Crowdfunding Conservation (and Other Public Goods)

Erik Ansink, Mark Koetse, Jetske Bouma, Dominic Hauck, Daan van Soest

Abstract: Crowdfunding has become an increasingly popular means to fund the provision of public goods and especially of nature conservation projects. We implement a lab-in-the-field experiment by setting up a web-based user interface, very similar to actual crowdfunding platforms, to test whether coordination mechanisms, like seed money and decoy projects, can increase the effectiveness of crowdfunding campaigns if multiple public goods projects are eligible for funding. We find some of our treatments to affect coordination especially via early contributions, but not always in an intuitive way. Our results are confirmed in a follow-up experiment with actual nature conservation projects.

JEL Codes: C93, H41, L31, Q57

Keywords: crowdfunding, nature conservation, lab-in-the-field experiment, threshold public goods, charitable giving

CROWDFUNDING WEBSITES ARE INCREASINGLY being used as a means to raise capital for the provision of public goods. Crowdfunding is the practice of a fundraiser

Erik Ansink is in the Department of Spatial Economics, Vrije Universiteit Amsterdam, and Tinbergen Institute (erik.ansink@vu.nl). Mark Koetse is at the Institute for Environmental Studies, Vrije Universiteit Amsterdam. Jetske Bouma is at the PBL Netherlands Environmental Assessment Agency. Dominic Hauck is at the Institute for Environmental Studies, Vrije Universiteit Amsterdam. Daan van Soest is in the Department of Economics, TSC and CentER, Tilburg University. We thank Harold Houba, Jos van Ommeren, seminar participants at various conferences, and two anonymous reviewers for comments. We thank Natuurmonumenten and Kantar TNS for excellent support and advice. We are grateful to the Netherlands Environmental Assessment Agency for financial support. Ansink, Bouma, and Hauck acknowledge financial support from the Dutch Research Council (NWO; grant no. 841.12.002). Koetse and van Soest acknowledge funding from the European Commission 7th Framework Programme (grants no. 308393 and 613420, respectively). None of the authors have relevant material or financial interests that relate to the research described in this paper.

Dataverse data: https://doi.org/10.7910/DVN/PB9DUS

Received June 9, 2020; Accepted November 18, 2021; Published online March 25, 2022.

Journal of the Association of Environmental and Resource Economists, volume 9, number 3, May 2022.

https://doi.org/10.1086/718280

attracting capital from many investors to fund a project through an online platform. Crowdfunding is relatively new, with well-known platforms such as FundRazr, GoFundMe, Kickstarter, and IndieGoGo having been launched only in the late 2000s. The crowdfunding of public goods, sometimes referred to as "civic" crowdfunding, has found its niche on these platforms (Stiver et al. 2015; Hudik and Chovanculiak 2018), but specialized civic crowdfunding websites (e.g., Fundly, IOBY, JustGiving, and GlobalGiving) have emerged as well. A large share of civic crowdfunding involves projects related to urban commons (e.g., gardens, parks, events), education, infrastructure, environment, and wildlife (Charbit and Desmoulins 2017).

One prominent example of crowdfunding conservation is the well-known Ocean Cleanup,[1] which raised USD 2 million from 38,000 donors to fund a series of expeditions to test high-seas plastics removal technologies. More generally, Gallo-Cajiao et al. (2018) made a global assessment of conservation crowdfunding on 72 crowdfunding platforms. They identified some 600 campaigns that have been implemented over the last decade, aimed at raising funds for protection of individual species or even of entire ecosystems. In addition to initiatives by citizens or nongovernmental organizations, many governments have also started to use crowdfunding to fund conservation. Examples include the US National Recreation and Park Association announcing "Fund Your Park" in 2016 and Australia's state of Victoria's "Threatened Species Collection" initiative to match revenues from selected crowdfunding campaigns in 2015.

The increasing interest for civic crowdfunding motivated us to study how conservation crowdfunding campaigns can be made more effective. Using crowdfunding may mitigate several obstacles that are typically associated with the private provision of public goods. For instance, crowdfunding allows cheap matching between projects and potential donors through low search costs, low risk exposure, low demand uncertainty, and transparent monitoring of progress (Agrawal et al. 2014; Strausz 2017). Yet, other obstacles remain, including cheap-riding and coordination problems. Cheap riding occurs whenever donors contribute to the public good but try to reach an outcome where their own relative contribution is low (Isaac et al. 1989). Coordination problems occur when the multiplicity of projects offered on a crowdfunding platform causes an "inefficient distribution of donations across projects" (Corazzini et al. 2015, 17), possibly leading to project failure and discouraging potential donors (Solomon et al. 2016). As sometimes half of the projects fail (see, e.g., Mollick [2014] for the case of Kickstarter), there is room for improvement—raising the total contributions to public goods projects or improving coordination on specific projects.

Most crowdfunding platforms offer an all-or-nothing design, where contributions are effectuated only if a prespecified funding threshold is met. This design thus mimics a threshold public goods game with full refund—a design that is known to boost donations

1. See https://theoceancleanup.com/milestones/crowd-funding-campaign/.

(Cadsby and Maynes 1999).[2] Yet, by itself this design does not solve all cheap-riding and coordination problems. Both problems, but particularly the coordination problem, could potentially be mitigated by signaling focal projects, for instance, by promoting such projects more prominently on the crowdfunding platform. Signaling has been demonstrated to increase welfare in coordination games (Schelling 1960; Mehta et al. 1994). In crowdfunding, signals used by fundraisers include short time windows and high thresholds while signals used by the platforms include all kinds of website design options such as the selection of featured projects (Mollick 2014; Belleflamme et al. 2015; Devaraj and Patel 2016).[3] Identification of the impact of such signals in practice is inevitably muddled by endogeneity problems.

In this paper, we test the impact of various crowdfunding designs, and specifically the use of focal projects, on project success. To control for possible endogeneity problems, we do so by means of a lab-in-the-field experiment that is designed to capture (most of) the key characteristics of real-world crowdfunding platforms. We investigate two mechanisms that aim to raise total contributions and/or facilitate coordination by signaling focal projects. The first is the provision of seed money (also known as "challenge gifts" in the literature on charitable giving; see List and Lucking-Reiley 2002; Rondeau and List 2008) by a donor or by the conservation agency itself. The second mechanism is the fundraiser advertising an additional project that closely resembles one of the existing projects, albeit slightly less attractive, and thereby serves as a decoy project. This decoy project, or seemingly irrelevant alternative, may trigger the so-called attraction effect (see Ariely and Wallsten 1995). Conservation agencies typically have a portfolio of projects they can implement so that they need to decide which (combination of) projects to advertise for crowdfunding, and when. When ranking all projects, some are likely to be dominated by others in the donors' perception (even though they all meet the conservation agency's efficiency criteria). In that sense, dominated projects arise naturally, and hence conservation agencies may consider harnessing the decoy effect to promote their target projects in the portfolio on offer.

We implement our analyses in the context of two or more public goods projects requiring funding. The benefit and cost parameters are chosen such that it is socially optimal to fund two projects, but the projects differ in both the bonus they pay when implemented as well as in the costs of funding them. In our baseline treatments we observe which project receives most funding and which one receives least. The two coordination mechanisms that we test, seed money and a decoy project, are then targeted at raising the

2. In game-theoretic terms, compared to a standard linear contribution structure the all-or-nothing design changes the nature of the game from a social dilemma into a coordination game, which is found to substantially increase contributions (Isaac et al. 1989).

3. Other aspects that affect crowdfunding success but are not under the control of the fundraiser include the number of potential donors, social norms, the characteristics of early donors, and herding (see Bøg et al. 2012; Crosetto and Regner 2014; Belleflamme et al. 2015).

contributions to the least-preferred (but still socially efficient) project. The possibility (and, in fact, desirability) of funding more than one project sets our study apart from the existing literature on crowdfunding, as we are interested in the possibility of how to raise contributions to multiple socially efficient public goods projects that may compete for funds. The key questions then are, first, whether these coordination mechanisms raise contributions to the project they are targeted at, and second, whether they are able to raise overall social welfare.

This setup is of specific relevance for conservation agencies as they typically have multiple projects available for implementation, while their preference ordering may not necessarily coincide with that of the general public. Consider the case of two projects, one aimed at conserving a charismatic species, and the other aimed at protecting biodiversity that is essential for an ecosystem's support system (think of "engineer species" like insects, worms, or algae). Or, as an alternative example, consider the case of a nature restoration project of an area close to an urban center and a similar project to be implemented in a more remote area. While in each of these two examples the first project is more likely to be funded by private individuals, the conservation agency may actually prioritize funding of the second. Preferably, the agency would like to raise funds for both projects. But if the second project fails to raise enough funds, does using seed money or decoys result in an increased likelihood of both projects ending up being funded, or do these instruments result in one project crowding out the other?

To bring the expanding literature on crowdfunding design closer to the challenges faced by conservation agencies, our lab-in-the-field experiment aims to extend the insights obtained by earlier crowdfunding experiments in the laboratory (Wash and Solomon 2014; Corazzini et al. 2015). Three differences stand out. First, we test two new crowdfunding designs that employ focal projects. Second, our lab-in-the-field experiment took place over a period of 4 days, with subjects being able to log in to the internet-based user interface of the game using any device at any moment from any location. This set-up is very different from the typical laboratory setting, where a session lasts for about 1 hour, and where multiple rounds are being played in an environment without distractions. Lab settings may cause subjects to be more sensitive to (subtle) differences in project characteristics than they would be in case of an actual crowdfunding event. Also, our set-up allows for continuous and sequential decision making—as opposed to simultaneous decision making in discrete rounds (cf. Bigoni et al. 2015). Third, we employ a representative sample of the relevant population as opposed to WEIRD students.[4]

Our lab-in-the-field approach has the advantage of increased realism and hence of better external validity. This is confirmed by the results of a follow-up framed field experiment that we implemented in cooperation with an environmental nongovernmental organization (NGO), in which participants were offered the option to invest in actual nature conservation projects. However, while lab-in-the-field experiments offer superior

4. WEIRD: Western, Educated, Industrialized, Rich, and Democratic; see Henrich et al. (2010).

external validity, they do come at the expense of less control, and especially so over subjects' understanding of the game. We will come back to these aspects in section 3.

Some of our results, reported in section 3, confirm earlier findings, such as the importance of coordination difficulties in general. Our analysis of the impact of seeding and seemingly irrelevant alternatives yields new findings. We find that such mechanisms affect coordination more than they affect cheap riding as measured by total amounts contributed. As expected, seeded projects serve as focal projects, receiving relatively more contributions and displaying a higher success rate. This effect is more pronounced when the number of available projects is large (six) than when it is small (two). This result is consistent with seeding being effective in mitigating coordination problems. Perhaps counterintuitively, adding a seemingly irrelevant project does not lead to an attraction effect. The decoy project is found to divert contributions away from its target project rather than to increase them and even to decrease the funding success rate of the nontargeted (yet socially efficient) project. An analysis of contribution dynamics reveals that both mechanisms affect project success through their differential impact on early contributions.

The paper that is closest to ours, Corazzini et al. (2015), tests whether the number and salience of projects affect project success. Varying project salience by changing merits or via random signals, Corazzini et al. demonstrate the existence of coordination failures when four projects are available compared to when there is just one project. Salience can overcome this failure, except when the salient project is dominated by other projects, in which case it is less successful. We find a different result with seeding being able to even increase the success rate of dominated goods. Our results also demonstrate that signals work in the context of crowdfunding, but not always as expected. Because signals may not work intuitively or may only work when coordination is particularly difficult, it is not straightforward to employ them in practice. Nevertheless, our results provide important insights for conservation agencies as to how to increase the success rate of their crowdfunding projects.

The setup of this paper is as follows. In section 1 we describe the experimental design, in section 2 we present our theory and hypotheses, and in section 3 we present the results of the lab-in-the-field experiment. We then move to presenting the results of our field experiment, implemented together with an environmental NGO, in section 4. Finally we present our conclusions in section 5.

1. EXPERIMENTAL DESIGN

The game that we implement is an online real-time threshold public goods game without rebate (if contributions exceed the threshold) but with full refund (if the threshold fails to be met). The linear version of the public goods game has been widely used to study efficiency and behavioral aspects of various mechanisms and design options in the provisioning of public goods (for overviews, see Ledyard 1995; Chaudhuri 2011). The modified version with a contribution threshold, the threshold public goods game, pays a bonus to each subject in the group, independent of whether she contributed or how much, if and only if the total amount contributed is equal to or larger than a predetermined threshold (van de Kragt et al. 1983; Palfrey and Rosenthal 1984; Bagnoli and Lipman 1989). The threshold public goods game captures the essence of the mechanism implemented by crowdfunding projects within the social domain (Andreoni 1998). In our setting, contributions are binding and cannot be withdrawn.[5] Full refund, a money-back guarantee in case the threshold is not reached, reduces the risk of contributing and hence is commonly implemented by major crowdfunding platforms to increase contributions (Isaac et al. 1989; Cadsby and Maynes 1999; Wash and Solomon 2014).[6] There is no possibility that any contributed tokens are forfeited, unless the threshold is reached and the corresponding bonus is paid out. Hence, an early small contribution as a signal that players should coordinate on a particular good is relatively cheap. The decision to offer no rebate in case of excess contributions (Marks and Croson 1998; Spencer et al. 2009) is motivated by our attempt to mimic real-world practice on most major crowdfunding platforms (e.g., Kickstarter, IndieGogo, Fundly, JustGiving, and GoFundMe). Offering no rebate also allows us to gauge our participants' understanding of the mechanism in place (see app. A; see also Cason and Zubrickas 2019).[7]

The subjects who participated in our experiment were a representative sample of the Dutch population (in terms of age, gender, and education; see also sec. 1.3). They played the game online using a web-based user interface that we developed, similar in style to conventional crowdfunding websites (see the online appendix). Our setting allowed for continuous interaction over a period of 4 consecutive days, from Thursday 8 a.m. to Sunday 10 p.m. As such, our game is a real-time threshold public goods game (Dorsey 1992; Kurzban et al. 2001; Duffy et al. 2007). The time span of 4 days was chosen as a compromise between the length of conventional crowdfunding campaigns on the one hand and subjects' attention and our intent to limit their time investment on the other. As

5. The no-withdrawal rule is fairly common on real-world crowdfunding platforms. It also prevents the experiment from becoming too time consuming for the participants, especially toward the end of the experiment.

6. Tabarrok (1998) and Zubrickas (2014) designed a modified version of full refund that includes an additional refund bonus. This mechanism is predicted to be even more effective in raising contributions than with just full refund. Cason and Zubrickas (2017, 2019) tested this prediction in a laboratory setting and found that, indeed, it outperforms full refund in terms of both project success and welfare. To the best of our knowledge the refund bonus has not been implemented yet on actual crowdfunding platforms, and hence we decided to just implement full refund without such additional bonuses.

7. Alternatively, we could have programmed the software such that contributions in excess of the threshold would be capped to just meeting the threshold. We would have still been able to detect irrational behavior, but more tokens would have remained available for additional investments. In retrospect, we regret not having implemented this, because the probability of groups meeting the thresholds would have been higher. In any case, the no rebate rule applied to all treatments, and hence it is unlikely to have affected the internal validity of our study.

argued by Levitt and List (2007), the choice of how to implement an experiment boils down to a trade-off between context (or "realism") and control. Laboratory experiments afford maximum control but may have limited external validity because of a lack of realism. Lack of realism may present itself in various forms. One example is that fairly subtle design features may be very visible (or salient) in the lab, whereas they would go largely unnoticed in real-world decision making situations—because subjects in a lab face fewer distractions from the experimental task than is the case in real-world decision-making situations. Similarly, in a lab setting subjects may tend to feel prompted to make an active decision in each decision round while in the real world people cannot be forced (or feel less obliged) to do so. Our 4-day lab-in-the-field experiment with "real" people as participants thus holds the promise of high external validity; a promise that we test in the follow-up experiment with real conservation projects; see section 4. Increased realism may come at the expense of control, however, especially of the subjects' understanding of the (consequences of the) various actions that they can take in the experiment.

When deciding about the choice between lab and lab in the field, our considerations were as follows. First, when implementing a lab-in-the-field experiment, we still have full control over the game's design and payoffs, while the factors that we cannot control (how often—if at all—people log in, whether they pay sufficient attention to the presence of either seed money or a seemingly irrelevant alternative) are likely to be important factors in the real world as well. We also reasoned that if fairly subtle changes in crowdfunding design (like adding the seemingly irrelevant alternative) would work in a lab environment, this would not guarantee that they would have the same impact in a lab-in-the-field type of environment, because real-life distractions may make it less likely that such changes are noticed. Second, and by definition, control is even worse in our follow-up experiment in which we, in cooperation with a nature conservation agency, offered participants the option to fund actual nature conservation projects. In this follow-up experiment we could control the costs of contributing as well as the available budgets, but not how people perceive "the bonus" (the environmental benefits obtained when the project is funded). Had we implemented our core experiment in the lab, we would not have been able to identify the cause of any possible differences between the core experiment and the follow-up experiment. Would they have been the result of a lack of understanding among the participants in the follow-up experiment, or rather because of the differences in the (perceived) cost-benefit ratio of contributing to the projects? Third, we reasoned that we would be able to mitigate possible misunderstandings by (i) rigorously pretesting our materials (instructions, the web-based user interface, etc.), (ii) making sure that the game instructions were accessible from the user interface throughout the experiment, and (iii) offering participants the option to contact us for clarification. We thus took great care to remove any source of confusion or misunderstanding. In appendix A we assess the extent to which possible confusion or misunderstandings may have caused seemingly irrational behavior. Although we find some differences in behavior depending on subjects' own evaluation of their understanding of the game, they are unlikely to have affected our outcomes.

Our analysis is based on a between-subjects design across five treatments. In our benchmark treatment, subjects have the option to contribute to two threshold public goods that differ in the contribution threshold that needs to be met for the project to go through, and in the bonus each subject receives if the threshold is reached.[8] Subsequent treatments differ from the benchmark treatment in terms of the number of goods and their characteristics. Each treatment was played by 90 subjects, distributed over 15 groups of six subjects each.[9] In total, across the five treatments, 450 subjects participated in the experiment.

1.1. Benchmark Treatment

In the benchmark treatment, which we will refer to as Ben, subjects were randomly assigned to groups of six. Upon logging in to the user interface, each subject received a one-time endowment of 34 tokens (worth EUR 0.20 per token) in her private account. Subjects could use this endowment to contribute to one or both of two threshold public goods by investing part or all of their endowment, either at once, or in smaller chunks. If the threshold for a public good (or project) is reached at the game deadline (i.e., Sunday 10 p.m.), a bonus is paid to each subject in the group regardless of their individual contributions to the public good.[10] Contributions to public goods in excess of the threshold are not returned ("no rebate"). If the threshold is not met at the game deadline, the project bonus is not paid out and any contributions to that public goods project are returned to the contributors' private accounts ("full refund"). Subjects' payoffs are the sum of what is left in their private accounts plus any bonus received from one or both projects.

The two Ben goods differ in threshold and bonus and we refer to them as Good 72_{18} and Good 84_{24}. Good 72_{18} has a threshold value of 72 and pays a bonus of 18 to each of the six subjects in the group, while Good 84_{24} has a threshold value of 84 and a per-subject bonus of 24. Hence, 72_{18} is easier to implement since it has a lower threshold (72 vs. 84) while 84_{24} offers a higher net payoff to the group as a whole (108 vs. 144). Group payoffs are maximized if both thresholds are met, but note that this requires (substantial) contributions by at least five players; the initial endowments are such that four or fewer subjects cannot finance both goods ($4 \times 34 < 72 + 84$).

8. Note that we use the terms "good" and "project" interchangeably.

9. Both the number of groups per treatment and the number of subjects per group were chosen based on the sample sizes used by Corazzini et al. (2015). They used 12 groups per treatment, with four subjects per group. We decided to use both more groups (15 rather than 12 per treatment) as well as more players per group (six rather than four) to make sure we would be at least as well powered as Corazzini et al. (2015).

10. Bonuses are symmetric for each subject to prevent coordination problems due to payoff asymmetry (see Wash and Solomon 2014).

1.2. Other Treatments

To test the impact of various crowdfunding design choices, we implement four additional treatments. These treatments differ from Ben in terms of the number of goods and their characteristics. Table 1 summarizes our treatments; for our motivation of the choice of treatment characteristics, see section 2.

In section 3, we show that, in Ben, 72_{18} is less successful than 84_{24}—both in terms of total contributions and in the number of times the threshold is reached. We developed two treatments, Seed72 and Sia72 (Sia = seemingly irrelevant alternative), to test whether, respectively, seeding and a decoy project can improve outcomes for the least successful Ben project, 72_{18}. Both 72_{18} and 84_{24} are socially efficient, and conservation agencies may want to stimulate especially the project that is least preferred by the donors—to increase the chances of both projects getting funded or maybe also because the preference ranking of the two projects by the agency differs from the preference ranking as perceived by the donors. Targeting the two mechanisms at stimulating the least successful Ben good provides a strong test for their effectiveness, especially because of Corazzini et al.'s finding that salience can only help overcome coordination problems if the salient good stands out based on merit. We now discuss, in turn, how we implemented each of the two mechanisms.

In Seed72 we add 20 seed tokens to 72_{18} at the start of the game. Different from papers analyzing the effects of seeding in the charitable contributions literature (List and Lucking-Reiley 2002), we raise the threshold by the same number to create 92_{18}. Because the bonus paid in case of success remains unchanged, 92_{18} under Seed72 is formally equivalent to 72_{18} under Ben, although subjects may perceive this differently. To highlight the seeding, subjects are informed in the instructions as well as in the relevant project description about the seeding amount made at the start of the game by "someone who does not participate in the game." Because of its formal equivalence, the main purpose of seeding is to signal a focal project.

Table 1. Overview of Projects in All Treatments

Treatment	Project Characteristics
Ben	72_{18}, 84_{24}
Seed72[a]	$\underline{92_{18}}$, 84_{24}
Sia72	72_{18}, 84_{24}, 72_{17}
Six	72_{18}, 84_{24}, 74_{16}, 78_{17}, 86_{22}, 90_{23}
SixSeed78[b]	72_{18}, 84_{24}, 74_{16}, $\underline{98_{17}}$, 86_{22}, 90_{23}

[a] The underlined project differs from its related project under Ben by an anonymous initial contribution of 20 tokens as well as an increase in its threshold of 20 tokens, making the two projects formally equivalent.

[b] The underlined project differs from its related project under Six by an anonymous initial contribution of 20 tokens as well as an increase in its threshold of 20 tokens, making the two projects formally equivalent.

In Sia72 we add a "seemingly irrelevant alternative" project, 72_{17}, to the two Ben projects. This project is aimed to serve as a decoy for 72_{18}. These two projects' thresholds are the same, while the bonus paid by 72_{17} is lower. That means that 72_{17} is strictly dominated by 72_{18}. If the decoy project triggers an attraction effect, then the target project 72_{18} is expected to be more successful than if the decoy project is not present, while 72_{17} itself should receive zero contributions. We are not aware of conservation agencies currently harnessing the decoy effect in their crowdfunding campaigns. However, implementing it is, in principle, a viable strategy. Conservation agencies typically have a portfolio of projects they can implement, and hence they need to decide which (combination of) projects to advertise for crowdfunding, and when. As stated in section 1.1, a group's joint endowments are large enough to fund two projects, but not three. Group welfare is maximized when both 84_{24} and 72_{18} get funded, and this is the case in both Ben and Sia72. Whether adding 72_{17} to the two Ben goods results in an increase in contributions to 72_{18} (preferably without decreasing the contributions to 84_{24}) is an open question. On the one hand, contributions may increase because of the attraction effect. On the other hand, adding the decoy project increases the number of advertised projects and hence may complicate (rather than facilitate) coordination.

The effectiveness of the two mechanisms may vary depending on the number of advertised projects. In the next two treatments, Six and SixSeed78, we increase the number of available goods from two to six. Coordination failures are more likely to arise when more projects are offered simultaneously. In Six, we add four goods that are dominated by the two Ben goods in terms of both their thresholds and bonuses; see table 1. In SixSeed78, similar to Seed72, we add 20 seed tokens to the least successful good under Six and raise its threshold by the same number. In section 3, we show that this least successful good is 78_{17}, and hence we create 98_{17} to be implemented in SixSeed78. A variation of treatment Six with an additional seemingly irrelevant project (SixSia78) would not make much sense, given that four of the six projects in this treatment are already dominated and relatively similar to one of the two Ben goods.

1.3. Sampling and Procedures

Subjects were sampled from a survey panel hosted by Kantar TNS, a Dutch survey consultancy. All communication with subjects was done via email by the consultancy. Subjects were invited to participate about 1 month in advance, asking them for their availability to participate in an online game and log in at least once a day over the course of 4 specified days. From the sample, subsamples for each treatment were formed, with each subsample being representative for the Dutch population in terms of gender, age, and education. Three days prior to the experiment, subjects were reminded of the experiment and given the opportunity to cancel their participation. One day before the start of the experiment, an email was sent to all participants with game instructions and a personal login URL. The game instructions were accessible from the user interface throughout the experiment. The experiment took place in three sessions (first Ben, then Sia72 and

Six, and finally Seed72 and SixSeed78), each lasting from Thursday 8 a.m. to Sunday 10 p.m., and were conducted in the period of October–November 2015. Each day of the 4-day session a reminder email was sent before noon to each subject with their personal login URL, irrespective of whether they had already logged in on that day.

Subjects who canceled prior to the start of the experiment were replaced by subjects from a standby sample. To assure full groups of active contributors we also allowed replacement on the first day of the game. We intended to replace any subject who did not log in on Thursday before 10 p.m. by a subject from a standby sample. As discussed in section 3, we were able to replace only a share of those subjects. Other subjects in the group were not informed of such replacements, nor could they derive this information through the user interface.

Instructions (see the online appendix) included the game description, the rules of play,[11] calculation of payoffs, an extensive example, and a FAQ list that was compiled based on the evaluation of a pilot study in October 2015. Instructions were kept as brief as possible since they had to be read on-screen. The specific wording was based on multiple rounds of testing during the pilot study. We framed the public goods as "projects' and contributions as "investments," without expecting significant impact of this framing on game behavior (Alekseev et al. 2017). We will use the terms interchangeably throughout the paper. Contact information to a designated employee at the research consultancy was provided in case anything was unclear (20 out of 450 subjects used this feature).

Upon logging in via their personal device (e.g., PC, laptop, tablet, smartphone), subjects entered the game's user interface, a web-based platform; a screenshot is displayed in the online appendix. The projects were identified by a number and were described in neutral terms. If more than two projects were presented, they were displayed with a maximum of two projects per row. General information on the game was displayed at the top of the screen and included (i) the number of remaining tokens in the subject's private account, (ii) the number of subjects in the group, and (iii) the game deadline (Sunday 10 p.m.). In addition, specific feedback is presented for each project, including (i) total contributions by all subjects in the group, (ii) the remaining contribution gap to the

11. A notable feature of our experimental setup was that subjects were only eligible to receive payments if they logged in at least once on each of the 4 days of the experiment. Lab-in-the-field experiments have the advantage of improved realism, but this comes at the expense of less control. When designing the experiment our main concern was that subjects would log on too infrequently for coordination to evolve. Imposing the rule that one needs to check daily up to the moment at which one has spent one's entire budget may have provided too strong incentives for subjects to spend all their tokens as quickly as possible. After all, while "getting rid of all one's tokens" is not a privately optimal strategy within the context of the experimental game, it may well be rational when taking into account one's opportunity costs of time in this multiple-day experiment. The rule of having to log in at least once on each of the 4 days thus seemed to be indispensable to induce active participation. The rule was emphasized in the instructions, and participants also received a daily reminder, via email, before noon.

threshold, (iii) the number of subjects in the group who have contributed, (iv) the subject's own total contribution and, in case of seeded projects under Seed72 and SixSeed98, (v) a statement on the number of seed tokens. This feedback was updated continuously; subjects were automatically logged off after 20 minutes of inactivity.

Within 3 days after the end of a 4-day session, payments were effectuated via a bank transfer by the research consultancy, conditional on having logged in at least once each day and filling out a short online survey (both conditions were announced in the invitation as well as in the experiment instructions). This survey contained items on attitude, game behavior, and game evaluation. The research consultancy provided us with additional information on a wide range of sociodemographic variables that had been collected prior to the start of the experiment. Regarding payoffs, each token was worth EUR 0.20. Starting with 34 tokens, subjects earned on average EUR 7.96 for an estimated time investment of, in total, about 17 minutes over 4 days (excluding the time spent on reading the instructions upon first login and any time spent thinking about possible strategies when not logged in).

2. THEORY AND PREDICTIONS

2.1. Theoretical Framework

Consider a setting with $g \in \{1, 2, \ldots, G\}$ public goods (or public goods projects) and $j \in \{1, 2, \ldots, J\}$ players. The game starts at time $t = 0$ and ends at time $t = T$. Each player has an endowment of e_j tokens in her private account, which is hers to keep, but which she can also use over the course of the game to contribute to one or more of the available G public goods projects. Denote a contribution by player j to good g at time t by c_{jg}^t. Total contributions by all players to good g up to time t are denoted by $C_g^t = \Sigma_j \Sigma_{s=1}^t c_{jg}^s$.

When a contribution threshold τ_g is reached at time T for good g, a bonus b_g is paid out to each player, independent of whether she contributed to the project or not. No rebate implies that excess contributions $C_g^T - \tau_g$ are wasted. Full refund implies that any contributions of each player in a group $(\Sigma_{s=1}^T c_{jg}^s)$ to a good that does not reach the threshold, that is, $C_g^T < \tau_g$, are returned to that player's private account. The following function denotes payoffs π_j:

$$\pi_j = e_j + \sum_g \left(\begin{cases} 0 & \text{when } C_g^T < \tau_g \\ b_g - c_{jg}^T & \text{when } C_g^T \geq \tau_g \end{cases} \right). \tag{1}$$

Payoffs are equal to the sum of endowment and good-specific payoffs. A subject's good-specific payoff is equal to the bonus she receives minus her own contributions to that good if the threshold is met, and zero otherwise, independent of own or others' contributions (an example calculation is provided in the instructions; see the online appendix).

Subjects have all necessary information available in the user interface. At each moment in the game, each player is informed about T (and hence $T - t$), the number of players who have contributed to each good, their total contributions to each good C_g^t, one's own (cumulative) contribution to each good $\Sigma_{s=1}^{t} c_{jg}^{s}$, as well as the number of tokens still available to her $(e_j - \Sigma_g(\Sigma_{s=1}^{t} c_{jg}^{s}))$. In addition, the game instructions are accessible from the user interface throughout the experiment.

In the static version of a single linear threshold public goods game with no rebate and full refund, there exist two sets of (pure-strategy) Nash equilibria. The first set contains inefficient equilibria. Such equilibria satisfy the *no deviation constraint* (Croson and Marks 2000). That is, the threshold is not met and no single subject can make the project succeed, considering the number of tokens still in her budget, the number of tokens that need to be invested to reach the threshold, and the bonus to be paid out upon success. Violation of the *no deviation constraint* implies that there is at least one subject who has enough remaining budget to make the project successful at an individual cost lower than the bonus received. Given full refund, any equilibrium with nonzero contributions in this set is payoff-equivalent to the zero-contributions equilibrium, which is always included in this set. For any player, contributing less would not affect payoffs while contributing more would either not affect payoffs (if such additional contribution does not fill the gap to the threshold) or would lead to a lower payoff (if the number of tokens that need to be invested is larger than the size of the bonus).

The second set contains efficient equilibria. In such equilibria, the threshold is exactly met with the sum of contributions being divided over the players with no player contributing more than the bonus from the good (cf. Bagnoli and Lipman 1989; Croson and Marks 2000). Because there is no rebate, contributing more than the threshold would be wasteful to the player making the investment, while contributing less involves not reaching the threshold and thereby forgoing the bonus.

Moving from one to multiple public goods in the static setting, contributions to each of the goods can be treated as separate games so that additional equilibria occur, each of which is a combination of the above equilibria for single goods. The number of equilibria is only constrained by the players' endowments, and this constraint will be more strict as the number of goods increases. Efficient equilibria correspond to those where, given endowment constraints, the sum of net payoffs is maximized. In our experiment, parameters for group size, endowments, and thresholds are selected such that each group can maximally reach two thresholds, which allows us to narrow down efficient equilibria to those where the two goods with the highest net payoff reach their threshold. Finally, moving from a static to a dynamic game allows all static equilibria to be supported in (perfect Bayesian) equilibrium (Marx and Matthews 2000).

2.2. Hypotheses

We now turn to motivating our treatments and formulate hypotheses based on expected game behavior. We compare outcomes between different treatments based on two criteria, measured at the group level. The weaker criterion is the average level of contributions to specific projects. The stronger criterion is the average number of projects for which the threshold is reached. We will refer to both criteria together as treatments being more "successful" (i.e., more thresholds reached and higher contributions) or less.

First, we present our hypothesis on overall game behavior, which combines three hypotheses as formulated by Bagnoli et al. (1992):

Hypothesis 1: (*a*) The thresholds of exactly two goods will be reached; (*b*) there will be no contributions in excess of the goods' thresholds; (*c*) all subjects' contributions will be individually rational, that is, no subject contributes more to a good than the level of its bonus.

Our hypotheses regarding the treatments are:

Hypothesis 2: Project success is lower under Six than under Ben.

Hypothesis 3: Seeded projects under Seed72 and SixSeed78 are more successful than their unseeded counterparts under Ben and Six.

Hypothesis 4: The target project (i.e., 72_{18}) is more successful under Sia72 than under Ben.

Regarding hypothesis 2, the Six treatment features an additional four goods, each of which is dominated by one of the two Ben goods. Increasing the number of available goods is expected to obstruct coordination (Corazzini et al. 2015), even though two goods dominate the other four.

Regarding hypothesis 3, the Seed72 treatment features Good 92_{18}. Compared to Ben Good 72_{18}, this project receives 20 seed tokens, with a concomitant increase in its threshold. Seeding a project is expected to signal that contributions should be made to the seeded good (List and Lucking-Reiley 2002; Rondeau and List 2008; Van de Rijt et al. 2014). Seeding, combined with full refund, is thus expected to simplify coordination on the seeded good, compared to Ben. Similar predictions hold for the seeded Good 98_{17} under SixSeed78 compared to 78_{17} under Six. One difference is, however, that the seeded project under SixSeed78 is a dominated project while this is not the case under Seed72; see table 1. Yet, recall that both seeded projects are the least successful projects in their unseeded counterpart treatments and, as such, both seeded projects clearly do not stand out based on their merit.

Regarding hypothesis 4, the Sia72 treatment features a "seemingly irrelevant alternative," Good 72_{17}, that is dominated by one of the other goods, 72_{18}, for which it serves as a decoy. Adding this decoy is hypothesized to lead to an "attraction effect," a signal that contributions should be made to the dominating good. This effect is described and tested in a large body of research in psychological and marketing science (see, e.g., Ariely and Wallsten 1995).[12] As is the case in standard decoy experiments, 72_{17} should not receive any contributions. Subjects in a group have enough budget to be able to jointly fund two projects, but not three. Obviously, funding (84_{24}, 72_{18}), (84_{24}, 72_{17}), and (72_{18}, 72_{17}) are all (nondegenerate) equilibria of the game and so is funding no projects. As we will show in section 3.1, we observe that in the Ben treatment $84_{24} \succ 72_{18}$, and hence we know that subjects' preference ordering in Sia72 is expected to be $84_{24} \succ 72_{18} \succ 72_{17}$. That means that it is optimal for subjects to coordinate on funding 84_{24} and 72_{18}. Indeed, compared to (72_{17}, 72_{18}) the group returns to coordinating on (84_{24}, 72_{18}) are substantial; for an investment of 12 additional tokens, the group's revenues increase by 42 tokens. In addition, coordinating on (84_{24}, 72_{18}) is not substantially more risky, not just because of the presence of the refund rule, but also because the game is dynamic (as opposed to one shot), so that only a few initial tokens invested in 84_{24} would substantially increase the success probability of that project; see also section 3.3.[13]

Whether adding the decoy is effective in raising contributions to 72_{18} is an open question. Adding the decoy necessarily increases the number of projects put up for funding (from two to three), and hence may actually complicate (rather than facilitate) coordination, resulting either in an increased probability of successfully funding 72_{18} at the detriment of the probability of successfully funding 84_{24}, or even in a decreased probability of successfully funding either of the Ben goods.

3. RESULTS

Analysis of the outcomes of the experiment is complicated because not all subjects actively participated in the experiment. Despite our efforts to ensure full groups of active contributors (as described in sec. 1.3), 44 out of 450 subjects did not log in even once, implying a nonparticipation rate of about 10%. In this section, we present two types of analyses. First, we analyze the behavior of individual subjects. For this participant-level analysis we drop all subjects who logged in on fewer than 4 days. Second, for the project

12. Note that Frederick et al. (2014) found that in less abstract choice situations, the attraction effect may disappear or even lead to a "repulsion effect"—i.e., resulting in an even higher probability of 84_{24} being funded compared to Ben. Since the choice situation in our experiment is characterized by a high level of abstraction, this outcome is not very likely. However, as suggested by one reviewer, it may be worthwhile to consider whether the repulsion effect can be harnessed by adding a project that is dominated by 84_{24}—think of 84_{20}. Whether this would improve coordination is left for future research.

13. If risk were an important consideration, the most preferred Ben project would have been 72_{18} rather than 84_{24}. Since the project with the largest net payoff turned out to be most preferred, it is unlikely that introduction of the decoy would result in subjects preferring 72_{17} over 84_{24}, while in the baseline treatment (Ben) they preferred 84_{24} over 72_{18}. This prediction is supported by the data: there were no groups in which both projects 72_{17} and 72_{18} were successful in terms of threshold reached, and there were also no groups in which these two projects jointly received more contributions than (84_{17}, 72_{18}).

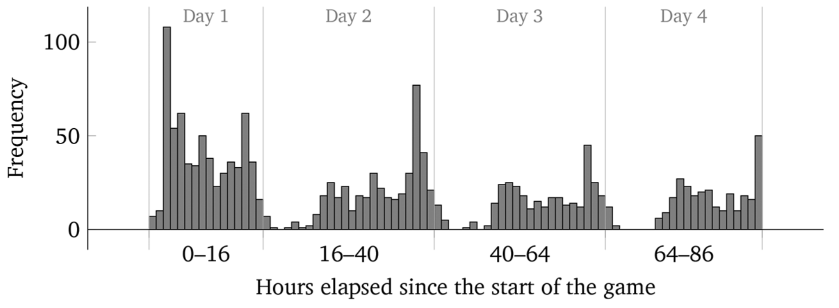

Figure 1. Investment frequency across all treatments (using the data of those 290 subjects who logged in at least once on each consecutive day) in 1-hour bins. Hour 0 corresponds to Thursday 8 a.m. and hour 86 corresponds to Sunday 10 p.m.

outcomes we use group data. Because at least five active group members were needed to be able to reach the thresholds, for the group- and project-level analyses we drop all groups that have two or more members who did not even log in once.[14] This results in dropping eight out of 75 groups: two for Seed72, three for Sia72, and three for Six. Group-level and project-level results and test statistics are provided for the remaining observations (67 groups in total, between 12 and 15 per treatment, contributing to 254 projects in total).

3.1. Overall Game Behavior (Hypothesis 1)

We combine the data of all five treatments to assess hypothesis 1. Before doing so, we first illustrate the size and timing of subjects' contributions in figures 1 and 2. The figures reveal several tendencies that make us confident that subjects took the game seriously, despite the game's length and the relatively low stakes.

Figure 1 shows the frequency of investments made across all treatments in 1-hour bins. Spikes can be observed at the start of the game, and every evening between 9 and 10 p.m. The figure also shows a modest decrease in investment frequency over the 4-day period, which is partly explained by subjects running out of tokens: at the end of the session, 63% of subjects (i.e., 188 of 290 subjects who logged in at least once on each consecutive day) had invested all of their tokens. Given this constraint, there appears to be no decline in interest in logging on as time progresses. Since successfully reaching a single project

14. Nonparticipation complicates the analysis but does not affect the internal validity of our study. Nonparticipation was not more prevalent in some treatments than in others (Kruskal-Wallis $\chi^2 = 0.27$, $p = .99$), and we also found no evidence of specific sociodemographic characteristics being predictive of nonparticipation (results available upon request). Of the remaining 406 subjects, 290 logged in on all 4 days, 51 on 3 days, 37 on 2 days, and 28 on only 1 day. In app. B we further compare game behavior between subjects who logged in each day with those who did not.

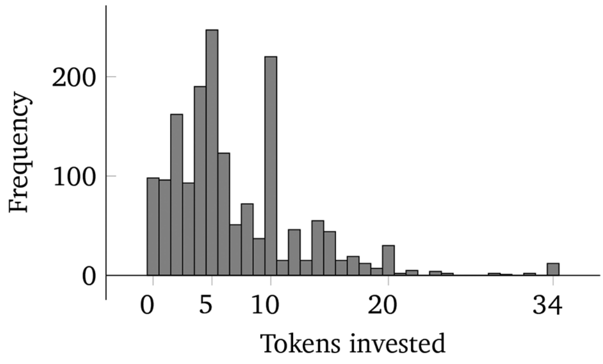

Figure 2. Frequency distribution of the number of tokens invested across all treatments, for all 290 subjects who logged in on each of the session's 4 days.

threshold required substantial contributions by the group, only 7 of the 77 successful projects reached their threshold on the first day. This number increases to 15 on day 2, 19 on day 3, and 36 on day 4. On this last day, 10 out of 36 projects reached their threshold only in the last hour, suggesting that subjects in these groups use cheap-riding strategies.

Figure 2 shows the distribution of tokens invested, averaged across all treatments. Focal contribution levels (multiples of five) are typically more often selected than other levels. The small spike at 34 indicates subjects who went all-in on one project, a seemingly irrational investment decision. The dominance of small contributions, less than five tokens, illustrates that subjects were signaling their interest in a specific project, were trying to cheap ride, or both. We interpret this as another indication that the game was taken seriously by most subjects. We now turn to testing hypothesis 1.

3.1.1. Hypothesis 1

We start with part *a* of hypothesis 1, which states that, independent of the treatment, groups will reach the thresholds of exactly two goods. We find very limited support for hypothesis 1*a*. Only 20 of 67 groups, or 30%, managed to coordinate on this efficient equilibrium. Another 37/67 (55%) reached the threshold of exactly one good, while 10/67 (15%) reached zero thresholds. Recall that outcomes with no or just one successful project may constitute (inefficient) equilibria, as long as the *no deviation constraint* is satisfied; more on this below. Bagnoli et al. (1992) found 48% of the groups in their study coordinating on efficient outcomes. The difference in success rate between their study and ours may be due to our study's field setting or possibly (also) because we do not allow for repetition (implying less scope for learning). The low number of groups that successfully fund two projects suggests that there is scope for improving overall welfare. This is also an important insight for conservation agencies. Given that in many instances cheap riding results in just one project ending up being funded, offering seed money or adding a decoy might allow the conservation agency to at least make sure that its focal project would be the one that is funded—and maybe these policies could even increase the probability of both key projects ending up getting funded. Whether this is indeed the case will be addressed in section 3.2.

Part *b* of hypothesis 1 states that there will be no contributions in excess of the project thresholds. We find moderate support for this part, as illustrated by figure 3. This figure displays the frequencies of project overinvestment (and hence also underinvestment, for negative numbers) across all 254 projects the 67 groups with five or more active group members could have invested in. The 39 projects that make up the spike at 0 were successful without wasting any tokens, implying efficient equilibrium behavior. There are 177 projects to the left of 0. These projects did not reach the threshold. Of these, 16 projects, marked with white bars, violate the *no deviation constraint*. These are projects where a single subject could have made the project successful (considering her remaining tokens, the remaining gap to the threshold, and the bonus upon success). Finally, strict overinvestment occurred in 38 projects. These are projects where one or more subjects displayed seemingly irrational behavior by contributing more than what was required to reach the threshold.

There are several possible explanations for overinvestments. One is the possibility of coordination failures in exactly reaching the threshold, for instance, when two contributions are made (virtually) simultaneously such that feedback on others' contributions was not yet visible on the user interface or possibly was overlooked. The data show that this could have happened on a few occasions at most. Other explanations include calculation errors, better-safe-than-sorry contributions, and the possibility that subjects derive utility other than the financial bonus from making a good successful. All in all, 200 of 254 projects, or 79%, are equilibrium outcomes. Related to part *b* of hypothesis 1,

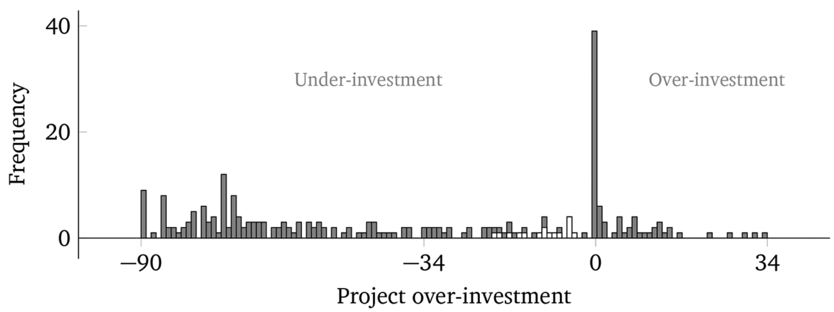

Figure 3. Frequencies of project overinvestment across all treatments, using all 254 projects the 67 groups with five or more active group members could have invested in. White bars mark the 16 projects for which one or more members could have unilaterally made the project successful but failed to do so (i.e., those projects that violated the so-called *no deviation constraint*).

we find that 38 of 254 projects, or 15%, feature contributions in excess of the projects' thresholds.

Part *c* of hypothesis 1 states that no subject contributes more to a good (or project) than the level of its bonus. We find weak support for this part. Only 68% of subjects (i.e., 196 of the 290 subjects who logged in each of the 4 consecutive days) display rational behavior in terms of contributions. However, as with part *b* of hypothesis 1, it is possible that part of the seemingly irrational behavior by the other 32% is driven by alternative explanations. Such explanations receive suggestive support from the contribution behavior of the 94 subjects who contributed more to a good than the level of its bonus. Of these, 60% overcontributed just four tokens or less while 14% overcontributed exactly 10 tokens. As stated above, calculation errors or better-safe-than-sorry contributions may have caused these overcontributions. In appendix A, we assess possible causes of three types of seemingly irrational behavior by linking our experimental outcomes to subjects' contribution behavior as well as to our survey results.

The results on individually rational contribution behavior are also illustrated by figure 4, which presents the distribution of payoffs (measured in tokens) across subjects. Recall that subjects' initial endowments equaled 34 tokens. Payoffs lower than 34 thus indicate that a seemingly irrational investment was made. Small spikes at 18 and 24 include subjects who went all-in on one successful Ben good and received the bonus of, respectively, 18 or 24 tokens. The spike at 34 includes subjects who did not invest or whose investments were not successful and got refunded. The spike at 42 includes subjects who went all-in on two successful Ben goods and received two bonuses that sum up to $18 + 24 = 42$ tokens. All payoffs higher than, approximately, $34 - (72 + 84)/6 + (18 + 24) = 50$ (and 45 for groups of five) include subjects who successfully engaged in cheap riding (or who were lucky with other group members contributing relatively more).

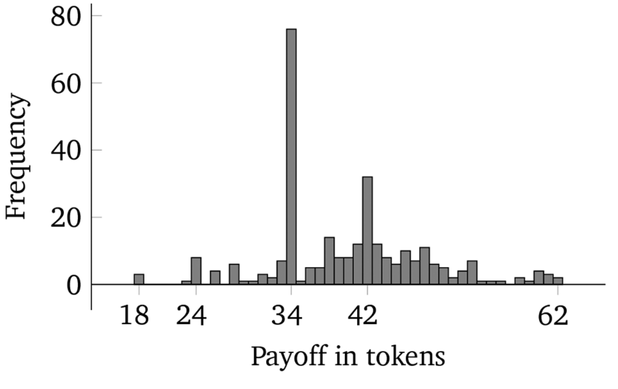

Figure 4. Payoff frequency in tokens across all treatments (using the data of those 290 subjects who logged in at least once on each consecutive day).

Combining the results for the three parts, we only find weak support for hypothesis 1. We do find evidence, however, that subjects took the game seriously, despite the low stakes and despite the length of the game.

3.2. Treatment Results (Hypotheses 2–4)

We now test hypotheses 2–4. We first provide graphical evidence of our treatment impacts in figure 5. The top panel of figure 5 shows the average number of project thresholds reached, while the bottom panel presents the related total contributions to each of the projects. As is clear from the outcomes of Ben, 84_{24} is preferred to 72_{18}. We now turn to testing whether contributions to the least-preferred (but still socially efficient) project can be increased by implementing treatments using seed money and the decoy project targeted at this least-preferred project. In addition, we will test whether or not these treatments resulted in an overall increase in social welfare, depending on whether total contributions were increased or whether the increased contributions to the one project occurred at the expense of contributions to the other.

Regarding the impact of seed money and the decoy in the context of just two projects eligible for funding, figure 5 shows that the differences are not so much in terms of the total number of successful projects but, rather, in the distribution of success over the different projects; compare the outcomes of Seed72 and Sia72 to those of Ben. Seed72 appears to increase the probability of 72_{18} getting funded. As total contributions to the two projects are roughly the same in Seed72 (see the bottom panel), seeding seems effective because it is able to improve coordination—but see the tests below.

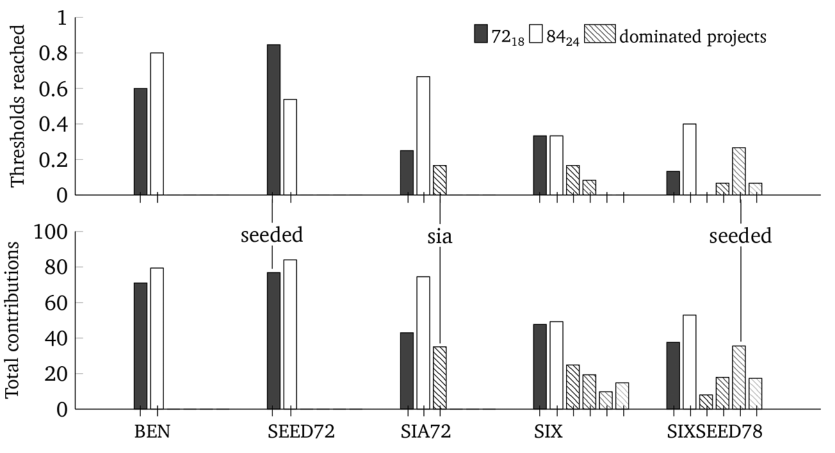

Figure 5. Average number of project thresholds reached (*top panel*) and total contributions (*bottom panel*) across all treatments (using those 67 groups with at least five members who logged in on each of the 4 days), separated by project.

We also find that seeding the public's least-preferred (but possibly the conservation agency's most-preferred) project does not raise overall welfare; the increased coordination on Seed72 results in a substantial decrease in the share of 84_{24} projects funded.

Next, we find that adding the decoy project causes overall contributions to the two Ben projects to decrease, and (surprisingly) more so for 72_{18} than for 84_{24}. Even though just two projects can be funded (implying that, from a welfare perspective, 72_{17} should not receive any contributions), Sia72 reveals that adding the decoy is counterproductive, both in terms of raising contributions to the least-preferred (yet socially efficient) Ben project (i.e., 72_{18}) and in terms of overall funding success. This result indicates that adding a third, strictly dominated, option simply increases the complexity of coordination as it reduces the probability of any project ending up being funded. As such, harnessing the decoy effect is likely to be counterproductive from the conservation agency's perspective.

Comparing Ben and Sia72, increasing the number of projects from two to three thus reduces efficiency, and the outcome is even worse when considering the outcomes of Six. While 84_{24} and 72_{18} continue to receive most contributions in Six, nonnegligible amounts of tokens are allocated to the other four projects. From a conservation agency's perspective, it is thus preferred to just advertise their most preferred conservation projects rather than their entire portfolio of conservation projects. Furthermore, figure 5 suggests that seeding is effective not just in the case of two projects on offer but also when there are six. In other words, if a conservation agency has a preferred project that may not be the public's favorite, seeding is helpful even if the number of projects on offer is large.

We now provide a series of formal statistical tests for the above results. We assess the impact of the various treatments on project success in terms of the number of thresholds reached and the total amount contributed, as well as on welfare; see table 2. To keep welfare comparisons across treatments straightforward, we present individual welfare as measured by received bonuses from successful projects, ignoring impacts from group size and excess contributions. This choice implies that welfare is strongly correlated with the number of thresholds reached; see the first row of table 2. Combined, the first two rows show that while contributions are fairly constant across treatments, coordination failures result in a lower number of thresholds reached for the three treatments with more than two projects as compared to Ben. Mann-Whitney rank-sum tests show that this difference

Table 2. Average Number of Thresholds Reached, Total Contributions, and Individual Welfare across All Treatments (Using Those 67 Groups with at Least Five Members That Logged in on Each of the Four Days)

	Ben	Seed72	Sia72	Six	SixSeed78
Thresholds reached	1.40	1.38	1.08	.92	.93
Total contributions	150.4	160.8	152.6	165.7	169.3
Individual welfare	30	28.2	23.3	18.5	19.5

with Ben is significant for Six ($z = 1.64$, $p = .10$) and SixSeed78 ($z = 2.24$, $p = .03$), but not for Sia72 ($z = 1.06$, $p = .29$). The last row shows that this effect on the number of thresholds reached translates into lower individual welfare. Again, Mann-Whitney rank-sum tests show that this difference with Ben is significant only for Six ($z = 1.83$, $p = .07$) and SixSeed78 ($z = 2.69$, $p = .01$), but not for Sia72 ($z = 1.30$, $p = .20$). Comparing Ben and Seed72, seeding did not significantly affect either the number of thresholds reached ($z = 0.21$, $p = .83$) or the amount of welfare obtained ($z = 0.83$, $p = .40$).

We thus conclude that while conservation agencies may use seeding and decoy projects to influence the choice of projects to be funded, they do not result in a significant increase of either aggregate contributions or of total welfare. In addition, increased complexity, as measured by the number of projects advertised, reduces both the number of successful projects and aggregate welfare. In the remainder of this section, we test hypotheses 2–4 using formal regression analyses.

3.2.1. Hypothesis 2

This hypothesis states that project success is lower under Six than under Ben. In table 3, we present ordinary least squares (OLS) regression estimates of the treatment effects on total contributions as well as on the number of thresholds reached.[15] The dependent variable in models 1 and 3 is total contributions, measured as the sum of contributions to all available projects. The dependent variable in models 2 and 4 is the number of project thresholds reached, measured as the sum of thresholds reached of all available projects. All analyses are implemented at the group level. Recall from section 2 that our parameter selection is such that each group can maximally reach two thresholds. Hence, the value of the dependent variable in these models is 0, 1, or 2. In all four models, we regress the dependent variable on treatment dummies as well as on the number of subjects per group (recall that we excluded all groups with four or fewer active members from the analysis). Models 1 and 2 consider all available projects while models 3 and 4 focus on 84_{24} and 72_{18} (the two Ben projects) only, since these are the two dominant projects across all treatments.

First, consider the outcomes of models 2 and 4. In these models, the sign and significance of the coefficients for the Six dummy indicate strong support for hypothesis 2, in terms of the number of thresholds reached. On average, groups under Six reach 0.5 fewer project thresholds than groups under Ben (see model 2). Model 3 shows that the total number of tokens contributed to the two key projects, 72_{18} and 84_{24}, is significantly smaller under Six than under Ben. Model 1—regressing total contributions to all available projects at the group level—is the exception, as here the coefficient on Six is positive. While surprising, it is important to stress that this did not result in more

15. All our results are robust to using alternative regression models, including (ordered) probit and Poisson models (with bootstrapping); results are available upon request.

Table 3. Project Success as Measured by Total Contributions and Number of Thresholds Reached

	All Projects		Ben Projects Only	
	Contributions (1)	Thresholds (2)	Contributions (3)	Thresholds (4)
Seed72	3.180	−.143	6.733	−.108
	(7.694)	(.225)	(8.696)	(.206)
Sia72	.643	−.344	−33.69***	−.503**
	(7.772)	(.227)	(8.785)	(.208)
Six	13.73*	−.510**	−54.27***	−.753***
	(7.772)	(.227)	(8.785)	(.208)
SixSeed78	20.92***	−.431**	−58.82***	−.840***
	(7.332)	(.214)	(8.287)	(.196)
No. of subjects/group	30.80***	.540***	15.74***	.393***
	(5.077)	(.148)	(5.738)	(.136)
Constant	−20.03	−1.585*	63.30*	−.776
	(28.57)	(.834)	(32.29)	(.765)
Observations	67	67	67	67

Note. Coefficient estimates from OLS regression models; standard errors in parentheses. All models are at the group level (using those 67 groups with at least five members who logged in on each of the 4 days). Contributions in models 1 and 3 equal the sum of contributions to all available projects (1) or Ben projects only (3). Similarly, thresholds in models 2 and 4 equal the sum of thresholds reached. Note that Ben is the omitted treatment category.

* $p < .10$.
** $p < .05$.
*** $p < .01$.

projects ending up being successful, as evidenced by model 2. Having more projects on offer thus increases total contributions but decreases the number of projects for which the threshold is met. In addition, the number of subjects per group is positively related to project success. We find that having a sixth group member makes a substantial difference in total contributions per group, which translates into 0.5 more project thresholds reached (see model 2).

The parametric results of table 3 are confirmed by Mann-Whitney rank-sum tests. Based on all available projects, total contributions under Six only just fail to be significantly higher compared to Ben ($z = -1.59$, $p = .11$), while the number of thresholds reached is lower under Six compared to Ben ($z = 1.64$, $p = .10$). Focusing on the two Ben projects, total contributions are lower under Six compared to Ben ($z = 4.05$, $p = .00$), while the number of thresholds reached is also lower under Six compared to Ben ($z = 2.83$, $p = .00$).

Going beyond hypothesis 2, table 3 shows that identical results occur for the other two treatments with more than two projects, Sia72 and SixSeed78. This result confirms the severity of coordination problem that occurs when multiple projects are available (Corazzini et al. 2015). Assessing Sia72, Six, and SixSeed78 combined, increasing the number of projects increases total contributions but decreases contributions to the two dominating projects. This dispersion of contributions leads to a decrease in the number of thresholds reached. Coordination problems lead to inefficiencies not only by decreasing the total number of thresholds reached but also by diverting investments away from the dominating projects.

3.2.2. Hypothesis 3

This hypothesis states that seeded projects under Seed72 and SixSeed78 are more successful than their unseeded counterparts in Ben and Six. We assess this hypothesis using linear regressions of project success focused on the target project. Recall from table 1 that the target project for Seed72 is project 72_{18} (or 92_{18} when seeded) and the target project for SixSeed78 is project 78_{17} (or 98_{17} when seeded). The results of the regression analyses are presented in table 4. Since the target project under Seed72 and Sia72 is identical, we also include a treatment dummy for Sia72 in the regressions of table 4 and refer

Table 4. Treatment Impacts on Project Success of the Target Project

	Project 72_{18}		Project 78_{17}
	Contributions (1)	Thresholds (2)	Contributions (3)
Seed72	5.937	.220	
	(6.751)	(.176)	
Sia72	−27.98***	−.356*	
	(6.762)	(.176)	
SixSeed78			26.38**
			(10.57)
No. of subjects/group	−.386	.110	5.115
	(5.828)	(.152)	(10.51)
Constant	73.14**	−.0106	−18.81
	(32.56)	(.848)	(59.20)
Observations	40	40	27

Note. Coefficient estimates from OLS regression models; standard errors in parentheses. All models are at the group level: 40 groups in models 1 and 2 based on target project 72_{18} (or 92_{18} when seeded) and 27 groups in model 3 based on target project 78_{17} (or 98_{17} when seeded), making 67 groups in total (all those groups with at least five members who logged in on each of the 4 days). Contributions in models 1 and 3 equal the sum of contributions to the target project. Thresholds in model 2 equal the sum of thresholds reached of the target project. Ben is the omitted reference treatment in models 1 and 2 and Six in model 3.

* $p < .10$.

** $p < .05$.

*** $p < .01$.

to the same table when assessing hypothesis 4 below. Total contributions and a dummy for threshold of the target project reached are regressed on the relevant treatment dummies as well as on the number of subjects per group, with Ben as the omitted reference treatment for models 1 and 2 and Six as the omitted reference treatment for model 3. Since SixSeed78 is a perfect predictor of the number of thresholds reached (see the top panel of fig. 5), we do not report the outcomes of the regression explaining the number of thresholds met for this treatment.

The coefficient on SixSeed78 in model 3 is both positive and highly significant. Together with the fact that SixSeed78 is a perfect predictor of successfully reaching the threshold, we thus find strong support for hypothesis 3. The coefficients for the Seed72 dummy in models 1 and 2 are, however, not significant, suggesting that seeding does not result in higher success of the seeded project if the number of projects offered is small.[16]

The results of table 4 are also confirmed by nonparametric Mann-Whitney rank-sum tests. For the seeded good under Seed72, we find that total contributions to 92_{18} under Seed72 do not statistically differ from those to 72_{18} under Ben ($z = -1.33$, $p = .18$) and the same holds for the number of thresholds reached ($z = -1.41$, $p = .16$). For the seeded good under SixSeed78, we find that total contributions to 98_{17} are higher than those to 78_{17} under Six ($z = -2.75$, $p = .01$) and the same holds for the number of thresholds reached ($z = -1.90$, $p = .06$). Combined, we find moderate support for a seeding effect: weaker under Seed72 and stronger under SixSeed78.

Seeding appears to work with six available projects but less so with two, despite the seeded project being dominated under SixSeed78 but not under Seed72. A straightforward explanation for this difference is that the multitude of available projects and their similarity in terms of characteristics (see table 1) may trigger boundedly rational behavior. If so, subjects may pay less or no attention to specific features of the problem (Gabaix 2014). Under these conditions, choosing a seeded project is an easy default (Carroll et al. 2009). Such default behavior is not induced when the decision problem is less complex, such as when only two projects have to be compared. In that case, subjects can make a fully informed comparison of potential payoffs. In line with results found by Corazzini et al. (2015), in doing so they ignore the possible coordination benefits offered by focusing on the seeded project. Subjects in our sample have a slight preference for the more efficient project 84_{24} (see the Ben results in fig. 5), which was the reason for seeding the less efficient project 72_{18} under Seed72. Because the seeded project is still the least efficient and because there are enough endowments to fund both, subjects may have paid less attention to the seeding.

3.2.3. Hypothesis 4

This hypothesis states that the targeted Good 72_{18} is more successful under Sia72 than under Ben. Again, we can use the regressions presented in table 4. The sign and

16. These results are robust to alternative specifications of the dependent variable; see the online appendix.

significance of the coefficient on Sia72 in models 1 and 2 of table 4 indicate a negative effect of adding a seemingly irrelevant project on project success of project 72_{18}. Similar to the impact of seeding, this result is robust to alternative specifications of the dependent variable (see the online appendix) as well as to the use of nonparametric Mann-Whitney rank-sum tests. The number of thresholds reached for 72_{18} under Sia72 is lower than under Ben ($z = 1.79$, $p = .07$), and the same holds for total contributions ($z = 2.85$, $p = .00$). Overall we find a strong negative effect of adding the seemingly irrelevant project; adding a third (and dominated) project complicates coordination without generating an attraction effect to the targeted project. This effect points, again, to the severity of coordination problems that occur when the number of available projects increases. Combined, we find no support for hypothesis 4.

Summarizing our results, we find that (*a*) adding projects strongly decreases overall project success, even when the added projects are dominated; (*b*) providing a project with seed money may strongly increase project success (but not unequivocally); and (*c*) adding a seemingly irrelevant project that serves as a decoy has a strong effect but not in the expected direction: the coordination problem worsens whereas an attraction effect was expected. This is a caution for conservation agencies considering using crowdfunding to finance their projects to put forward only their key projects, but also that seeding can be used to affect the probability with which individual projects are funded.

3.3. Contribution Dynamics

Having documented how the various crowdfunding designs affect project success, we now examine the role of early contributions. If we find that a specific design is conducive to project success, is it because it (among others) increases early contributions, which subsequently result in higher project success rates? In this section, we exploit variation in the amount of early contributions to answer this question.

Evidence that early contributions have a substantial effect on eventual project success is shown in figure 6. In this figure, "early contributions" are measured as the sum of contributions made to a project in the first 5 hours, 10 hours, or 15 hours of the 4-day session. They comprise, respectively, 19%, 32%, and 47% of total contributions over the course of the experiment. We find a strong positive correlation between early contributions and project success. To test whether there is a causal impact of early contributions on project success, we estimate a two-stage least squares model. We implement the following identification strategy. First, we use project and treatment characteristics as instrumental variables to generate exogenous variation in early contributions in the various projects. Next, we use the predicted value of early contributions to explain project success.

Key to any instrumental variable analysis is whether the instrumental variables affect the outcome variable of interest (in our case project success) directly, or only indirectly via the potentially endogenous variable (in our case early contributions). If we cannot reject

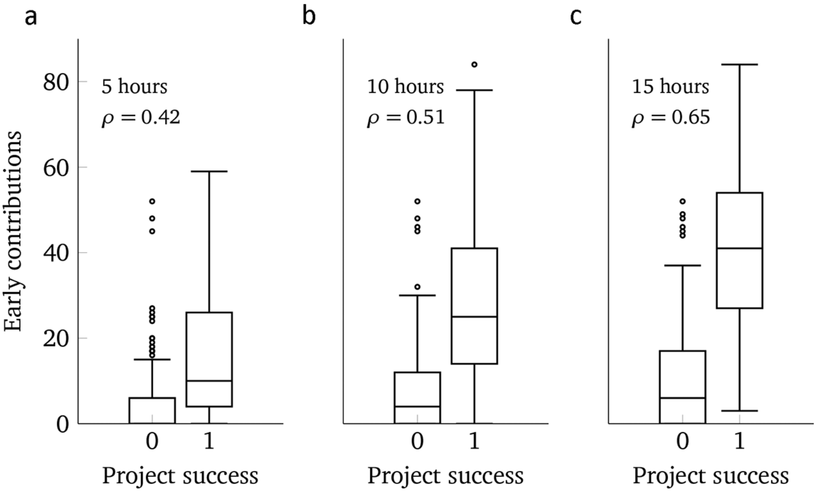

Figure 6. Box plots of early contributions against project success for all 254 projects (all those projects eligible for funding in the 67 groups of which at least five members logged in on each of the 4 days), as well as the related Spearman's ρ, for three different measures of "early."

the null of no direct influence, we conclude that project and treatment characteristics have a differential impact on early contributions and that higher early contributions have a strong impact on ultimate project success. If, however, we reject the null of no direct influence, we learn that project and treatment characteristics influence project success via mechanisms other than just their impact on early contributions.[17]

Using observations from Ben, Seed72, Six, and SixSeed78, we have exogenous variation in (1) the bonus/threshold ratio, (2) the number of projects in the treatment, and (3) whether the project is seeded. All three characteristics are expected to affect early contributions. First, the bonus/threshold ratio is a measure of project efficiency, and we expect more efficient projects to attract more early contributions. Second, with six projects to choose from, coordination is more difficult, and in this case we expect fewer early contributions (as subjects may decide to wait longer). Third, allocating seed money to a specific project is expected to facilitate coordination and may hence increase early contributions. We instrument early contributions by the bonus/threshold ratio and dummies for the number of projects in the treatment (Many = 1 for treatments with six projects) and whether the project is seeded (Seed = 1), as well as their interaction. The results of the two-stage least squares (2SLS) model are presented in table 5.

17. We admit that this test is less clean than generating differential early contribution rates using an additional treatment. The analysis in this subsection should thus be viewed as exploratory.

Table 5. Impact of Early Contributions on Project Success, Measured as the Number of Thresholds Reached

	First 5 Hours (1)	First 10 Hours (2)	First 15 Hours (3)
Instrumented early contributions	.0399***	.0299***	.0206***
	(.00615)	(.00407)	(.00220)
Constant	−.00717	−.102**	−.104***
	(.0427)	(.0519)	(.0378)
First stage (early contributions):			
Many	−12.08***	−14.51***	−21.46***
	(2.586)	(3.395)	(3.793)
Seed	6.647	14.71**	19.23***
	(4.898)	(6.090)	(6.792)
Many × Seed	−4.558	−6.314	−2.632
	(4.650)	(5.792)	(6.535)
Bonus/threshold	38.38	99.64*	179.9***
	(29.44)	(50.92)	(58.24)
Constant	6.537	−1.972	−11.29
	(8.316)	(14.01)	(16.05)
Observations	218	218	218

Note. Coefficient estimates from 2SLS regression models for three cut-off points for "early" contributions (and robust standard errors in parentheses). All models are at the project level, excluding projects under treatment Sia72 (i.e., 218 projects). Many and Seed represent dummy variables referring to whether a project is in a treatment with six projects (Many = 1) and whether a project is seeded (Seed = 1).

* $p < .10$.

** $p < .05$.

*** $p < .01$.

We document the following results. First, we find that having competition from many projects significantly decreases early contributions, while both the bonus/threshold ratio and seeding significantly increase early contributions—after more than 5 hours have passed. Second, we find that the project characteristics are jointly highly significant as predictors of early contributions. The associated F-test has a value of 12.21 for 5 hours, 15.04 for 10 hours, and 25.82 for 15 hours (each with $p = .00$). The values of the Kleibergen-Paap statistics are very high too: $\chi^2 = 31.55$, for 5 hours, $\chi^2 = 32.42$, for 10 hours, and $\chi^2 = 47.71$, for 15 hours (each with $p = .00$). Next, as shown in the second-stage regressions, early contributions positively affect project success. The coefficients for instrumented early contributions should be interpreted as the change in probability of project success associated with one additional token contributed early. Hence, the impact of early contributions in the first 5 hours, increasing the probability of project success by 4%, is substantial. The Hansen-J tests are all insignificant ($p = .23$ for 5 hours, $p = .86$ for 10 hours, and $p = .82$ for 15 hours), suggesting that, indeed, treatments have a differential impact on project success through their differential effect on early contributions.[18]

In sum, we find that project success is driven by early contributions. Projects with characteristics that induce early contributions, such as seeded projects, are therefore more likely to ultimately become successful.

4. A CONSERVATION EXPERIMENT

To verify our results in a setting (even) closer to the field, we conducted a follow-up framed field experiment. We teamed up with Natuurmonumenten, the Dutch Society for Nature Conservation (see http://www.natuurmonumenten.nl/), and ran three more sessions in which we replaced the abstract projects from our core experiment with actual nature conservation projects. Upon reaching a project threshold, the bonus of the project was not paid out in tokens to the group members but rather the tokens were transferred to money and paid out to the charity project. Hence, we replaced the local public good in the core experiment—from which only group members benefit—to a global public good that benefits society as a whole (Blackwell and McKee 2003). The projects' environmental characteristics were communicated in the form of a "nature development score," a proxy for improvements in biodiversity realized in the respective conservation projects. To avoid level effects, we scaled the nature development scores such that they were identical to the values of the bonuses in the benchmark treatment of the core experiment: 18 and 24, respectively. While the environmental impacts were thus described using this simple metric, subjects themselves needed to decide how to evaluate those impacts. After all, some may attach higher (use and/or nonuse) values to a specific conservation outcome than others. While the lab-in-the-field experiment afforded us control of both benefits and costs of donors' contributions, in this follow-up experiment we can only control costs.

The conservation projects were described as "stimulating the development of new nature and contributing to the overall quality of nature in the Netherlands." The two conservation projects, Soesterveen and Vlijmers Ven, were selected by Natuurmonumenten as two projects that were similar in terms of their characteristics except, importantly, for their contribution to biodiversity, with Soesterveen scoring 33% higher, consistent with the difference in bonus between the two Ben projects. In addition, costs for development in Soesterveen were higher, consistent with the difference in thresholds between projects. To keep differences in instructions and user interface as small as possible, the conservation projects were described in general terms, without revealing their name or location, while parameter values for projects' threshold and bonus were kept constant between

18. These results are robust to (i) using a probit regression in the second stage; (ii) adding "subjects per group" in the first stage; (iii) replacing the dependent variable "thresholds reached" by the alternative measure of project success as used in sec. 3.2, "Total Contributions"; and (iv) using disjoint intervals 0–5, 5–10, and 10–15 hours. All results are available upon request.

the two experiments, with the bonus now referring to the "nature development score" (see the online appendix).

Three treatments were repeated in this conservation experiment: Ben, Seed72, and Sia72. The two treatments with six projects could not be rerun because of a lack of ready-to-be-implemented nature conservation projects. Sampling and procedures were identical to those in the core experiment. The experiment took place in three sessions in the period of May–June 2017. The conservation experiment had a lower share of active subjects; 224 of the 270 subjects logged in at least once. As a result, 10 out of 45 groups had two or more subjects who did not log in even once, and these groups were dropped from the analysis: four for Ben, four for Seed72, and two for Sia72.

Overall game behavior was similar to that in our core experiment, albeit that funding success is lower, see the online appendix. This is probably due to the fact that the bonus is now truly an environmental return that also benefits nonparticipants, rather than a monetary reward paid out to each member of the group. The top panel of figure 7 shows the average number of thresholds reached per group across all treatments, while the bottom panel shows the related average total contributions per group across treatments. Comparing figure 5 and figure 7, total contributions in the conservation experiment are slightly lower (on average −17%, Mann-Whitney $z = 3.55$, $p = .00$) while the number of thresholds reached is much lower (on average −58%, Mann-Whitney $z = 4.29$, $p = .00$).

We reran models 1 and 2 of table 4 for the conservation experiment; see table 6. We analyze the treatment effects on total contributions and number of thresholds reached, focusing on project 72_{18} (or 92_{18} when seeded), the target project for Seed72 and Sia72. We find qualitatively similar results of project success in the conservation experiment. Seeding is found to be ineffective in raising contributions (see the coefficient on Seed72 in model 1, and the coefficient in model 2 suggests that it is even counterproductive—albeit that this result is only weakly significant). Sign and significance of the coefficient for the Sia72 dummy in models 1 and 2 indicate a negative effect of adding a seemingly irrelevant project on project success of project 72_{18}, confirming our result for hypothesis 4 of our core experiment. These results are again robust to alternative specifications of the dependent variable; see the online appendix. Overall, the conservation experiment confirms the results of our core experiment in terms of general investment behavior, the impact of seeding and adding a seemingly irrelevant project.

5. CONCLUSIONS

In this study, we assess the impact of crowdfunding design on the success of crowdfunded public goods using both a lab-in-the-field experiment and a framed field experiment. Specifically, we analyze whether mechanisms that signal focal projects can be used to increase project success. We do so since private citizens, NGOs, and governments are increasingly using crowdfunding as a conservation funding tool, while results also apply to other public goods. We find that (*a*) coordination is worse the larger the number of

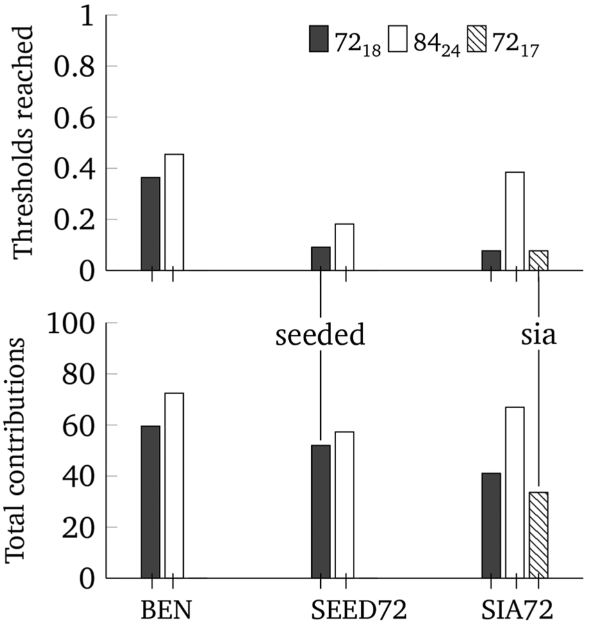

Figure 7. Results of the conservation experiment: Average number of thresholds reached (*top panel*) and total contributions (*bottom panel*) across all treatments (i.e., 35 groups), separated by project.

projects advertised, so that adding additional projects decreases the target projects' success rates, even when the added projects are dominated; (*b*) providing a project with seed money may strongly increase project success (but not unequivocally); and (*c*) adding a seemingly irrelevant project that serves as a decoy has a strong effect, but not in the expected direction: coordination problems worsen whereas an attraction effect was expected.

Mechanisms such as seed money and seemingly irrelevant projects may be used to mitigate coordination problems in civic crowdfunding. Our results demonstrate, however, that the signals given by such mechanism may not work intuitively or may only work when coordination is particularly difficult. Hence it is not straightforward to employ these mechanisms to mitigate coordination failure and cheap riding. This result is relevant both for the crowdfunding platforms that seek to maximize success rates and for fundraisers that aim to amass investments to their own project. Regarding contribution dynamics, we find that project success is driven by early contributions and that our treatments affect project success through their differential impact on early contributions.

Taken together, our results partly confirm earlier lab findings and point to the scope for design alternatives that potentially mitigate coordination problems in the crowdfunding of conservation and other public goods. Key decisions involve the number of projects

Table 6. Results of the Conservation Experiment: Treatment Impacts on Project Success of the Target Project

	Project 72_{18}	
	Contributions (1)	Thresholds (2)
Seed72	−7.545	−.273*
	(8.491)	(.147)
Sia72	−20.30**	−.321**
	(8.199)	(.142)
No. of subjects/group	16.39**	.303**
	(7.301)	(.127)
Constant	−26.89	−1.232*
	(38.96)	(.676)
Observations	35	35

Note. Coefficient estimates from OLS regression models (and standard errors in parentheses). All models are at the group level: 35 groups with target project 72_{18} (or 92_{18} when seeded). Contributions in model 1 equal the sum of contributions to the target project. Thresholds in model 2 equal the sum of thresholds reached of the target project. Treatment names represent dummy variables, with Ben as the omitted reference treatment.

* $p < .10$.

** $p < .05$.

*** $p < .01$.

to put forward simultaneously and the choice for a coordination mechanism to signal focal projects. Our results show that the number of projects should be kept low, while seeding can be used to affect the shares of contributions received by specific projects without reducing overall donations.

APPENDIX A
SEEMINGLY IRRATIONAL BEHAVIOR

We assess possible causes of three types of seemingly irrational behavior by linking our experimental outcomes to subjects' contribution behavior as well as to the survey results. We distinguish three types of seemingly irrational behavior (see also sec. 3.1): (i) individual investments in a project that exceed its bonus (94 subjects), (ii) individual investments beyond the project threshold (37 subjects), and (iii) violation of the no deviation constraint (13 subjects). We test for differences between two subsamples: subjects who made one or more of these seemingly irrational decisions and subjects who did not. Results are reported in table A1.

Table A1. Mean Values by Seemingly Irrational Behavior across All Treatments (Using the Data of Those 290 Subjects Who Logged in at Least Once on Each Consecutive Day)

	Rational (Mean)	Irrational (Mean)	Mann-Whitney (*p*-Value)
Age (years)	49.51	55.39	.00
Education (8 category scale)	4.82	4.60	.16
Gender (male = 0, female = 1)	.49	.47	.77
Household income (20 category scale)	14.40	14.64	.80
No. of logins	6.78	6.07	.49
I give to society (1–5 Likert scale)	3.33	3.48	.29
I put family first (1–5 Likert scale)	3.60	3.57	.61
I find money decisions hard (1–5 Likert scale)	2.63	2.56	.70
Goal: Fair Play (1–5 Likert scale)	3.44	3.59	.29
Goal: Earn Tokens (1–5 Likert scale)	4.11	3.87	.01
Goal: Do as Others (1–5 Likert scale)	2.23	2.36	.38
Goal: Two Projects (1–5 Likert scale)	3.99	3.84	.19
Game: Instructions Not Clear (1–5 Likert scale)	2.12	2.46	.00
Game: Easy to Play (1–4 Likert scale)	3.23	2.96	.00
Game: Wanted to Play More (no = 0, yes = 1)	.81	.77	.45

Note. Mean values of scores on selected survey items separated by whether or not subjects have displayed any of the three types of seemingly irrational behavior as described in the main text. 1–5 Likert scale items range from 1 = completely disagree to 5 = completely agree. Note that "Game: Easy to Play" was measured using a 1–4 Likert scale ranging from 1 = disagree to 4 = agree. The third column displays the *p*-value of the associated Mann-Whitney rank-sum tests.

The relevant Mann-Whitney tests indicate that the two subsamples differ in four variables: Age, Goal: Earn Tokens, Game: Instructions Not Clear, and Game: Easy to Play. Subjects displaying seemingly irrational behavior were older (by approximately 6 years), were less focused on earning tokens, and had more difficulties in understanding the game instructions and playing the game. Note that the last three variables were measured using Likert scales, and although these differences are intuitive and statistically significant, the ordinal nature of Likert scales makes their economic impact, roughly 0.3-point differences in means on a 4- or 5-point scale, hard to interpret. Importantly, the two subsamples do not differ in terms of Education nor the remaining survey items on game behavior and attitude. To assess the possible impact of subjects who did not fully comprehend the game on group behavior, we repeated all regressions of tables 4–6, while controlling for the number of subjects in each group that scored a 4 or 5 on Game: Instructions Not Clear. These robustness checks, available upon request, did not reveal any substantive differences with the results as reported in the paper. Overall, our interpretation of these results is that there is no strong evidence of a systematic cause of seemingly irrational behavior in our experiment.

Importantly, compared to a laboratory experiment, the reduced level of control in our lab-in-the-field setting implies that we risked that imperfect understanding of the game's design could affect our results. As explained in section 1, we were very careful in removing any source of confusion or misunderstanding. We argue, however, that any remaining misunderstandings are unlikely to have affected the (internal) validity of our experiment. There is no a priori reason to expect these scores to differ between the treatments. We tested for such differences in game understanding based on the scores on Game: Instructions Not Clear and find no differences (Kruskal-Wallis $\chi^2 = 5.75$, $p = .21$). While imperfect understanding is likely to have increased the noise in our experimental outcomes, we find no evidence for this to have affected our treatment differences.

Finally, one may wonder whether imperfect participation (see app. B) has an impact on seemingly irrational behavior. In the full sample of 290 subjects who logged in on each day, the percentage who behaved irrationally (i.e., one or more of three types of irrational behavior), was 42%. If we exclude all subjects from all groups where not every member logged in each day, we find that 38% of the remaining sample behaved irrationally. Irrational behavior is thus by and large unaffected by other group members logging in less.

APPENDIX B
FULL VERSUS INCOMPLETE PARTICIPATION

Despite all our efforts and design choices aimed at inducing all subjects to actively participate in the experiment, not all subjects logged in on each of the 4 days of the experiment. This is a potential concern because solving a coordination problem requires subjects' monitoring their fellow group members' behavior—and this holds especially true for our dynamic game.

Table B1 shows the breakdown of subjects by the number of separate days they logged in on our platform. Of the 450 participants, 290 logged in on each of the 4 days the experiment lasted; those who did not are by and large equally distributed over having participated 0, 1, 2, or 3 days.

Table B1. Frequency of Days Logged in across All Treatments (Using the Data of All 450 Subjects in Our Experiment)

Number of Separate Login Days	Frequency
0	44
1	28
2	37
3	51
4	290

Of course we can only speculate about the mechanism that gave rise to this participation pattern, and hence we also looked at whether there are important differences in behavior between those who logged in every day, and those who failed to do so on one or more days. For brevity, we will refer to these groups as, respectively, 4 Day subjects and 1–3 Day subjects. More specifically, we looked at differences in (i) the number of times subjects made a contribution, (iii) the total amount contributed, and (iii) the total number of tokens received by the subjects. The results are presented in table B2.

Table B2. Mean Values by Login Days across All Treatments (Using the Data of All 406 Subjects Who Logged in at Least Once)

	1–3 Days (Mean)	4 Days (Mean)	Mann-Whitney (p-Value)
Times contributed	2.93	4.61	.00
Total amount contributed	25.78	29.72	.02
Total no. tokens	40.65	39.80	.76

We find that 4 Day subjects contributed more often (4.61 vs. 2.93 times) and contributed a larger number of tokens (29.7 vs. 25.8) than 1–3 Day subjects. While these differences are statistically significant, in absolute terms they are not very large, and there is also no significant difference in the total number of tokens earned. In fact, from this table we derive that 1–3 Day subjects contributed more per donation (8.80 tokens on average) compared to 4 Day participants (6.45 tokens). Combined with the fact that the share of 1–3 Day subjects does not really vary with the number of days they logged in (as shown in table B1), this contribution pattern suggests that 1–3 Day subjects may have reasoned that they made their fair share of contributions already and were less eager to monitor how play developed over the remainder of the game, ignoring the implications for their payoff.

Imperfect participation is a concern for the assessment of the absolute effectiveness of the various mechanisms. From an experimental perspective, the treatment differences are more important; if we find no significant differences in nonparticipation rates across our treatments, nonparticipation may have negatively impacted the number of times thresholds were met but are unlikely to have negatively affected the internal validity of our experiment. We implemented three types of tests. First, as already stated in section 3, we did not find any significant differences in sociodemographic characteristics between treatments among the 1–3 Day subjects. Second, we also tested for differences in the overall imperfect participation distributions between the five treatments, but we failed to find any significant results (Kruskal-Wallis $\chi^2 = 2.94$, $p = .57$). Nonparticipation is thus not more prevalent in some treatments than in others. Third, we also tested whether the average numbers presented in table B2 hide important differences between treatments. Again, we find no evidence for this either (the respective test outcomes are Kruskal-Wallis

$\chi^2 = 0.78$, 0.27, and 0.08, with p-values of respectively .38, .60, and .78). Hence, we do not find any evidence of treatment-induced differences in either the pattern or in the amounts and frequencies subjects contributed.

As a final robustness check, we reran the regressions reported in tables 3 and 4 using only those groups that consist of only 4 Day subjects. That means that the number of groups falls from 67 (those with at least five members having logged in on each consecutive day) to just 26. When rerunning tables 3 and 4, the results obtained are by and large unaffected, at least not in terms of the sign and magnitude of the coefficients. However, standard errors are about 50% larger because of the smaller number of observations, and hence some of the coefficients become insignificant. In terms of reliability, the estimates are more precise with the weaker participation criterion (i.e., at least five members having logged in on each of the 4 days), while the nonparticipation rates do not differ between treatments. For this reason we decided to focus our analysis on the sample of groups with five or more active members; results for the stricter definition of participation are available upon request.

REFERENCES

Agrawal, Ajay, Christian Catalini, and Avi Goldfarb. 2014. Some simple economics of crowdfunding. In *Innovation policy and the economy*, ed. J. Lerner and S. Stern. Chicago: University of Chicago Press.

Alekseev, Aleksandr, Gary Charness, and Uri Gneezy. 2017. Experimental methods: When and why contextual instructions are important. *Journal of Economic Behavior and Organization* 134:48–59.

Andreoni, James. 1998. Toward a theory of charitable fund-raising. *Journal of Political Economy* 106 (6): 1186–1213.

Ariely, Dan, and Thomas S. Wallsten. 1995. Seeking subjective dominance in multidimensional space: An explanation of the asymmetric dominance effect. *Organizational Behavior and Human Decision Processes* 63 (3): 223–32.

Bagnoli, Mark, Shaul Ben-David, and Michael McKee. 1992. Voluntary provision of public goods: The multiple unit case. *Journal of Public Economics* 47 (1): 85–106.

Bagnoli, Mark, and Barton L. Lipman. 1989. Provision of public goods: Fully implementing the core through private contributions. *Review of Economic Studies* 56 (4): 583–601.

Belleflamme, Paul, Nessrine Omrani, and Martin Peitz. 2015. The economics of crowdfunding platforms. *Information Economics and Policy* 33:11–28.

Bigoni, Maria, Marco Casari, Andrzej Skrzypacz, and Giancarlo Spagnolo. 2015. Time horizon and cooperation in continuous time. *Econometrica* 83 (2): 587–616.

Blackwell, Calvin, and Michael McKee. 2003. Only for my own neighborhood? Preferences and voluntary provision of local and global public goods. *Journal of Economic Behavior and Organization* 52 (1): 115–31.

Bøg, Martin, Heike Harmgart, Steffen Huck, and Abra M. Jeffers. 2012. Fundraising on the internet. *Kyklos* 65 (1): 18–30.

Cadsby, Charles B., and Elizabeth Maynes. 1999. Voluntary provision of threshold public goods with continuous contributions: Experimental evidence. *Journal of Public Economics* 71 (1): 53–73.

Carroll, Gabriel D., James J. Choi, David Laibson, Brigitte C. Madrian, and Andrew Metrick. 2009. Optimal defaults and active decisions. *Quarterly Journal of Economics* 124 (4): 1639–74.

Cason, Timothy N., and Robertas Zubrickas. 2017. Enhancing fundraising with refund bonuses. *Games and Economic Behavior* 101:218–33.

———. 2019. Donation-based crowdfunding with refund bonuses. *European Economic Review* 119:452–71.

Charbit, Claire, and Guillaume Desmoulins. 2017. Civic crowdfunding: A collective option for local public goods? OECD Regional Development Working paper 2017/02, Organisation for Economic Cooperation and Development, Paris.

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

Corazzini, Luca, Christopher Cotton, and Paola Valbonesi. 2015. Donor coordination in project funding: Evidence from a threshold public goods experiment. *Journal of Public Economics* 128:16–29.

Crosetto, Paolo, and Tobias Regner. 2014. Crowdfunding: Determinants of success and funding dynamics. Jena Economic Research Papers 2014-035, Friedrich-Schiller-University Jena.

Croson, Rachel T. A., and Melanie B. Marks. 2000. Step returns in threshold public goods: A meta- and experimental analysis. *Experimental Economics* 2 (3): 239–59.

Devaraj, Srikant, and Pankaj C. Patel. 2016. Influence of number of backers, goal amount, and project duration on meeting funding goals of crowdfunding projects. *Economics Bulletin* 36 (2): 1242–49.

Dorsey, Robert E. 1992. The voluntary contributions mechanism with real time revisions. *Public Choice* 73 (3): 261–82.

Duffy, John, Jack Ochs, and Lise Vesterlund. 2007. Giving little by little: Dynamic voluntary contribution games. *Journal of Public Economics* 91 (9): 1708–30.

Frederick, Shane, Leonard Lee, and Ernest Baskin. 2014. The limits of attraction. *Journal of Marketing Research* 51 (4): 487–507.

Gabaix, Xavier. 2014. A sparsity-based model of bounded rationality. *Quarterly Journal of Economics* 129 (4): 1661–1710.

Gallo-Cajiao, Eduardo, Carla Archibald, Rachel Friedman, Rochelle Steven, Richard A. Fuller, Edward T. Game, Tiffany H. Morrison, and Euan G. Ritchie. 2018. Crowdfunding biodiversity conservation. *Conservation Biology* 32 (6): 1426–35.

Henrich, Joseph, Steven J. Heine, and Ara Norenzayan. 2010. The weirdest people in the world? *Behavioral and Brain Sciences* 33 (2–3): 61–83.

Hudik, Marek, and Robert Chovanculiak. 2018. Private provision of public goods via crowdfunding. *Journal of Institutional Economics* 14 (1): 23–44.

Isaac, R. Mark, David Schmidtz, and James M. Walker. 1989. The assurance problem in a laboratory market. *Public Choice* 62 (3): 217–36.

Kurzban, Robert, Kevin McCabe, Vernon L. Smith, and Bart J. Wilson. 2001. Incremental commitment and reciprocity in a real-time public goods game. *Personality and Social Psychology Bulletin* 27 (12): 1662–73.

Ledyard, John O. 1995. Public goods: A survey of experimental research. In *The handbook of experimental economics*, ed. John H. Kagel and Alvin E. Roth. Princeton, NJ: Princeton University Press.

Levitt, Steven D., and John A. List. 2007. What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives* 21 (2): 153–74.

List, John A., and David Lucking-Reiley. 2002. The effects of seed money and refunds on charitable giving: Experimental evidence from a university capital campaign. *Journal of Political Economy* 110 (1): 215–33.

Marks, Melanie, and Rachel Croson. 1998. Alternative rebate rules in the provision of a threshold public good: An experimental investigation. *Journal of Public Economics* 67 (2): 195–220.

Marx, Leslie M., and Steven A. Matthews. 2000. Dynamic voluntary contribution to a public project. *Review of Economic Studies* 67 (2): 327–58.

Mehta, Judith, Chris Starmer, and Robert Sugden. 1994. The nature of salience: An experimental investigation of pure coordination games. *American Economic Review* 84 (3): 658–73.

Mollick, Ethan. 2014. The dynamics of crowdfunding: An exploratory study. *Journal of Business Venturing* 29 (1): 1–16.

Palfrey, Thomas R., and Howard Rosenthal. 1984. Participation and the provision of discrete public goods: A strategic analysis. *Journal of Public Economics* 24 (2): 171–93.

Rondeau, Daniel, and John A. List. 2008. Matching and challenge gifts to charity: Evidence from laboratory and natural field experiments. *Experimental Economics* 11 (3): 253–67.

Schelling, Thomas C. 1960. *The strategy of conflict*. Cambridge, MA: Harvard University Press.

Solomon, Jacob, Wenjuan Ma, and Rick Wash. 2016. Highly successful projects inhibit coordination on crowdfunding sites. In *Proceedings of the ACM Conference on Human Factors in Computing*, 4568–72. San Jose, CA: ACM.

Spencer, Michael E., Stephen K. Swallow, Jason F. Shogren, and John A. List. 2009. Rebate rules in threshold public good provision. *Journal of Public Economics* 93 (5): 798–806.

Stiver, Alexandra, Leonor Barroca, Shailey Minocha, Mike Richards, and Dave Roberts. 2015. Civic crowdfunding research: Challenges, opportunities, and future agenda. *New Media and Society* 17 (2): 249–71.

Strausz, Roland. 2017. A theory of crowdfunding: A mechanism design approach with demand uncertainty and moral hazard. *American Economic Review* 107 (6): 1430–76.

Tabarrok, Alexander. 1998. The private provision of public goods via dominant assurance contracts. *Public Choice* 96 (3–4): 345–62.

van de Kragt, Alphons J. C., John M. Orbell, and Robyn M. Dawes. 1983. The minimal contributing set as a solution to public goods problems. *American Political Science Review* 77 (1): 112–22.

Van de Rijt, Arnout, Soong M. Kang, Michael Restivo, and Akshay Patil. 2014. Field experiments of success-breeds-success dynamics. *Proceedings of the National Academy of Sciences* 111 (19): 6934–39.

Wash, Rick, and Jacob Solomon. 2014. Coordinating donors on crowdfunding websites. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing*, 38–48. Baltimore: ACM.

Zubrickas, Robertas. 2014. The provision point mechanism with refund bonuses. *Journal of Public Economics* 120:231–34.