An Economic Approach to Alleviate the Crises of Confidence in Science: With An Application to the Public Goods Game *

Luigi Butera †

John A. List ‡

April 5, 2017

Abstract

Novel empirical insights by their very nature tend to be unanticipated, and in some cases at odds with the current state of knowledge on the topic. The mechanics of statistical inference suggest that such initial findings, even when robust and statistically significant within the study, should not appreciably move priors about the phenomenon under investigation. Yet, a few well-conceived independent replications dramatically improve the reliability of novel findings. Nevertheless, the incentives to replicate are seldom in place in the sciences, especially within the social sciences. We propose a simple incentive-compatible mechanism to promote replications, and use experimental economics to highlight our approach. We begin by reporting results from an experiment in which we investigate how cooperation in allocation games is affected by the presence of Knightian uncertainty (ambiguity), a pervasive and yet unexplored characteristic of most public goods. Unexpectedly, we find that adding uncertainty *enhances* cooperation. This surprising result serves as a test case for our mechanism: instead of sending this paper to a peer-reviewed journal, we make it available online as a working paper, but we commit never to submit it to a journal for publication. We instead offered co-authorship for a second, yet to be written, paper to other scholars willing to independently replicate our study. That second paper will reference this working paper, will include all replications, and will be submitted to a peer-reviewed journal for publication. Our mechanism allows mutually-beneficial gains from trade between the original investigators and other scholars, alleviates the publication bias problem that often surrounds novel experimental results, and accelerates the advancement of economic science by leveraging the mechanics of statistical inference.

JEL-Classification: A11, C91, H40, D64
Keywords: Replications in Economics, Knightian Uncertainty, Public Goods, Quality of Charitable Investments.

1 Introduction

Returns from public goods and social programs are, more often than not, intrinsically uncertain. As such, it is often difficult to precisely quantify ex-ante what benefits these investments accrue to society. Yet, people are typically called to voluntarily contribute to public goods before uncertainty unfolds. Individuals may hold information that can help narrow down the possible returns of a public goods investment, but a significant degree of uncertainty may still persist. Thus, in addition to not knowing whether other people will contribute or free ride – the classic strategic uncertainty problem – individuals make decisions while aware that the information they hold is imprecise, and knowing that other people as well hold imprecise information, which could potentially differ from their own.

One prominent example is giving to charity (see List, 2011, for a review). Returns from charitable investments are hard to quantify, and donors may have different opinions, information and beliefs about the value of a given social service or program. Despite these obstacles, and despite the absence of formal incentives and institutional rules to alleviate free-riding, sustained cooperation is often observed.

Knightian uncertainty represents a pervasive feature of public goods investments, but little is known about how it affects cooperation in social dilemmas. This paper addresses this gap using a laboratory experiment in which participants play a repeated public goods game where the returns from investing in the good are observed with noise. We vary the level of uncertainty about quality and whether each participant only observes his own noisy signal, or observes everyone's signals. Previous experimental work has studied public goods provision in a world of certainty: while there is uncertainty about other people's behavior, everybody knows what cooperation buys.

Here we take a step closer to reality and ask: can cooperation be sustained over time when the quality of a given public good cannot be precisely estimated? We find, surprisingly, that uncertainty has a positive effect on cooperation. For any given level of quality, cooperation breaks down significantly less when quality is observed with noise compared to the baseline case in which quality is perfectly observed. The economic magnitude of these effects is large: on average, the rate of decline in contributions is halved in the presence of uncertainty.

We show that our results are unlikely driven by confusion, since cooperation when noisy signals are publicly observed is inversely correlated with the informativeness of the signals. Otherwise said, as we reduce uncertainty, cooperation decreases. In the limiting case where public signals fully resolve uncertainty, cooperation rates revert back to those observed in the baseline. We argue that the presence of Knightian uncertainty fosters conditional cooperation by generating ambiguity around the determinants of players' payoffs. When the returns from public goods contributions are perfectly observed, any reduction in payoffs can only be attributed to other players free-riding. When the exact quality of a public good is unobserved however, lower returns from a public good may be driven in part by a lower-than-expected quality of the good itself. While uncertainty has no effect on the Nash

equilibrium outcome, it does affect decisions of conditional cooperators who may become more tolerant to payoffs' reductions, effectively limiting the "snowball effect" of free-riding on conditional cooperation. An alternative and related explanation is that the presence of uncertainty facilitates cooperation among betrayal averse individuals (Bohnet et al. 2008, Aimone and Houser 2012).

This paper makes two key contributions: first, we show that uncertainty about public goods' returns represents one of the mechanisms through which provision can be sustained in absence of incentives mechanisms or institutional rules. The role of Knightian uncertainty in public goods decisions, and in particular its coordination properties, has received no attention in previous work.

Second, and perhaps most importantly, we propose a novel and concrete pathway to tackle untraveled empirical research avenues. Experimental studies sometimes produce surprising results that challenge the status-quo of economic knowledge. Such studies face a binary fate: either their relevance and validity is dismissed, or they are embraced as important and insightful findings. As Maniadis et. al (2014) point out, however, such judgments commonly rely on statistical significance as the main criterion for acceptance. As such, many new and surprising empirical results may be false positives simply because of the mechanics of statistical inference. Similarly, new surprising results that suffer from low power, or marginal statistical significance, may sometimes be dismissed even though they point toward an economic association that is ultimately true. In the first instance, false positives may generate entirely new research efforts that are based on a false premise. In the second instance, false negatives may delay economic inquiries of otherwise fruitful and important research avenues.

This study, like many others, represents an example of this inference problem. Our results reject the hypothesis that uncertainty reduces public goods contributions at the p<.05 level, and our design has power of about 80%. Are our surprising results capturing a true association? Given the mechanics of inference, it is difficult to provide a definite answer based on one single exploration. Even if we had very low priors about the likelihood of our results being true, our findings would lead us to update our priors upward using Bayes rule. For example, suppose that before running our experiment we had a prior about the likelihood that uncertainty increases cooperation of only $\pi=1\%$. After running the experiment and observing the results, we would update this prior to 13.9%, a very large increase relative to the initial beliefs.[1] However, if we were to (fail to) replicate our study once, the post study probability that our results capture a true association would fall to 3%, and further fall to 0.007% with two (unsuccessful) replications. This exercise highlights how further replications reduce the probability that an initial finding is true, but at the same time shows how few replications can dramatically increase the reliability and credibility of a novel experimental finding. With priors as low as 1%, just three

---

[1] As shown in Maniadis et al. (2014), the post-study probability that a result captures a true association can be calculated as follows: $\text{PSP} = \frac{(1-\beta)\cdot\pi}{(1-\beta)\cdot\pi+\alpha(1-\pi)}$.

additional successful replications would lead to update the post-study probability at above 80%, establishing thus robust evidence around a novel economic insight. While in theory replicating novel findings is desirable, in practice the incentives for replication are seldom in place. Once a study has been published, the original investigators have little incentive to replicate their own findings, and have only to lose from others doing so. On the other hand, the returns from replicating a published study are generally low, discouraging other scholars to replicate the original study.

This paper proposes and puts into practice a novel and simple mechanism that allows mutually beneficial gains from trade between original investigators and other researchers. In our mechanism, the original investigators, upon completing their initial study, write a working paper version of their study. While they do share their working paper online, they do however commit not to submit it to any journal for publication, ever. The original investigators instead offer co-authorship of a second paper to other researchers who are willing to independently replicate the experimental protocol in their own research facilities.[2] Once the team is established, but before beginning replications, the replication protocol is pre-registered at the AEA experimental registry, and referenced in the first working paper. This is to guarantee that all replications, both successful and failed, are properly accounted for, eliminating any concerns about publication biases. The team of researchers composed by the original investigators and the other scholars will then write and coauthor a second paper, which will reference the original unpublished working paper, and submit it to an academic journal. Under such an approach, the original investigators accept to publish their work with several other coauthors, a feature that is typically unattractive to economists, but in turn gain a dramatic increase in the credibility and robustness of their results, should they replicate. Further, the referenced working paper would provide a credible signal about the ownership of the initial research design and idea, a feature that is particularly desirable for junior scholars. On the other hand, other researchers would face the monetary cost of replicating the original study, but would in turn benefit from coauthoring a novel study, and share the related payoffs. Overall, our mechanism could critically strengthen the reliability of novel experimental results and facilitate the advancement of scientific knowledge.

The remainder of our paper is organized as follows. Section 2 provides a concise review of the experimental work on public goods provision. Section 3 presents the experimental design. Section 4 presents the results. Section 5 presents a simple, incentive-compatible framework for incentivizing replications in Economics. Section 6 presents the replication plan for this paper. Finally, section 7 concludes.

---

[2]Replications would use the original program, instructions, and payoffs of the original paper, as well as the same sample size.

2　Background

Public goods provision represents one of the most studied decision-making environments in the experimental economics literature. A comprehensive review of this literature goes beyond the scope of this paper, but the interested reader may refer to Ledyard (1995) and Chaudhuri (2011). Three stylized facts that are relevant for this paper emerge from the literature on public goods contributions: first, initial contributions to public goods typically exceed zero, what the Nash equilibrium would predict.[3] Second, cooperation tends to unfold over time (Andreoni 1995), a tendency linked to the presence of heterogeneous preferences such as self-interest, altruism, and conditional cooperation (Brandts and Schram 2001; Fehr et al. 2001; Bowles and Gintis 2002; Frey and Meier 2004; Fischbacher and Gatcher 2010; Ambrus and Pathak 2011; Fischbacher et al. 2014). Third, centralized institutions, such as taxation, competition, and voting rules (e.g. Falkinger et al. 2000; Kosfeld et al. 2009; Dannenberg et al. 2010; Reuben and Tyran 2010; McEvoy et al. 2011; Putterman et al. 2011; Kesternich et al. 2014), and decentralized institutions, such as communication, moral and monetary sanctioning and rewards (Fehr and Gachter 2000; Masclet et al. 2003; Bochet et al. 2006; Sefton et al. 2007; Gachter et al. 2008; Bochet and Putterman 2009; Nikiforakis 2010), all contribute to various degrees to promote cooperation.

A feature common to all these studies is the absence of risk and Knightian uncertainty about the value of the public good: the marginal per capita return from investing in the public good (hereinafter MPCR) is perfectly observed by all participants. Only a handful of papers depart from certainty about the value of the public good. Fisher et al. (1995) introduce heterogeneity and incomplete information about other players' returns by allowing different players to have different MPCR values. In their setting, while there is incomplete information about other players' MPCRs, each player knows how much the public good is worth to him. The authors find that subjects' behavior seem to be only affected by their own MPCR, showing thus a negligible effect of information about other people's returns.[4] Levati et al. (2009) study contributions under imperfect information about the MPCR: in their setting, the MPCR can take one of two (equiprobable) values, and each subject knows these values and their probability distribution.[5] The parameters are set such that the behavior of a risk-neutral individual should not be affected by the

---

[3]Various factors contribute to higher-than-predicted contributions, such as kindness (Andreoni 1995), confusion and decision errors (Anderson et al. 1998; Kurzban and Houser 2002), warm-glow (Andreoni 1990, Palfrey and Prisbey 1997), strategic play (Andreoni 1988), distributional concerns (Fehr and Schmidt 1999; Bolton and Ockenfels 2000), and intentions' signaling (Rabin 1993; Charness and Rabin 2002; Dufwenberg and Kirchsteiger 2004; Cox et al. 2007; Cox et al. 2008).

[4]For a study on the role of uncertainty and incomplete information in leader-follower environments see Komai et al. 2011.

[5]Notice that their public goods game is a social dilemma only in expectations, since the low value of the MPCR is such that it is not socially optimal to contribute, and the high value is such that it is both socially optimal and a Nash equilibrium to contribute everything.

presence of risk. They find that under imperfect information, overall contributions decrease, a result solely driven by lower contributions in the first period of the repeated public goods game. Boulu-Reshef et al. (2016) modify the design of Levati and colleagues by maintaining each possible scenario as a social dilemma, and allowing contributors to receive different MPCRs. The authors find little effect of risk on contributions.

The only study that we are aware of that uses Knightian uncertainty is Gangadhran and Nemes (2009), in which the authors employ a within-subject design to sequentially introduce Knightian uncertainty over the private account and the public good account, and allow participants to pay a fee to convert uncertainty over the two possible values of the MPCR (zero or the MPCR) into risk. Similarly to Levati et al. (2009), the game is a social dilemma only in expectations, since there is a positive probability that participants receive a zero payoff even when everyone fully contributes.

Our paper departs from previous studies in a number of important ways: first, our parameter space allows us to investigate the effect of Knightian uncertainty over a rich set of situations, from social dilemmas, to situations where it might be socially optimal not to fund the public good, and to cases where fully contributing might be a Nash equilibrium. Second, in our setting Knightian uncertainty cannot be converted into risk, an experimental feature that confounds the effect of Knightian uncertainty with strategic uncertainty.[6] Third, our design introduces (and varies) Knightian uncertainty by providing subjects with a noisy signal (drawn from a known support but unknown distribution) about the true value of the MPCR, an approach similar in spirit to what has been done in the common value auction literature (see Harrison and List 2008). This structure allows us to capture a critical feature of real-life public goods: when choosing whether and how much to contribute, individuals have to take into account that other contributors, like themselves, may hold over-optimistic or under-optimistic beliefs about the value of the public good.

3 Experiment design

3.1 A public goods game with environmental uncertainty: definitions

In a standard linear public goods game, participants are randomly assigned to groups of size N, and are endowed with M tokens that they can either invest in a private account that accrues only to their own payoff, or contribute $g \in [0, M]$ tokens to a group project that pays a return θ to all group members. There is no Knightian uncertainty in this game, as θ is perfectly observed by all members. In this game thus, each player's decision is characterized by the following general payoff function:

---

[6]In such setting, subjects not only face strategic uncertainty over other people's contributions, but also over others willingness to pay to eliminate Knightian uncertainty.

$$\pi_i = M - g_i + \theta \cdot \sum_{j=1}^{N} g_j \tag{1}$$

In our experiment we introduce Knightian uncertainty in the public goods game in the following simple way. Instead of observing θ, each participant receives a noisy signal $s_i = \theta + \epsilon_i$, where ϵ_i is distributed according to an unknown distribution with mean zero and standard deviation σ. It is common knowledge that all signals are drawn from the same distribution. Depending on treatments however, participants either observe only their own signal (private signal), or observe their own signal and the signals of all other group members (public signals).

When signals are privately observed, the payoff function takes the form:

$$\pi_i = M - g_i + \mathbf{E}[\theta|s_i] \cdot \sum_{j=1}^{N} g_j \tag{2}$$

When signals are publicly observed instead, the payoff function becomes:

$$\pi_i = M - g_i + \mathbf{E}[\theta|s_i \bigcap \mathbf{s}_j] \cdot \sum_{j=1}^{N} g_j \tag{3}$$

where $si \bigcap \mathbf{s}_j$ is the intersection between a player's own signal and the vector of signals $\mathbf{s}_j$ received by other group members. This simply means that the true θ has to be compatible with all signals. Equation 3 shows that public signals can vary in how informative they are about the underlying true value of θ: if at least two group members receive opposite "extreme signals", then the true θ is perfectly identified, and uncertainty is fully resolved. The opposite situation is when $s_j = s \ \forall j$ (e.g. everyone receives the same signal), in which case observing other people's signals does not add any useful information.

3.2 Experiment details

We first describe the general procedure, and then provide details about our treatments. In each session, 16 participants play 4 repeated public goods games in groups of 4 players. Each game consists of 8 rounds. In each round, participants choose how to allocate 10 tokens between a private account and a group account.[7] After each game, groups are reformed using a stranger matching procedure. Participants are only identified by a randomly generated ID number. It is common knowledge since the beginning that only one of the 4 games will be randomly selected for payments, and that each player will be paid

---

[7]20 tokens are worth US$1. Each token placed in the private account is worth 1 token only to the subject. After each round, participants are informed about their own payoff for that round, but are not told how many tokens other players have invested in the group account.

6

the sum of earnings made in the 8 rounds that constitute that game. In all treatments, the instructions specify which are possible values that the MPCR can take. The minimum possible value of the MPCR is 0.05 and the maximum is 1.25, with increments of 0.1. There are therefore 13 possible values that the MPCR can take. In all treatments, subjects are told that in 3 out of 4 games the true MPCR is constant within each game (e.g. the MPCR does not vary between rounds); instead, in one of the 4 games the true MPCR is randomly drawn every round (with replacement) from the 13 possible values. In all treatments, the 3 games with constant MPCR have always the following (predetermined) MPCR values: 0.25, 0.55, and 0.95. We have two sessions per treatment, and we (partially) vary the order in which games are played : in one session the order of games is: 0.25, 0.55, 0.95, VARIABLE; in the other the order is 0.95, 0.55, 0.25, VARIABLE. Before the beginning of each game, participants are informed about whether the game has a constant or variable MPCR. To control for risk and ambiguity preferences, at the end of the experiment all participants play an incentivized Eckel-Grossman risk task (Eckel and Grossman 2002), and an ambiguity task. This basic structure is common to all treatments.

We have a total of four treatments in our experiment plus a baseline. We have thus a total of 160 subjects, equally balanced across treatments. The baseline treatment *Base-VCM* is a standard public goods game without Knightian uncertainty.

We have two private signal treatments in which participants only observe their own signal. In treatment *Private Thin* each participant receives a private signal known to be drawn from the interval: true MPCR ±0.1. So for instance if a participant receives a private signal of 0.55, he knows that the true MPCR can either be 0.45, 0.55, or 0.65. He also knows that if the true MPCR is, for instance, 0.65, another player might have received a signal of 0.55, 0.65, or 0.75. Differently, in treatment *Private Thick* participants receive a private signal known to be drawn from the interval: true MPCR ±0.2. So for instance if a participant receives a private signal of 0.55, he knows that the true MPCR can either be 0.35, 0.45, 0.55, 0.65, or 0.75.

We have two public signals treatments, *Public Thin* and *Public Thick*, that have the same parameters of the private conditions, but differ in the fact that participants also observe the signals of other group members.[8]

The experiment was conducted at the ExCEN experimental laboratory at Georgia State University, and was programmed using O-Tree (Chen et al. 2016). Participants received a show-up fee of $10.

---

[8]In the three constant MPCR games, participants receive only one signal per game.

4 Results

We present our results in two steps. We first summarize a simple non-parametric analysis of the effect of uncertainty on public goods contributions.[9] We show how our baseline results compare to previous standard public goods games without uncertainty. We then compare aggregate behavior in the baseline to our two treatments in which participants receive a private signal about the true MPCR. Finally we compare results in baseline and private signal conditions to treatments in which signals are publicly observed.[10] In the second step we analyze our data using econometric models that fully take into account group-specific dynamics and the panel nature of our data.

Table 1 sets the stage for our non-parametric analysis, showing average round contributions aggregated by rounds for all treatments and periods with constant MPCR levels.[11]

---

[9]Unless otherwise noted, all results from pairwise comparisons are from two-sided non-parametric Wilcoxon Mann-Whitney tests. Table 1 also reports p-values corrected for multiple hypothesis testing (see List et. al 2016).

[10]Notice how our public signals treatments serve two purposes: first, they represent an approximation of institutions that reduce uncertainty about public goods' quality by aggregating private information; second they provide us with a control for whether participants understood our experimental setup, and in particular the relationship between signals and returns from the public good.

[11]Table 1 averages round contributions between the two orders in which participants can experience the periods. Participants are randomly assigned in each treatment to two sequences of periods (sequence 1: {0.25, 0.55, 0.95, Variable period}. Sequence 2: {0.95, 0.55, 0.25, Variable period}. In each treatment, 16 participants experience Sequence 1 and 16 sequence 2. For each treatment and MPCR level, we test the equality of distributions across orders using two-sample Kolmogorov-Smirnov tests. In all treatments with private or public uncertainty (both thin and thick), we find that the distributions across orders are not statistically different for periods with MPCR equal to 0.25, 0.55, while are different at 1% level when MPCR equals 0.95, with average contributions always higher under order 2. A similar exercise on baseline Base-VCM reveals that distributions with MPCR equal 0.25 are different at 10% level, and at 1% level when MPCR equals 0.55 and 0.95. In all cases, average contributions are higher under order 2. We believe averaging contributions across orders is appropriate for two reasons: First, while under some conditions the order clearly affects the *level* of contributions, the rate of decay of contributions over time is nevertheless the same between the two orders for all treatments and MPCRs. Second, the differences across orders are always in the same direction for all treatments and MPCR (e.g. order 2 has higher average contributions). Thus, for the non-parametric section of our results we pool the two orders together, while we explicitly take order into account in our econometric analysis.

Table 1: Percentage of endowment contributed to the public account by treatment and MPCR level.

MPCR= 0.25	Round								
Treatment	1	2	3	4	5	6	7	8	Total
Baseline VCM	40.9	30.6	26.9	27.8	28.4	21.9	21.2	13.8	26.4
Private Thin	31.6	38.1	27.8	32.8	35.3	25.9	29.4	22.8	30.5
Public Thin	33.8	33.1	34.4	25.6	22.2	15.3	17.2	21.6	25.4
Private Thick	34.7	32.2	25	23.4	22.8	20.3	14.1	18.1	23.8
Public Thick	33.4	29.7	31.2	40.6	30.6	25.3	24.1	27.5	30.3
Baseline - Private Thin	9.3	-7.5	-0.9	-5	-6.9	-4	-8.2	-9	-4.1*[ns]
Baseline - Private Thick	6.2	-1.6	1.9	4.4	5.6	1.6	7.1	-4.3	2.6[ns][ns]
Baseline - Public Thin	7.1	-2.5	-7.5	2.2	6.2	6.6	4	-7.8	1[ns][ns]
Baseline - Public Thick	7.5	0.9	-4.3	-12.8	-2.2	-3.4	-2.9	-13.7	-3.9**[ns]
MPCR= 0.55	Round								
Treatment	1	2	3	4	5	6	7	8	Total
Baseline VCM	49.4	51.9	47.8	46.2	37.5	37.2	36.6	33.4	42.5
Private Thin	50.9	55	54.1	48.8	49.7	48.1	49.7	42.8	49.9
Public Thin	45	46.9	42.2	47.5	39.1	39.7	35	37.5	41.6
Private Thick	51.2	53.4	55.3	52.8	44.4	35	42.8	37.8	46.6
Public Thick	52.5	51.2	50.6	51.2	37.2	34.1	38.4	35	43.8
Baseline - Private Thin	-1.5	-3.1	-6.3	-2.6	-12.2	-10.9	-13.1	-9.4	-7.4***[††]
Baseline - Private Thick	-1.8	-1.5	-7.5	-6.6	-6.9	2.2	-6.2	-4.4	-4.1*[ns]
Baseline - Public Thin	4.4	5	5.6	-1.3	-1.6	-2.5	1.6	-4.1	0.9[ns][ns]
Baseline - Public Thick	-3.1	0.7	-2.8	-5	0.3	3.1	-1.8	-1.6	-1.3[ns][ns]
MPCR= 0.95	Round								
Treatment	1	2	3	4	5	6	7	8	Total
Baseline VCM	68.1	68.1	66.9	57.2	63.1	57.8	52.5	49.7	60.4
Private Thin	76.2	75.3	75	77.5	66.6	70.3	65.9	72.2	72.4
Public Thin	66.2	70.9	70.6	64.4	62.5	66.6	65.3	55.6	65.3
Private Thick	73.8	73.8	67.8	66.9	67.8	66.2	71.2	68.4	69.5
Public Thick	65.3	68.4	69.1	73.8	65.3	58.4	60.3	60.9	65.2
Baseline - Private Thin	-8.1	-7.2	-8.1	-20.3	-3.5	-12.5	-13.4	-22.5	-12***[†††]
Baseline - Private Thick	-5.7	-5.7	-0.9	-9.7	-4.7	-8.4	-18.7	-18.7	-9.1***[††]
Baseline - Public Thin	1.9	-2.8	-3.7	-7.2	0.6	-8.8	-12.8	-5.9	-4.9[ns][ns]
Baseline - Public Thick	2.8	-0.3	-2.2	-16.6	-2.2	-0.6	-7.8	-11.2	-4.8[ns][ns]

Note: The last column reports significance level from two-sided pairwise WMW tests of differences, both uncorrected and corrected for multiple hypothesis testing using the correction procedure of List et al. (2016). We report uncorrected p-values using the following notation: *** $p<0.01$, ** $p<0.05$, * $p<0.1$, ns, $p\geq0.1$. We report (in brackets) the significance level corrected for multiplicity of groups (e.g. multiple MPCRs) and multiplicity of treatments (see List et al. 2016). We report p-values corrected for multiple hypothesis testing using the following notation: ††† $p<0.01$, †† $p<0.05$, † $p<0.1$, ns, $p\geq0.1$

Figure 1 conveys three clear messages. First, contribution patterns in our baseline are comparable to previous public goods experiments. Second, contributions under private uncertainty (either thin or thick, except for Private Thick with MPCR=0.25) are always higher than the baseline, and cooperation breaks down less frequently when the quality of the public good is uncertain. Third, contribution and cooperation levels under public signals are similar to the corresponding baseline cases.

We first look at baseline contributions. For all values of the MPCR of the public good, we observe that cooperation breaks down over time.[12] When MPCR=0.25, average contributions start at 40.9%, and progressively decline to 13.8% in the final round. Similarly, when the MPCR equals 0.55, subjects start by contributing about half of their endowment, ending up with average contributions of about one third (33.4%) in the last round. The majority of previous public goods experiments use an MPCR of 0.5, and our results fully compare. For instance, in Andreoni (1995) average contributions start at 56% in round 1 and decline to 26.5% in round 10 (in their round 8 contributions are 35.1%). Even when the MPCR level is almost such that fully contributing is a Nash equilibrium (e.g. MPCR=0.95), average contributions start at 68%, only to decline to 49.7% in the last round.

Observing higher contributions and more stable cooperation under environmental uncertainty is striking. For MPCR equal to 0.55 and 0.95, average contributions to (private) uncertain public goods are significantly higher than the corresponding certain public goods. Even when the MPCR is almost such that contributing is not socially optimal (MPCR=0.25), we find that average contributions are 4% higher under thin uncertainty than the baseline (no difference under thick private uncertainty). More importantly, we find that for almost all levels of quality of the public good and level of (private) uncertainty, cooperation breaks down less than in the baseline case. When uncertainty is thin, the difference between first and last periods' average contributions are never statistically significantly different, no matter the level of the MPCR. As uncertainty grows (Private thick treatment), we observe a statistically significant decline in cooperation when MPCR=0.25 (z=0.23,p=0.021) and MPCR=0.55 (z=1.85,p=0.063) (but not when MPCR=0.95). While the latter decays of contributions are statistically significant, the magnitude of such declines is small: 16.5% when MPCR equals 0.25, and 13.4% when MPCR equals 0.55.[13]

Clearly, as every subject receives a (random) private signal, these results could be driven by subjects receiving upward biased signals. Figure 1 reports average round contributions for Private Thin and private Thick divided by whether subjects received a signal above, equal, or below the true MPCR, and compares them with average round contributions in Baseline VCM. Averaging contributions by the type of signal received clearly abstracts from other group members' signals and behavior, but provide a first important indication for how uncertainty affects cooperation.

---

[12]For all MPCR levels, first and last periods' average contributions are significantly different (MPCR=0.25: z=3.5,p=0.000; MPCR=0.55: z=2.1,p=0.035; MPCR=0.95: z=1.9, p=0.048).

[13]In the baseline, the decline for MPCR equal to 0.25 is 27.2% and 16% when MPCR equals 0.55.

Figure 1: Average contributions in Baseline and Private thin or thick treatments by the type of signal received
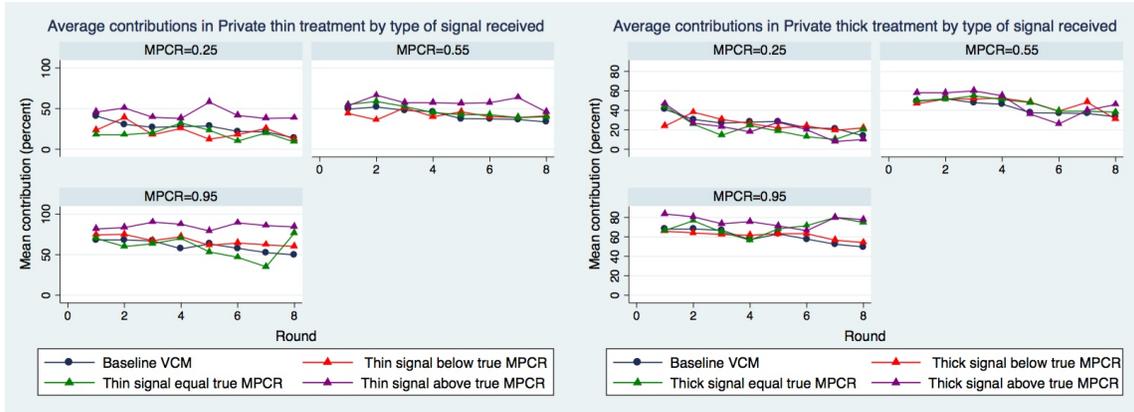

Figure 1A shows that in treatment Private Thin, subjects who receive upward biased signals tend to contribute more than their baseline counterparts. The figure also importantly shows that subjects who receive a downward biased signal do not differ in their behavior from baseline subjects. This is surprising because for those subjects the best possible outcome is indeed the true MPCR observed by baseline subjects. Figure 1B shows contributions from subjects facing a more highly uncertain public good. The pattern is similar to Private Thin, with the exception that as uncertainty grows, also subjects receiving a signal equal to the true MPCR tend to contribute (slightly) more than the baseline.

So far we have abstracted from the fact that any given individual i is repeatedly observed over time t (32 rounds) and makes decisions in four separate groups g (8 sequential decisions in each group), which likely differ in the composition of signals received, value of the public good, and other group characteristics and dynamics. To account for these differences, we adopt the following econometric strategy. For each set of results, we estimate highly dimensional fixed effects models with standard errors clustered both at the group and individual level (see Cameron et al. 2008; Correia 2017). As our observations are censored by design (about 17% are left-censored), we also estimate the coefficients of each regression using random effects panel tobit models with group dummies.[14]

Before turning to the econometric analysis of our main data, we first address the possibility that our subjects were confused by the experimental design. To address this concern we turn to our public treatments, which provide a useful test for confusion and errors. If participants did not understand the experimental procedure, then contributions in public treatments groups in which public signals fully resolve uncertainty should differ from base-

---

[14]While tobit models are reported as a robustness check, boundaries exercises show low sensitivity of our estimates to changes in the censoring threshold, suggesting that linear estimators are unbiased.

line contributions. For instance, this could happen if subjects failed to take into account others members' signals, or did not understand that the true MPCR must be compatible with all signals. For similar reasons, the level of informativeness of the public signals (i.e. how many values of MPCR are admissible given the set of public signals) should have little predictive power on players' contributions if our results are driven by confusion or errors.

Table 2 shows that this is not the case. Columns 1 and 2 compare round contributions from baseline subjects, who perfectly observe the true MPCR (called henceforth θ), and round contributions from subjects in our two public treatments who are in groups in which public signals fully resolve uncertainty. Conditionally on receiving fully informative public signals, contributions are statistically indistinguishable from baseline contributions (coefficient "public signals treatments") under both model 1 and 2. Columns 3 and 4 include all subjects from public treatments and baseline, and explore how the level of noise of public signals affect contributions. Our main variable of interest in column 3 and 4, "number of possible θs", counts the number of values that the true MPCR can take given the observed public signals. For example, if the number of possible θs is equal to 3 in a given group, then subjects know than any of those 3 values could be the true MPCR. Under both econometric specifications, and for any given value of the true MPCR, noisier public signals induce higher contributions. For instance, In model 4 a 1 unit increase in the number of possible MPCRs compatible with all signals induce an average increase in contributions of 1.081 tokens for any given level of the true value of θ (p=0.067).

Table 2: Comparing baseline and public signals treatments contributions

Dependent variable	(1) RE panel tobit Tokens to PG	(2) HDFE linear model Tokens to PG	(3) RE panel tobit Tokens to PG	(4) HDFE linear model Tokens to PG
Public signal treatments (0=no; 1=yes)	0.751	0.326	-0.591	
	(1.938)	(0.369)	(1.262)	
Round number (by period - 1 to 8)	-0.314***	-0.229***	-0.555***	-0.383***
	(0.0332)	(0.0355)	(0.137)	(0.129)
Period (1 to 4)	-1.685*	-0.262***	-1.667*	
	(0.891)	(0.0974)	(0.861)	
Order of periods (1= 0.25, 0.55, 0.95, Var.; 2= 0.95, 0.55, 0.25, Var.)	3.269	0.702*	-3.038**	
	(2.031)	(0.373)	(1.355)	
True θ	7.168***	5.745***	4.790***	3.657***
	(0.526)	(0.489)	(1.219)	(1.060)
Number of possible θs compatible with all signals			-0.575	-0.233
			(2.202)	(1.565)
Number possible $(\theta s)^2$			-0.0165	-0.0550
			(0.356)	(0.202)
Only one possible θ (0=no; 1=yes)			1.317	0.787
			(1.209)	(0.967)
True θ X n. possible θs			1.731**	1.081*
			(0.753)	(0.586)
True θ X Round			0.267	0.176
			(0.199)	(0.164)
Round X n. possible θ			0.154*	0.105
			(0.0841)	(0.0769)
Round X n. possible θs X true θ			-0.121	-0.0820
			(0.122)	(0.0905)
Constant	1.960		14.18**	
	(5.160)		(5.705)	
sigma u	2.172***		2.317***	
	(0.205)		(0.189)	
sigma e	3.149***		3.084***	
	(0.0664)		(0.0516)	
Observations	1,960	1,960	3,072	3,072
R-squared		0.293		0.559
Number of groups		69		96
Number of subjects	92	92	96	96

Column 1 and 2 only include observations from the baseline treatment and from groups within the public signals treatments (both thin and thick) in which public signals uniquely identify the true MPCR. Column 3 and 4 include the baseline treatment and all observations from public signals treatments (both thin and thick). Columns 1 and 3 reports estimates from random effects panel tobit models with dummies for groups (dummies not displayed in the table). Columns 2 and 4 reports estimates from highly dimensional fixed effects lineal models with standard errors clustered both at the group and individual level. Groups and individual fixed effects are absorbed in model 4 but not in model 1, so that the coefficient for "Public signal treatment" can be estimated. Variable "True θ" identifies the true MPCR for the round. Variable "Number of possible θs compatible with all signals" counts the number of values that are compatible with the true θ given the public signals. The dummy "Only one possible θ (0=no; 1=yes)" takes value 0 when the public signals do not uniquely identify the true θ, and 1 when they do (and in all observations in the baseline). (Robust) standard errors in parentheses (for models 2 and 4). *** p<0.01, ** p<0.05, * p<0.1.

Having shown that confusion has difficulty explaining our results, we turn to an econometric analysis of cooperation under Knightian uncertainty. In particular, we are interested

in how uncertainty affects cooperation over rounds, and how uncertainty affects cooperation both within and between different values of the true MPCR. We first assess the effect of cooperation between MPCR and over time.

We estimate the following general model (for our highly dimensional linear model):

$$y_{i,g,t} = \alpha + \beta_1 \mathbf{T} + \beta_2 \mathbf{X}_{it} + \beta_3 \mathbf{Y}_{gt} + \beta_4 \mathbf{X}_{it} \cdot \mathbf{T} + \beta_5 \mathbf{X}_{gt} \cdot \mathbf{T} + \epsilon_{igt} \tag{4}$$

Where our dependent variable, $y_{i,g,t}$, is the contribution to the public good made by i in time t while in group g. $\mathbf{T}$ is our treatment, baseline vs. uncertainty under private signals. $\mathbf{X}_{it}$, with a slight abuse of notation, is a vector of individual fixed effects plus a set of individual information received in time t, such as the type of signal received. $\mathbf{X}_{gt}$ is a vector of group characteristics, such as the true value of the MPCR for that group in time t, the contributions made by other group members, and the types of signals received by other group members.

Table 3 reports estimates from a random effects tobit model and a highly dimensional fixed effect model with standard errors clustered at the group and individual level. Under the latter specification we find, unsurprisingly, that contributions decrease over time (0.27 less tokens invested in each subsequent round, p=0.000). However, when the value of the public good is observed with noise, contributions decrease significantly less over time (interaction between time and treatment equal to 0.128 tokens, p=0.072). The effect is stronger and more statistically significant under our tobit specification. We also find evidence of a strong and positive effect of uncertainty on between-MPCR contributions (2.96 tokens, p=0.008). Finally, we do not find evidence that other group members' contributions in previous rounds have a statistically significant effect on contributions, and neither does its interaction with our uncertainty treatment. We also do not find evidence of a significant effect of the composition of other members' signals. We construct two dummy variables for whether at least one other group member has received an upward/downward biased signal, and neither is statistically significant.[15]

---

[15]The same results are obtained if we include the number of other members who received an upward/downward biased signal instead of these 2 dummies.

14

Table 3: Effect of uncertainty on contributions over time and between levels of MPCR

Dependent variable	(1) RE panel tobit Tokens to PG	(4) HDFE linear model Tokens to PG
MPCR type (1=0.25; 2=0.55; 3=0.95; 4=Var.)	-0.141	
	(0.799)	
Round number	-0.394***	-0.278***
	(0.0625)	(0.0627)
Private signal received	3.634***	2.654**
	(1.251)	(1.139)
True θ (true MPCR)	-0.0570	-0.498
	(1.242)	(0.997)
Uncertainty (0=baseline; 1=private signals treatments)	-3.659	
	(2.730)	
True θ X private signal received	4.085***	2.968***
	(1.213)	(1.091)
Round number X uncertainty	0.158**	0.128*
	(0.0754)	(0.0715)
Other group members' contributions ($t-1$)	-0.0310	-0.0229
	(0.0250)	(0.0172)
Others' contribution ($t-1$) X Uncertainty	-0.0120	-0.00895
	(0.0298)	(0.0225)
Order of periods (1= 0.25, 0.55, 0.95, Var.; 2= 0.95, 0.55, 0.25, Var.)	1.956	
	(1.324)	
Period (1 to 4)	2.002	
	(1.736)	
At least 1 other member received private signal$> \theta$	0.0964	0.206
	(0.338)	(0.264)
At least 1 other member received private signal$< \theta$	0.0969	0.0110
	(0.323)	(0.208)
Sigma u	2.227***	
	(0.183)	
Sigma e	3.268***	
	(0.0607)	
Constant	1.320	
	(4.789)	
Observations	2,688	2,688
R-squared		0.578
Number of subjects	96	96

Sample includes all observations from baseline treatment, and both private signals treatments (thin and thick, pooled). Column 1 reports estimates from a random effects panel tobit model. Column 2 reports estimates from a highly dimensional fixed effect linear model with standard errors clustered at the group and individual level. (Robust) standard errors in parenthesis (for model 2). *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

We conclude our econometric analysis by looking at how uncertainty affects within-MPCRs contributions, that is, how uncertainty affects cooperation at any given level of quality of the public good.

Table 4 explores how receiving a private signal greater or lower than the true MPCR θ affects contributions at any given level of true MPCR. Under our highly dimensional linear model specification, we find suggestive evidence of an effect of upward biased signals within levels of MPCR, but no effect of downward biased signals, a result in line with the trends seen in figure 1. We construct two dummies that capture whether an individual has received a signal above or below the true MPCR θ, and we interact them with our true MPCR levels (0.55, 0.95. 0.25 is the omitted MPCR level). When the MPCR is equal to 0.55, receiving an upward biased signal increases contributions by 1.7 tokens (p=0.062) compared to when MPCR equals 0.25. We do not find a statistically significant effect for an MPCR equal to 0.95. In our linear specification, we do not find significance for the interaction of different MPCRs with signals lower than the true MPCR. The results for our tobit specification are consistent, with the exception of a statistically significant effect on the interaction between MPCR=0.95 and receiving a signal lower than the true MPCR (p<0.05).

Table 4: Effect of uncertainty on contributions within levels of MPCR θ

	(1)	(2)
Dependent variable	Tokens to PG	Tokens to PG
Round number	-0.296***	-0.197***
	-0.0352	-0.0313
True θ	10.85***	5.411**
	-2.842	-2.397
Uncertainty (0=Baseline)	0.593	
	-1.494	
MPCR type (1=0.25; 2=0.55; 3=0.95)	0.143	
	-0.349	
Signal $> \theta$	0.813	0.000371
	-0.726	-0.594
Signal $< \theta$	0.445	-0.00401
	-0.742	-0.669
Private signal received	-1.885	0.493
	-2.796	-2.232
Period with fix MPCR=0.55	-2.273**	
	-0.973	
Period with fix MPCR=0.95	-1.934	
	-1.233	
MPCR fix=0.55 X Signal$> \theta$	1.838**	1.701*
	-0.797	-0.902
MPCR fix=0.55 X Signal$< \theta$	0.476	0.629
	-0.831	-0.765
MPCR fix=0.95 X Signal$> \theta$	1.014	1.109
	-0.968	-0.991
MPCR fix=0.95 X Signal$< \theta$	-1.978**	-0.57
	-0.908	-0.84
Order of periods (1= 0.25, 0.55, 0.95, Var.; 2= 0.95, 0.55, 0.25, Var.)	1.35	
	-1.537	
N. other members receiving signal$> \theta$	-0.197	-0.0974
	-0.237	-0.128
N. other members receiving signal$= \theta$	-0.187	-0.0325
	-0.265	-0.191
Other members contributions $(t-1)$	-0.00917	-0.00717
	-0.0238	-0.0195
Others contributions $(t-1)$ X Uncertainty	-0.028	-0.0216
	-0.0282	-0.0233
Sigma u	2.247***	
	-0.185	
Sigma e	3.257***	
	-0.0605	
Constant	5.442*	
	-2.998	
Observations	2,688	2,688
R-squared		0.579
Number of subjects	96	96

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

17

Sample includes all observations from baseline treatment, and both private signals treatments (thin and thick, pooled). Column 1 reports estimates from a random effects panel tobit model. Column 2 reports estimates from a highly dimensional fixed effect linear model with standard errors clustered at the group and individual level. The omitted MPCR level is 0.25. Dummies "Signal $> \theta$" and "Signal $< \theta$" indicate whether an individual has received a signal above or below the true MPCR θ. (Robust) standard errors in parenthesis (for model 2). *** p<0.01, ** p<0.05, * p<0.1.

5 A simple incentive-compatible framework for replication in economics

Current publication practices in Economics provide little incentives to replicate existing studies. As Hamermesh (2007) puts it, "Economists treat replication the way teenagers treat chastity — as an ideal to be professed but not to be practiced". Once a paper is published, the original authors have little incentive to replicate their own work. Similarly, the incentives for other scholars to closely replicate existing work are typically very low: replication studies rarely attract the attention of major journals, thus providing very low expected returns on this dimension from such investments. While in this paper we focus on the close replication of an existing experimental design, we note that other types of replications are rare as well, such as obtaining published datasets to replicate the results, or investigating a research question using a different design and setting.[16]

The lack of close replication of existing experiments is particularly problematic with studies that find novel and surprising results. Building on a formal methodology developed in the health sciences literature (Wacholder et al. 2004; Ioannidis 2005; Moonesinghe et al. 2007), Maniadis et al. (2014) propose a simple framework that highlights how to use a Bayesian approach to evaluate how a novel result should move scholars' priors about a given phenomenon. We focus here on the simplest case of updating after observing results from one study. Let π be the prior that a given scholar has about a given scientific relationship. Call α the significance level of an experiment investigating such relationship, and $(1 - \beta)$ the power of the experiment. The post-study probability (PSP) that a given scientific association is true can be computed using the following formula:

$$\text{Post-Study Probability} = \frac{(1 - \beta) \cdot \pi}{(1 - \beta) \cdot \pi + \alpha(1 - \pi)} \tag{5}$$

where $(1 - \beta) \cdot \pi$ represents the probability that a true result is declared true for any given prior π, and the denominator represents the probability that *any* result is declared true (e.g. $\alpha(1 - \pi)$ is the probability of a type I error given prior π). So for instance if a given scholar belied that a certain scientific result had a 1% chance of being true at $\alpha = 5\%$ level and power $(1 - \beta) = 80\%$, after observing one study confirming that result he would update his priors to 13.9%. Even more dramatically, a scholar holding priors of 10% would update the post-study probability to 64%. This example highlights how volatile low priors are when they only depend on evidence provided by a single study.

---

[16]Hamermesh (2007) surveyed authors of 139 empirical studies published between 2002 and 2004 in Industrial and Labor Relations Review (ILRR) and the Journal of Human Resources (JHR), both journals with open data access policies, and found that the mean number of requests for data in each of these two specialized journals was just one, and 60.5% of authors of these papers never received a request to share the data. Hamermesh conducted a similar survey with authors who published on the American Economic Review between 1999 and 2000, and found that the median request for data was three, with 22% of authors never receiving a request.

18

More importantly however, the post-study probability of novel and surprising results is very sensitive to failed replications. Figure 2 shows how scholars holding initial priors $\pi = 1\%$ and 10% should update their priors based on subsequent replications that fail to replicate the results at $\alpha = 5\%$ with power $(1 - \beta) = 80\%$. As seen, when only the original study is available, priors would move from 1% to 13.9% with initial priors $\pi = 1\%$, and from 10% to 64% when initial priors are $\pi = 10\%$. With one additional failed replication, the PSP given $\pi = 1\%$ would fall from 13.9% to about 3%, and to 0.07% with two failed replications. Assuming $\pi = 10\%$, one failed replication would lower the PSP to 27.2%, while two would further reduce the PSP to 7.3% It can be easily seen that with three or more failed replications, the post-study probability converges to zero regardless of the initial priors.

Figure 2: Post-study probability of a given result being true as a function of the number of failed replications and priors $\pi = \{1\%, 10\%\}$(assuming $\alpha = 0.05, \beta = 0.8$)
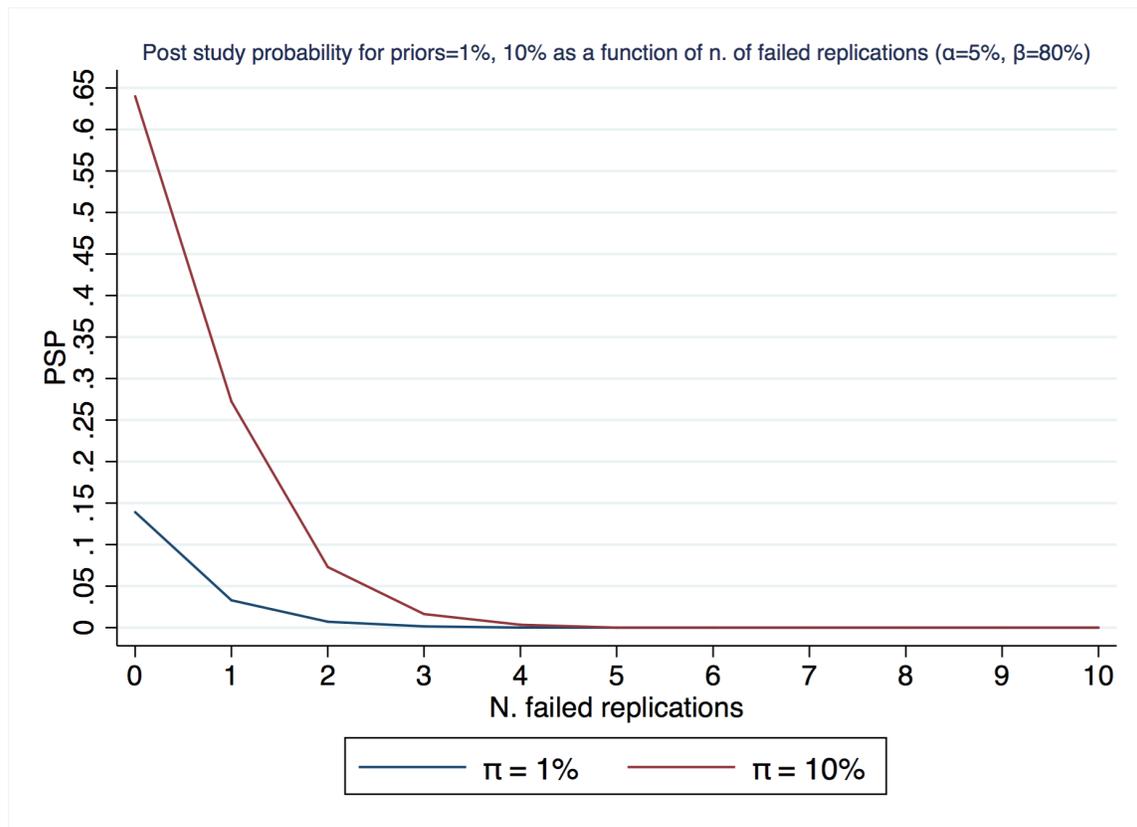
Figure 2 shows how a few failed replications can rapidly undermine the credibility of a novel finding, but overall fail to account for the benefits of replicating. Figure 3 shows how the post-study probability (assuming $\pi = 1\%$) varies based on the number of successful replications out of five and out of ten total replications. The message is clear: just a few successful replications allow robust convergence of post-study probabilities at above 80%.

Figure 3: Post-study probability of a given result being true as a function of the number of successful replications (assuming $\pi = 0.01, \alpha = 0.05, \beta = 0.8$)
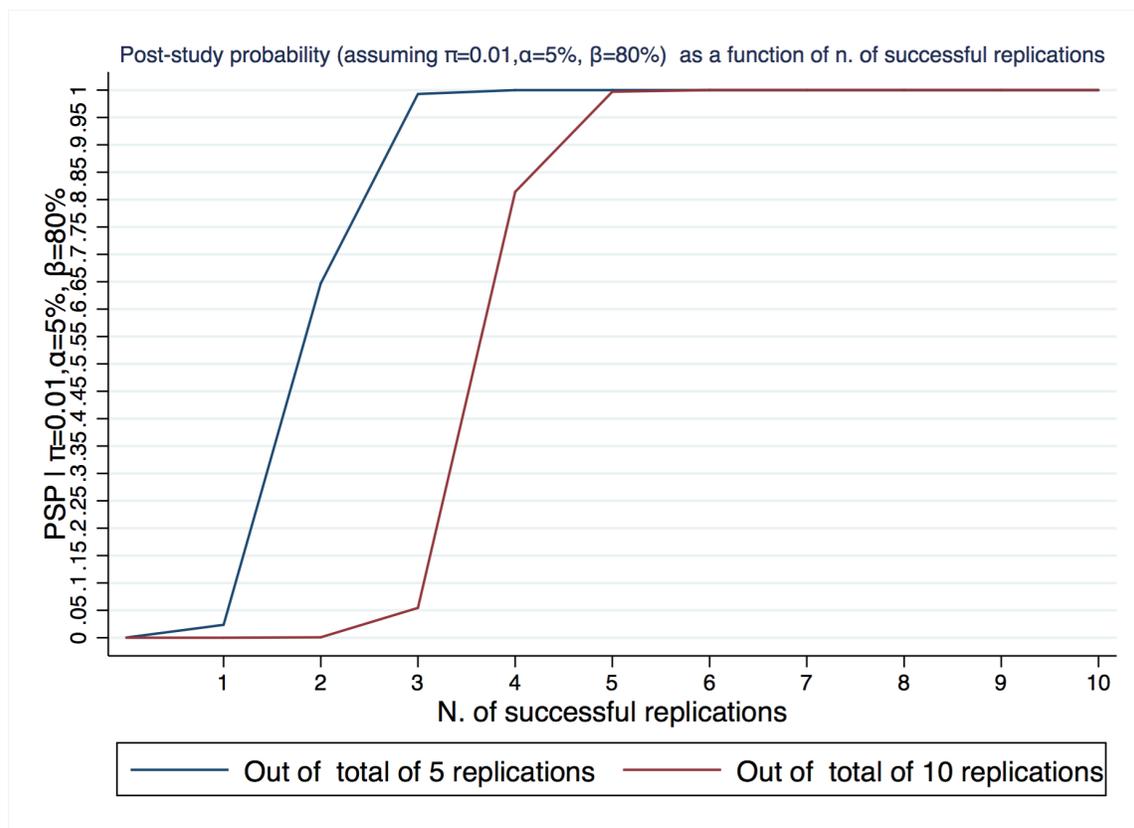
As we have seen however, there are little to no incentives in place for a wider and more systematic use of replications. In the next subsection we propose a simple incentive-compatible mechanism to promote replications.

5.1 The replication mechanism

We propose and put into practice a simple mechanism based on what we believe to be mutual gains from trade between the original investigators and other scholars interested in the same research topics. Our mechanism is as follows for experimentation but it could easily be adapted to more general empirical exercises:

Step 1: Upon completion of data collection and analysis of a new experiment, the original investigators find a significant result. The original investigators commit to write a working paper out of the original study, but commit not to send it to a refereed journal, ever. The working paper, as explained below, should be posted online on an academic repository (e.g. SSRN, working paper series etc.). After calculating the minimum number of replications necessary to substantiate their results given their design, the original researchers offer co-authorship of a second paper to other scholars who are willing to replicate independently the exact experimental protocol at their own institution using their own financial resources. It is a mutual understanding that the second paper is the only paper that will be sent to refereed journals upon completion of all replications, and will include an analysis of the original dataset and of all replication datasets. It is also a mutual understanding that the second paper will reference the first working paper, and that the latter will be coauthored only by the original investigators. The reference to the first working paper serves a dual purpose: it enables the original investigators to credibly signal the paternity of the original research idea, and, as explained below, it provides a binding commitment device for original investigators and other scholars alike that increases the credibility of the replication strategy.

Step 2: Once the original investigators find an agreement with scholars willing to commit to replicate the original study, the original authors pre-register the replication protocol at the American Economic Association RCT registry. The registered protocol includes details about the experimental protocol and materials (e.g. instructions), the data analysis and findings of the original study, lists the names and affiliations of the scholars who will replicate the study, and provides a tentative timeline for replications. All parties agree that only the replications listed in the AEA pre-registration will be included in the second paper.[17]

Step 3: Once step 3 is completed, the original investigators include in the first working paper a section describing the replication protocol, including the list of scholars who will replicate and the reference number for the AEA pre-registration. The original authors then post their first working paper online.

---

[17]The reason for listing the replications and the replications' team in the pre-registration is twofold: first, it provides a commitment device for all scholars involved in the project. Second, and most importantly, it provides a credible signal about the total number of replications that will be conducted. This is critical to avoid any publication bias (e.g. only successful replications are included in the final second paper).

Step 4: Replications are conducted, data is collected, and the second working paper is written and submitted to a refereed journal by the original investigators and the other participating scholars.

6 Replication plan

This section outlines our replication plan. Below is the list of scholars who have agreed to replicate our study and co-author the second paper which will be sent out to a refereed journal.

1. Phillip Grossman - Monash University.

2. Daniel Houser - George Mason University.

3. Marie Claire Villeval - CNRS and University of Lyon.

The replication plan and the original study are registered at the AEA RCT Registry and the unique identifying number is: "AEARCTR-0002142". The public URL of the trial is: www.socialscienceregistry.org/trials/2142.

7 Conclusion and discussion

We set out to investigate how the presence of Knightian uncertainty about the value of a public good affects cooperation, a pervasive and yet unexplored characteristic of most public goods. In a laboratory experiment, participants receive noisy signals about the value of the public good. We vary whether individual signals are privately or publicly observed and the degree of uncertainty. We find, somewhat surprisingly, that introducing noise about the returns from contributing to a public good increases cooperation. We suggest that uncertainty may increase the tolerance for payoffs' reductions of conditional cooperators, since in this environment low returns from a public good may be driven in part by a lower-than-expected quality of the good, rather than just by the presence of free-riders.

Our study, like many others, explores a novel research question and finds results that may look quite surprising to many scholars. Instead of arguing for the veracity of our results, and sending this article to a peer-reviewed journal, we take a step back and propose to put into practice a simple incentive-compatible mechanism to promote replications of novel research findings. We commit never to publish this paper, but we instead offered co-authorship for a second paper to scholars interested in independently replicating this study. Regardless of whether the results of this paper will replicate or ultimately prove being false positives, we believe that our mechanism represents a valuable approach for experimental economists to enhance – and not undermine – the impact of novel studies

exploring uncharted territories, and it is our hope that other scholars will follow our path. More generally, we believe that this approach has viability for any empirical exercises, whether it is an experiment, an empirical exercise using naturally-occurring data, or a meta-analysis of previous studies. Replication exercises should be a pillar of deepening our scientific understanding, and should play a critical role in any model of building scientific knowledge.

References

[1] Jason A Aimone and Daniel Houser. What you dont know wont hurt you: a laboratory analysis of betrayal aversion. *Experimental Economics*, 15(4):571–588, 2012.

[2] Attila Ambrus and Parag A Pathak. Cooperation over finite horizons: A theory and experiments. *Journal of Public Economics*, 95(7):500–512, 2011.

[3] Simon P Anderson, Jacob K Goeree, and Charles A Holt. A theoretical analysis of altruism and decision error in public goods games. *Journal of Public Economics*, 70(2):297–323, 1998.

[4] James Andreoni. Why free ride?: Strategies and learning in public goods experiments. *Journal of public Economics*, 37(3):291–304, 1988.

[5] James Andreoni. Impure altruism and donations to public goods: A theory of warm-glow giving. *The economic journal*, 100(401):464–477, 1990.

[6] James Andreoni. Cooperation in public-goods experiments: kindness or confusion? *The American Economic Review*, pages 891–904, 1995.

[7] Olivier Bochet, Talbot Page, and Louis Putterman. Communication and punishment in voluntary contribution experiments. *Journal of Economic Behavior & Organization*, 60(1):11–26, 2006.

[8] Olivier Bochet and Louis Putterman. Not just babble: Opening the black box of communication in a voluntary contribution experiment. *European Economic Review*, 53(3):309–326, 2009.

[9] Iris Bohnet, Fiona Greig, Benedikt Herrmann, and Richard Zeckhauser. Betrayal aversion: Evidence from brazil, china, oman, switzerland, turkey, and the united states. *The American Economic Review*, 98(1):294–310, 2008.

[10] Gary E Bolton and Axel Ockenfels. Erc: A theory of equity, reciprocity, and competition. *American economic review*, pages 166–193, 2000.

[11] Béatrice Boulu-Reshef, Samuel Brott, Adam Zylbersztejn, et al. Does uncertainty deter provision to the public good? *Revue Economique*, 2016.

[12] Samuel Bowles and Herbert Gintis. Behavioural science: homo reciprocans. *Nature*, 415(6868):125–128, 2002.

[13] Jordi Brandts and Arthur Schram. Cooperation and noise in public goods experiments: applying the contribution function approach. *Journal of Public Economics*, 79(2):399–427, 2001.

[14] A Colin Cameron, Jonah B Gelbach, and Douglas L Miller. Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3):414–427, 2008.

[15] Gary Charness and Matthew Rabin. Understanding social preferences with simple tests. *Quarterly journal of Economics*, pages 817–869, 2002.

[16] Ananish Chaudhuri. Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics*, 14(1):47–83, 2011.

[17] Daniel L Chen, Martin Schonger, and Chris Wickens. otreean open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97, 2016.

[18] Sergio Correia. Linear models with high-dimensional fixed effects: An efficient and feasible estimator. Technical report, 2017.

[19] James C Cox, Daniel Friedman, and Steven Gjerstad. A tractable model of reciprocity and fairness. *Games and Economic Behavior*, 59(1):17–45, 2007.

[20] James C Cox, Daniel Friedman, and Vjollca Sadiraj. Revealed altruism. *Econometrica*, 76(1):31–69, 2008.

[21] Astrid Dannenberg, Andreas Lange, and Bodo Sturm. On the formation of coalitions to provide public goods-experimental evidence from the lab. Technical report, National Bureau of Economic Research, 2010.

[22] Martin Dufwenberg and Georg Kirchsteiger. A theory of sequential reciprocity. *Games and economic behavior*, 47(2):268–298, 2004.

[23] Catherine C Eckel and Philip J Grossman. Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and human behavior*, 23(4):281–295, 2002.

[24] Josef Falkinger, Ernst Fehr, Simon Gächter, and Rudolf Winter-Ebmer. A simple mechanism for the efficient provision of public goods: Experimental evidence. *The American Economic Review*, 90(1):247–264, 2000.

[25] Ernst Fehr and Simon Gächter. Cooperation and punishment in public goods experiments. *The American Economic Review*, 90(4):980–994, 2000.

[26] Ernst Fehr and Klaus M Schmidt. A theory of fairness, competition, and cooperation. *Quarterly journal of Economics*, pages 817–868, 1999.

[27] Urs Fischbacher and Simon Gächter. Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *The American economic review*, 100(1):541–556, 2010.

[28] Urs Fischbacher, Simon Gächter, and Ernst Fehr. Are people conditionally cooperative? evidence from a public goods experiment. *Economics letters*, 71(3):397–404, 2001.

[29] Urs Fischbacher, Simeon Schudy, and Sabrina Teyssier. Heterogeneous reactions to heterogeneity in returns from public goods. *Social Choice and Welfare*, 43(1):195–217, 2014.

[30] Joseph Fisher, R Mark Isaac, Jeffrey W Schatzberg, and James M Walker. Heterogenous demand for public goods: Behavior in the voluntary contributions mechanism. *Public Choice*, 85(3-4):249–266, 1995.

[31] Bruno S Frey and Stephan Meier. Social comparisons and pro-social behavior: Testing" conditional cooperation" in a field experiment. *The American Economic Review*, 94(5):1717–1722, 2004.

[32] Simon Gächter, Elke Renner, and Martin Sefton. The long-run benefits of punishment. *Science*, 322(5907):1510–1510, 2008.

[33] Lata Gangadharan and Veronika Nemes. Experimental analysis of risk and uncertainty in provisioning private and public goods. *Economic Inquiry*, 47(1):146–164, 2009.

[34] Daniel S Hamermesh. Viewpoint: Replication in economics. *Canadian Journal of Economics/Revue canadienne d'économique*, 40(3):715–733, 2007.

[35] Glenn W Harrison and John A List. Naturally occurring markets and exogenous laboratory experiments: A case study of the winner's curse. *The Economic Journal*, 118(528):822–843, 2008.

[36] Daniel Houser and Robert Kurzban. Revisiting kindness and confusion in public goods experiments. *The American Economic Review*, 92(4):1062–1069, 2002.

[37] John PA Ioannidis. Why most published research findings are false. *PLoS Med*, 2(8):e124, 2005.

[38] Martin Kesternich, Andreas Lange, and Bodo Sturm. The impact of burden sharing rules on the voluntary provision of public goods. *Journal of Economic Behavior & Organization*, 105:107–123, 2014.

[39] Mana Komai, Philip J Grossman, and Travis Deters. Leadership and information in a single-shot collective action game: An experimental study. *Managerial and Decision Economics*, 32(2):119–134, 2011.

[40] Michael Kosfeld, Akira Okada, and Arno Riedl. Institution formation in public goods games. *The American Economic Review*, 99(4):1335–1355, 2009.

[41] O Ledyard. Public goods: some experimental results. *Handbook of Experimental Economics Kagel, J. & Roth, A.(ed.)(Princeton Univ. Press, Princeton, 1995)*, 1995.

[42] M Vittoria Levati, Andrea Morone, and Annamaria Fiore. Voluntary contributions with imperfect information: An experimental study. *Public Choice*, 138(1-2):199–216, 2009.

[43] John A List. The market for charitable giving. *The Journal of Economic Perspectives*, 25(2):157–180, 2011.

[44] John A List, Azeem M Shaikh, and Yang Xu. Multiple hypothesis testing in experimental economics. 2016.

[45] Zacharias Maniadis, Fabio Tufano, and John A List. One swallow doesn't make a summer: New evidence on anchoring effects. *The American Economic Review*, 104(1):277–290, 2014.

[46] David Masclet, Charles Noussair, Steven Tucker, and Marie-Claire Villeval. Monetary and nonmonetary punishment in the voluntary contributions mechanism. *The American Economic Review*, 93(1):366–380, 2003.

[47] David M McEvoy, Todd L Cherry, and John Stranlund. The endogenous formation of coalitions to provide public goods: Theory and experimental evidence. *Available at SSRN 1802703*, 2011.

[48] Ramal Moonesinghe, Muin J Khoury, and A Cecile JW Janssens. Most published research findings are falsebut a little replication goes a long way. *PLoS Med*, 4(2):e28, 2007.

[49] Nikos Nikiforakis. Feedback, punishment and cooperation in public good experiments. *Games and Economic Behavior*, 68(2):689–702, 2010.

[50] Thomas R Palfrey and Jeffrey E Prisbrey. Anomalous behavior in public goods experiments: How much and why? *The American Economic Review*, pages 829–846, 1997.

[51] Louis Putterman, Jean-Robert Tyran, and Kenju Kamei. Public goods and voting on formal sanction schemes. *Journal of Public Economics*, 95(9):1213–1222, 2011.

[52] Matthew Rabin. Incorporating fairness into game theory and economics. *The American economic review*, pages 1281–1302, 1993.

[53] Ernesto Reuben and Jean-Robert Tyran. Everyone is a winner: Promoting cooperation through all-can-win intergroup competition. *European Journal of Political Economy*, 26(1):25–35, 2010.

[54] Martin Sefton, Robert Shupp, and James M Walker. The effect of rewards and sanctions in provision of public goods. *Economic inquiry*, 45(4):671–690, 2007.

[55] Sholom Wacholder, Stephen Chanock, Montserrat Garcia-Closas, Nathaniel Rothman, et al. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *Journal of the National Cancer Institute*, 96(6):434–442, 2004.