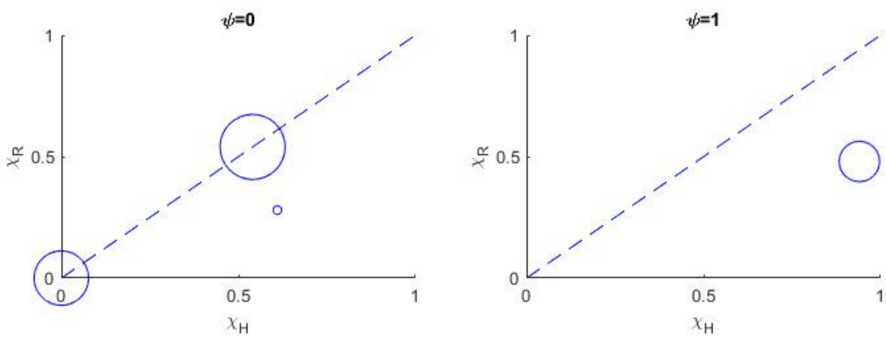

Figure 6: Results from the clustering algorithm. Cluster size is proportional to the number of subjects in each cluster.

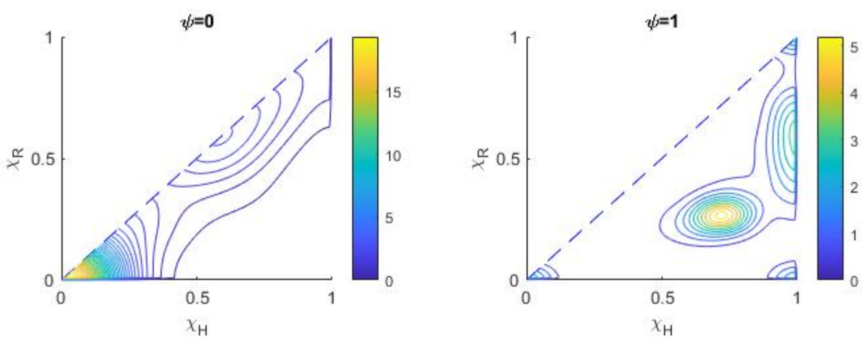

Figure 7: Contour map showing the probability density of clusters calculated from 800 bootstrapped samples. Contour lines are spaced 0.5 units apart. Note that the color scale differs between the left and right panels.

from previously realized decisions, this type exhibits only partial cursedness ($\chi_R = 0.48$). While a strict application of the ideas in Esponda and Vespa (2014) would imply $\chi_H = 1$ and $\chi_R = 0$, the subjects in this cluster still exhibit behavior that is consistent with Esponda-Vespa reasoning: they are substantially better at extracting information from realized, as compared to hypothetical, events. In effect, these subjects appear to ignore all information that is available from either concurrent or future decisions, but extract partial information from past decisions.

The final cluster, consisting of only 5% of subjects, contains subjects who are apparently not playing cutoff strategies. The structural model finds that no parameter combination fits the behavior of this group of subjects well, and they have a goodness-of-fit parameter of $\lambda \approx 0$. Thus, the position of this cluster in our three dimensional parameter space has little interpretable meaning.

A bootstrapping procedure provides evidence for the consistency of this structural model and clustering procedure. For each bootstrap we draw a new sample randomly from the data at the group-round level, and then re-estimate the model and clustering algorithm.[38] Figure 7 displays a contour map of the probability density function of clusters given 800 bootstraps (i.e. 3200 clusters).[39] The largest density occurs at $\chi_R = \chi_H = 0$ and $\psi = 0$, demonstrating the robustness of the Nash equilibrium cluster across bootstraps. While greater variation exists in other regions of the parameter space, there is a clear mass of density along the Eyster and Rabin (2005) partially cursed diagonal. There are also two masses of density with unawareness ($\psi = 1$), with the Esponda and Vespa (2014) cluster situated closer to the mass with $\chi_H = 1$ and $\psi = 1$.

[38]For computational reasons, we assume $K = 4$ clusters for all bootstraps rather than allowing the number of cluster to be determined endogenously as in the original data analysis.

[39]We use a kernel density estimator that corrects for boundary effects using the reflection method.

5 Discussion

In the previous section we applied our structural model to a particular experimental design. It is, however, conceptually straightforward to apply the same three parameter structural model to other dynamic games of private information with interdependent values.[40] It remains an open question, for future research, whether the specific parameter estimates found here are directly applicable to other environments. For example, estimates of χ_H and χ_R may be higher in more complex decision making contexts, and lower in environments where subjects find it easier to process information. In particular, considering the results of Martinez-Marquina et al. (2019), it would be reasonable to expect the difference $\chi_H - \chi_R$ would be greater in decision making environments where the state space is larger (or otherwise more complex). The awareness of subjects, as measured by ψ, might also vary across environments. In less complex decision making environments subjects might realize that future decisions of others are important, while in more complex environments this dependency may be shrouded.

The results from our clustering algorithm (Figure 6) are generally supportive of the two models in the literature that are most naturally applied to our environment: Nash equilibrium and Cursed equilibrium. Fully 71% of the subjects are assigned to clusters that coincide exactly to either Nash equilibrium or Cursed equilibrium behavior. An alternative interpretation of this result is that 71% of the subjects do not exhibit a distinction between contingent thinking that is applied to hypothetical or realized opponent behavior.

The third largest cluster, consisting of 24% of the subjects, does process information from hypothetical and realized behavior differently. The source of the failure of hypothetical contingent reasoning is *both* unawareness and complexity. This cluster exhibits unawareness ($\psi = 1$) and faces a moderate complexity penalty ($\chi_H - \chi_R = 0.46$). Notably, this cluster of subjects remains partially cursed even when considering realized events ($\chi_R = 0.48$). Thus, our

[40]Identification of the structural model requires both static and dynamic environments. We are not aware of any previous work that includes a data set rich enough to estimate our model.

results are related to, but distinct from, those of Esponda and Vespa (2014).[41]

Also of interest is that we detect only weak time trends in aggregate outcomes, which suggests limited learning. Moreover, the individual level strategy estimation does not present substantial evidence of learning.[42] In particular, it might appear surprising that a substantial proportion of our subjects are identified as being unaware of inference problems in the first stage of the dynamic treatment, yet at least partially resolve the inference problem in later stages, without realizing their mistake after repeated plays of the game. This is reminiscent of a surprising finding in Esponda and Vespa (2014). Through repeated play in their sequential voting treatment subjects displayed an understanding of what to do in each realized contingency–but when they subsequently faced the simultaneous version they behaved as if they did not have this experience.

There are two plausible explanations for this lack of learning among unaware subjects in our experiment. First, subjects may become aware, through experience, that they *should* consider extracting information from the hypothetical behavior of their opponents yet remain unable to figure out *how* to do so and, therefore, do not alter their behavior. Alternatively, a subject may never realize that they should extract information from the hypothetical behavior of opponents because they do not pay attention to the feedback provided at the end of each round. Such an outcome would be consistent with models of rational inattention, which have recently found experimental support in Dean and Neligh (2019) and Martin (2016).

Finally, note that the model's measure of complexity, the difference $\chi_H - \chi_R$, is a behavioral proxy for the underlying complexity of contingent thinking. Primarily, we consider the underlying source of complexity to be a complexity of calculation. Even when a person knows that they should condition a calculation on particular state(s) of the world, the mere existence, and potential

[41]This difference is likely a function of the different games used in the two studies. In the voting game of Esponda and Vespa (2014) it is, arguably, harder to recognize the inference problem in the hypothetical context but easier to perform the inference conditional on having recognized it.

[42]If subjects were learning, or adjusting their intended cutoff targets over time, then we would expect to observe greater error indexes when estimating subject level cutoffs in Section 3.2.

realization, of non-payoff relevant states makes the calculations more difficult to perform.

This intuition is also reflected in recent work by Martinez-Marquina et al. (2019), who decompose the complexity of contingent thinking into two components: a complexity induced by the number of potential outcomes ("states") that must be considered, and a complexity induced by a lack of certainty. Thus, our notion of complexity incorporates the computational complexity of Martinez-Marquina et al. (2019) but does not include the power of certainty, and we do not attempt to decompose complexity into its component pieces.

Related, yet distinct, notions of complexity are found in the recent literature. Oprea (2020) has studied complexity as it relates to the implementation of strategies where, among other conditions, the implementation of a strategy is perceived as more complex the greater the number of states it must be conditioned on. Li (2017) has studied complexity in the context of mechanism design, where a strategy is defined to be "obviously strategy proof" if its lowest possible payoff is greater than the largest payoff of any other strategy. A mechanism that has an obviously strategy proof strategy can be considered to be less complex than one that does not.

6 Conclusion

As noted in the introduction, it is uncommon for threshold PG environments to involve simultaneous decision making. Nevertheless, the incentives for a threshold PG game can be modified to reflect either the Dynamic or Static treatments by revealing, or not revealing, respectively, the current level of contributions in real time. The choice of information structure is therefore an important and easily manipulable policy variable for the designer of a threshold PG mechanism.

Our theoretical analysis suggests that there is an important tradeoff between the Static and Dynamic mechanisms. The Static mechanism generates a higher rate of PG provision, but achieves this, in part, by increasing the proportion of times that the PG is provisioned inefficiently. The Dynamic mechanism can improve PG provision choices, but introduces greater complexity. For a crowdfunding company such as Kickstarter this implies a tradeoff between revenue (which is a function of the number of projects that are financed) and long term reputation (which is harmed when consumers purchase or support a poor product).

The experimental results do not support this theoretical tradeoff. The rate of PG provision in the Dynamic treatment is, if anything, slightly *higher* than in the static treatment and we do not find a difference in the rate of "mistakenly" provisioned PG. Further, because the threshold for commiting to the PG is substantially higher in the first stage of the Dynamic treatment than the Static treatment, there will be fewer near-misses (where a PG almost, but not quite, reaches the funding threshold) in the Dynamic mechanism. Each of these properties suggests that the Dynamic mechanism is likely to be more desirable from a practical standpoint. It is therefore perhaps no accident that Kickstarter and other crowdfunding sites typically update previous contributions continuously to promote information dissemination in their versions of a dynamic mechanism.

Our experiment introduced a dynamic treatment with an endogenous choice of public good contribution timing in order to distinguish unawareness from previously studied types of complexity leading to failures of contingent reasoning. We find evidence, consistent with the existing literature, for failures of contingent thinking: decisions in the Static treatment are biased towards Cursed equilibrium. We also find that strategies in the first stage of the Dynamic treatment differ significantly from strategies in the Static treatment. This suggests that, in aggregate, unawareness is not a primary determinant of behavior. Unawareness implies ignorance about the option value of waiting, and our structural analysis classifies only a minority (about one-quarter) of subjects as unaware. We also find that *unawareness* and *complexity* effects of hypothetical contingent reasoning are correlated across individuals. Although human beings may be innately aware of the need for contingent thinking, and that they should take actions that allow their future selves to make use of valuable contingent information, it appears that many have difficulty in effectively solving contingent thinking problems optimally.

A recommendation that arises in more natural and less structured environments is that agents who are able to effectively solve the contingent thinking problem could, to improve the decision making of others, propose explicitly conditional strategies.[43] For example, an agent who announces "I will do X if, and only if, you do Y" simultaneously draws attention too, and reduces the complexity of, the contingent thinking problem for others. We are starting to see examples of this sort of innovative strategic thinking for policies to address climate change. Crampton (2021), for instance, recently proposed that New Zealand should adapt their current Emissions Trading Scheme price cap, which is currently a fixed cap, to float with the carbon price in other jurisdictions. A floating price cap will tie the cost of emissions abatement in New Zealand, and therefore the amount of emissions abatement, to the strength of emissions targets set in other countries (most notably, the European Union). Thus, contingent actions are built into the regulation directly.

References

Ali, S., Mihm, M., Siga, L., and Tergiman, C. (2021). Adverse and advantageous selection in the laboratory. *American Economic Review*, 111(7):2152–2178.

Ali Bchir, M. and Willinger, M. (2013). Does a membership fee fost successful public good provision? an experimental investigation of the provision of a step-level collective good. *Public Choice*, 157:25–39.

Araujo, F. A., Wang, S. W., and Wilson, A. J. (2021). The times the are a-changing: Experimenting with dynamic adverse selection. *American Economic Journal: Microeconomics*, 13(4):1–22.

Bagnoli, M. and Lipman, B. L. (1989). Private provision of public goods: Fully implementing the core through private contributions. *The Review of Financial Studies*, 56(4):583–601.

Bazerman, M. H. and Samuelson, W. F. (1983). I won the auction but don't want the prize. *Journal of Conflict Resoution*, 27(4):618–634.

Bochet, O. and Siegenthaler, S. (2021). Competition and price transparency in the mar-

[43] While nudging the behavior of others towards Nash equilibrium is not beneficial in all environments, it is beneficial in the type of common value environments studied in this paper.

ket for lemons: Experimental evidence. *American Economic Journal: Microeconomics*, 13:113–140.

Brocas, I. and Carrillo, J. (2020). Adverse selection and contingent reasoning in preadolescents and teenagers.

Capen, E., Clapp, R., and Campbell, W. (1971). Competitive bidding in high-risk situations. *Journal of Petroleum Technology*, 23:641–653.

Carrillo, J. D. and Palfrey, T. R. (2011). No trade. *Games and Economic Behavior*, 71(1):66–87.

Charness, G. and Levin, D. (2009). The origin of the winner's curse: A laboratory study. *American Economic Journal: Microeconomics*, 1(1):207–236.

Chen, D. L., Schonger, M., and Wickens, C. (2016). otree – an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Cox, C. (2015). Cursed beliefs with common-value public goods. *Journal of Public Economics*, 121:52–65.

Cox, C. and Stoddard, B. (2021). Common-value public goods and informational social dilemmas. *American Economic Journal: Microeconomics*, 13:343–369.

Crampton, E. (2021). An alternative approach to setting NZ's emissions trading scheme price cap. `https://www.nzinitiative.org.nz/reports-and-media/opinion/new-opinion-45/`.

Croson, R., Fatas, E., Neugebauer, T., and Morales, A. (2015). Excludability: A laboratory study on forced ranking in team production. *Journal of Economic Behavior and Organization*, 114:13–26.

Dean, M. and Neligh, N. (2019). Experimental tests of rational inattention. Mimeo.

Dickinson, D. (1998). The voluntary contributions mechanism with uncertain group payoffs. *Journal of Economic Behavior and Organization*, 35:517–533.

Esponda, I. and Vespa, E. (2014). Hypothetical thinking and information extraction in the laboratory. *American Economic Journal: Microeconomics*, 6(4):180–202.

Esponda, I. and Vespa, E. (2021). Contingent preferences and the sure-thing principle: Revisiting classic anomalies in the laboratory.

Eyster, E. (2019). Chapter 3 - errors in stratgeic reasoning. *Handbook of Behavioral Economics: Applications and Foundations 1*, 2:187–259.

Eyster, E. and Rabin, M. (2005). Cursed equilibrium. *Econometrica*, 73(5):1623–1672.

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *Quarterly Journal of Economics*, 129:1661–1710.

Gailmard, S. and Palfrey, T. (2005). An experimental comparison of collective choice procedures for excludable public goods. *Journal of Public Economics*, 89:1361–1398.

Gangadharan, L. and Nemes, V. (2009). Experimental analysis if risk and uncertainty in provisioning private and public goods. *Economic Inquiry*, 47:146–164.

Gillen, B., Snowberg, E., and Yariv, L. (2019). Experimenting with measurement error: Techniques with applications to the caltech cohort study. *Journal of Political Economy*, 127(4):1826–1863.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with orsee. *Journal of the Economic Science Association*, 1(1):114–125.

Ivanov, A., Levin, D., and Peck, J. (2009). Hindsight, foresight, and insight: An experimental study of a small-market investment game with common and private values. *American Economic Review*, 99:1484–1507.

Kagel, J. H. and Levin, D. (1986). The winner's curse and public information in common value auctions. *American Economic Review*, 76:894–920.

Kagel, J. H. and Levin, D. (2002). *Common Value Auctions and the Winner's Curse*. Princeton University Press, Princeton, NJ.

Levati, M. V., Marone, A., and Fiore, A. (2009). Voluntary contributions with imperfect information: An experimental study. *Public Choice*, 138:199–216.

Levin, D., Peck, J., and Ivanov, A. (2016). Separating bayesian updating from non-probabilistic reasoning: An experimental investigation. *American Economic Journal: Microeconomics*, 8:39–60.

Li, S. (2017). Obviously strategy-proof mechanisms. *American Economic Review*, 107(11):3257–3287.

Magnani, J. and Oprea, R. (2017). Why do people violate no-trade theorems? a diagnostic test. University of California at Santa Barbara, Economics Working Paper Series.

Martin, D. (2016). Rational inattention in games: Experimental evidence.

Martinez-Marquina, A., Niederle, M., and Vespa, E. (2019). Failures in contingent reasoning: The role of uncertainty. *American Economic Review*, 109(10):3437–3474.

Ngangoue, M. K. and Weizsacker, G. (2021). Learning from unrealized versus realized prices. *American Economic Journal: Microeconomics*, 13:174–201.

Oprea, R. (2020). What makes a rule complex? *American Economic Review*, 110(12):3913–3951.

Stoddard, B. (2017). Risk in payoff-equivalent appropriation and provision games. *Journal of Behavioral and Experimental Economics*, 69:78–82.

Swope, K. (2002). An experimental investigation of excludable public goods. *Experimental Economics*, 5:209–222.