Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Economia

Ciclo XXVII

Settore Concorsuale di afferenza: 13/A1

Settore Scientifico disciplinare: SECS-P01

Experiments on Pro-Sociality and Environmental Issues

Presentata da: Riccardo Ghidoni

Coordinatore Dottorato

Matteo Cervellati

Relatore

Marco Casari

Esame finale anno 2015

When do the Expectations of Others Matter?

An Experiment on Distributional Justice and Guilt Aversion*

Riccardo Ghidoni† Matteo Ploner‡

Abstract

In a modified dictator game experiment, we study how distributional justice, measured by the proportionality between effort exerted and rewards obtained, and guilt feelings triggered by others' expectations affect dictator's choices. We consider these two sources of behavior in isolation and in interaction. Our results suggest that both justice concerns and guilt aversion are important drivers of behavior. However, the expectations of others are more relevant when the choice environment is likely to induce less equitable outcomes.

Keywords: Justice; Guilt Aversion; Entitlement Rights; Beliefs; Experiments.

JEL: A13; C91; D03

*Financial support was provided by a grant from the Einaudi Institute for Economics and Finance (EIEF).

†Department of Economics, University of Bologna. Email: `riccardo.ghidoni2@unibo.it`. Web: `https://sites.google.com/site/riccardoghidoni/`

‡Department of Economics and Management, University of Trento. Email: `matteo.ploner@unitn.it`. Web: `https://sites.google.com/site/mploner/`

1 Introduction

In a well-known passage of his Second Treatise on Government, John Locke writes: "I think, it is very easy to conceive, without any difficulty, how labour could at first begin a title of property in the common things of nature [...]" (Chapter 5, Section 51). This sentence captures the essence of what Locke labels a law of nature, i.e. that property rights on goods originate directly from effort exerted to generate them. The Lockean law of nature is grounded on a basic justice principle according to which outcomes should be related to actions (desert). This general distributional principle has been subject to extensive philosophical debate and has attracted the attention of experimental research, both in social psychology (e.g. Leventhal and Michaels, 1969) and in economics (e.g. Hoffman and Spitzer, 1985; Konow, 2000).

We aim at assessing the relevance of the "hard-wired" justice principle that relates actions and consequences against an alternative measure of justice driven by emotions originating in social interactions. Emotions have indeed been recognized to have a major influence on economic behavior (Elster, 1998). In the last decade, the emotion of guilt has received attention both from a theoretical and experimental perspective, due also to the theory of *guilt aversion* (e.g. Dufwenberg and Gneezy, 2000; Charness and Dufwenberg, 2006; Battigalli and Dufwenberg, 2007; Ellingsen et al., 2010).[1] According to this theory, decision-makers can experience a negative feeling, i.e. guilt, whenever they believe that their action will contribute to let their counterpart down.

Relying on a within-subjects experimental design, we endeavor to understand if and how justice concern à la Locke and guilt aversion interact in shaping the behavior of decision-makers. Our main objective is to test whether expectations of the counterpart about the behavior of the decision-makers affect decision-makers' behavior only when these expectations are not in conflict with justice considerations. In other words, justice considerations may be key to understanding when others' expectation are seen as *legitimate* by the decision-maker and, thus, worth taking into account.

Our data show that outcome-based social preferences like altruism (e.g. Cox et al., 2008), inequity-aversion (e.g. Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999), or

[1] Although the focus of the present paper is on the theory of guilt aversion as formalized by Charness and Dufwenberg (2006) and Battigalli and Dufwenberg (2007), we are aware that the emotion of guilt has received attention in other settings (e.g. Lindbeck and Nyberg, 2006; Cervellati et al., 2010).

efficiency (e.g. Charness and Rabin, 2002) play a modest role in explaining the behavior of the subjects. Instead, guilt feelings and justice considerations are key in driving the choices of decision-makers. However, quite unexpectedly, others' expectations become more relevant when the choice environment offers less protection to entitlement rights originated by effort and, thus, it is less likely to produce equitable outcomes. In the concluding section, we provide a possible explanation for this result which contradicts our main research hypothesis. We also draw attention to how further research on the interaction between institutional protection of entitlement rights and subjective expectations about standards of behavior is needed.

The rest of the paper proceeds as follows: in Section 2, we present a literature review of desert and of guilt aversion theories; Section 3 describes the experimental game and the experimental procedures; Section 4 reports the research hypotheses; Section 5 illustrates the experimental results; in Section 6, we discuss the results and provide final remarks.

2 Related Literature

2.1 Distributional Justice

Konow (2003) presents an extensive review of justice concepts that are relevant to economics and provides the empirical basis to build a descriptive theory of justice. The author highlights the relevance of theories that appeal to desert and relate fair allocations to individual actions. Within this class of models, equity theory (e.g. Adams, 1963) provides guidance to assess the fairness of allocations in which a production stage is involved. The basic tenet of the theory is that an equitable allocation should preserve the proportionality of resources invested (input) and rewards obtained (output) across individuals. Thus, those investing more resources in the production of the output should obtain more out of it than those investing less. Hoffman and Spitzer (1985) point out how the proportionality principle underlying equity theory captures the essence of the Lockean natural law. Empirical support for the relevance of this justice principle is provided by experimental studies in social psychology (e.g. Leventhal and Michaels, 1969; Mikula, 1974) and economics (e.g. Konow, 2000). Relying on empirical evidence accumulated, Gill and Stone (2010) provide a theoretical account of desert concerns in tournaments and shows how concerns of this kind may substantially alter outcomes in strategic interactions.

To establish whether an allocation is fair, it is crucial to define the nature of the input against which the output is evaluated. According to attribution theory (e.g. Weiner, 1985), only factors that are directly controlled by individuals qualify to establish the fairness of an allocation. This principle of justice is captured also by the accountability principle of Konow (1996), which distinguishes between discretionary variables and exogenous variables. According to Konow, only the former should be taken into account when assessing the fairness of allocations. The recent experimental work by Becker (2013) shows that normative beliefs are strongly influenced by accountability considerations, but individuals often violate the principle and selfishly enjoy rewards that originate from pure luck. These results are also supported in the structural estimation presented by Cappelen et al. (2007), when comparing alternative fairness ideals. The authors identify a plurality of fairness ideals in the population, but also show that a non negligible share of the population maintains a libertarian view according to which individuals deserve what they produce, irrespective of the control exerted over the production factors.

2.2 Guilt Aversion

The idea of guilt aversion understood as the aversion to letting others down was first introduced by Dufwenberg and Gneezy (2000), who observe, both in a lost wallet game and in a dictator game, a positive correlation between trustees' (dictators') transfers and their second-order beliefs elicited after the play.[2] In a later experiment, also Bacharach et al. (2007) find evidence supporting this definition of guilt aversion in three modified trust games. Charness and Dufwenberg (2006) and Battigalli and Dufwenberg (2007) develop a formal model of guilt aversion based on the analytical framework of the *psychological games* (Geanakoplos et al., 1989; Battigalli and Dufwenberg, 2009).[3]

An important application of guilt aversion is provided by Charness and Dufwenberg (2006), who exploit this theory to explain why individuals in experiments tend to keep non-binding promises even when they would earn more by not doing so. Specifically, Charness and Dufwenberg propose a modified trust game that allows for moral hazard by trustees (so-called trust game *with hidden action*), where subjects can also exchange free

[2] The lost wallet game is a modified trust game which allows for several wealth multipliers, and where the trustor only faces a dichotomous choice between trusting the trustee or not.

[3] Battigalli and Dufwenberg (2007) introduce a distinction between *simple guilt* and *guilt from blame*, where the main difference between the two is in the fact that in the second form of guilt aversion, the decision-maker dislikes being blamed, rather than simply letting the counterpart down.

form messages before the play. Charness and Dufwenberg first observe a positive correlation between promises and trustworthiness. Moreover, consistently with guilt aversion, they find that trustworthy choices are positively correlated with trustees' second-order beliefs elicited after the play. Thus, Charness and Dufwenberg conclude that promises foster trustees' second-order beliefs about trustors' first-order belief on trustworthiness, triggering a sense of guilt in those trustees who let trustors down.

Charness and Dufwenberg's conclusions are questioned by Vanberg (2008), who tests whether individuals keep their promises because they dislike letting others down, as predicted by guilt aversion, or rather because they have a taste for keeping their own word (so-called "commitment-based explanation"), as modeled by Ellingsen and Johannesson (2004) and Kartik (2009). Using a design that closely resembles that of Charness and Dufwenberg's, Vanberg finds support for the commitment-base explanation. In a trust game with hidden-action, where subjects could only send bare promises, Charness and Dufwenberg (2010) test whether "truth-value [is] all we need to capture an important aspect of human motivation, or does the context in which the statement was made matter". These authors find limited support both for a commitment-based explanation and for guilt aversion.

Ellingsen et al. (2010) challenge Charness and Dufwenberg's conclusions from another perspective arguing that the positive correlation between trustees' second-order beliefs and their back-transfers is mainly due to the so-called "false consensus" effect (Engelmann and Strobel, 2000, 2011), rather than to guilt aversion.[4] According to this alternative explanation, trustees may consider their own back-transfer choice as the most representative, and hence believe that trustors' first-order beliefs will coincide with their own. Thus, the same trustees who prefer to make a large back-transfer could be those who believe that trustors expect a large back-transfer. To test this hypothesis, instead of eliciting second-order beliefs after the play, Ellingsen et al. reveal their partner's first-order belief to those subjects who were supposed to suffer from guilt. Ellingsen et al. do not detect any evidence of guilt aversion in any of their experimental games, i.e. dictator game, lost-wallet game, and trust game. Reuben et al. (2009), however, proposes a cleaner version of the design by Ellingsen et al. and find that trustees are less trustworthy when trustors

[4]Vanberg also remarks that the correlation between second-order beliefs and actions could be driven by the false consensus effects.

have low expectations, concluding that mis-trust is self-fulfilling.[5] Additional evidence in support of guilt aversion in a different experimental setting, is offered by Ockenfels and Werner (2014) who study how different scales with which beliefs are elicited before the game can subsequently affect subjects' behavior. Their findings also seem to suggest that, in dictator games, dictators are motivated by not letting recipients' down rather than by conforming to a norm regarding dictators' behavior. Bellemare et al. (2011) estimate structural models of guilt from a large scale sequential game, where subjects are willing to pay between forty to eighty cents to avoid letting their partner down by one euro.

Recently, Battigalli et al. (2013) proved how guilt aversion can offer an explanation for the findings of the seminal experiment on deception by Gneezy (2005). Dufwenberg et al. (2011) show how framing effect can influence contributions in a framed public goods game through changes in subjects' second-order beliefs. Finally, Babcock et al. (2014) study the social effects of incentives in a series field experiments. One of the finding is that subjects prefer individual incentives to team incentives, even if the latter are more effective. Authors show that this result is consistent with the hypothesis that team effects operate through guilt rather than pure altruism.

3 Experimental Design

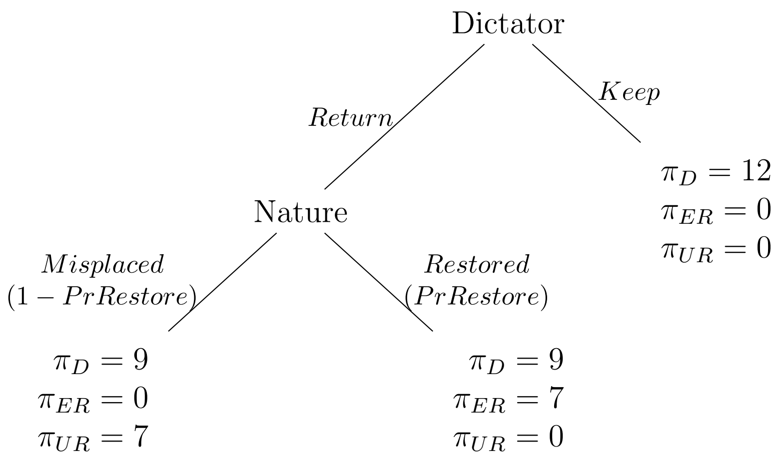
The Modified Dictator Game. In our dictator game (Figure 1), a *dictator* is endowed with two wallets. The dictator can either keep both wallets, or return one wallet to the *entitled recipient*, who lost it. Both players have exerted an effort to earn their own wallet. Thus, if the dictator keeps both wallets, the entitled recipient remains with no reward for the effort done. If the dictator chooses to return the wallet, the entitled recipient will restore it only with a certain "restoring probability", which can take value 4/6, 5/6, or 6/6. Otherwise, the wallet is misplaced by Nature to an *unentitled recipient* (who exerted no effort), with probability 2/6, 1/6, or 0/6, respectively.[6] The restoring probability is

[5] Reuben et al. operate three major changes on the Ellingsen et al.'s design: (1) they provide all subjects with the same instructions; (2) all subjects play both as trustor and trustee sequentially, and beliefs are elicited when subjects play the role of senders, in a round when beliefs are unused; (3) they increase the reward for the accuracy of beliefs and the game payoffs.

[6] The payoffs ensure that neither efficiency nor surplus distribution change under alternative restoring probabilities. Choosing *Keep* is inefficient, since the payoff obtained by the dictator is lower than the sum of two wallets, i.e. 12 ECU instead of 14 ECU.

private information of the dictator, and the recipients do not observe the action of the dictator. So, the entitled recipient cannot infer whether the dictator kept her wallet, or Nature misplaced it.[7] The restoring probability captures (exogenous) institutional aspects, known only to the decision-maker, which ensure different degrees of protection to the entitlement right of the entitled recipient.

Figure 1: Modified Dictator Game.



Note: D: dictator; ER: entitled recipient; UR: unentitled recipient.

The Experimental Session. 15 subjects participates in each session, divided into two groups, A (10 subjects) and B (5 subjects). Group A members actively participate in all the three stages of the experiment (Figure 2); group B members actively participate only in the second stage, i.e. beliefs elicitation. Group B members enter the laboratory after Stage 1 and are allowed to surf the Internet during Stage 3.[8]

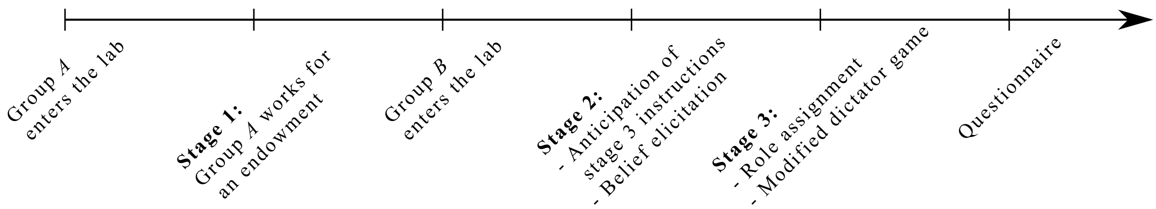
In Stage 1, each group A member earns her wallet, by counting the number of zeros in seven 15×8 tables that sequentially appear on computer screens, and contain 0 and 1 digits in random proportions.[9] For each table solved, the subject earns 1 ECU (Experimental

[7] The presence of the Nature and the unobservability of the dictator's action are common features to the games of Charness and Dufwenberg (2006) and Vanberg (2008).

[8] Interactions between group A and group B were avoided: group B had to show up 10 minutes later than group A, and group A was allowed to leave the lab only after group B had left. To ease the understanding of instructions, group A and B were identified as green and red, respectively. We provided subjects with intuitive slides and checked comprehension using a quiz (Bigoni and Dragone, 2012).

[9] We borrow this task from Abeler et al. (2011) to induce a sense of entitlement on wealth in the experiment. Previous works have shown the relevance of asset legitimacy in simple bargaining situations. As an example, Hoffman et al. (1994) and Cherry et al. (2002) administered a knowledge questionnaire

Figure 2: Timeline of the Experimental Session.



Currency Unit, 1 ECU=€1). Subjects are not time constrained, and can make mistakes. At the end of Stage 1, each group A member virtually owns a wallet of 7 ECU.[10]

In Stage 2, all subjects are explained the modified dictator game, in order to collect their expectations about the dictator's behavior. Before the play, half of the group A members randomly "lose" their wallet, and become entitled recipients. Lost wallets are "found" by the other half of group A members, who become dictators. Group B members are the unentitled recipient. The game is played repeatedly with a *perfect stranger matching* for 3 rounds. Each dictator faces the three restoring probabilities in random order (unknown to the recipients). No feedback is given to the players at the end of each round. At the end of the session, one round is randomly drawn for the payment: dictators are informed about the outcome of the payment round, while recipients only learn the amount of ECU they obtained. Unlike Charness and Dufwenberg (2006), we do not elicit second-order beliefs after the play. In this way, we avoid identification problems due to false consensus effects (Engelmann and Strobel, 2000, 2011; Ellingsen et al., 2010). Instead, we induce guilt feelings by providing dictators with the first-order belief of the entitled recipient.[11]

Belief elicitation is key to our design. In Stage 2, before role assignment, we ask *all* subjects to anticipate how many times out of the three, they expect a *generic* dictator to return the wallet (Table 1). The reward for the belief accuracy depends on a randomly

to participants before taking part in an ultimatum and a dictator game respectively, and show that this strongly reduces other-regarding concerns. We chose the Abeler et al. task because it provides a more direct measure of effort exerted relative to a knowledge test in which human capital and luck seem to be more important than effort.

[10]In the instructions, we do not use the word "wallet", but we stress that each subject earns her own endowment by performing the task.

[11]A similar solution to false consensus effect is used by Ellingsen et al. (2010) and Reuben et al. (2009), with the former finding no correlation between decision-maker choices and counterparts' expectations, while the latter find evidence of guilt aversion.

selected choice of a dictator: subjects know that the relevant dictator for the payment of belief's accuracy is not the same dictator whose choice is selected for the payment of the dictator game.[12] A *quadratic scoring rule* is used to incentivize subjects to truthfully report their belief. To avoid a methodologically questionable omission of relevant information, we inform group A members that their beliefs may be disclosed to dictators. Knowing that with 50% probability their beliefs are reported to dictators, group A members might strategically manipulate their beliefs.[13] We control this issue *ex-post*, by comparing the belief distributions of group A and group B, whose members know that their belief are not disclosed to anyone.

Table 1: Elicitation of First–Order Beliefs

	Dictator will choose *Return*...			
	0 out of 3	1 out of 3	2 out of 3	3 out of 3
Your guess...	☐	☐	☐	☐
Your earnings if in the drawn choice...				
Dictator chose *Return*	€0	€2.80	€4.40	€5
Dictator chose *Keep*	€5	€4.40	€2.80	€0

Procedures. The experiment was implemented at the Cognitive and Experimental Economics Laboratory (CEEL) of the University of Trento. 180 students took part in 12 sessions. Subjects were recruited using dedicated software. The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007). All subjects received a show-up fee of €3.

4 Behavioral Predictions

Under standard assumptions of selfish rationality, dictators in our game should never choose to return the wallet to an entitled recipient, irrespective of both the probability with which the wallet gets misplaced and the entitled recipient's expectation. However, a large body of literature has highlighted how social preferences may play an important role

[12]This way, payoffs consequences of a dictator's choice do not extend to the belief elicitation. To inhibit recipients' belief updates, we do not provide any feedback during the three rounds of play.

[13]Roles in the game are not yet assigned when beliefs are elicited. This should discourage potential strategic manipulation of reported beliefs by prospective group A members.

in determining the behavior of individuals participating in experiments. In our context, outcome-based social preferences like altruism (e.g. Cox et al., 2008), inequity-aversion (e.g. Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999), or efficiency (e.g. Charness and Rabin, 2002) may induce the dictator to return the wallet. Thus, unlike under assumptions of selfish rationality, a large class of social preferences predicts that dictators may choose to return the wallet. However, like under selfish rationality, outcome-based social preferences predict that dictators decisions should be neither affected by different restoring probabilities nor by the expectations of the entitled recipient.

While selfishness and outcome-based social preferences provide us with useful benchmark behavioral predictions, the main focus of our inquiry is on how justice concerns linked to the protection of entitlement rights (proxied by restoring probabilities) and counterparts' expectations (measured by recipients' first-order beliefs) affect dictators' decisions. Accordingly, we present some testable hypotheses which refer to the impact of these potential sources of behavior.

When a dictator "finds" the wallet lost by an entitled recipient, a strongly unfair and inequitable allocation is exogenously induced because both subjects have exerted the same effort, but (almost) all the surplus generated is given to the dictator. Dictators can restore justice by returning the wallet for which the entitled recipient worked. However, when the returned wallet is misplaced because of an interference of Nature, an even less equitable allocation is in place as the wallet is given to someone who did not exert any effort at all. This is due to the fact that the welfare loss when one receives more than what deserved is likely to be smaller than the welfare loss when another party receives more than what deserved.[14] Therefore, alternative restoring probabilities impact on the anticipated proportionality between efforts exerted to "generate" the wallet and rewards obtained when the wallet is returned.

To illustrate our behavioral predictions, we consider a simple model in which the utility of the dictator i is given by $U_i = \pi_i + \theta\phi(\alpha_j, \lambda, \pi_j)$, where π_i and π_j are the

[14] According to Gill and Stone (2010), when one gets more than what deserved, she can either experience a welfare gain (*desert elation*) or a welfare loss (*desert guilt*). However, when the other gets more than what deserved, a loss of welfare is unambiguously experienced (*desert loss*). Moreover, the change in welfare experienced when receiving less than what deserved is assumed to be larger than the change in welfare when receiving more, consistently with the loss aversion assumption of Prospect Theory (Kahneman and Tversky, 1979).

monetary payoffs of i and j, α_j captures j's beliefs about wallet return, and λ is the restoring probability faced by the dictator. The function $\phi(\cdot)$ captures welfare gains from returning the wallet, with $\theta > 0$ measuring how much this is valued relative to increases in own monetary payoff. We allow $\phi(\cdot)$ to change according to motivations of dictators and identify three main sources of motivation: pure justice concerns, pure guilt aversion, and reasonable guilt aversion. In the following, we discuss qualitative predictions obtained for distinct motives.[15]

4.1 Pure Justice Concerns

According to equity theory (Adams, 1963) and to the accountability principle (Konow, 2000, 2003), individuals investing more resources in the production of a certain surplus should receive a larger share of it than those investing less. In our context, justice concerns based on input/output proportionality translate into a higher likelihood of returning the wallet when the chances of a misplaced return are lower.

In terms of the utility function presented above, we capture pure justice concerns by assuming that $\phi(\cdot) = \lambda\pi_j$, so that $U_i = \pi_i + \theta\lambda\pi_j$. Under this representation of U_i, we have that the utility of returning the wallet is increasing in λ. Accordingly, we predict that

Hypothesis 1. *the higher the restoring probability (i.e. the lower the likelihood of a misplaced return), the more likely dictators are to choose to return the wallet.*

4.2 Pure Guilt Aversion

According to the theory of guilt aversion (Charness and Dufwenberg, 2006; Battigalli and Dufwenberg, 2007), an individual will be disappointed when the final outcome of the game does not match expectations. Thus, decision-makers who can influence the outcome of the game may experience a negative emotion of guilt whenever their action contributes to letting their counterpart down. A guilt sensitive decision-maker may hence prefer to renounce part of her material payoff to avoid the psychological cost of guilt. Given that in our experiment recipients are kept blind about the restoring probability faced by the

[15] For the sake of exposition, we neglect here that the three motivations are simultaneously at work. However, the analysis of experimental outcomes reported below allows for this possibility.

dictator, more optimistic recipients will always be more disappointed by realizing that they did not receive their wallet back. The feeling of guilt experienced when retaining the wallet should hence be stronger when facing an optimistic recipient than when facing a less optimistic one, irrespective of the restoring probability.

More formally, we capture pure guilt aversion by assuming that $\phi(\cdot) = \alpha_j \pi_i$, which implies that $U_i = \pi_i + \theta \alpha_j \pi_i$. The utility of returning the wallet is higher for higher values of α_j. Since in our context dictators freely access the expectations of the entitled recipient[16], we predict that

Hypothesis 2. *the more optimistic the recipients' expectations about the return of the wallet, the more likely dictators are to return it..*

4.3 Reasonable Guilt Aversion

Predictions based on pure justice considerations and pure guilt aversion provide us with a guidance to evaluate the impact of equity considerations and of recipient's expectation in isolation. However, these two sources of behavior are likely to interact in shaping dictator's behavior. Even though guilt aversion predicts dictators to suffer the same amount of guilt irrespective to the restoring probability, we expect dictators to attach a different weight to the same expectation when facing different restoring probabilities. The intuition here is that the impact of counterpart's expectation on dictator's choices is conditional upon environmental features, which may remain unknown to the former but not to the latter. We rely on the assumption that beliefs of j, reflect justice considerations. This implies that if $\alpha_j = 1/3$ (seldom), j expects i to return the wallet only when $\lambda = 6/6$. By the same token, when $\alpha_j = 2/3$ (often), j expects the dictator i to return the wallet when $\lambda = 6/6$ and $\lambda = 5/6$, and when $\alpha_j = 3/3$ (always), j expects the dictator i to return the wallet when $\lambda = 6/6$, $\lambda = 5/6$, and $\lambda = 4/6$.

When a dictator i learns α_j, he will check the consistency between the restoring probability λ in that specific round and the belief α_j. A belief is reasonable only if it

[16] Similar to Charness and Dufwenberg (2006), we implicitly assume that "players do not coordinate on some "equilibrium"; it refers only to the individual player and properties of his/her utility." In particular, we assume that recipients do not perfectly anticipate the degree of guilt sensitivity of the dictators they meet across the three rounds. This assumption seems reasonable since guilt sensitivity can differ across individuals (Tangney, 1995). Moreover, we *never* provide recipients with feedback about dictators' choices in order to avoid any belief update.

reflects the actual restoring probability. Accordingly, the set R of reasonable beliefs is given by the following ordered pairs (λ, α_j):

$$R = \{(6/6, 1/3),\ (6/6, 2/3),\ (6/6, 3/3),\ (5/6, 2/3),\ (5/6, 3/3),\ (4/6, 3/3)\}.$$

In terms of the utility function presented above, we assume that $\phi(\cdot) = I_R \alpha_j \pi_j$, where I_R is equal to 1 if $\alpha_j \in R$, and equal to 0 otherwise. Thus, we predict that

Hypothesis 3. *conditional upon beliefs being reasonable, the more optimistic the recipients' expectations about the return of the wallet, the more likely dictators are to return it.*

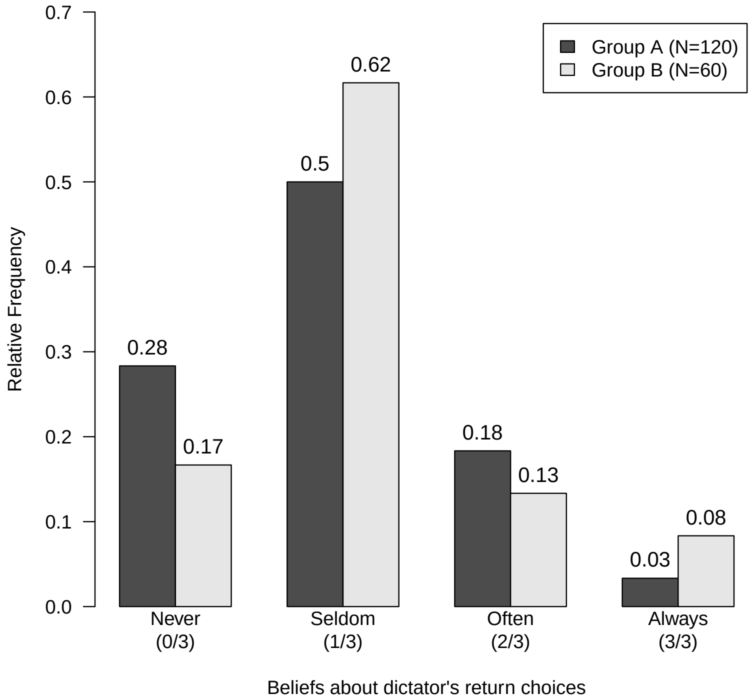
5 Results

5.1 Descriptive Statistics

Figure 3 shows the separate distribution of the first-order beliefs about dictators' return choices for the A and B groups.[17] When self-reporting their beliefs, individuals in the A group are aware that this piece of information could be potentially disclosed to dictators. On the other hand, individuals in the B group know that their beliefs are not reported to dictators. Thus, the comparison between the self-reported beliefs of the two groups provides us with a control of the potential strategic manipulation of beliefs by prospective recipients. We adopt the labels "Never", "Seldom", "Often", and "Always" to identify beliefs that range from 0 returns out of 3 choices made, to 3 returns out of 3 choices made.

[17] When beliefs are elicited, members of group A do not already know to which role they will be assigned in the modified dictator game.

Figure 3: First-Order Beliefs about Dictator's Return

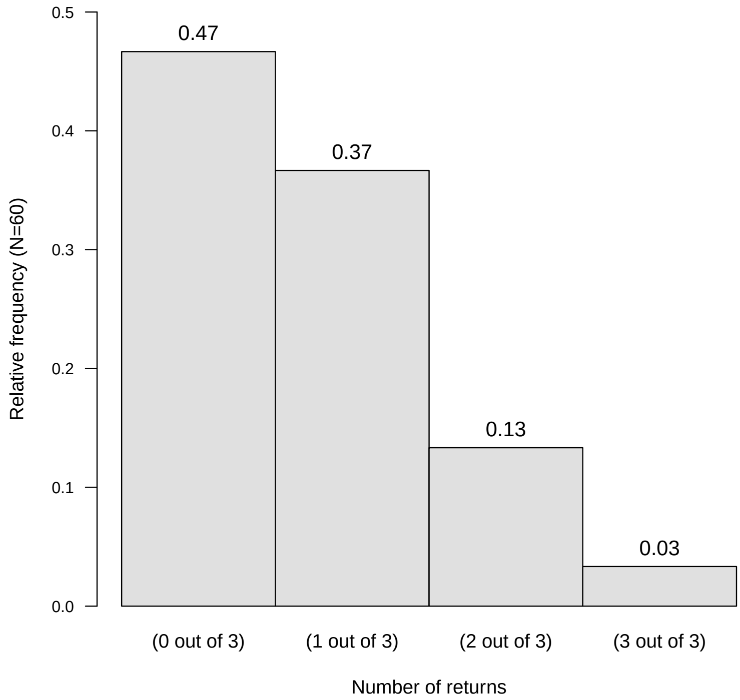


In both group A and B, a large majority of the participants (about 80%) expects a generic dictator to return the wallet less than 2 times out of the 3 total choices, with the mode of the distribution corresponding to the "Seldom" return frequency (1 out of 3). Although members of group A are slightly more pessimistic than group B members (0.32 vs. 0.37 respectively), when comparing the distributions of beliefs in the two groups, no statistically significant difference is observed (Fisher's Exact Test, p-value=0.117). Thus, we can safely conclude that those in the A group do not strategically manipulate their beliefs in a significant way.

Earnings in the belief task provide us with a direct measure of belief accuracy. Median earnings in the task are equal to €4.40 both for the red and the green group, just one step away from the maximum earnings of €5. Thus, self-reported beliefs seem to be overall quite accurate.

Figure 4 displays the relative frequency of individual-level return choices.

Figure 4: Return Choices per Dictator



As shown above, 47.0% of the dictators choose to never return. Among those choosing to return, the majority chooses to return the wallet only once and only 2 dictators out of 60 choose to always return the wallet. Overall, half of the dictators choose differently across the three rounds.

Figure 4 clearly shows that about half of collect choices are compatible with selfishness. Concerning the conspicuous share of non-selfish dictators, very few choices are compatible with standard outcome-based social preferences. As pointed out in Section 4, non-selfish individuals concerned only with the payoff consequences of their actions are likely to always return the wallet, irrespective of the restoring probability and of beliefs of the counterpart. The behavior of non-selfish dictators seems to be largely conditional upon the contextual elements in which the choice is made for different probabilities of success, with more dictators returning for the highest probability (30.0%) than for the other two probabilities of success (21.7%). Table 2 reports the percentage returns observed for alternative levels of restore probability and of counterpart's beliefs, both when considered alone and when jointly taken. As an example, the first cell in the table shows that out of the 19 dictators who jointly faced a probability of success of 4/6 and a recipient's belief of zero returns in a round, only 1 of them decided to return the wallet (i.e. 5.3%).

As shown by the last column in Table 2 (*Overall*), in correspondence to the highest

Table 2: Return Choices conditional upon Recipient's Beliefs and Restoring Probabilities

	Recipient's First-Order Belief				
PrRestore	Never (N=19)	Seldom (N=27)	Often (N=13)	Always (N=1)	*Overall*
4/6	5.3%	25.9%	38.5%	0.0%	21.7%
5/6	15.8%	25.9%	23.1%	0.0%	21.7%
6/6	31.6%	37.0%	15.4%	0.0%	30.0%
Overall	17.5%	29.6%	25.6%	0.0%	24.4%

probability of restore we can observe the highest percentage in returns, which is in line with Hypothesis 1. However, the percentage of returns in the intermediate probability of restore and in the lowest do not differ. The bottom row in Table 2 provides evidence about the impact of a counterpart's beliefs and shows that the lower percentage of returns is observed in correspondence to the lowest level of beliefs, in line with Hypothesis 2, but the maximum share of returns is observed in correspondence to "Seldom" rather than to "Often".[18]

When taking into account the joint effect of beliefs and restore probabilities, the frequency of returns monotonically increases in the beliefs of the counterpart, in correspondence to $PrRestore = 4/6$. A Fisher's exact test (FET) shows that there is a statistically significant difference between return choices in correspondence to the "Never" and "Often" belief levels (p-value$= 0.029$). Moreover, the return rates observed in correspondence to these two belief levels are the overall lowest and highest respectively. In contrast, for $PrRestore = 5/6$ and $PrRestore = 6/6$ the impact of beliefs of the counterpart seems to be more erratic, with no significant differences in return choices in correspondence to the "Never" and "Often" belief levels (FET, p-values≥ 0.420).[19]

The descriptive analysis reported above suggests that both restore probabilities and beliefs of the counterpart have a positive impact on return decisions even though the impact of these variable is not fully in line with Hypothesis 1 and 2. Moreover, beliefs and probabilities seem to interact in shaping dictator choices in a way that is in contrast to our Hypothesis 3. Below we report a regression analysis that provides us with a further

[18]The belief level "Always" is neglected in the discussion because only one observation is available.

[19]For any $PrRestore$, the comparisons of "Never" and "Seldom" and of "Seldom" and "Often" do not reveal any significant difference (all p-values≥ 0.115)

test of our three main hypotheses and which casts further light on the behavior in the experiment.

5.2 Regression Analysis

Table 3 reports the outcomes of a series of Generalized Linear Mixed Logit (GLM Logit) models controlling for repeated choices at the individual level through random effects. The dependent variable in all regressions is the dictator's decision to return the wallet (*Return*), assuming value 1 when the dictator returns and 0 otherwise.

Four distinct estimates are presented in Table 3. In Model 1, we only consider first-order beliefs of the recipient (*RecBel*) as the main explanatory variable, with *RecBel* assuming values $\{\frac{0}{3}, \frac{1}{3}, \frac{2}{3}, \frac{3}{3}\}$ for beliefs equal to "Never", "Seldom", "Often", and "Always", respectively. In Model 2, we also control for the alternative probabilities with which the entitled recipient restores her wallet once the dictator chose to return it ($PrRestore$), i.e. $\frac{4}{6}, \frac{5}{6}$, or $\frac{6}{6}$. In Model 3, we also include the interaction between recipients' beliefs and a dummy variable capturing the "reasonableness" of beliefs, according to the definition given above (see Hypothesis 3).[20] Model 4 introduces, as an explanatory variable, the dictator's first-order belief about the behavior of other dictators (*DictBel*) and how these belief stand relative to the belief of the recipient, with a specific focus on the case in which beliefs of the recipient are more optimistic than those of the dictator ($RecBel > DictBel$).[21]

Finally, for each estimate we considered a set of control variables regarding gender ($Male$), employment status ($Worker$), citizenship ($Italian$), and previous participation in experiments (*Experienced*).

[20]Beliefs are deemed reasonable if they reflect the actual probability of restoring faced by the dictator, under the assumption that more optimistic beliefs are associated to higher restoring probabilities. Specifically, the dummy variable *Reason* is equal to 1: when $PrRestore = \frac{6}{6}$ and $RecBel = \frac{1}{3}$ or $RecBel = \frac{2}{3}$ or $RecBel = \frac{3}{3}$; when $PrRestore = \frac{5}{6}$ and $RecBel = \frac{2}{3}$ or $RecBel = \frac{3}{3}$; when $PrRestore = \frac{4}{6}$ and $RecBel = \frac{3}{3}$. Otherwise, the dummy *Reason* is equal to 0.

[21]The beliefs of the dictator are collected before knowing the actual role in the game. Thus, *DictBel* and *RecBel* refer to the same set of beliefs.

Table 3: Determinants of return choices.

	Model 1	Model 2	Model 3	Model 4
(Intercept)	-0.750 (0.730)	-2.040 (1.377)	-5.138 (1.983)**	-5.595 (2.021)**
RecBel	0.695 (0.756)	0.718 (0.766)	3.254 (1.325)*	4.254 (1.456)**
PrRestore		1.531 (1.366)	4.936 (2.054)*	4.675 (2.080)*
RecBel $\times$ *Reason*			-3.450 (1.414)*	-3.283 (1.453)*
DictBel				2.983 (0.985)**
RecBel > *DictBel*				-1.217 (0.725)$^{\circ}$
Male	-0.289 (0.420)	-0.290 (0.425)	-0.526 (0.440)	-0.521 (0.410)
Worker	-0.019 (0.225)	-0.019 (0.228)	-0.010 (0.231)	-0.064 (0.222)
Italian	-0.865 (0.602)	-0.877 (0.613)	-0.915 (0.615)	-0.750 (0.543)
Experienced	0.248 (0.505)	0.247 (0.512)	0.414 (0.518)	-0.245 (0.516)
AIC	209.519	210.241	205.661	184.078
BIC	231.870	235.784	234.398	219.200
Num. obs.	180	180	180	180
Num. groups: ID	60	60	60	60

Note: Standard errors in parentheses. All estimates are from GLM Logit models including random effects at subject-level. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\circ}p < 0.1$.

The information criteria reported in the lower panel of Table 3 (AIC and BIC) unanimously suggest that Model 4 best exploits the information available and, thus, we focus on this specification when commenting on our results. However, as shown by the comparison of Models 3 and 4, the main results are robust across alternative specifications.[22]

Model 4 shows that a higher probability of restoring the wallet to the entitled recipient induces more returns on the side of dictators, providing support to the hypothesis of justice concern (Hypothesis 1). In other words, when the probability of a misplaced return is lower, dictators are more likely to choose to return.

The positive and statistically significant coefficient of *RecBel* suggests that more optimistic beliefs of the recipient are more likely to trigger a return than less optimistic ones. This provides support to the hypothesis of guilt aversion (Hypothesis 2).

The estimated coefficient of the interaction *RecBel* $\times$ *Reason* is negative and statistically significant. Thus, in contrast to our Hypothesis 3, facing recipients that maintain reasonable beliefs discourages returns. This surprising result seems to be largely due to the frequent return choices observed when restoring probabilities are low, as highlighted

[22]Results are robust to specifications where controls and/or random effects are omitted.

also by Table 2.

Finally, controls on dictator's first-order beliefs suggest that expectations about behavior of others in the same role are positively correlated to dictators' choices (*DictBel*). Furthermore, when the beliefs of the recipient are too optimistic relative to those of the dictator ($RecBel > DictBel$), returns are less likely.

6 Discussion and Conclusions

Results from our experiment demonstrate that justice concerns, originating in the balance between effort and rewards, play an important role in shaping dictators' choices in our modified dictator game. When there is a higher probability that the endowment is returned to the individual who did not work to generate it, fewer dictators decide to return. This result hence adds to previous works highlighting the relevance of the proportionality between inputs and outputs in a production process (e.g. Adams, 1963; Konow, 2000).

Our experiment also shows that dictators react to the expectations of their counterpart, by being more likely to return when facing optimistic recipients. As a result of our chosen experimental design, it is possible to causally interpret the effect that recipients' beliefs have on dictators choices. After ruling out confounds as the false *consensus effect* and verifying the absence of strategic manipulation of expectations, we can safely interpret this finding as an attempt of dictators to avoid the negative emotion of guilt. While evidence on guilt aversion is still mixed, our findings are partially consistent with those experiments providing empirical support to this theory.

Our results clearly show that both justice concerns and counterpart's expectations affect dictators. We further contribute to the literature by showing how these two relevant sources of behavior interact. The intuition driving our research hypothesis is that counterpart's expectations are more likely to affect behavior when they reflect justice considerations. To our surprise, however, the expectations of the entitled recipients are more likely to be fulfilled when the likelihood of a misplaced return is higher. In other words, a lower protection of entitlement rights crowds-in the role of others' expectations. We believe that this result may deserve further attention as it provides interesting insights into the working of institutions, which are proxied by different restoring probabilities in our experiment. Data collected suggest that when institutions are weaker, i.e. they offer less protection to legitimate property rights, individuals may rely more on a subjective measures of justice, as captured by the expectations of the counterpart which trigger a sense of guilt when disappointed.

Finally, we observe that, together with beliefs of the counterpart, the beliefs of the decision-maker also play a crucial role in explaining behavior. We hence attempt to rationalize the evidence from Model 4 in Table 3 in the light of the theoretical framework proposed by Bicchieri (2006), according to which individuals are more inclined to comply to a norm when, (1) they expect others to follow it, and (2) they are expected by others to follow it, conditional upon others' expectations to be legitimate. In line with the first part of Bicchieri's argument, we find a strong positive correlation between dictators' beliefs about a generic dictator choosing to return the wallet and their own decision to return. At the same time, we see that dictators facing over-optimistic beliefs (with respect to their own) tend to react in the opposite way, suggesting that for others' expectations to be effective these must appear appropriate.

References

Abeler, J., Falk, A., Goette, L., and Huffman, D. (2011). Reference points and effort provision. *The American Economic Review*, 101(2):470–492.

Adams, S. J. (1963). Toward an Understanding of Inequity. *Journal of Abnormal and Social Psychology*, 67(5):422–436.

Bacharach, M., Guerra, G., and Zizzo, D. (2007). The self-fulfilling property of trust: An experimental study. *Theory and Decision*, 63(4):349–388.

Battigalli, P., Charness, G., and Dufwenberg, M. (2013). Deception: The role of guilt. *Journal of Economic Behavior & Organization*, 93:227–232.

Battigalli, P. and Dufwenberg, M. (2007). Guilt in games. *The American Economic Review: Papers and Proceedings*, 97(2):170–176.

Babcock, P., Bedard, K., Charness, G., Hartman, J., and Royer, H. (2014). Letting down the team? Social effects of team incentives. *Journal of the European Economic Association*, *forthcoming*.

Battigalli, P. and Dufwenberg, M. (2009). Dynamic psychological games. *Journal of Economic Theory*, 144(1):1–35.

Becker, A. (2013). Accountability and the fairness bias: The effects of effort vs. luck. *Social Choice and Welfare*, 41(3):685–699.

Bellemare, C., Sebald, A., and Strobel, M. (2011). Measuring the willingness to pay to avoid guilt: Estimation using equilibrium and stated belief models. *Journal of Applied Econometrics*, 26(3):437–453.

Bicchieri, C. (2006). *The grammar of society: the nature and dynamics of social norms.* Cambridge University Press, New York.

Bigoni, M. and Dragone, D. (2012). Effective and efficient experimental instructions. *Economics Letters*, 117(2):460–463.

Bolton, G. E. and Ockenfels, A. (2000). Erc: A theory of equity, reciprocity and competition. *The American Economic Review*, 90(1):166–193.

Cappelen, A. W., Hole, A. D., Sorensen, E. O., and Tungoddedn, B. (2007). The pluralism of fairness ideals: An experimental approach. *The American Economic Review*, 97(3):818–827.

Cervellati, M., Esteban, J., and Kranich, L. (2000). Work values, endogenous sentiments redistribution. *Journal of Public Economics*, 94(9):612–627.

Charness, G. and Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, 74(6):1579–1601.

Charness, G. and Dufwenberg, M. (2010). Bare promises: An experiment. *Economics Letters*, 107(2):281–283.

Charness, G. and Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, 117(3):817–869.

Cherry, T. L., Frykblom, P., and Shogren, J. F. (2002). Hardnose the dictator. *The American Economic Review*, 92(4):1218–1221.

Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58(6):1015–1026.

Cox, J. C., Friedman, D., and Sadiraj, V. (2008). Revealed altruism. *Econometrica*, 76(1):31–69.

Dufwenberg, M., Gächter, S., and Hennig-Schmidt, H. (2011). The framing of games and the psychology of play. *Games and Economic Behavior*, 73(2):459–478.

Dufwenberg, M. and Gneezy, U. (2000). Measuring beliefs in an experimental lost wallet game. *Games and Economic Behavior*, 30(2):163–182.

Ellingsen, T. and Johannesson, M. (2004). Promises, threats and fairness. *The Economic Journal*, 114(495):397–420.

Ellingsen, T., Johannesson, M., Tjøtta, S., and Torsvik, G. (2010). Testing guilt aversion. *Games and Economic Behavior*, 68(1):95–107.

Elster, J. (1998). Emotions and economic theory. *Journal of Economic Literature*, 36(1):47–74.

Engelmann, D. and Strobel, M. (2000). The false consensus effect disappears if representative information and monetary incentives are given. *Experimental Economics*, 3(3):241–260.

Engelmann, D. and Strobel, M. (2011). Deconstruction and reconstruction of an anomaly. *Games and Economic Behavior*, 76(2):678–689.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

Geanakoplos, J., Pearce, D., and Stacchetti, E. (1989). Psychological games and sequential rationality. *Games and Economic Behavior*, 1(1):60–79.

Gill, D. and Stone, R. (2010). Fairness and desert in tournaments. *Games and Economic Behavior*, 69(2):346–364.

Gneezy, U. (2005). Deception: The role of consequences. *The American Economic Review*, 95(1):384–394.

Hoffman, E., McCabe, K., Shachat, K., & Smith, V. (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior*, 7(3): 346–380.

Hoffman, E. and Spitzer, M. L. (1985). Entitlements, rights, and fairness: An experimental examination of subjects' concepts of distributive justice. *The Journal of Legal Studies*, pages 259–297.

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–292.

Kartik, N. (2009). Strategic communication with lying costs. *Review of Economic Studies*, 76(4):1359–1395.

Konow, J. (1996). A positive theory of economic fairness. *Journal of Economic Behavior & Organization*, 31(1):13–35.

Konow, J. (2000). Fair shares: Accountability and cognitive dissonance in allocation decisions. *The American Economic Review*, 90(4):1072–1091.

Konow, J. (2003). Which is the fairest one of all? A positive analysis of justice theories. *Journal of Economic Literature*, XLI:1188–1239.

Leventhal, G. S. and Michaels, J. W. (1969). Extending the equity model: Perceptions of inputs and allocation of reward as a function of duration and quantity of performance. *Journal of Personality and Social Psychology*, 12(4):303–309.

Lindbeck, A. and Nyberg, S. (2000). Raising children to work hard: Altruism, work norms, and social insurance. *The Quarterly Journal of Economics*, 121(4):1473–1503.

Locke, J. (2011). *Second Treatise of Government*. Simon & Brown.

Mikula, G. (1974). Nationality, performance, and sex as determinants of reward allocation. *Journal of Personality and Social Psychology*, (4):425–440.

Ockenfels, A. and Werner, P. (2014). Scale manipulation in dictator games. *Journal of Economic Behavior & Organization*, 97:138–142.

Reuben, E., Sapienza, P., and Zingales, L. (2009). Is mistrust self-fulfilling? *Economics Letters*, 104(2):89–91.

Tangney, J. (1995). Recent Advances in the Empirical-Study of Shame and Guilt. *American Behavioral Science*, 38(8):1132–1145.

Vanberg, C. (2008). Why do people keep their promises? an experimental test of two explanations. *Econometrica*, 76(6):1467–1480.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, 92(4):548.

Appendices

A Instructions

[ONLY THE FIRST 10 SUBJECTS ARE IN THE LAB]

Welcome!

For showing up you have earned €3. During the session you can earn more money. The entire amount will be paid to you in private at the end of this session. Please, follow the instructions carefully and do not speak to the other participants. If you have questions, raise your hand and one of the experimenters will answer to you in private.

The session consists of 4 stages. You are identified as the *Green Group* and you are going to complete the first stage. After the first stage, another group of participants will enter this room,and will be identified as the *Red Group*.

Stage 1 – Work. In this stage you have to count the number of zeros within seven tables that will consecutively appear on your screen. For each table you will earn 1 token once you have inputed the correct number of zeros.

In a following stage, something unexpected could occur which implies the loss of the tokens you have earned. At the end of the session, each token you own will be converted into €1.

The Red Group can now come in.

[RED GROUP ENTERS THE LAB]

Welcome!

For showing up you have earned €3. During the session you can earn more money. Please, follow the instructions carefully and do not speak to the other participants.

The 10 participants who were already in this room are identified as the *Green Group*. The 5 participants that have just entered are instead identified as the *Red Group*.

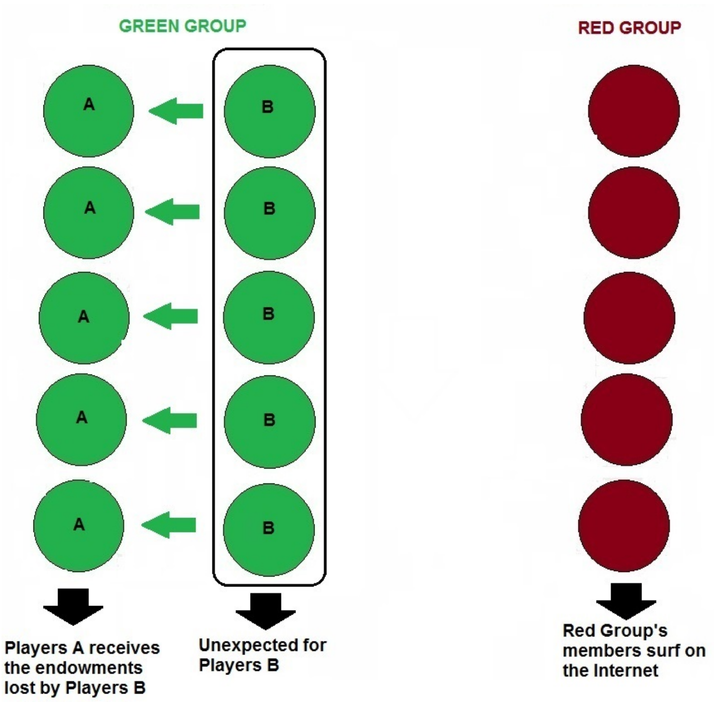
The members of the Green Group have already been here for about 15 minutes, and have just completed Stage 1 of the session, which consisted of counting the number of zeros in a sequence of tables. For doing the work, every member of the Green Group has earned 7 tokens. At the end of the session, each token you own will be converted into €1.

The members of the Red Group will participate only in the next stage of the session, and then they will be allowed to freely surf on the Internet, but the members of the Green Group will have to participate actively in the whole session. Before being allowed to leave, they will have to wait until all the members of the Red Group are paid.

To understand the Stage 2 (*estimation*) you need to know about Stage 3 (*unexpected*) and Stage 4 (*decision*). Before proceeding to Stage 2, we will give you the instructions for stage 3 and 4. *Everyone,* please carefully follow the instructions.

Stage 3 – Unexpected. In this stage none of the participants are asked to take any decisions. The computer will execute all the procedures automatically.

If you are a member of the Green Group, you could suffer an *unexpected*: one person every two, randomly selected by the computer, will lose the 7 tokens earned from the work in Stage 1 and will become *Participant B* (see the picture). The lost earnings of Participant B will be given to another person, *Participant A*, randomly selected from among the 5 members of the Green Group who did not suffer the unexpected (see the picture).



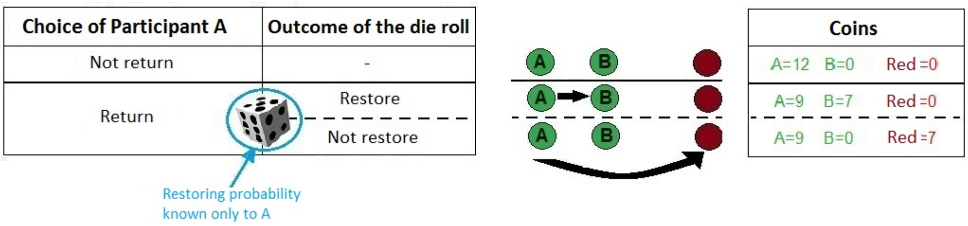
If you are a Green Group member, on your screen you will see the role that was assigned to you. If you are a Red Group member, you will be allowed to surf the Internet.

Stage 4 – Decision. Every Participant A is paired with the Participant B from whom he/she has received the gain. Each Participant A will have to decide whether to return the gain to Participant B:

- If he/she decides to *not return*, A earns a total of 12 tokens and B remains with 0 tokens.

- If he/she decides to *return*, A earns a total of 9 tokens and B restores his/her initial earning of 7 coins with a probability between 67% and 100%. In the cases where B does not restore his/her earnings, the 7 tokens are transferred to a member of the Red Group, selected at random. The outcome of the restitution is determined with a die roll (performed by the computer).

Participants' total earnings are summarized in the following diagram:



The situation just described will be repeated three times. In each repetition, each Participant B will be paired again with a new Participant A, different from the one previously met (i.e. you will never be paired with the same person for more than one repetition). Are there any questions about this?

There are still three important things to say about Stage 4:

1. In each of the three repetitions, the restoring probability is different and randomly changes between 67% and 100% in the following ways:

Important: Only Participant A knows the exact probability with which the 7 tokens will be restored by B. Participant B will *never* know, because the order with which the different restoring probabilities appear is random and remains unknown to B.

2. All the decisions will be taken sequentially, without receiving feedback on their outcome. At the end of Stage 4, every participants will see the number of tokens earned. Only Participant A will be informed about the outcome of his/her choice in the repetition randomly selected for the payment (see below).

3. One of the three repetitions, selected at random by the computer, will be paid at the end of the session.

<u>Before proceeding, we will ask you to answer some control questions about the instructions. We can now return to Stage 2.</u>

Stage 2 – Estimation. You are now asked to report what decisions you expect from Participants A in Stage 4 (decision). In particular, you have to estimate how many times you expect a generic Participant A will choose to return the gain to the Participant B across the three repetitions (from 0 out of 3, to 3 out of 3). Is that clear?

The more your estimate is accurate, the more money you can earn. To evalauate the accuracy of your estimate, the computer will match you to a Participant A selected at random. The Participant A randomly selected to determine your earning for the Stage 2 (estimate) will be different from the Participant A who will be selected at random to determine your earning in the stage 4 (decision). At the end of the session, the computer will draw at random one of the three decisions made by Participant A you are matched to. The computer will compare your estimate with the choice of A in the drawn decision (the choice could be either “Return” or “Not return”).

In the table you can see how much you will earn given your estimate and the choice of Participant A.

	Dictator will choose *Return*...			
	0 out of 3	1 out of 3	2 out of 3	3 out of 3
Your guess...	☐	☐	☐	☐
Your earnings if in the drawn choice...				
Dictator chose *Return*	0 €	2.80 €	4.40 €	5 €
Dictator chose *Keep*	5 €	4.40 €	2.80 €	0 €

Note: During Stage 4, before each decision, Partecipants A will be also informed about the estimate made in Stage 2 by Participant B with whom they are paired.

Payment: The total amout paid at the end of the session (in addition to the €3 for showing up) is the sum of the gain for the accuracy of your estimate in Stage 2, and your earnings in the repetition randomly selected from Stage 4.

From here on, Red Group members can surf the Internet.

B Slides and Quiz

All the slides used in the experiment are based on picture depicted in the experimental instructions. Subjects observed all the possible outcomes of the game.

Translation of the quiz:

1. During Stage 3, how are roles assigned to Green Group members?
   - At random
   - Those who have not worked well suffer the unexpected and become Participants B
   - Those who have not worked well suffer the unexpected and become Participants A
2. During Stage 4, every Red Group member...
   - chooses whether or not to return the gain lost by Participant A
   - leaves the room
   - waits for all Participants A to complete their choices, surfing the Internet
3. During Stage 4, every Participant B...
   - chooses whether or not to return the gain lost by Participant A
   - waits for all Participants A to complete their choices
   - waits for all Participants A to complete their choices, surfing the Internet
4. During Stage 4, every Participant A...
   - chooses whether or not to return the gain lost by Participant B
   - waits for all Participants A to complete their choices
   - waits for all Participants A to complete their choices, surfing the Internet

5. During Stage 4, what happens when a Participant A chooses to return the gain to Participant B?
    - The gain is always restored by Participant B (the original owner)
    - A die is rolled to establish whether the gain actually is restored to B. The restoring probability is always different
    - A die is rolled to establish whether the gain actually is restored to B. The restoring probability is always the same
6. During Stage 4, what happens when a Participant A chooses to return the gain to Participant B but the roll of the die determines that the gain is not restored?
    - The gain is not returned to anybody and thus it is lost
    - The gain remains to Participant A
    - The gain is transfered to a member of the Red Group chosen at random
7. During Stage 4, what happens when Participant A chooses to return the earnings to Participant B, but the roll of the die determines that the earning is not restored?
    - The earning is not returned to anybody and thus it is lost
    - The earning remains to Participant A
    - The earning is transfered to a member of the Red Group chosen at random
8. In each repetition, the restoring probability...
    - is known to all participants
    - is known only to Participant B
    - is known only to Participant A
9. When is the die rolled?
    - Whenever Participant A chooses to not return
    - Whenever Participant A chooses to return
    - Always, irrespective to the choice of A
10. Participant A...

- cannot change her decision (“Return” / “Not Return”) across the 3 repetitions
- can change her decision (“Return” / “Not Return”) across the 3 repetitions
- cannot change her decision (“Return” / “Not Return”) in only one repetition

Contested Public Projects: the Role of Information and Transfers *

Riccardo Ghidoni†

Abstract

A key driver of the opposition to public projects which can be perceived as noxious (e.g. landfills, power plants, or incinerators) is the mistrust of local communities towards the proposer of the project. In a novel laboratory experiment, I study how to promote a fruitful participation of communities in the realization of public projects, by testing the effectiveness of two tools available to policy–makers: *information disclosure* about the project, and *compensatory side-transfers*. Both tools display critical aspects. Endogenous information disclosure is threatened by the existence of an equilibrium in which actors do not exert the cooperative effort necessary to make information available. Monetary side-transfers might crowd-out the intrinsic motivation of the community members, which would lead them to host a socially desirable project. I find that both tools significantly increase the number carried out projects, creating positive social surplus. However, the tools have different impacts on the types of projects which are realized.

Keywords: Experiment; Public projects; Trust; NIMBY; Information disclosure; Compensatory transfers; Side-payments

*I am grateful to Marco Casari, Maria Bigoni, and Stefania Bortolotti for their constant support and their advices throughout this research. This paper also benefited from discussions with Peter Martinsson, Giuseppe Pignataro, Jan Potters, Florian Schuett, Sigrid Suetens, Matthias Sutter, Paolo Vanin, and seminars participants at the University of Bologna and Tilburg University.

†Department of Economics, University of Bologna. Email: `riccardo.ghidoni2@unibo.it`. Web: `https://sites.google.com/site/riccardoghidoni/`

1 Introduction

The opposition to public projects that might be socially desirable is widespread problem in the field. Typical examples of contested projects are power plants, landfills, airports, incinerators, and clinics for mental health problems. Previous economic studies have focused on the specific case in which the contested project *always* imposes a cost on the community living in a perspective host site, while it is *always* beneficial to the other actors, i.e. the non-host communities and the proposer of the project (e.g. Frey and Oberholzer-Gee, 1997; Perez-Castrillo and Wettstein, 2002; Feinerman et al., 2004; Sakai, 2012). This type of projects is indeed defined as "Not In My Backyard" or NIMBY, because everyone would want the project to be realized in a community other than his or her own.[1]

While the NIMBY problem is a relevant source of opposition to public projects, several case studies suggest that the opposition is frequently triggered by a lack of trust of local communities in the proposer of the project (e.g. Petts, 1995; Baxter et al., 1999). Often, proposers argue that their project is safe and that costs and benefits are balanced among the actors. Instead, host communities tend to be skeptical about it. In some cases, communities question the social desirability of the project, in other cases they oppose to it because they believe they will bare its entire cost, just as in the NIMBY case. When the host community members are uncertain about the *type* of the project, they may even oppose to projects from which they would benefit, for instance because a careful redistribution of the surplus generated by the project was designed, or because the project would have a positive return in terms of working places.

In an environment characterized by asymmetric information about the project's type, there are at least five major issues: (1) the proposer's choice of the project's type; (2) the selection of the host site; (3) the social surplus generated by the project; (4) information disclosure about the project's type; (5) compensatory side-transfers to the individuals living in the host site. Since the purpose of this study is to provide insights on which tools could mitigate oppositions, I will focus on issues (4) and (5), leaving the remaining for future research. In this regard, I developed a novel experimental game that captures the problem of mistrust between the community and the proposer of the project. In the game, a proposer endowed with different types of project can either try to carry it out, or

[1]Frey et al. (1996) define NIMBY projects as "all undertakings that increase overall welfare but impose net costs on the individuals living in the host community".

renounce to it. All types of project generate the same positive social surplus, but differ in how costs and benefits are allocated among the actors. The type of the project is private information of the proposer. In the baseline, if a proposer chooses to carry out the project, a randomly selected host site will decide whether or not to accept to realize the unknown project.

When the project's type is uncertain, the process of *information disclosure and assimilation* is crucial to its realization. I investigate this aspect in a first treatment, by allowing actors to disclose information about the project's type through a costly joint effort. This treatment relies on the model of communication with moral hazard in teams of Dewatripont and Tirole (2005). Since the game has two equilibria, one in which no effort is exerted and another in which all actors exert their effort, the experiment answers to the empirical question of which equilibrium will be actually played. I find that subjects generally coordinate on the Pareto superior equilibrium, where the information disclosure tool is exploited. Thus, the tool is effective in increasing the number of projects realized. In a second treatment, I test another tool available for policy-makers, namely *monetary side-transfers* to compensate the host. Standard economic theory predicts that these side-transfers should mitigate the opposition of the host community, as long as they are fully compensatory. However, the evidence from non-incentivized surveys suggests that side-transfers can increase the opposition (e.g. Kunreuther and Easterling, 1990; Frey et al., 1996). In particular, Frey and Oberholzer-Gee (1997) show that in perspective host communities where the civic sense prevails, monetary side-transfers crowd out the intrinsic motivation of the inhabitants, which would lead them to bear a cost for the common good. In the experiment (which shares similar aspects with the mentioned surveys), I find that side-transfers have a positive effect on the likelihood of realizing a project. The data are consistent with the idea that when the civic sense is low and the intrinsic motivation is weak, side-transfers can still be effective (Frey and Oberholzer-Gee, 1997). Indeed, the neutral framing used in the experiment may induce subjects to reason in terms of payoffs rather than in terms of behaving in a civic way. In addition, the experiment reveals that subjects tend to endorse the realization of noxious projects even when they are only partially compensated for hosing it. I show that this result can be rationalized with preferences for social-welfare à la Charness and Rabin (2002). This behavioral trait has been previously overlooked by the literature on public projects and side-transfers.

While both the tools studied in this paper significantly enhance the projects' realization, side-transfers and information disclosure have different implications for the type of projects that are implemented. In the treatment with endogenous information disclosure, almost only balanced projects go through. Instead, in the treatment with side-transfers, the number of noxious projects that goes through is much higher. In the latter treatment, if sufficiently compensated, hosts agree to implement projects knowing that these will damage the non-host.

This experiment aims at providing complementary evidence to the few empirical studies existing in the literature on contested public projects. The experimental technique allows to overcome some identification challenges that characterize empirical studies. For instance, it may be difficult to precisely assess the payoffs of the actors involved in a field interaction. Another challenge is the endogeneity of the choice of the host sites. Moreover, most of the empirical studies rely on surveys carried out in perspective host-sites, and respondents might have an incentive to answer strategically (Frey and Oberholzer-Gee, 1997). Finally, with a focus on side-transfers, experiments can offer a valuable tool to detect any hypothetical bias in the self-reported answers (Harrison, 2006).

The paper proceeds as follows: Section 2 presents a short literature review; Section 3 describes the experimental games and the experimental procedures; Sections 4 provides the theoretical analysis; Section 5 presents the results; and Section 6 concludes.

2 Literature Review

Different from this paper, previous studies have investigated the issue of the opposition to public projects by assuming NIMBY projects.[2] Therefore, in this section, I review the literature on the NIMBY phenomenon in economics.

A long standing body of literature takes a normative stance, proposing mechanisms which allow to identify an optimal host site (i.e. a site characterized by the lowest cost to become host of the project) and induce it to accept the project through side-transfers. A seminal contribution in this perspective is the one of Kunreuther and Kleindorfer (1986). Other important theoretical contributions to this literature are the mechanisms of Perez-Castrillo and Wettstein (2002), Minehart and Neeman (2002), and Laurent-Lucchetti and Leroux (2011). Recently, Sakai (2012) provided an axiomatic analysis of the NIMBY problem, characterizing a *fair pricing rule*. Some of these mechanisms have also been tested with lab experiments (Kunreuther et al., 1987; Pérez-Castrillo and Veszteg, 2007). What is common to these mechanisms is the assumption that the evaluation of the project is private information of each community. My experimental setup starts from a different

[2]Some of the reported studies extend their analysis also to the opposite case, in which the project is only beneficial to the host.

perspective: the proposer of the project knows the type of the project (and the payoffs generated). Communities, instead, know the different project's types, but they are uncertain about which type of project is proposed. My setup is motivated by the case studies of Petts (1995) and Baxter et al. (1999), in which the community members fear that the project will be noxious to them, while proposers try to argue that this will not be the case.

Another stream of the theoretical literature studies the NIMBY problem in a framework of political economy, and focuses on issues related to lobbying behavior by communities involved in the process of siting of the public project. Feinerman et al. (2004) study a situation in which landowners of the possible host sites exert pressures on politicians. The authors find that when the distribution of landownership is perfectly equitable and all communities participate in the political arena, the government locates the facility at the socially optimal site. Bellettini and Kempf (2013) include the size of the project in their analysis, and find that the uniqueness of optimal site is not guaranteed. Moreover, under lobbying behavior of households they find that inefficient provision of the public good is possible. An important aspect that sets my experiment apart previous studies, is the uniqueness of the host site. While the multiple possible host sites are possible in many cases, there are also situations in which geographical constraints limit the possible candidates to a single host site (e.g. Petts, 1995).

Besides the above mentioned experiments and a large number of case studies, the empirical evidence on the opposition to public projects is rather small. To some extent, my experiment shares some similarities with the questions of the surveys conducted by Kunreuther and Easterling (1990) and Frey et al. (1996). In these face-to-face surveys, individuals living in a perspective host community were offered with different hypothetical monetary transfers, aimed at compensating them for hosting a project. These individuals were hence asked whether they were willing or not to accept both the project and the transfer. Both the studies find a negative correlation between the size of side-transfers offered and the propensity to host the public project. Frey and Oberholzer-Gee (1997) show that this evidence is consistent with the hypothesis that monetary side-transfers crowd out the intrinsic motivation of the community members.

3 Experimental Design

In this experiment there are three treatments, that differ in the possibility to endogenously disclose information about the project, or to offer compensatory side-transfers to the host. All the other features of the design are identical across the treatments (Table 1).

Table 1: Experimental Treatments.

	Baseline	Information	Side-transfer
Information disclosure about the project's type	None	Endogenous	None
Side-transfer to the host	None	None	Possible
Matching protocol within a cluster	Random	Random	Random
Anonymity	No subject IDs	No subject IDs	No subject IDs
Number of sessions	4	4	4
Number of matching clusters	8	8	8
Show-up fee	2€	2€	2€
Number of periods	20	20	20

The Baseline Game. In the baseline treatment, I employ the game depicted in Figure 1. The game includes the Nature and three other actors: a *Proposer* (P), a *Host site* (H), and a *Non-Host site* (NH).[3]

The Nature endows P with one of three equally likely types of project: *extracting* (e), *NIMBY* (n), or *balanced* (b) project. All projects can generate the same positive social surplus, but they differ in how costs and benefits are shared among the actors. The project's type is private information of P, who must choose whether to *renounce* (Out) or to *offer* it to H (In). If P chooses Out, the game ends in a *status quo*, where every actor gets π_0. If P chooses In, H must decide whether to *accept* (a) or *reject* (r) to host the project. If H chooses r, P suffers an opportunity cost obtaining $\underline{\pi} < \pi_0$, while H and NH get π_0. If H chooses a, the actors' payoffs depend on the type of the project realized.[4] The payoff from project $k \in \{e, n, b\}$ to actor $i \in \{P, H, NH\}$ is denoted by π_i^k.

- Project e is beneficial to P, and is harmful for H and NH ($\pi_P^e > \pi_0 > \pi_{NH}^e = \pi_H^e$);

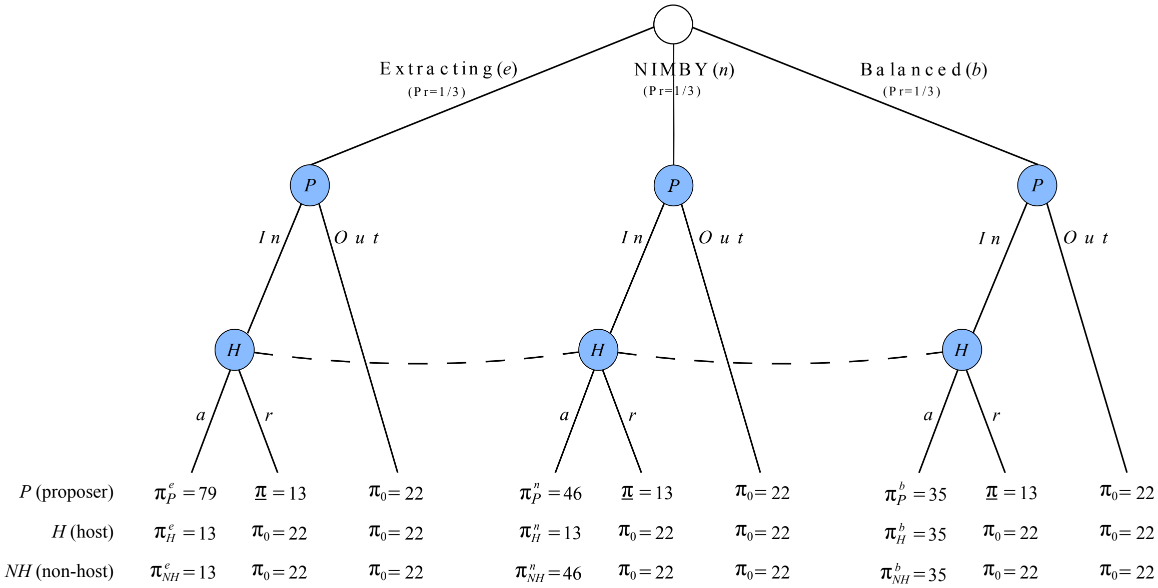
[3]In the field, several public and private agents concur to the siting a public project (e.g. central government, government siting agencies, and private construction firms). These agents, who have the common goal to successfully realize the project, are here grouped into a unique agent, i.e. the "proposer". The "host site", instead, can be interpreted as single member of a community who votes for or against hosting a public project in a referendum.

[4]In the field, there are cases in which the project is realized despite the opposition. This game best resembles situations in which the endorsement of the host community (and of the local authorities) is binding to realize the project (e.g. Petts, 1995).

- Project n is beneficial to P and NH, and is harmful for H ($\pi_P^n = \pi_{NH}^n > \pi_0 > \pi_H^n$);
- Project b is beneficial to all three actors ($\pi_P^b = \pi_{NH}^b = \pi_H^b > \pi_0$).

For simplicity, I assume $\pi_H^n = \pi_H^e = \pi_{NH}^e = \underline{\pi}$. In the baseline treatment, NH takes no action.

Figure 1: Baseline Game.



Treatment with Information Disclosure. This treatment relies on the model of collaborative communication of Dewatripont and Tirole (2005), "in which the sender's and receiver's motivations endogenously determine the transfer of knowledge". The model captures two important obstacles to information disclosure: (1) the possibility of imperfect alignment in actors' interests in the project, and (2) moral hazard in teams.[5]

All actors can help H to learn the type of the project offered by P through a joint effort. In particular, actors' efforts determine the probability d, with which H learns the project's type (successful disclosure). With probability $(1 - d)$, the disclosure fails, and H does not learn the project's type. Information disclosure is described by Equation 1:

$$d \equiv e_P(e_H + e_{NH}) \tag{1}$$

[5]Communication is often characterized by a private costs both for the sender, who must formulate a clear message, and for the recipient, who must stay focused to absorb the content of the message. As a consequence, when information disclosure fails, the sender can complain that the receiver did not listen carefully, and the receiver can argue that the message was not clear.

where e_P, e_H, and e_{NH} are the efforts of P, H, and NH, respectively. Each actor can either exert no effort or positive effort which costs c. Two further features are borrowed from the model of Dewatripont and Tirole (2005). First, the successful disclosure is never certain: if all actors exert their effort, H learns the project's type with probability 0.90; if only P and H (or NH) exert their effort, H learns the project's type with probability 0.45; in all other cases, H does not learn the project's type.[6] Second, effort choices are private information.

Equation 1 implies that P's effort is strategic complement to the efforts of H and NH. It is reasonable to assume that the more "cooperative" is P, the more profitable is for community members to listen, and the other way around. The tool also implies strategic substitutability between H's and NH's efforts. Again seems plausible that the more communities' members participate to public meetings, the more likely is knowledge to be transmitted. However, this can induce inhabitants to free-ride on others' effort.

This game does not directly allow P to lie about his project's type. However, P's choice to propose a harmful project without exerting the effort can be seen as an attempt to lie to H: since efforts are private information and there is always a positive probability with which the disclosure fails, P can try to "hide" his responsibility for an unsuccessful disclosure behind the chance.

Treatment with Side-transfers. If P chooses to invest, he can offer or not a compensatory transfer t to H facing a cost $c(t)$.[7] The transfer is only implemented if H accepts the project. P can choose between two levels of transfers:

- Low transfer ($\underline{t}$):partly compensates H for hosting a harmful project ($\underline{\pi} + t_l < \pi_0$);
- High transfer ($\bar{t}$): fully compensates H for hosting a harmful project ($\underline{\pi} + t_l \geq \pi_0$).

P endowed with project e or n can sustain the cost $c(t_l)$ without incurring in a loss ($\overline{\pi} - c(t_l) > \pi_0$). Only P with project e can sustain the cost $c(t_h)$ without incurring in a loss ($\pi_P^e - c(t_h) > \pi_0$). If P has project b, he always incurs in a loss if offers a positive transfer ($\pi_P^b - c(t_l) < \pi_0$). I also impose the deadway loss due to $c(t)$ to be small enough to make the projects' realization less efficient than the *status quo* ($\sum_i \pi_i^k - c(t) + t > 3\pi_0$).

[6] The positive effort is equal to 0.63. The effort cost function postulated by Dewatripont and Tirole (2005) is such that the cost of reaching $d = 1$ is equal to $+\infty$.

[7] In the experiment, P cannot finance the side-transfer by "taxing" NH as it can happen in the field. Since the interest is here mainly on the reaction of H to different levels of side-transfer, this simplification does not undermine the relevance of the treatment.

The Experimental Session. 18 subjects participated in each session. At the beginning of the session, roles were randomly assigned: 6 subjects acted as P for the entire session; the remaining 12 subjects played 10 round as H, and 10 round as NH in unpredictable order.[8] Subjects were randomly divided into two independent clusters. Within each cluster, *stranger matching protocol* was used to match subject in groups of three across the 20 periods of play. Each group included one P, one H, and one NH. At the end of every period, P and H received feedback about their earning in that period, while the NH was only informed about earnings in each period at the end of the session. At the end of the session, one period was randomly drawn, and subjects were paid according to their payoff in the drawn period, with a conversion rate of €0.30 for each token.

Procedures. I implemented four sessions for each treatment. I recruited 216 subjects (72 for each treatment) from the database of the Bologna Laboratory for Social Sciences (BLESS), who were invited via ORSEE (Greiner, 2004). The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007).

4 Theoretical Predictions

Since H cannot observe the project of P, H's beliefs about the project offered by P are key to determine the equilibria of the games. The solution concept is perfect Bayesian equilibrium (PBE): all actors' strategies must be optimal given their beliefs, and beliefs are derived from equilibrium strategies through Bayes' rule whenever possible. I assume that actors share a common prior about the project's type equal to 1/3, and that they are risk neutral. The derivations of the equilibria are in Appendix A.

4.1 Baseline Game

Proposition 1. *Even though H finds it profitable to host the balanced project, in equilibrium no project is ever realized.*

In the baseline, P is unable to credibly convince H that the offered project is balanced. Indeed, if H would accept an unknown project, P would invest even when endowed with harmful projects. The unique BNE of the baseline game is indeed that P never invests, because he correctly anticipate that H would always reject.

[8] The order was the same across sessions. During a session, P, H, and NH were labeled A, $B1$ and $B2$, respectively.

The BNE is a PBE, if there exist a range of off-equilibrium belief which supports it. Let q^b be the off-equilibrium belief with which H expects the offered project to be balanced, i.e. $q^b = \Pr\{b|In\}$. Then, H chooses to accept the project if and only if the expected payoff from accepting is greater than the payoff from choosing to reject:

$$q^b \pi_H^b + (1 - q^b)\underline{\pi} > \pi_0$$

$$q^b > q^* \equiv \frac{\pi_0 - \underline{\pi}}{\pi_H^b - \underline{\pi}} \tag{2}$$

So, for $q^b < q^*$, the baseline game has a unique PBE, in which P never invests because he correctly anticipates that H would always reject. For $q^b > q^*$, no PBE exists.

4.2 Game with Information Disclosure

Proposition 2. *Even if there is the possibility to disclose information about the project's type, no project should go through.*

Conditional upon a joint effort by the actors, the information disclosure tool (Equation 1) enables P to credibly inform H about the type of his project. If the disclosure succeeds, it is strictly dominant for H to accept only project b. By backward induction, it is strictly dominant for P to renounce to harmful projects. It is also weakly dominant for H to exert her effort only if she rejects whenever information disclosure fails.[9] So, if P chooses to invest, actors' expected utilities are

$$U_P = d\left(\pi_P^b\right) + (1 - d)(\underline{\pi}) - C(e)$$

$$U_H = d\left(\pi_H^b\right) + (1 - d)(\pi_0) - C(e)$$

$$U_{NH} = d\left(\pi_{NH}^b\right) + (1 - d)(\pi_0) - C(e)$$

The condition under which exerting the effort generates a higher payoff than the *status quo* is $d > \frac{C(e)}{\pi_P^b - \underline{\pi}}$ for P, and $d > \frac{C(e)}{\pi_{H(NH)}^b - \pi_0}$ for H and NH. Given the parameters of this experiment, both conditions can be satisfied if P and H or NH exert their effort. However, since P is supposed to renounce to harmful projects, H finds it even more profitable to always accept. This generates an incentive for P to propose also harmful projects. Anticipating this, H prefers to always reject. As consequence, P never invests.

[9]This strategy is weakly dominant because when P chooses *Out*, H is indifferent between her off equilibrium path strategies. However, its elimination does not affect the equilibrium outcome.

The unique equilibrium of this game remains that P renounce to all types of project, because he correctly anticipate that H would reject without exerting the effort.

4.3 Game with Side-transfer

Different assumptions on H's preferences lead to different predictions. I focus on three behavioral types. *Self-regarding* H is the standard benchmark. *Civic* H obtains an intrinsic benefit for behaving in a civic way (Frey and Oberholzer-Gee, 1997). *Social-welfare concerned* H has a taste for maximizing social surplus accounting for its distribution (Charness and Rabin, 2002). Predictions are summarized in Proposition 3.

Proposition 3. .

- *Self-regarding H: Only extracting projects are realized because only P endowed with the extracting project can offer a fully compensatory transfer.*
- *Civic H: If the off-equilibrium belief about the offered project being balanced or NIMBY are high enough, all types of project can go through and no transfer is ever implemented, because transfers would crowd out H's intrinsic motivation.*
- *Social-welfare concerned H: NIMBY and extracting projects go through, because in both P can offer a partially compensatory transfer, which is enough to convince H to host the project.*

Self-regarding Host. As said, P can sustain the cost of positive transfers, $\underline{t}$ or $\bar{t}$, only if endowed with the project e or n. So, $\underline{t}$ and $\bar{t}$ signal to H that the project is harmful. H never accepts if offered with $\underline{t}$, because she infers that her final payoff would be of $(\underline{\pi} + t_l)$ which is lower than π_0 by definition. Instead, H always accepts the $\bar{t}$ (which is affordable only by P with project e), since $(\underline{\pi} + t_h)$ is higher than π_0 by definition. Although H is aware that P with project b cannot afford a positive transfer without incurring in a loss, H will reject if no transfer is offered: H anticipates that if she accepts a null transfer, P endowed with a harmful project will mimic P endowed with project b.

The game has two BNE. In the first BNE, P never invests, because H always rejects. This BNE is not a PBE, because for any off-equilibrium belief, H would accept if offered with $\bar{t}$. In the second BNE, P endowed with the extracting project invests and offers the fully compensatory transfer; if endowed with the NIMBY or the Balanced project, P renounces to it. This BNE is a PBE as long as H's off-equilibrium belief is lower than q^* (Condition 2).

Civic Host. Even if this experiment may be hostile to intrinsic motivation, it is interesting to establish on a theoretical basis what is the effect of transfers when H is intrinsically motivated. I adapt to my setup the model of Frey and Oberholzer-Gee (1997), who define intrinsic motivations as related to "activities one simply undertakes because one likes to do them or because the individual derives some satisfaction from doing his or her duty". H's expected utility from accepting a project is given by the sum of her expected monetary return and the expected intrinsic benefit for hosting the project:

$$\begin{aligned} U_H(A|t) = &\Pr(b|t)(\pi_H^b + t) + (1 - \Pr(b|t))(\underline{\pi} + t) \\ &+ \Pr(e|t)i(e,t) + \Pr(n|t)i(n,t) + \Pr(b|t)i(b,t) \end{aligned} \tag{3}$$

where D measures the intrinsic benefit from choosing to host the project, with D decreasing in t ($D_t < 0$). Consistent with the definition of intrinsic motivation of Frey and Oberholzer-Gee, I assume H only obtains an intrinsic benefit if she either hosts projects n or b; if she hosts project e, the intrinsic benefit is always null. For simplicity, I also assume that, for a given transfer, hosting project b or n generates the same intrinsic benefit ($i(n|t) = i(b|t) > 0$).[10]

Considering that only P with harmful projects can offer a positive transfer, it can be shown that, if D does not decreases too slowly in t, H's expected utility from accepting an unknown project without a positive transfer is larger than the utility from accepting the project and a partially compensating side-transfer (Appendix A.3.2). This result is consistent with the crowding-out theory, and it implies that H should be more likely to accept the project when t is equal to 0 rather then when t is equal to 6. By backward induction, if H's intrinsic motivation is strong enough, i.e. $i(t = 0) \geq 2.55$, a self-interested P always invests (irrespective to his project's type), without offering any side-transfer. If H's intrinsic motivation is weak, i.e. $i(t = 0) < 2.55$, P endowed with the extracting project always invests and transfers 12 tokens, as long as the $i(\text{Extracting}, t = 12)$ is non-negative. P endowed with either the NIMBY or the balanced project always opts for the *status quo*.

Social-Welfare Concerned Host. I model social-welfare concern following Charness and Rabin (2002, Appendix 1). First, I define the *disinterested social-welfare criterion* as

[10] The model departs from the one of Frey and Oberholzer-Gee (1997) in two main respects: (1) while Frey and Oberholzer-Gee assume the project to be NIMBY, I allow for uncertainty on the project's types; (2) H chooses whether or not to accept the project, instead of choosing a level of endorsement.

follows:

$$W(\pi_P, \pi_H, \pi_{NH}) = \delta \min\{\pi_P, \pi_H, \pi_{NH}\} + (1-\delta)[\pi_P + \pi_H + \pi_{NH}] \tag{4}$$

where π indicates a monetary payoff, and $\delta \in (0,1)$ measures the concern for helping the worst-off actor versus maximizing the total social surplus. H's utility is then defined as the weighted sum of her payoff and the social-welfare criterion:

$$V_H(\pi_P, \pi_H, \pi_{NH}) \equiv (1-\lambda)\pi_H + \lambda\left[\delta \min\{\pi_P, \pi_H, \pi_{NH}\} + (1-\delta)[\pi_P + \pi_H + \pi_{NH}]\right] \tag{5}$$

with $\lambda \in [0,1]$ measuring how much H cares about pursuing the social-welfare versus her own self-interest.[11]

Under a wide range of parameters δ and λ, H's expected utility from accepting the partially compensating transfer is larger than the expected utility derived from rejecting (Figure 2). Allowing for social-welfare concern increases the number of BNE with respect to the case in which H is selfish (Appendix A.3.3). However, the game only admits two PBE. In the first PBE, for any off-equilibrium belief, H accepts the project if offered with a partially compensating transfer ($t = 6$). By backward induction, P proposes both the NIMBY and the extracting project, but always offers the partially compensating transfer; P renounces to the balanced project. Under the same parameter restrictions, if H's off-equilibrium belief about the offered project being balanced is high enough, a second PBE exists. In this equilibrium, P offers all the types of project without offering a transfer, because H finds it optimal to always accept.

5 Results

I first present aggregate results (Results 1–3) and then I focus on the channels behind those results (Results 4–5). The analysis exploits the data from the last 10 periods of play, in which one can be more confident that subjects have understood the game.[12]

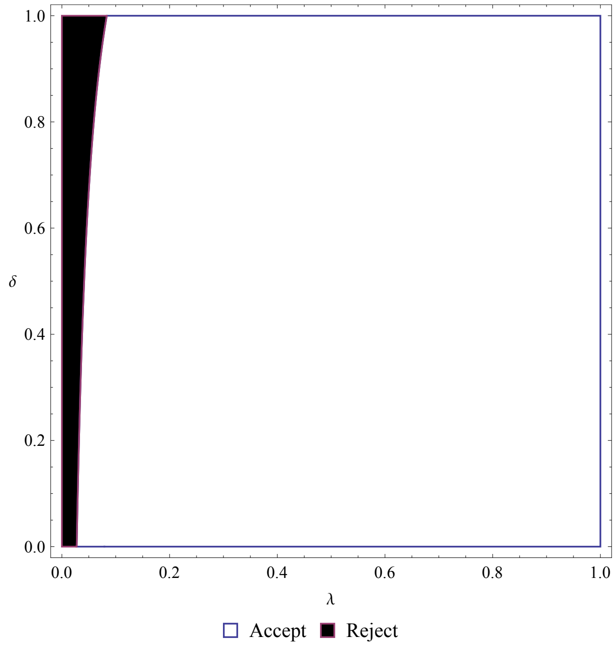
Result 1. *Both information disclosure and side-transfers have a positive and significant effect on the number of realized projects.*

Out of 240 investment opportunities in each treatment, 51 projects were realized in the baseline, 78 in the information treatment, and 104 in the side-transfer treatment. Panel A of Table 2 reports the estimates from OLS regressions comparing the percentage

[11] The reciprocity component is omitted because I consider a one-shot game.

[12] The same results hold true over the 20 periods, but the evidence is more noisy.

Figure 2: Social-Welfare Parameters and the Decision to Accept.



Notes. The graph depicts the parameters' space ($\delta \times \lambda$). In white is the region where parameters are such that H prefers to accept both partially and fully compensating side-transfers, rather than reject the project.

of projects realized in the baseline with percentage of projects realized in the information treatment (column 1) and in the side-transfer treatment (column 2), respectively. The treatment effect is measured by the variable *Treatment dummy*, which takes value 0 in the baseline, and value 1 to indicate either the information treatment (column 1), or the side-transfer treatment (column 2). Period dummies from 12 to 20 are included to account for learning. Since period dummies' coefficients are largely insignificant, I did not report them. The estimated coefficients of *Treatment dummy* are positive and significant at the 1% significance level, indicating that both tools foster the projects' realization. The regressions in panels B and C of Table 2 allow to understand whether the increase in the number of realized projects is driven by more investments (panel B) and/or higher acceptance (panel C). The positive and significant coefficients of *Treatment dummy* in column 2 of panels B and C show that the possibility to transfer tokens to the host increases both the number of offered and accepted projects. Instead, the information tool has a significant effect only on the number of accepted projects (first columns of panels B and C).

Relative to the information treatment, Result 1 does not confirm Proposition 2, which predicts that no project should be carried out because the information disclosure tool is not employed by the parties. The positive effect of side-transfers diverges from previous survey evidence (Kunreuther and Easterling, 1990; Frey et al., 1996). The evidence from this treatment indeed appears more in line with the predictions of standard economic theory, or with the hypothesis that hosts are concerned about increasing the social-welfare, rather than with the hypothesis of motivational crowding-out. Still, the data do not allow to disentangle whether subjects whether subject were not intrinsic motivated, or if their intrinsic motivation was not strong enough to lead them to realize the project without the side-transfer.[13]

[13]Frey and Oberholzer-Gee (1997) argue that side-transfers only worse off the outcome of the interaction only if individuals are intrinsic motivated. If individuals are not intrinsic motivated, the authors do not exclude the possibility that side-transfers can be beneficial. This issue if further developed in Section 6.

Table 2: Regressions on the Effect of Tools on the Projects' Realization.

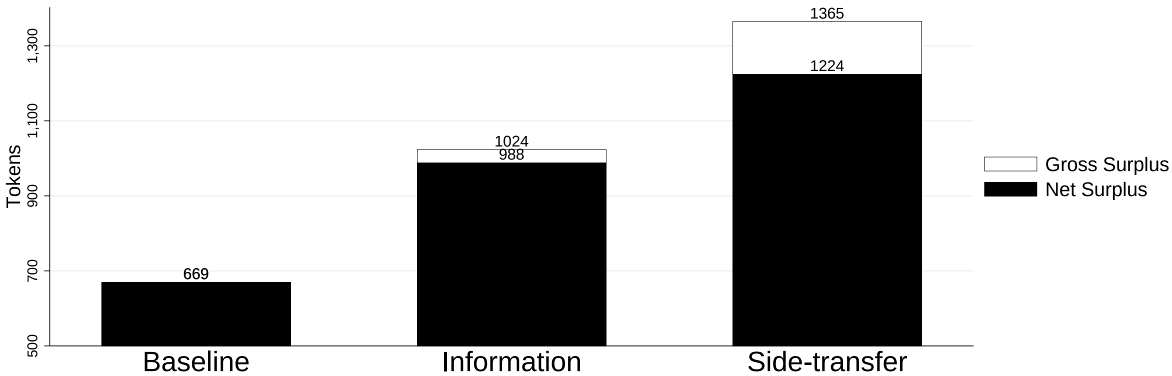
	Baseline vs. Information (OLS)	*Baseline vs. Side-transfer* (OLS)
A. **Dependent Variable:**	**% of realized projects**	
Treatment dummy	11.2500***	22.0833***
	(3.3724)	(3.6797)
Constant	19.3750***	20.2083***
	(5.9479)	(6.1752)
Period dummies	Yes	Yes
Observations	160	160
Number of clusters	16	16
R-squared	0.1158	0.2177
B. **Dependent Variable:**	**% of invest choices**	
Treatment dummy	5.8333	17.9167***
	(5.8337)	(3.8861)
Constant	51.2500***	51.4583***
	(8.0100)	(7.7210)
Period dummies	Yes	Yes
Observations	160	160
Number of clusters	16	16
R-squared	0.0922	0.1605
C. **Dependent Variable:**	**% of accept choices**	
Treatment dummy	17.8151***	19.6490***
	(6.3926)	(5.9373)
Constant	42.7208***	45.0761***
	(11.7542)	(11.042)
Period dummies	Yes	Yes
Observations	155	156
Number of clusters	16	16
R-squared	0.0891	0.1022

Notes. The dependent variable is the percentage of the projects realized / proposed / accepted within a matching cluster, in each period. The variable Treatment dummy (0,1) measures the treatment effect when information disclosure / side-transfer is possible or impossible. Dummies from period 12 to 20 are included to account for learning. Robust standard errors in parentheses. One matching cluster in a period is one unit of observation. The difference in the number of observation in panel *C* is due to the fact that there were periods in which no investment was made within the matching cluster. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Result 2. *Information disclosure and side-transfers significantly increase the social surplus generated in the experiment.*

Figure 3 suggests that the higher number of realized projects in the information and the side-transfer treatments (Result 1) translated into higher generated social surplus. Social surplus is defined as the total amount of tokens earned by all actors from the projects' realization.

Figure 3: Gross and Net Surplus across Treatments.



Notes. The graph reports the average of matching groups' gross (white) / net (black) surplus generated across the last 10 periods of play. In the gross surplus, the deadway loss related to the use of the tools is neglected. In baseline, gross and net surplus overlap, as no tool is available.

The positive effect of the introduction of both tools on the social surplus is statistically significant at the 1% significance level according to the regressions in Table 3, where I control for possible learning effects. Consistent with the theory, Result 2 holds true both when considering the gross surplus (panel A) and the surplus net of the deadway loss related to the use of the tools (panel B).

Table 3: Regressions on Gross and Net Surplus Realized across Treatments.

	Baseline vs. Information (OLS)	*Baseline vs. Side-transfer* (OLS)
A. **Dependent Variable:**	**Gross Surplus**	
Treatment dummy	35.4375***	69.5625***
	(10.6232)	(11.5912)
Constant	61.0313***	63.6563***
	(18.7360)	(19.4520)
Period dummies	Yes	Yes
Observations	160	160
Number of clusters	16	16
R-squared	0.1158	0.2177
B. **Dependent Variable:**	**Net Surplus**	
Treatment dummy	31.8625***	55.4625***
	(10.4565)	(11.1767)
Constant	60.6938***	67.7063***
	(18.4197)	(19.6581)
Period dummies	Yes	Yes
Observations	160	160
Number of clusters	16	16
R-squared	0.1058	0.1679

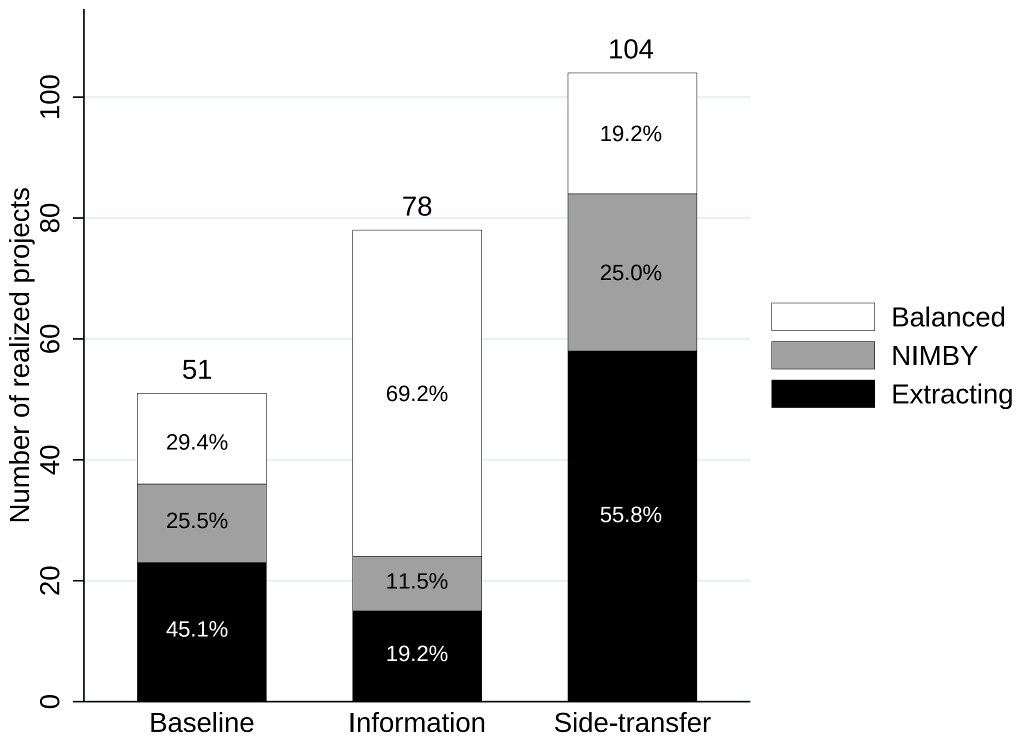
Notes. The dependent variable is the sum of the surplus generated through project's realization within a matching cluster in each period. The variable Treatment dummy (0,1) measures the treatment effect when information disclosure / side-transfer is possible or impossible. Dummies from period 12 to 20 are included to account for learning. Robust standard errors in parentheses. One matching cluster in a period is one unit of observation. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Result 3. *Information disclosure and side-transfers have different impacts on the type of projects realized and on how the surplus is shared across the actors.*

Figure 4 illustrates the number and the composition of the realized projects, across the three treatments. In the baseline, 45.1% of the 51 realized projects were extracting. NIMBY and balanced projects represent the 25.5% and 29.4% of the realized projects, respectively. In the information treatment, the percentages of extracting and NIMBY projects decrease (19.2% and 11.5%, respectively), in favor of an increase in the percentage of balanced projects (69.2%). The differences between the composition of the realized projects in the baseline and in the information treatments are all statistically significant, according to OLS regressions (Table 4, first column).

The composition of the projects realized in the treatment with side-transfers is almost the opposite of the one information treatment: 55.8% of the realized projects were extracting, while NIMBY and balanced projects were only the 25.0% and 19.2%, respectively. Although the differences in the composition are not statistically significant (Table 4, second column).[14] Although in a non-statistical way, the introduction of side-transfers increased the share of extracting projects and decreased the share of balanced projects with respect to the baseline. This evidence is hence quite consistent with the assumption of self-regarding or social-surplus concerned hosts.

Figure 4: Realized Projects and Project's Types across Treatments.



Notes. The graph summarize both the number of projects realized in each treatment (outside the bars), and the percentages of each project's type (inside the bars).

[14]The differences in the shares of both balanced and extracting projects are close to be significant at the 10% confidence level, with $p-$values of 0.108 and 0.106, respectively.

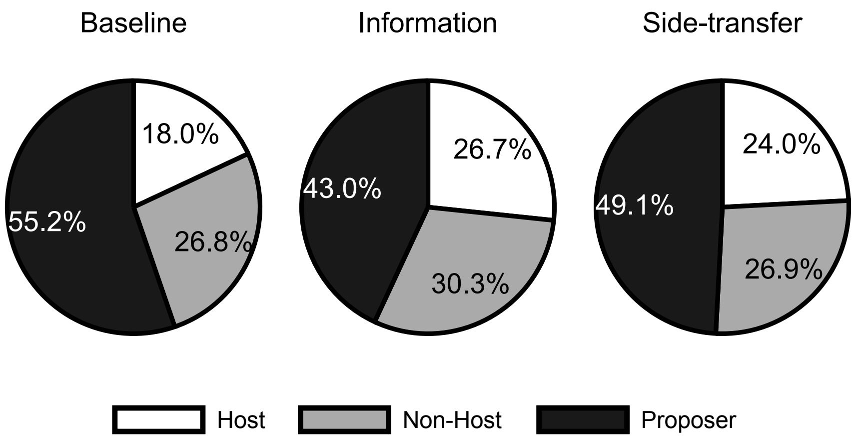
Table 4: Regressions on the Realization of each Type of Project.

	Baseline vs. Information (OLS)	*Baseline vs. Side-transfer* (OLS)
A. **Dependent Variable:**	**Share of extracting projects**	
Treatment dummy	-0.2836***	0.1444
	(0.0862)	(0.0886)
Constant	0.3701***	0.4990***
	(0.1354)	(0.1595)
Period dummies	Yes	Yes
Observations	107	114
Number of clusters	16	16
R-squared	0.1428	0.0697
B. **Dependent Variable:**	**Share of NIMBY projects**	
Treatment dummy	-0.1657**	-0.2495
	(0.0546)	(0.078)
Constant	0.2494**	0.2412**
	(0.2403)	(0.1118)
Period dummies	Yes	Yes
Observations	107	114
Number of clusters	16	16
R-squared	0.0697	0.1138
C. **Dependent Variable:**	**Share of balanced projects**	
Treatment dummy	0.4493***	-0.1225
	(0.0841)	(0.0756)
Constant	0.3804***	0.2598**
	(0.1411)	(0.1148)
Period dummies	Yes	Yes
Observations	107	114
Number of clusters	16	16
R-squared	0.2393	0.0758

Notes. The variable Treatment dummy (0,1) measures the treatment effect when information disclosure / side-transfer is possible or impossible. Dummies from period 12 to 20 are included to account for learning. Robust standard errors in parentheses. One matching cluster in a period is one unit of observation. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The different compositions of the realized projects across the treatments reflect into differences in the allocations of surplus among the actors (Figure 5). According to OLS regressions (Table 5), both in the information and in the side-transfer treatments, the share of the surplus obtained by proposers is significantly lower than in the baseline. However, side-transfers have a weak effect, in terms of size and significance. The treatments have a positive and strongly significant effect on the share of the surplus obtained by the hosts. Finally, the surplus share of the non-host significantly increases with the introduction of the information disclosure tool; their situation is unchanged when side-transfers are introduced. While treatment effects on proposers' and hosts' shares are consistent with standard economic theory, the effect on the non-hosts share in the side-transfer treatment is counter intuitive. In the side-transfer treatment, only extracting projects should to go through. Thus, one would expect a strong negative effect on the surplus share of the non-hosts. Why is this effect not observed? The explanation behind the evidence relies on the fact that, in the side-transfer treatment, the number of realized NIMBY projects doubled with respect to the baseline (26 projects versus 13 projects). As is clarified by Result 5, this unexpected increase in the number of NIMBY project is consistent with the assumption of social-surplus concerned hosts.

Figure 5: Surplus Allocation across Treatments.



Note. The graph reports the percentage of the net surplus obtained by each actor in the last 10 periods of play, across the three treatments. Average net surplus was 669 tokens in the baseline, 988 tokens in the information treatment, and 1224 in the side-transfer treatments.

Table 5: Regressions on the Surplus Allocation.

	Baseline vs. Information (OLS)	*Baseline vs. Side-transfer* (OLS)
A. **Dependent Variable:**	**Share of the surplus to proposers**	
Treatment dummy	-0.137***	-0.0542*
	(0.0336)	(0.0318)
Constant	0.5148***	0.5979***
	(0.0538)	(0.0614)
Period dummies	Yes	Yes
Observations	107	114
Number of clusters	16	16
R-squared	0.1807	0.1396
B. **Dependent Variable:**	**Share of the surplus to hosts**	
Treatment dummy	0.0956***	0.061***
	(0.018)	(0.0164)
Constant	0.2012***	0.1545***
	(0.0305)	(0.0290)
Period dummies	Yes	Yes
Observations	107	114
Number of clusters	16	16
R-squared	0.2324	0.1775
C. **Dependent Variable:**	**Share of the surplus to non-hosts**	
Treatment dummy	0.0414*	-0.0068
	(0.0228)	(0.0246)
Constant	0.2840***	0.2476***
	(0.0349)	(0.0437)
Period dummies	Yes	Yes
Observations	107	114
Number of clusters	16	16
R-squared	0.0771	0.1061

Notes. The variable Treatment dummy (0,1) measures the treatment effect when information disclosure / side-transfer is possible or impossible. Dummies from period 12 to 20 are included to account for learning. Robust standard errors in parentheses. One matching cluster in a period is one unit of observation. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Result 4. *Subjects exploit the information disclosure tool.*

In the information treatment, there always exists a Pareto inferior equilibrium, in which actors do not make use of the information disclosure tool (Section 4.2). However, the experimental data show that subjects frequently exploit the possibility to disclose information about the project's type. Table 6 reports the average effort rate for all the three actors in the information treatment. The average effort rate of each actor is strongly different from 0. As expected, in 95.5% of the cases, proposers endowed with a balanced project chose to exert a positive effort. This is not statistically different from 100% of the cases according to a Wilcoxon signed-rank test ($Z = -1.000$, p-value $= 0.3173$). The percentage of proposers' positive efforts is not statistically different from 0 when they are endowed with noxious projects according to Wilcoxon signed-rank tests ($Z = 1.414$, p-value $= 0.1575$ for extracting; $Z = 1.413$, p-value $= 0.1575$ for NIMBY). The

Table 6: Effort Decisions of the Actors.

	Proposer	Host	Non-Host
Average individual effort rate (%)	54.6	62.7	58.1
	(21.6)	(43.0)	(40.7)
Average number of effort decisions per subject	6.2	3.1	3.1
	(2.8)	(1.2)	(1.1)
Number of subjects	22	48	48

Notes. The average individual effort rate is computed by taking the average of effort decisions for each individual and than computing the average effort rate for each actor. Standard errors in parentheses.

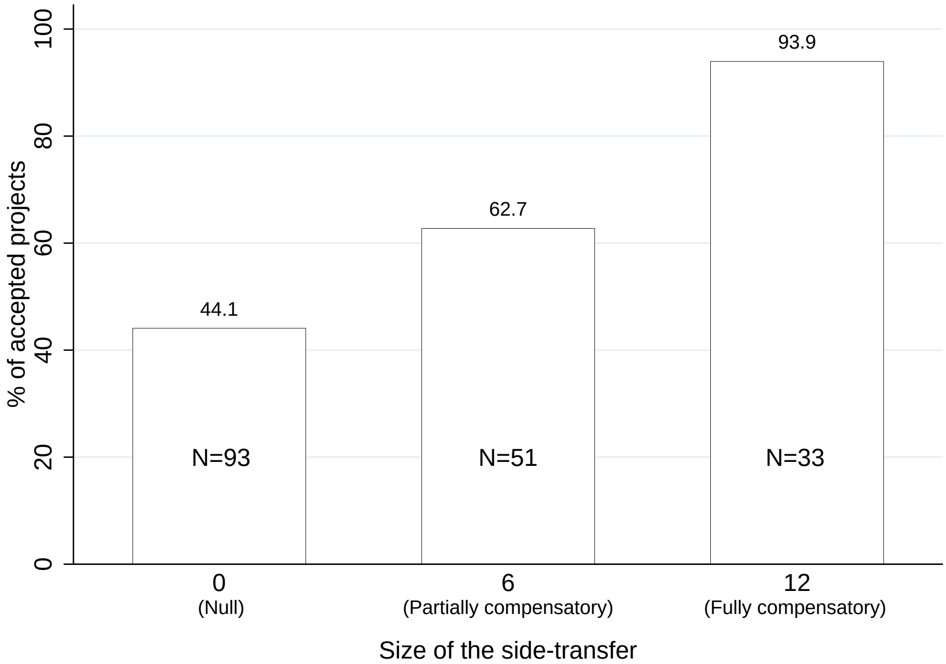
substitutability between the efforts of the host and the non-host should could lead to a free-riding problem. The effort rates of host and non-host subject are statistically different when controlling for repeated choices.

Result 5. *Side-transfers are effective in increasing the number of projects realized, also when they do not fully compensate the cost for hosting a noxious project.*

The graph depicted in Figure 6 shows that host subjects offered with positive side-transfers tend to accept the proje ct more often. In particular, the average acceptance when offered a side-transfer of 12 tokens is not statistically different from 1 (Wilcoxon signed-rank test: $Z = -1.414$, p-value $= 0.1573$). Subjects also tend to accept when they receive a transfer of only 6 tokens, which does not fully compensate them for hosting a noxious project. This results are consistent with the hypothesis that hosts are not intrinsically motivated, but they are concern about maximizing the social surplus. Finally, the acceptance rate is statistically different from 0 also when no side-transfer is offered

(Wilcoxon signed-rank test: $Z = 6.403$, p-value $= 0.0000$). However, according to the Logit regressions reported in Table 7 (column 1), the likelihood with which the host accepts the project when offered with a null transfer is not statistically different from the likelihood of accepting an unknown project in the baseline. This is not the case when

Figure 6: Effectiveness of Side-transfers.



Notes. The graph reports the average of individual acceptance rates across the last 10 periods of play. Overall, there were 93 transfers of 0 tokens, 51 transfers of 6 tokens, and 33 transfers of 12 tokens.

6 Final Remarks

This paper develops a novel experimental setup to provide new insights on the long standing problem of communities opposition to the realization of public projects which can be perceived as noxious (e.g. landfills, incinerators). Different from previous literature, I stress that a key driver of the opposition is the mistrust of host community towards the proposer of the project. Such mistrust emerges as a consequence of an informational asymmetry about the type of the project. This study focuses on types of project which produce the same social surplus, but differ in how the surplus is allocated among the actors, leaving for future research cases in which projects also differ in the generated surplus.

The paper also wants to shed light on how policy-makers can promote a fruitful participation of local communities in the realization of public projects. To this purpose, two

Table 7: Regressions on Accept Decisions under Alternative Transfers.

Dependent Variable:	**% of accept choices**		
	$t = 0$	$t = 6$	$\bar{t}$
	(Logit)	(Logit)	(Logit)
Treatment dummy	-0.137***	-0.0542*	
	(0.0336)	(0.0318)	
Constant	0.5148***	0.5979***	
	(0.0538)	(0.0614)	
Period dummies	Yes	Yes	
Observations	107	114	
Number of clusters	16	16	
R-squared	0.1807	0.1396	

Notes. The average individual effort rate is computed by taking the average of effort decisions for each individual and than computing the average effort rate for each actor. Robust standard errors in parentheses.

tools are studied. The fist tool is *endogenous information disclosure* about the project's type. According to the model of Dewatripont and Tirole (2005), information can be credibly disclosed only if both the sender and the receiver of the information exert a costly effort in communication. However, the entry stage that characterize the game studied in this experiment the tool not exploited in equilibrium. However, subjects tend to exploit the tool in the experiment, leading to an increase in the number of realized projects. This result emerges as the combined effect of a reduction in the number of noxious projects realized and an increase in the realization of balanced projects. The unexpected use of the tool may be rationalize by subjects' concern for social surplus.

The second tool available to policy-makers tested in the experiment, is *monetary side-transfers* to compensate the host community. Previous empirical studies have found that monetary transfer offered inhabitants of perspective host site reduce their endorsement towards the project (Kunreuther and Easterling, 1990; Frey et al., 1996). Frey and Oberholzer-Gee (1997) explain this evidence through the theory of the motivational crowding-out: the introduction of a monetary incentive crowds out the intrinsic benefit that community members experience when they do their civic duty by hosting a socially desirable project. In contrast, the present experiment outlines a significant increase in the number realized projects due to the introduction of side-transfers. Indeed, subjects are likely to accept to host a noxious project whenever they are at least partially compensated. The evidence is hence more consistent with the hypothesis that subjects are concerned about social surplus (Charness and Rabin, 2002). This different result can be explained by a variety of aspects. First, the neutral framing used in the experiment is hostile to intrinsic motivation, since subjects are induced to think in terms of payoffs. Moreover, Kunreuther and Easterling (1990) and Frey et al. (1996) ran their surveys on a representative sample of the population, while in my experiment subjects are mostly bachelor and master students: besides differences in the age of the subjects, there might as well be significant differences in their level of education. Risk aversion is another possible driver of the difference in the results. In general, subjects are moderately risk averse in the lab (e.g. Holt and Laury, 2005). In the field, where stakes are much higher, the assumption of risk neutrality is improper (e.g. Hartley et al., 2014). Finally, I do not claim that the evidence from my experiment rejects the hypothesis of motivational crowding-out. Rather, it offers an empirical support to the conclusion of Frey and Oberholzer-Gee (1997, p. 753): when the public spirit is not prevailing, side-transfers can enhance the realization of public projects. In the experiment this seems to be the case, since host subjects are ready to harm non-host subjects by accepting extracting projects. The experiment shows that, when the public spirit is weak, side-transfers can be particularly helpful, because individuals still care about social-welfare. In other words, subjects are prone to accept partially compensating side-transfers, renouncing to part of their payoff, to increase the social surplus.

References

Baxter, J. W., J. D. Eyles, and E. S. J. (1999). From siting principles to siting practices: a case study of discord among trust, equity and community participation. *Journal of Environmental Planning and Management 42*(4), 501–525.

Bellettini, G. and H. Kempf (2013). Why not in your backyard? on the location and size of a public facility. *Regional Science and Urban Economics 43*(1), 22–30.

Charness, G. and M. Rabin (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics 117*(3), 817–869.

Dewatripont, M. and J. Tirole (2005). Modes of communication. *Journal of Political Economy 113*(6), 1217–1238.

Feinerman, E., I. Finkelshtain, and I. Kan (2004). On a political solution to the nimby conflict. *The American Economic Review 94*(1), 369–381.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics 10*(2), 171–178.

Frey, B. S. and F. Oberholzer-Gee (1997). The cost of price incentives: An empirical analysis of motivation crowding-out. *The American Economic Review 87*(4), 746–55.

Frey, B. S., F. Oberholzer-Gee, and R. Eichenberger (1996). The old lady visits your backyard: A tale of morals and markets. *Journal of Political Economy 104*(6), 1297–1313.

Greiner, B. (2004). An online recruitment system for economic experiments.

Harrison, G. W. (2006). Experimental evidence on alternative environmental valuation methods. *Environmental and Resource Economics 34*(1), 125–162.

Hartley, R., G. Lanot, and I. Walker (2014). Who really wants to be a millionaire? estimates of risk aversion from gameshow data. *Journal of Applied Econometrics 29*(6), 861–879.

Holt, C. A. and S. K. Laury (2005). Risk aversion and incentive effects: New data without order effects. *The American Economic Review 95*(3), 902–904.

Kunreuther, H. and D. Easterling (1990). Are risk-benefit tradeoffs possible in siting hazardous facilities? *The American Economic Review 80*(2), 252–56.

Kunreuther, H., P. Kleindorfer, P. J. Knez, and R. Yaksick (1987). A compensation mechanism for siting noxious facilities: Theory and experimental design. *Journal of Environmental Economics and Management 14*(4), 371–383.

Kunreuther, H. and P. R. Kleindorfer (1986). A sealed-bid auction mechanism for siting noxious facilities. *The American Economic Review 76*(2), 295–99.

Laurent-Lucchetti, J. and J. Leroux (2011). Choosing and sharing. *Games and Economic Behavior 73*(1), 296–300.

Minehart, D. and Z. Neeman (2002). Effective siting of waste treatment facilities. *Journal of Environmental Economics and Management 43*(2), 303–324.

Pérez-Castrillo, D. and R. F. Veszteg (2007). Choosing a common project: Experimental evidence on the multibidding mechanism. *Journal of Economic Behavior & Organization 63*(3), 394–411.

Perez-Castrillo, D. and D. Wettstein (2002). Choosing wisely: A multibidding approach. *The American Economic Review 92*(5), 1577–1587.

Petts, J. (1995). Waste management strategy development: a case study of community involvement and consensus-building in hampshire. *Journal of Environmental Planning and Management 38*(4), 519–536.

Sakai, T. (2012). Fair waste pricing: an axiomatic analysis to the nimby problem. *Economic Theory 50*(2), 499–521.

Appendices

A Derivation of the Equilibria

The PBE of all games are found as it follows. By representing the game in normal form, I first find all the profiles of strategies that are BNE. Next, for each BNE, I verify whether there exists a system of beliefs such that together they constitute a PBE. I focus only on pure strategies.

A.1 Baseline game

A strategy of P specifies for each project's type (extracting, NIMBY, and balanced, respectively) whether P renounces to the project (Out) or invests (In). H has only two pure strategies: to accept (A) or reject (R) the unknown project. In the normal-form game (Figure 7) are reported the expected payoffs of H, P, and NH, respectively.[15] Best responses are underlined.

Figure 7: Normal-form baseline game.

Host \ Proposer	$OutOutOut$	$OutOutIn$	$OutInOut$	$OutInIn$	$InOutOut$	$InOutIn$	$InInOut$	$InInIn$
A	$\underline{22}$, 22, 22	$\underline{26.3}$, 26.3, 26.3	19, 30, 30	$\underline{23.3}$, 34.3, 34.3	19, 41, 19	$\underline{23.3}$, 45.3, 23.3	16, 49, 27	20.3, $\underline{53.3}$, 31.3
R	$\underline{22}$, $\underline{22}$, 22	22, 13, 22	$\underline{22}$, 13, 22	22, 13, 22	$\underline{22}$, 13, 22	22, 13, 22	$\underline{22}$, 13, 22	$\underline{22}$, 13, 22

In the baseline game, there exists a unique BNE, which is ($OutOutOut, R$). If H's off-equilibrium path belief about the offered project being balanced is $q_B < 9/22$ (Equation 2), the BNE is a PBE.

A.2 Information treatment

A strategy of P specifies for each project's type (extracting, NIMBY, and balanced, respectively) whether P renounces to the project (Out), or invests exerting the effort (In_e) or not (In_n). A strategy of H specifies whether H exerts her effort (e) or not (n), and if accepts (A) or not (R) in the case of a successful disclosure for each project's type

[15] NH's expected payoff is reported for completeness, even though it does not affect the outcome of the game when actors are characterized by self-regarding preferences.

(extracting, NIMBY, and balanced, respectively), and in the case of a disclosure failure. Finally, a strategy of NH specifies whether she exerts the effort (e) or not (n).

Figure 8 reports the game in strategic form after the elimination of dominated strategies: in Matrix A. NH does not exert the effort; in Matrix B. NH exerts it. There exist three BNE:

1. $(OutOutOut, RRA_nR_n, n)$,
2. $(OutOutOut, RRA_nR_n, e)$, and
3. $(OutOutOut, RRA_eR_e, n)$.

BNE 1. is a PBE if H's (and NH's) off-equilibrium belief about the offered project being balanced is so low that H (and NH) prefers to reject rather than exert her effort and accept if the project is balanced even if every other actor exerts his/her effort. Formally,

$$d[q^b\pi_H^b + (1-q^b)\pi_0] + (1-d)\pi_0 - C(e) \leq \pi_0$$

$$q^b \leq \frac{C(e)}{d(\pi_H^b - \pi_0)} \equiv \underline{q}$$

When $q^b < \underline{q}$, BNE 2. is not a PBE, because playing e is a dominated strategy for NH. If q^b is high enough, no BNE is PBE because H would accept also if the disclosure fail. So, H and NH do not have any incentive to exert their effort (i.e. H plays RRA_nA_n, and NH plays n). Formally, this happens when the H's expected payoff from accepting without exerting the effort is higher than the expect payoff from exerting the effort and rejecting when disclosure either fails or the project is not balanced:

$$d[q^b\pi_H^b + (1-q^b)\pi_0] + (1-d)\pi_0 - C(e) \leq q^b\pi_H^b + (1-q^b)\underline{\pi}$$

$$q_B \geq \frac{70}{103} \equiv \bar{q}$$

BNE 3. is never a PBE because if H's and NH's off-equilibrium belief is $\bar{q} > q_B > \underline{q}$, NH plays e, and H free-rides on NH's effort by playing RRA_nA_n.

Figure 8: Normal-form game with endogenous information disclosure.

Matrix A.

Non-host plays n

		Proposer											
		$OutOutOut$	$OutOutIn_n$	$OutOutIn_e$	$OutIn_nOut$	$OutIn_nIn_n$	$OutIn_nIn_e$	$In_nOutOut$	In_nOutIn_n	In_nOutIn_e	In_nIn_nOut	$In_nIn_nIn_n$	$In_nIn_nIn_e$
	RRA_nA_n	$\underline{22}$, 22, $\underline{22}$	$\underline{26.3}$, 26.3, $\underline{26.3}$	$\underline{26.3}$, 25.7, $\underline{26.3}$	19, 30, $\underline{30}$	$\underline{23.3}$, 34.3, $\underline{34.3}$	$\underline{23.3}$, 33.7, $\underline{34.3}$	19, 41, $\underline{19}$	$\underline{23.3}$, 45.3, $\underline{23.3}$	$\underline{23.3}$, 44.7, $\underline{23.3}$	16, 49, $\underline{27}$	20.3, $\underline{53.3}$, $\underline{31.3}$	20.3, 52.7, $\underline{31.3}$
Host	RRA_nR_n	$\underline{22}$, $\underline{22}$, $\underline{22}$	22, 19, $\underline{22}$	22, 18.3, 22	$\underline{22}$, 19, $\underline{22}$	22, 16, $\underline{22}$	22, 15.3, 22	$\underline{22}$, 19, $\underline{22}$	22, 16, $\underline{22}$	22, 15.3, 22	$\underline{22}$, 16, $\underline{22}$	$\underline{22}$, 13, $\underline{22}$	$\underline{22}$, 12.3, $\underline{22}$
	RRA_eR_e	$\underline{22}$, $\underline{22}$, $\underline{22}$	22, 19, $\underline{22}$	23.3, 21.6, 24	21.3, 19, $\underline{22}$	20.7, 16, $\underline{22}$	22.6, 18.6, 24	21.3, 19, $\underline{22}$	20.7, 16, $\underline{22}$	22.6, 18.6, 24	20.7, 16, $\underline{22}$	20, 13, $\underline{22}$	$\underline{22}$, 15.6, $\underline{24}$

Matrix B.

Non-host plays e

		Proposer											
		$OutOutOut$	$OutOutIn_n$	$OutOutIn_e$	$OutIn_nOut$	$OutIn_nIn_n$	$OutIn_nIn_e$	$In_nOutOut$	In_nOutIn_n	In_nOutIn_e	In_nIn_nOut	$In_nIn_nIn_n$	$In_nIn_nIn_e$
	RRA_nA_n	$\underline{22}$, 22, $\underline{22}$	$\underline{26.3}$, 26.3, 25.7	$\underline{26.3}$, 25.7, 25.7	19, 30, 29.3	$\underline{23.3}$, 34.3, 33	23.3, 33.7, 33	19, 41, 17.7	$\underline{23.3}$, 45.3, 22	23.3, 44.7, 22	16, 49, 25.7	20.3, $\underline{53.3}$, 29.3	20.3, 52.7, 29.3
Host	RRA_nR_n	$\underline{22}$, $\underline{22}$, $\underline{22}$	22, 19, 21.3	24, 21.6, $\underline{23.3}$	$\underline{22}$, 19, 21.3	22, 16, 20.7	24, 18.6, $\underline{22.6}$	$\underline{22}$, 19, 20.7	22, 16, 20.7	24, 18.6, $\underline{22.6}$	$\underline{22}$, 16, 20.7	$\underline{22}$, 13, 20	$\underline{24}$, 15.6, $\underline{22}$
	RRA_eR_e	$\underline{22}$, 22, $\underline{22}$	21.3, 19, 21.3	25.2, $\underline{24.9}$, $\underline{25.2}$	21.3, 19, 21.3	20.7, 16, 20.7	$\underline{24.6}$, 21.9, $\underline{24.6}$	21.3, 19, 21.3	20.7, 16, 20.7	$\underline{24.6}$, 21.9, $\underline{24.6}$	20.7, 16, 20.7	20, 13, 20	23.9, 18.9, 23.9

A.3 Side-transfers treatment

A.3.1 Selfish Host

It can be shown that there exist two BNEs:

1. $(OutOutOut, R_{t_0}R_{\underline{t}}R_{t_h})$, and

2. $(In_{\bar{t}}OutOut, R_{t_0}R_{\underline{t}}A_{\bar{t}})$.

where the subscripts denote the size of the side-transfer offered by P.

BNE 1. is not a PBE, because if H is offered $\bar{t}$ it is strictly dominant for her to play $A_{\underline{t}}$ for any out of equilibrium belief. The BNE 2. is a PBE as long as H's out of equilibrium belief are below q^* (Equation 2).

A.3.2 Civic Host

As Frey and Oberholzer-Gee (1997), I restrict the analysis to the case in which the intrinsic benefit from hosting either project n or b when no transfer (t_0) is offered is large enough to make H's expected utility of accepting the project larger than the expected utility of rejecting it. Formally,

$$\frac{1}{3}\underline{\pi} + \frac{1}{3}(\underline{\pi} + i(t_0)) + \frac{1}{3}(\pi_H^b + i(t_0)) > \pi_0$$

$$i(t_0) > \frac{1}{2}(3\pi_0 - 2\underline{\pi} - \pi_H^b) \tag{6}$$

H's expected utility of accepting an unknown project with t_0 is larger than the expected utility of accepting a project with a $\underline{t}$ if

$$\frac{1}{3}(2\underline{\pi} + \pi_H^b + 2i(t_0)) > \frac{1}{2}(2\underline{\pi} + 2t_l + i(t_l))$$

$$i(t_0) > \frac{1}{2}(\underline{\pi} + 3t_l - \pi_H^b) + \frac{3}{4}i(t_l) \tag{7}$$

Note that when Condition 6 holds, $i(t_0) > 1/2(\underline{\pi} + 3t_l - \pi_H^b)$. So, Condition 7 is fulfilled when D does not decrease too slowly in t. If Condition 7 holds, there exist six BNE:

1. $(OutOutOut, R_{t_0}R_{\underline{t}}R_{\bar{t}})$,

2. $(In_{t_0}In_{t_0}In_{t_0}, A_{t_0}R_{\underline{t}}R_{\bar{t}})$,

3. $(In_{t_0}In_{t_0}In_{t_0}, A_{t_0}A_{\underline{t}}R_{\bar{t}})$,

4. $(In_{t_0}In_{t_0}In_{t_0}, A_{t_0}R_{\underline{t}}A_{\bar{t}})$.

5. $(In_{t_0}In_{t_0}In_{t_0}, A_{t_0}A_{\underline{t}}A_{\overline{t}})$, and

6. $(In_{\overline{t}}OutOut, R_{t_0}R_{\underline{t}}A_{\overline{t}})$.

BNE 1. to 3. are not PBE, because if H is offered $\overline{t}$, it is strictly dominant for her to play $A_{\overline{t}}$ for any out of equilibrium belief.[16] Similarly, also BNE 6. cannot be a PBE, because, knowing that for the balanced P it is strictly dominated to offer $\underline{t}$, it is always optimal for H to choose $R_{\underline{t}}$ which delivers π_0, instead of choosing $A_{\underline{t}}$, which delivers only 19. BNE 5. is a PBE, if the out of equilibrium beliefs of H satisfy the following condition:

$$(1 - q^b - q^n)\underline{\pi} + q^n_H(\underline{\pi} + i(t_0)) + q^b_H(\pi^b_H + i(t_0)) > \pi_0$$

$$q^b > \frac{\pi_0 - \underline{\pi} - q^n_H i(t_0)}{\pi^b_H - \underline{\pi} + i(t_0)}$$

A.3.3 Social-Surplus Concerned Host

Under the parameter restrictions depicted in Figure 2, there exist eight BNE:

1. $(OutOutOut, R_0R_{\underline{t}}R_{\overline{t}})$,

2. $(In_{\underline{t}}In_{\underline{t}}Out, R_0A_{\underline{t}}R_{\overline{t}})$,

3. $(In_{\underline{t}}In_{\underline{t}}Out, R_0A_{\underline{t}}R_{\overline{t}})$,

4. $(In_{\underline{t}}In_{\underline{t}}Out, R_0A_{\underline{t}}A_{\overline{t}})$,

5. $(In_{\overline{t}}OutOut, R_0R_{\underline{t}}A_{\overline{t}})$,

6. $(In_0In_0In_0, A_0R_{\underline{t}}R_{\overline{t}})$,

7. $(In_0In_0In_0, A_0A_{\underline{t}}R_{\overline{t}})$, and

8. $(In_0In_0In_0, A_0A_{\underline{t}}A_{\overline{t}})$.

BNE 1. to 3., 6., and 7. are not PBE, because if H is offered $\overline{t}$, it is strictly dominant for her to play $A_{\overline{t}}$ for any off-equilibrium path belief. Similarly BNE 5. is not PBE, because if a social-welfare concerned H is offered $\underline{t}$, it is strictly dominant for her to play $A_{\underline{t}}$ for any off-equilibrium path belief. BNE 4. is a PBE for any out of equilibrium belief. BNE

[16] This is true as long as accepting $\overline{t}$ does impose a psychological cost on H. For consistency with the model of Frey and Oberholzer-Gee (1997), I rule out this possibility.

8. is a PBE if the H's off-equilibrium path beliefs are such that the expected benefit by accepting the unknown project are larger than the those of rejecting it. Formally,

$$q^b\left((1-\lambda)\pi_H^b+\lambda[\delta\pi_H^b+(1-\delta)\overline{S}]\right)+(1-q^b)\left((1-\lambda)\underline{\pi}+\lambda[\delta\underline{\pi}+(1-\delta)\overline{S}]\right)$$
$$>(1-\lambda)\pi_0+\lambda[\delta\underline{\pi}+(1-\delta)\underline{S}]$$

where $\overline{S}$ and $\underline{S}$ are the social surplus when a project is accepted or rejected, respectively.

$$q^b>\frac{(1-\lambda)(\pi_0-\underline{\pi})+\lambda(1-\delta)(\underline{S}-\overline{S})}{(\pi_H^b-\underline{\pi})(1-\lambda+\lambda\delta)}$$

Cooperation and Climate Change: a Dynamic Experiment*

Marco Casari[†]   Riccardo Ghidoni[‡]

Abstract

This paper studies what factors can favor or hamper cooperation given the specificities of mitigating climate change. Climate negotiations are a social dilemma with a "public bad" that presents a dynamic aspect: the damage is proportional to the stock of heat-trapping gases in the atmosphere that comes from the accumulation of present and past emissions. This paper designs an experiment tailored on the field salient characteristics of the climate change challenge and studies the dynamic of cooperation in limiting noxious emissions. We compare a static vs. a dynamic versions of the problem to assess its impact on the ability of societies to cooperate.

*We are thankful to Giacomo Calzolari, who provided great support for the theoretical backbone of the paper.

[†]Department of Economics, University of Bologna. Email: marco.casari@unibo.it.
Web: http://www2.dse.unibo.it/casari

[‡]Department of Economics, University of Bologna. Email: riccardo.ghidoni2@unibo.it.
Web: https://sites.google.com/site/riccardoghidoni/

1 Introduction

The vast majority of the scientific studies provides unequivocal evidence of a sharp rise both of land and ocean surface temperature anomalies since 1850. An undisputed driver of the global warming is the increased concentration of greenhouse gases (GHG) in the atmosphere. Human activity is responsible for most of GHG emissions, which reached their highest peak in human history in 2010 (49 $GtCO_2eq/yr$). Without adequate mitigation policies, the temperature increase in 2100 is expected to range from 3.7°C to 4.8°C compared to pre-industrial levels.

It is difficult to assess all the possible impacts of global warming and estimate their magnitude. However, scholars agree that its net economic effect will be negative for the majority of the countries (Tol, 2009). Moreover, the climatic change may lead to catastrophic events (e.g. the disappearance of entire ecosystems), which are difficult to quantify in economic terms.

To prevent the occurrence of these negative scenarios, all countries should significantly abate their GHG emissions so that gases concentration in the atmosphere can be slowly disposed, or at least maintained under control. However, since every country will benefit from the mitigation effort the other countries, there is room for opportunistic behaviors which may undermine the international cooperation required to avoid the global warming. These features make climate negotiations similar to a *common-pool problem*.

In this paper, we develop an experimental platform to study which are the specific aspects of the "climate change game" that may hinder a cooperative reduction of the emissions. In particular, we investigate whether the dynamic nature of this problem makes it behaviorally difficult to cooperate. In every round of our infinite horizon game, countries individually benefit from their own emission choice, but the sum of all countries' emissions (i.e. global emissions) generates a damage which harms everyone. Across three treatments, we vary the persistence of the damage due to the global emissions of certain round. This is equivalent to say that we vary the rate at which the stock of global emissions deteriorates. In the *long persistence* treatment, the whole global emissions of one round survives to the next. In this scenario, past emissions choices will keep damaging the countries in the future. In the *no persistence* treatment, instead, the stock of emissions clears in every round. Finally, in the *partial persistence* treatment, we study an intermediate scenario, in which half of the global emissions survives from one round to the other. The relevant equilibrium concept is the Constant Action Markov equilibrium, which prescribes that countries will choose their emission level without taking into account the exernality that they will impose on every other country, thus every country will emit above the social optimum level. Cooperation among countries can be sustained as an equilibrium outcome in all treatments. The damage coefficient is adjusted in every treatment to ensure that the equilibrium predictions remain the same across treatments.

To add credibility to our setup, we distinguish between *rich* and *poor* countries. Rich and poor countries have the same preferences over the per capita output they produce through their own emissions. However, the same level of emissions generates a lower per capita output in poor countries, because poor countries are generally characterized by a larger population than rich countries.[1] As a consequence, for the same level of output, rich countries will obtain a lager benefit than poor countries. While standard economic theory predicts that this form of heterogeneity should not affect the play, a large body of experimental literature has shown that inequality can be an important driver of behavior. In particular, there is experimental evidence that inequality can impede cooperation (e.g. Tavoni et al., 2011; Anderson et al., 2008).

This study does not investigate which tools can help to enhance international cooperation (e.g. communication, voting, or direct forms of punishments). We leave this issue for future research.

2 Literature Review

This paper contributes to the growing literature that study climate negotiations through the use of laboratory experiments. Milinski et al. (2008) are the first to modify a PGG to capture the aspect of collective risk that characterize the climate change game. In their experiment, subjects can invest their private endowment to prevent the climatic catastrophe, i.e. the loss of the private savings of all the 6 subjects matched in a group. At the end of the 10 rounds of play, if the money invested in environment protection by a group is below a known threshold, the catastrophe takes place with a certain probability. Experimental treatments differ in the likelihood of the catastrophe: 90% chances, 50% chances, and 10% chances. Authors find that the catastrophe is likely to be avoided when its likelihood is higher: with 90% chances, almost half of the groups mange to avoid it. In the other two treatments, 80% up to the 100% of the groups failed in preventing the catastrophe. This findings are consistent with the theoretical benchmark of rational selfish subjects (Barrett, 2011).

Building on the design of Milinski et al. (2008), Tavoni et al. (2011) study whether

[1]Of curse many other dimensions of heterogeneity across countries characterize the context of interest.

communication can help cooperation. Tavoni et al. argue that the big problem behind this type of interaction is the multiplicity of equilibria. Communication in the form of non-binding pledges on future investment decisions can help countries to coordinate on cooperative equilibria. The authors also consider that countries are asymmetric, and distinguish between poor and rich countries. Poor countries begin the game with an endowment equivalent to 70% of the endowment of rich countries, because poor countries are forced to invest to the environment protection in the first 3 rounds, while rich countries (forcedly) invest 0. This type of inequality is different from the one of our experiment, which is based on the fact that poor countries can achieve less *per capita* output through their emissions, rather than on the idea of "historic responsibility" implemented by Tavoni et al. (2011). The data show that, while inequality has a strong negative effect on the ability to avoid the catastrophe, communication helps cooperation a lot also in the case of asymmetric countries. Communication did not only improved coordination. It also allowed rich countries to communicate to poor countries their willingness to invest more than poor countries in environment protection, in order to balanced the effect of the first three rounds played by the computer.

As noticed by Barrett (2011), Tavoni et al. (2011) try to add realism to the evidence of Milinski et al. (2008). However, Barrett (2011) criticize the absence of scientific uncertainty: subjects know the exact threshold they must reach to avoid the catastrophe. The issue of climate negotiation under scientific uncertainty is tackled by Barrett and Dannenberg (2012), who study how uncertainty of impacts and uncertainty on the threshold affect the play in a modified PGG. The authors show both theoretically and experimentally that the fear of crossing a dangerous threshold can turn climate negotiations into a coordination game where cooperative outcome can be reached especially when communication among players is allowed (and this holds even if there is uncertainty about the magnitude of the catastrophe). However, the uncertainty on the value of the dangerous threshold turn the negotiation back into prisoner dilemma, where cooperation collapse.

The experiments presented above differ from ours because they focus on a catastrophic risk which can be avoided if countries' investments in environment protection are above a certain threshold. In this regard, Barrett (2011) argues that it does not seem realistic that all the efforts below or above the threshold are wasted. In reality, every additional ton of emission can exacerbate the global warming. In our experiment, indeed, the damage is gradually increasing in the stock of the emissions. Another important difference is that all these studies are characterized by finite horizon.

There are however other experiments which investigate the emergence of cooperative behaviors in games characterized by dynamic externalities. Though not all these experiments explicitly refer to the climate change problem, they are close to our study. Previous experiments in this literature only consider the case of symmetric players. Thus, the asymmetry among the players sets our study aside from the rest of this literature. In a finite horizon common-pool resource game, Herr et al. (1997) analyze cooperation when the current extraction of one subject increases the extraction costs of others in the current period only (time-independent externalities), or both in the current period and in future periods (time-dependent externalities). Cooperative outcomes are poor predictors of behavior for both types of externalities. Moreover, lower payoffs are reported in the time-dependent treatments, in part due to myopic behavior. An important difference between this experiment and ours is the time horizon, since we study an infinite horizon game.

Vespa (2013) studies cooperation in an infinite horizon stylized common-pool game (Levhari and Mirman, 1980). Specifically, Vespa (2013) studies whether the reproduction rate of the natural resource helps or not cooperation. He finds that most of the observed behavior is consistent with Markov strategies. However, when the reproduction rate of the resource is high, most diffused strategy tit-for-tat. As a consequence, it seems that when the reproduction rate is low, cooperation is more fragile: subjects do not return to cooperate upon any deviation. Moreover, if subjects cannot choose the maximal level of extraction from the common resource, cooperation increases.

Pevnitskaya and Ryvkin (2013) develop a dynamic public bad experiment based on the model of Dutta and Radner (2004), to explore how time horizon influence cooperation between pairs of subjects. For the majority of the rounds, the authors do not detect significant differences between the finite and infinite horizon games, with the exception of the few last rounds in the finite treatment. However, within the finite horizon scenario there is a positive effect of learning: subjects learn how to cooperate by reducing their emissions. This effect is not registered in the infinite horizon scenario. Moreover, authors find that when the experiment has an environmental framing subjects are less likely to choose to pollute.

Another close paper to our experiment is the study on climate change by Sherstyuk et al. (2014). The authors experimentally study an infinite horizon game characterized by dynamic negative externalities (Dutta and Radner, 2004). Different from us, Sherstyuk et al. (2014) mostly focus on understanding whether overlapping generations can reach similar cooperation outcomes of a long-lived decision-maker. The authors also allow for the use of non-strategic instruments which can enhance cooperation (advices to followers, access to past history). The data show that socially optimal outcomes are often achieved and sustained by long-lived players but not by overlapping generations. Under overlapping generations observed actions tend to become more myopic.[2] The issue of intergenerational decision-making in common-pool problems has also been studied by Chermak and Krause (2002), and by Fischer et al. (2004). The first document scarce levels of intergenerational cooperation in a finite horizon setting. The second find that subjects expects others to care about future generations, but actions are not consistent with such an expectation.

3 The Model

We consider a set I consisting of four countries. Countries are of two types: two countries are rich (r), and two countries are poor (p). So, $I = \{r, r, p, p\}$. In every round t, each country $i \in I$ chooses its level of GHG emissions $e_i(t)$ from an interval ranging from 0 to a common upper-bound e, i.e. $e_i \in [0, e]$. Countries use their emissions to produce output. For the same level of emissions, p achieves a lower output per capita than r, because p's population is bigger. All countries benefit in the same way from their output per capita, but they are equally damaged by the stock of pollution accumulated. These features are captured by the following instantaneous payoff:

$$u_i(e_i, E) \equiv \gamma Ln(a_i e_i) - (c \times E)/4,$$

where $\gamma > 0$, E is the current stock of pollution and $c/4 > 0$ is the magnitude of the damage related to it. The parameters a_r and a_p capture the differences in population between r and p, so $a_p \leq a_r$. The stock of pollution E accumulates according to the following dynamic equation:

$$E(t) = \sigma E(t-1) + \sum_{i \in I} e_i(t),$$

where $\sigma \in [0, 1]$ is the persistence rate of the emissions, and $E(0) = E_0 \geq 0$ is the initial level of pollution.

Let us define $u_i(t) = u(e_i(t), E(t))$, where $e_i(t)$ is the action taken by country i at time

[2] Besides the research question and the symmetry among players, several other aspects differ between the experiment of Sherstyuk et al. (2014) and our study. First, they consider a unique, low (equal to 0.3) persistence rate of the stock of emissions. In our experiment we compare three possible persistence rates (0, 0.5, and 1). Second, the game is played in groups of 3, while our groups consist of 4 subjects. Third, the number continuation probability is much lower than the one we use (0.75 against 0.92). Fourth, they consider a quadratic utility function while we employ a logarithmic utility function.

t. The total payoff of country i is

$$v_i = \sum_{t=0}^{\infty} \delta^t u_i(t)$$

where $\delta \in [0, 1)$ is the discount factor.

All countries observe all their individual actions $e_i(t)$ at any t. A strategy of a country i is a mappings from H, the set of all possible histories, into actions.

Global Optimum. The global objective is an un-weighted sum of individual payoffs:

$$v = 2(v_r + v_p).$$

Proposition 1. *The efficient actions are constant and, for any player, equal to*

$$e^* = \gamma \frac{1 - \sigma\delta}{c}.$$

Proposition 1 simply states that the marginal benefit $1/e_i$ is equal to the marginal "over-time" (total) global cost $4 \times (c/4)/(1 - \delta\sigma)$.

Constant-Actions Markov Perfect Equilibria. In these equilibria, strategies do not depend on E_0, neither on the equilibrium path nor out-of-equilibrium.

Proposition 2. *Constant-MPE contemplate a constant action for any player in any period equal to*

$$e^{CMP} = 4\gamma \frac{1 - \sigma\delta}{c}$$

Notice this is simply the marginal benefit γ/e_i equal to marginal "over-time" (total) cost

$$\frac{c}{4}\left[1 + \delta \frac{\sigma}{(1 - \delta\sigma)}\right] = \frac{c}{4\,(1 - \delta\sigma)}$$

Clearly, $e^{CMP} = 4e^*$ and the difference

$$e^{CMP} - e^* = 3e^*$$

is decreasing in σ. For future reference, as shown in the proof, we can write the associated value function as

$$V_i^{CMP}(E_0) = U_i^{CMP} - \frac{c}{4(1 - \delta\sigma)} E_0,$$

where

$$U_i^{CMP} = \frac{1}{1-\delta}\left[\gamma Ln(a_i\ e^{CMP}) - \frac{c}{1-\delta\sigma}e^{CMP}\right].$$

We now show the difference between the individual (present-valued) payoff with the global optimum solution and the constant-MPE:

$$V_i^*(E_0) - V_i^{CMP}(E_0) = \frac{\gamma}{1-\delta}\left(3 - Ln(4)\right) \geq 0.$$

Finally, notice that when $\sigma = 0$ so that the game is in fact a repeated one, then e^{CMP} is equal to the Nash equilibrium action of the repeated game.

Trigger Strategies Equilibria. We focus on trigger strategies equilibria which rely on the threat to revert to the constant-actions MPE e^{CMP} discussed above. We name these "Constant actions trigger equilibria".

Result *[To be proven] Constant actions trigger equilibria do exist.*

Proposition 3. *If $\delta \geq \bar{\delta}$ there exists a constant-actions trigger equilibrium (CT equilibrium) with a constant action $\hat{e}_i = e^* = \gamma\frac{1-\ \delta}{c}$ for any player i, where*

$$\bar{\delta} = \frac{1}{9}(8Ln(2) - 3) \simeq 0.28.$$

Notice that incentive compatibility is uniquely affected by δ: none of the other parameters affects incentive compatibility.

References

Anderson, L. R., J. M. Mellor, and J. Milyo (2008). Inequality and public good provision: An experimental analysis. *The Journal of Socio-Economics 37*(3), 1010–1028.

Barrett, S. (2011). Avoiding disastrous climate change is possible but not inevitable. *Proceedings of the National Academy of Sciences 108*(29), 11733–11734.

Barrett, S. and A. Dannenberg (2012). Climate negotiations under scientific uncertainty. *Proceedings of the National Academy of Sciences 109*(43), 17372–17376.

Chermak, J. M. and K. Krause (2002). Individual response, information, and intergenerational common pool problems. *Journal of Environmental Economics and Management 43*(1), 47–70.

Dutta, P. K. and R. Radner (2004). Self-enforcing climate-change treaties. *Proceedings of the National Academy of Sciences of the United States of America 101*(14), 5174–5179.

Fischer, M.-E., B. Irlenbusch, and A. Sadrieh (2004). An intergenerational common pool resource experiment. *Journal of Environmental Economics and Management 48*(2), 811–836.

Herr, A., R. Gardner, and J. M. Walker (1997). An experimental study of time-independent and time-dependent externalities in the commons. *Games and Economic Behavior 19*(1), 77–96.

Levhari, D. and L. J. Mirman (1980). The great fish war: an example using a dynamic cournot-nash solution. *The Bell Journal of Economics*, 322–334.

Milinski, M., R. D. Sommerfeld, H.-J. Krambeck, F. A. Reed, and J. Marotzke (2008). The collective-risk social dilemma and the prevention of simulated dangerous climate change. *Proceedings of the National Academy of Sciences 105*(7), 2291–2294.

Pevnitskaya, S. and D. Ryvkin (2013). Environmental context and termination uncertainty in games with a dynamic public bad. *Environment and Development Economics 18*(01), 27–49.

Sherstyuk, K., N. Tarui, M.-L. V. Ravago, and T. Saijo (2014). Intergenerational games with dynamic externalities and climate change experiments.

Tavoni, A., A. Dannenberg, G. Kallis, and A. Löschel (2011). Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *Proceedings of the National Academy of Sciences 108*(29), 11825–11829.

Tol, R. S. (2009). The economic effects of climate change. *The Journal of Economic Perspectives*, 29–51.

Vespa, E. I. (2013). Cooperation in dynamic games: An experimental investigation.

Appendices

A Proofs

Proof of Proposition 1. At any date t, let E be any initial value of emission and $V(E)$ the associated value function, i.e. the solution of the dynamic programming problem (the Hamilton Jacobi Bellman equation)

$$V(E) = Max_{e_r,e_p}\{2\gamma\,[Ln(a_r e_r) + Ln(a_p e_p)] - c\times(\sigma E + 2e_r + 2e_p)] + \delta V(\sigma E + 2e_r + 2e_p)\}.$$

Let $e_p(E), e_r(E)$ be solutions to the previous maximization. Plugging these into the previous equation we obtain a functional equation in $V(E)$ to be solved. We guess that $V(E)$ takes the following form:

$$V(E) = 2(w_p + w_r) - 2(k_p + k_r)E.$$

We now have to verify if these four parameters w_i and k_r exist that satisfy the HJB equation and to identify them. From the HJB equation, applying our guess for the value function we obtain

$$\begin{aligned} 2(w_p + w_r) - 2(k_p + k_r)E \;\; &= \;\; Max_{e_r,e_p}\{2\gamma\,[Ln(a_r e_r) + Ln(a_p e_p)] - c\times(\sigma E + 2e_r + 2e_p) + \\ &\quad + \delta[2(w_p + w_r) - 2(k_p + k_r)(\sigma E + 2e_r + 2e_p)]\}. \end{aligned}$$

The necessary conditions on e_r, e_p are

$$\frac{\gamma}{e_i} = c + 2\delta(k_p + k_r)$$

or

$$\bar{e}_i = \frac{\gamma}{c + 2\delta(k_p + k_r)}$$

which is independent of E. Plugging into the HJB Equation, we have

$$\begin{aligned} 2(w_p + w_r) - 2(k_p + k_r)E \;\; &= \;\; 2\gamma\,[Ln(a_r\bar{e}_r) + Ln(a_p\bar{e}_p)] - c\times(\sigma E + 2\bar{e}_r + 2\bar{e}_p) + \\ &\quad + \delta[2(w_p + w_r) - 2(k_p + k_r)(\sigma E + 2\bar{e}_r + 2\bar{e}_p)] \end{aligned}$$

Solving for $w_p + w_r$

$$w_p+w_r = \frac{1}{1-\delta}\left\{(k_p+k_r)E + \gamma\left[Ln(a_r\bar{e}_r) + Ln(a_p\bar{e}_p)\right] - \frac{1}{2}c \times (\sigma E + 2\bar{e}_r + 2\bar{e}_p) + \delta[-(k_p+k_r)(\sigma E + 2\right.$$

in order for $w_p + w_r$ to be independent of E it must be

$$k_p + k_r - \frac{1}{2}c\sigma - \delta(k_p + k_r)\sigma = 0$$

that is

$$k_p + k_r = \frac{1}{2}\frac{c\sigma}{1-\sigma\delta}$$

which shows $\bar{e}_i = e^*$. Finally, we also notice that

$$\begin{aligned} w_p + w_r &= \frac{1}{1-\delta}\left\{\gamma\left[Ln(a_r\bar{e}_r) + Ln(a_p\bar{e}_p)\right] - \frac{1}{2}c \times (2\bar{e}_r + 2\bar{e}_p) + \delta[-\frac{1}{2}\frac{c\sigma}{1-\sigma\delta}(2\bar{e}_r + 2\bar{e}_p)]\right\} \\ &= \frac{1}{1-\delta}\left[\gamma\left[Ln(a_r\bar{e}_r) + Ln(a_p\bar{e}_p)\right] - \frac{c}{2(1-\sigma\delta)}(2\bar{e}_r + 2\bar{e}_p)\right]. \end{aligned}$$

QED

Proof of Proposition 2. We show that if all players $j \neq i$ play the constant action e^{CMP} then the best response for player i is $e_i = e^{CMP}$ which leads to a value function of the type

$$V(E) = w - k\sigma E.$$

(notice that the stock of pollution inherited from the past is σE). Indeed with this guess on the value function we can write

$$w - k\sigma E = Max_{e_i}\{\gamma Ln(a_i e_i) - \frac{c}{4} \times \left(\sigma E + e_i + 3e^{CMP}\right) + \delta\left[w - k\sigma(\sigma E + e_i + 3e^{CMP})\right]\}.$$

The maximizer must satisfy

$$\frac{\gamma}{e_i} = \frac{c}{4} + \delta\sigma k \Longleftrightarrow e_i = \frac{4\gamma}{c + 4\delta\sigma k}$$

and, subsisting back, we have that it must be

$$-k\sigma = -\frac{c}{4}\sigma - \delta k\sigma^2$$

or, equivalently

$$k = \frac{c}{4\left(1-\delta\sigma\right)}$$

the previous HJB equation does not depend on E. It then follows that the best response is indeed

$$e_i = \frac{4\gamma}{c + 4\delta\sigma k} = \frac{4\gamma}{c + 4\delta\sigma\frac{c}{4(1-\delta\sigma)}} = e^{CMP}.$$

It is also useful to notice that with this result we can write

$$w = \frac{1}{1-\delta}\left[\gamma Ln(a_i e^{CMP}) - \frac{c}{4\left(1-\delta\sigma\right)}4e^{CMP}\right].$$

QED

Proof of Proposition 3. For any E, it can be shown that the incentive compatibility constraint for any player i with a (candidate) CT equilibrium with actions $\hat{e}_i$ is

$$\begin{aligned}
&\gamma Ln(a_i\ \hat{e}_i) - \frac{c}{4}\times\left(\sigma E + 2\hat{e}_r + 2\hat{e}_p\right) + \delta\left\{\hat{U}_i - \frac{c}{4(1-\delta\sigma)}\sigma\left(\sigma E + 2\hat{e}_r + 2\hat{e}_p\right)\right\} \\
\geq\ &\gamma Ln(a_i\ \tilde{e}_i) - \frac{c}{4}\times\left(\sigma E + \tilde{e}_i + \hat{e}_{i'} + 2\hat{e}_j\right) + \delta\left\{U_i^{CMP} - \frac{c}{4(1-\delta\sigma)}\sigma\left(\sigma E + \tilde{e}_i + \hat{e}_{i'} + 2\hat{e}_j\right)\right\}
\end{aligned}$$

Clearly, the optimal deviation $\tilde{e}_i$ requires

$$\tilde{e}_i = e^{CMP}$$

and the constraint becomes

$$\begin{aligned}
&\gamma Ln(a_i\ \hat{e}_i) - \frac{c}{4}\times\hat{e}_i + \delta\left\{\hat{U}_i - \frac{c}{4(1-\delta\sigma)}\sigma\hat{e}_i\right\} \\
\geq\ &\gamma Ln(a_i\ e^{CMP}) - \frac{c}{4}\times e^{CMP} + \delta\left\{U_i^{CMP} - \frac{c}{4(1-\delta\sigma)}\sigma e^{CMP}\right\}
\end{aligned}$$

Using the definition of

$$\hat{U}_i = \frac{1}{1-\delta}\left[\gamma Ln(a_i\ \hat{e}_i) - \frac{c}{4(1-\delta\sigma)}(2\hat{e}_r + 2\hat{e}_p)\right]$$

and of

$$U_i^{CMP} = \frac{1}{1-\delta}\left[\gamma Ln(a_i\ e^{CMP}) - \frac{c}{1-\delta\sigma}e^{CMP}\right]$$

the constraint can be finally rewritten as (using $\hat{e}_i = \hat{e}_r = \hat{e}_p$)

$$\frac{1+3\delta}{4(1-\delta\sigma)} c \left(e^{CMP} - \hat{e}_i\right) \geq \gamma Ln(a_i\ e^{CMP}) - \gamma Ln(a_i\ \hat{e}_i)$$

Now we consider in particular the efficient $\hat{e}_i$ and the constraint finally becomes

$$\frac{\gamma}{4}(3 + 9\delta - 8Ln(2)) \geq 0.$$

QED