When Citizens Judge Science:
Evaluations of Social Impact and Support for Research

Chiara Franzoni
Politecnico di Milano

Henry Sauermann
ESMT Berlin

Diletta Di Marco
Politecnico di Milano

Abstract

The direction of science has for a long time been set by professional scientists, funding agencies, and policy makers. New mechanisms enable "citizens" or "lay people" without a professional background in science to shape research agendas by voting on funding proposals or allocating their own resources through crowdfunding. Although there are hopes that such mechanisms can democratize science and steer it towards projects with greater social impact, there are also concerns that they give greater weight to the preferences and needs of selected parts of the population, most notably groups with above-average income and education. Moreover, citizens' personal experience with certain problems may bias their assessments of projects addressing such problems, potentially leading to inefficient resource allocation and outsized influence of self-selected groups of stakeholders. To assess such concerns, we study over 2,300 evaluations of research proposals made by members of the general public. We find significant differences in project support by income and education, but only if supporting a project imposes a personal cost upon evaluators. Evaluators' personal experience with a particular problem is associated with greater support but does not appear to inflate evaluations of a project's social impact. When evaluating a project's social impact, however, citizens focus on the importance of the problem and pay less attention to whether or not the project is able to deliver a solution. We discuss implications for the literature on the science of science and innovation as well as for funding agencies and policy makers.

Key words: Science funding, agenda setting, social impact, crowd science

Corresponding author: henry.sauermann@esmt.org. We thank Cindy Wu, Denny Luan, and David Lang for comments and for sharing insights from the platform Experiment.com. This project was supported by the Science of Science Funding initiative with support from the Alfred P. Sloan Foundation.

1 Introduction

Members of the general public are gaining increasing influence over the direction of scientific research. In many countries, “citizens” or “lay people” who are not professional scientists are involved in consultations over research priorities (Briley & Singh, 2017; de Graaf et al., 2017; Jasanoff, 2003; Pidgeon et al., 2014) and patients without formal scientific training serve as panel members or reviewers for grant proposals (Den Oudendammer et al., 2019; INVOLVE, 2006). A more recent development is that members of the public get involved in selecting research proposals directly and without the traditional facilitation processes. For example, lay people have directly funded thousands of research projects on crowdfunding platforms such as Experiment.com or Crowd.science (Sauermann et al., 2019). The Canadian Fathom Fund relies on citizens to evaluate the societal relevance of proposals by matching funds that projects have raised through crowdfunding by up to 300%.[1] A Danish local government recently asked citizens to vote over the internet which medical research projects should be funded.[2]

Advocates argue that such mechanisms can steer science towards projects with greater social impact. One hope is that public participation can aggregate diverse perspectives and preferences, especially if evaluations of proposals are made by a representative cross-section of society (Irwin, 1995; Pidgeon et al., 2014). Another argument is that citizens who have personal experience with problems may possess experiential knowledge that can be valuable in assessing the importance of different problems or the merits of proposed solutions (Briley & Singh, 2017; Caron-Flinterman et al., 2005; Collyar, 2005; Epstein, 1995).

But there are also important concerns. One is that the citizens who participate in setting research agendas are not representative of the general population, often self-selecting based on income, education, or interest in science. As such, research priorities emerging from public involvement may not reflect the needs and preferences of broader society (Rothwell, 2001; West & Pateman, 2016). A related concern is that the personal experience with particular problems may lead citizens to over-estimate the social relevance of those problems, or to steer resources towards projects that serve primarily their own personal needs (Edlin et al., 2007; Hogg & Williamson, 2001; Shapiro & Bloch-Elkon, 2008). Concerns such as these have led to vocal opposition against efforts to involve the public more heavily in agenda setting, especially among professional scientists (Golumbic et al., 2017).[3]

Despite considerable public debate, there is little evidence on how members of the general public judge scientific research projects, especially when doing so independently and without facilitation by panel leaders or administrators. Focusing on potential concerns regarding non-representative participation as well as the role of prior experience, we provide empirical evidence using data from over 2,300 study participants recruited on Amazon Mechanical Turk. These individuals served as lay-evaluators of real research proposals that were raising funds on the platform Experiment.com. Evaluators first scored the proposals with respect to three criteria commonly used in project

---

[1] https://fathom.fund/ Accessed June 8, 2021.

[2] https://www.sdu.dk/en/forskning/forskningsformidling/citizenscience/et+sundere+syddanmark Accessed June 8, 2021.

[3] There are several additional arguments in favor and against public involvement in setting research agendas. Among others, public involvement may increase the legitimacy of the resulting decisions (Landemore, 2012), but there are also concerns that lay-people decide based on inaccurate information or outdated beliefs (Kahne & Bowyer, 2017; Scheufele & Krause, 2019).

evaluations: scientific merit, team capabilities, and social impact (see Appendix A for an overview of criteria used by funding agencies). They could then indicate their support for the proposals by (1) *recommending* projects for funding and (2) *donating* part of their participant payment to fund the project. We use a number of additional measures for supplementary analyses and robustness checks.

Our analysis shows three key results. First, we show that the level of education and income are not related to funding recommendations, but people with lower levels of income or education are less likely to support the project with their private funds. This suggests that even small personal costs induce potentially strong selection of participants, limiting the intended benefits of inclusion and representativeness when involving lay evaluators. Policy makers and funding agencies may be able to increase diversity by actively addressing such costs and barriers.

Second, we find that lay-evaluators who have a personal experience with the topic of a research proposal (e.g., citizens who have a family member affected by Alzheimer's disease and evaluate a project about Alzheimer's disease) are significantly more likely to support the proposal with both recommendations and donations. Contrary to concerns about biases due to over-generalization or wishful thinking, however, this additional support does not result from an inflated expectation of the social impact of the project. Funding agencies and policy makers should recognize that lay-evaluators' personal experience has different implications for assessments of social impact versus overall support for a project. Depending on the purpose of involving the public, as well as the risk of non-representative participation, they should choose carefully between different mechanisms to elicit citizens' preferences and experiential knowledge.

Third, we confirm that evaluations of social impact are a strong predictor of funding recommendations and especially the decision to support a project with private funds. However, our analysis of qualitative responses highlights a point of attention: when evaluating social impact, citizens tend to focus on the perceived importance of the problem and pay little attention to the projects' ability to actually solve the problem. Consistent with that observation, higher perceived social impact strongly predicts support for a project even if a projects' scientific merit or team capabilities are assessed to be low. In light of recent concerns that scientists in peer review panels may place too much emphasis on things that could go wrong, and too little on potential gains (e.g., (Franzoni et al., 2021), the evaluation of lay people and scientists may be complementary.

In addition to suggesting implications for policy and practice, this study also contributes to research in the broader domain of science and innovation. First, we complement prior work on peer-evaluations in science (Boudreau et al., 2014; Franzoni & Stephan, 2021; Gallo et al., 2016; Hug & Aeschbach, 2020; Li, 2017) with initial insights into project evaluations made by non-professional citizens. Although our study does not directly compare crowd and expert evaluations, the results suggest potential benefits from involving citizen evaluators, but also from employing mechanisms that combine respective strengths of crowds and experts. Second, we complement a recent stream of work showing that scientists' and innovators' concerns about "social impact" can have important implications for the rate and direction of their efforts (Cohen et al., 2020; Ganguli et al., 2021; Guzman et al., 2020). We complement this prior work by studying the social impact motives of project evaluators (rather than scientists or innovators). Moreover, rich individual-level data allow us to go beyond establishing average effects to provide a deeper understanding of the factors that shape social impact evaluations – including characteristics of projects as well as evaluators.

The rest of the paper is organized as follows. Section 2 defines key concepts and provides conceptual background to guide our empirical analysis. We describe our empirical strategy and measurement in Section 3. Section 4 reports results and Section 5 concludes with a discussion of implications and contributions to the literature.

2 Background

2.1 Citizen engagement, social impact, and problem importance

Consistent with the literature on public participation in science, we use the terms "citizen" or "lay people" to refer to individuals who are not professional scientists. These terms are used in a very inclusive sense, comprising individuals with different levels of socio-economic status, education, race and gender, as well as legal status (Bonney et al., 2014; ECSA, 2020; US Congress, 2016).

A key rationale for involving lay people in the allocation of resources is to increase the societal relevance and impact of research (de Graaf et al., 2017; Lloyd & White, 2011; Pidgeon et al., 2014). Considering that scientific research spans different fields in the social, natural, and medical sciences, we conceptualize "social impact" to include benefits to people (e.g., related to health and poverty reduction), but also nature (e.g., environmental conservation).[4] There is no consensus as to how (potential) social impact should be evaluated, but the importance of the problem that projects address is arguably an important factor (Davis & Laas, 2014). Judgments of problem importance may reflect factual information, such as the number of people affected by particular diseases or the change in average temperature experienced in a region during a decade. But such evaluations also reflect inherently subjective elements and personal convictions, such as the importance of living longer, or the perceived need of preserving biodiversity, which the scientists are in no better position to assess than everyone else. It is this subjective and "trans-scientific" nature of problems (Weinberg, 1972; Eyal, 2019) that motivates the involvement of citizens, who may not be experts in science but who are affected by the outcomes that science produces (Irwin, 2001).[5]

As noted in the introduction, there is a range of different approaches to involve citizens in agenda setting, such as public consultations about research priorities, inclusion of patients in funding panels, asking citizens to vote on research proposals, as well as crowdfunding. These mechanisms differ with respect to who is allowed to participate (e.g., selected by panel organizers vs. self-selected in response to an open call). They also elicit citizens' evaluations and decisions using different mechanisms (e.g., input into collaborative panel discussions vs. independent funding decisions). In this paper, we focus on decisions where citizens respond to an open call and evaluate specific research projects independently from each other.[6]

---

[4] This broad conceptualization is consistent with expectations of social relevance and "broader impacts" expressed by funding agencies such as the NIH, NSF, or ERC (Davis & Laas, 2014) and also resonates with the wide scope of the Sustainable Development Goals (United Nations, 2018).

[6] By focusing on independent evaluations, the paper abstracts from potential interactions among evaluators, e.g., in discussions on funding panels or by seeing each other's decisions in crowdfunding settings.

[6] By focusing on independent evaluations, the paper abstracts from potential interactions among evaluators, e.g., in discussions on funding panels or by seeing each other's decisions in crowdfunding settings.

2.2 Non-representative preferences: The role of income and education

Proponents of public involvement in agenda setting argue that it can give decision makers a better understanding of societal needs (Irwin, 1995; Lehner, 2013; Pidgeon et al., 2014). An important premise of this rationale is that the participants who help shape research agendas are representative of broader society, bringing to the table a wide range of perspectives and preferences.

However, representativeness may be jeopardized if certain groups of citizens are more likely to agree to participate in evaluation exercises, or self-select in response to open calls for participation.[7] There is evidence of such selection effects from the context of crowd and citizen science, where non-professionals perform research tasks such as collecting and analyzing data (Bonney et al., 2014; Franzoni & Sauermann, 2014). In particular, participants tend to have much higher levels of education and socio-economic status than the general population and are less likely to come from under-represented minorities (Geoghegan et al., 2016; Haklay, 2015; Raddick et al., 2013). A likely driver of self-selection is that project participation imposes a considerable cost on participants (in terms of both time and resources) and that participants with higher income are better able to afford those costs (Haklay et al., 2021; Sauermann et al., 2020). Selection effects with respect to education may reflect that more highly educated individuals have greater trust in science and expect greater benefits from new scientific knowledge (Füchslin et al., 2019).

Putting aside such selection effects with respect to participation in agenda setting exercises, we suggest that similar mechanisms may come into play as those who participate in evaluation exercises decide whether to support a particular project. First, to the extent that supporting a project imposes a personal financial cost on participants (e.g., requiring the donation of personal money), individuals with higher income should be more likely to express support, holding all else equal. Income should be less predictive of project support if support is costless to the individual (e.g., making a funding recommendation). Second, to the extent that higher levels of education are associated with greater trust in science or expected gains from scientific research, individuals with higher levels of education should be more willing to incur personal costs of supporting a project. Education should be less predictive of support if support is costless to the individual.

Hypothesis 1: Evaluators' levels of income and education are positively associated with support for a research project. These relationships are stronger if support imposes additional personal costs on the evaluator.

Empirical support for H1 would raise important concerns regarding the "democratization" benefits of public involvement in agenda setting. Even if participants are a representative cross-section of society, mechanisms that impose a personal cost on evaluators may deter those with low income or education from supporting a project, leading to research agendas that are biased towards the preferences of wealthy and educated citizens.

2.3 Impact evaluations and project support: The role of personal experience with a problem

A second rationale for public participation in agenda setting is that some citizens have personal experience in a problem domain. This experience may allow them to identify the most pressing

[7] Facilitated mechanisms such as funding panels or public deliberations often seek to ensure diversity by hand-selecting participants with different backgrounds and perspectives, although not all invitees may agree to participate.

questions for research or to evaluate potential solutions to important problems (see also Beck et al., 2020; Caron-Flinterman et al., 2005; Von Hippel & Von Krogh, 2015). Reflecting this rationale, funding panels in the medical sciences solicit input from affected patients and caretakers rather than the general population (Briley & Singh, 2017; Den Oudendammer et al., 2019).

Although experiential knowledge may allow citizens to identify novel problems or judge potential solutions, there is the concern that people with experience in a problem domain also have a personal interest in seeing that problem solved. As such, vocal minorities may advocate for research that addresses their particular problem regardless of its broader social relevance, and self-selection of evaluators with respect to prior experience may lead to biased research agendas. This may not be problematic if advocates supply their own resources to support research (e.g., Callon & Rabeharisoa, 2008; Van Brussel & Huyse, 2018) but is concerning if public resources are diverted from possibly more important issues. In this context, we examine two questions.

First, we examine whether evaluators' personal experience with the problem addressed by a research project is indeed associated with greater support. The answer is quite clear in related contexts such as open-source software or crowdfunding of new products, where "self-use" motives were shown to be very important (Lakhani & Wolf, 2006; Von Krogh et al., 2012). However, scientific research differs from those contexts in that project results are more uncertain and distant in the future. As such, even people for whom a problem is personally relevant may see little personal benefit from supporting a research project addressing that problem. Nevertheless, we still expect:

Hypothesis 2a: Citizens who have experience with a particular problem are more likely to support a project addressing this problem.

The argument leading to H2a assumed that personal experience with a problem increases support for a project because of expected (long-term) private benefits. An alternative channel is that personal experience with a problem leads evaluators to assess more positively the potential social impact of a research project, which in turn increases project support. One reason is that people who are personally affected by a problem may believe it also more important for society at large, e.g., because people tend to generalize from themselves to others (Edlin et al., 2007). For example, a citizen with cancer may rate the social impact of a cancer project more highly than a citizen without cancer because she believes that cancer is a more important problem for society broadly. Another reasons is that the personal relevance of a problem leads evaluators to be overly optimistic that a project addressing this problem will be able to find a solution. Such *desirability biases* (also called *wishful thinking*) in the forecasting of successful outcomes have been documented in other context, using samples of both experts and the general population (Blind et al., 2001; Lench et al., 2016). Thus, we expect what follows.

Hypothesis 2b: The relationship between personal experience and project support is mediated by greater perceived social impact of a project.

In both cases, a strong role of personal experience in shaping funding decisions reinforces concerns about selection effects (see also section 2.2): If experience with a particular problem is over-represented among evaluators, research on that problem will be over-funded, relative to research on problems that are under-represented among evaluators.[8]

2.4 The moderating role of scientific merit and team capabilities

Although our focus is on social impact, evaluators' decisions to support a project may also depend on evaluations of other project characteristics, most notably scientific merit and team capabilities (see Appendix and Table A1).

One possibility is that these characteristics are already reflected in assessments of a project's potential social impact: If the science is not sound or the team lacks the necessary skills to perform the research, then a project will yield no results and will have little social impact. Alternatively, evaluators may consider scientific merit, team capabilities, and social impact to be distinct dimensions (see Hug & Aeschbach, 2020; Lamont, 2009), with the latter reflecting the perceived importance of the problem. In that case, scientific merit and team capabilities may influence project support even controlling for social impact. Whether and how much evaluations of these other criteria predict citizens' support is an interesting empirical question. Moreover, it is important to know whether the criteria interact, i.e., whether assessments of social impact are more strongly associated with project support if the project is also rated as having high scientific merit and strong capabilities. In consumer decisions, different product attributes are often substitutes and higher values on one attribute can compensate for lower levels on others (Payne et al., 1993). In "production" settings such as science, however, the three aspects are likely complementary such that the importance of the problem should matter more if evaluators believe that the science is sound and the team is capable of completing the project:

Hypothesis 3: Evaluators' assessments of social impact have a stronger association with support for a project if scientific merit and researcher capabilities are evaluated to be high.

3 Data and Measures

3.1 Scientific research proposals

Citizens typically evaluate project proposals that are shorter and less technical than those reviewed by traditional peer-reviewers (Sauermann et al., 2019). To make evaluations as realistic as possible, we selected four real research proposals that were posted on the science crowdfunding platform Experiment.com. These proposals span the area of social sciences, biology/environmental sciences, and health and were posted during the year 2020. More specifically, Project 1 proposed to examine the relationship between sexual orientation and economic preferences to understand wage gaps; Project 2 proposed to examine the influence of rabies and otter behavior on human-otter conflict in Florida; Project 3 proposed targeted drug discovery for Covid-19; and Project 4 proposed to study bacterial DNA as a potential trigger for Alzheimer's disease. Table A2 in the Appendix provides additional detail on the proposals.

[8] In a related vein, studies have documented the over-representation of certain demographic groups among research subjects in clinical trials as well as among innovators. This over-representation may shape the direction of science and innovation in a way that is biased against the need of underrepresented groups (Dresser, 1992; Koning et al., 2020).

3.2 Evaluators and evaluation process

We recruited citizen evaluators on Amazon Mechanical Turk (Mturk).[9] Mturk is one of the most commonly used platforms for research in economics, management, and the behavioral sciences (Aguinis et al., 2020; Buhrmester et al., 2018). Although care has to be taken regarding the representativeness of samples and attention checks (Paolacci & Chandler, 2014), studies show that Mturkers are comparable to other samples with respect to economic behaviors or social preferences, and results from Mturk studies are also stable across time (Arechar et al., 2017; Buhrmester et al., 2016). A key advantage of Mturk for our purposes is that it gives access to a very heterogeneous sample of individuals, which is important given our interest in the role of individual-level factors such as education, income, and personal experience with different research topics.

We limited participation to US-based individuals but did not screen participants based on their prior involvement in science or research evaluation. The recruiting message stated that the task was to evaluate scientific research projects and that participants should have a general interest in science, but that prior experience with science was not required. Consistent with recommendations in the literature, we set minimum requirements regarding the number of prior Mturk engagements (100) as well as approval rates (98%), but kept these thresholds relatively low in order to keep the sample diverse (Aguinis et al., 2020; Robinson et al., 2019). To obtain sufficient variation in respondents' educational background, we oversampled individuals with graduate degrees using Mturk qualification settings. Participants were compensated with a fixed payment between $1.50-2.00 and could also receive an additional bonus (see variable descriptions below).

The evaluation task was embedded in a Qualtrics questionnaire. Respondents first answered a few warm-up questions and were then asked to click on a link that led to the live project on the platform Experiment.com. After reading through the project, participants returned to Qualtrics to complete the evaluation task. Towards the end, respondents answered questions on their background and demographic characteristics. Respones were given individually with no interaction among participants. We also collected all evaluations for a given Experiment.com project within a few days such that all evaluators saw virtually the same proposal (with only minor changes in funding amounts). We additionally controlled for day fixed effects in our regression analyses.

Over 73% of Mturkers who started the questionnaire also completed it. We dropped from the analysis approxately 20% of completed tasks because they did not meet our pre-established accuracy criteria with respect to correct answers on attention check questions, time spent answering the questionnaire, responses to open ended questions (see below), as well as uniqueness of IP addresses (Aguinis et al., 2020). Our final sample includes 2,350 completed evaluations (528 for Project 1; 585 for Project 2; 684 for Project 3; 553 for Project 4). Table 1 shows descriptive statistics, including demographic characteristics of the evaluators.

3.3 Measures

3.3.1 Criteria evaluations

After participants returned from reading the project proposal and completed attention checks, they scored the project with respect to potential social impact, scientific merit, and team capabilities using

[9] This study was approved by the Institutional Review Board of the National Bureau of Economic Research (NBER).

5-point scales. We constructed the respective questions based on the criteria used by funding agencies and peer-evaluators (see Appendix and Table A1) while also considering that our evaluators are a diverse sample of lay people. In particular, we asked:

- "Social impact describes how much a research project can influence and benefit society, health, or the natural environment. In your opinion, how large is the social impact of this project?" (*social impact*, from "No social impact at all" to "Very high social impact")
- "Scientific merit describes how much a research project can advance academic knowledge in a particular domain. In your opinion, how large is the scientific merit of this project?" (*scientific merit*, from "No scientific merit at all" to "Great scientific merit")
- How would you rate the capabilities of the research team to carry out this project? (*team capabilities*, from "Very low capabilities" to "Very high capabilities")

The order of these three questions was randomized. To make coefficient estimates more comparable, we estimate regressions using standardized z-scores of these measures (mean=0, SD=1 computed at the level of individual projects). Table 1 shows means for both the original scales and the standardized measures.

3.3.2 Funding recommendations and donations

After participants had scored social impact, scientific merit, and team capabilities, we measured support for a research project using two different approaches. First, we asked "Considering everything, do you think this project should be funded?". Respondents indicated their recommendation on a 5-point scale ranging from "Absolutely not!" to "Absolutely yes!" (*recommendation*). This measure allowed respondents to support a project without incurring additional personal cost, i.e., it is similar in spirit to approaches that ask citizens for funding recommendations. The average score was 3.88 and the share of evaluators giving a score of 5 was 31.62%.

The second measure involved an additional personal cost to the respondent. In particular, we told the respondent "We will give you a budget of $1, in addition to your current payment for taking this survey. If you would like to support the research project you just saw, you can donate this extra $1 to it. In this case we will match your donation with an additional $1 and will donate a total of $2 to the project by the end of the day. If you would not like to support the research project you just say, you can take the extra $1 as a bonus for yourself".[10] Respondents then chose to either donate to the project or take the bonus (*donate*). These choices were real and we donated the collected funds as promised on the Experiment.com website. The share of respondents who decided to support projects with a personal donation was 39%. The two measures of project support are highly correlated: Among those who did not recommend funding (scores of 1 or 2), the share of donors is less than 2%; among those who did recommend funding (scores of 4 or 5), the share of donors is 52%.

[10] For the first two projects, we offered a stronger match ($0.5 bonus vs. 2$ donation). As such, the levels of donations are not directly comparable across projects. To account for this difference, all regressions are estimated with project fixed effects. Although the donations in this study were smaller than typical contributions to scientific crowdfunding (Sauermann et al., 2019), our main goal was to create a cost difference between the two measures of project support. Even the relatively small amount of money was meaningful: $1 represented a roughly 50% bonus on top of the compensation individuals received for participating in the study.

3.3.3 *Individual characteristics*

Education. Respondents indicated their highest level of *education* (up to high school; up to bachelor's degree; graduate degree).

Income. Respondents indicated their annual household *income* using categories (up to \$25,000; \$25,001-\$50,000; \$51,000-\$75,000; \$75,001-\$100,00; more than \$100,000).

Personal experience with problem. One of our key constructs is the personal experience evaluators have with the problem addressed by a particular research proposal. Rather than asking for a subjective assessment, we created a more "objective" measure by asking about relevant background characteristics. For the Experiment.com project on sexual orientation and economic preferences, we asked how many gay or lesbian people the respondent personally knew. For the project on human-otter conflict in Florida, we asked whether or not the respondent lived in Florida. For the project on Covid-19 drug discovery, we asked whether the respondent had been diagnosed with Covid-19 or had experienced job loss, reduced income, or challenges with care for others due to Covid. For the project on Alzheimer's disease, we asked whether the respondent personally knew anyone who has or had Alzheimer's (including self, family members, or friends). We use the respective responses to create a dichotomous measure that captures whether the evaluator had high vs. low personal experience with the problem addressed by the research proposal (*personal experience*).[11]

3.3.4 *Individual and project-level controls*

Gender. Respondents indicated their gender (male, female, other/don't want to answer). The dummy variable *female* equals 1 for respondents who self-identified as female and 0 for males.

Age. Respondents indicated their *age* using five categories (18-30; 31-40; 41-50; 51-60; 60+).

Crowdfunding experience. Respondents indicated whether they had contributed to crowdfunding campaigns in the past (*donated before*). An open-ended follow-up question shows that the most common platform used was GoFundMe. We use this variable as a rough proxy for prior experience with crowdfunding campaigns as well as for individual differences in the willingness to donate.

Project clarity. After respondents returned from reviewing the project proposal, we asked "How clear is what the researchers are planning to do in this Experiment.com project"?, using a 5-point scale. We include the measure *project clarity* to account for potential differences in understanding on the part of evaluators. This measure may also help account for individual differences in response behavior (e.g., the tendency of some respondents to generally give more positive evaluations than others).

Reward. Participants were recruited in several batches per project. The large majority of participants (69%) were offered a compensation of \$1.80 for their participation, but we also included some batches with slightly different reward levels (ranging from \$1.50 to \$2) in order to balance labor supply and to explore potential differences in donation behavior. The reward level was not randomly assigned and we control for it using the variable *reward* (in US cents).

Project and day fixed effects. The four projects differ with respect to factors such as the scientific field and topic, the time at which they were online, as well as the funding goal, which may

[11] Most of the questions were mandatory. However, we did not force responses for potentially sensitive questions such as LGBT acquaintances or personal experience with diseases. In addition, we gave respondents the option to check "Don't want to answer". Only few respondents chose this option, and these cases are treated as missing data.

lead to project-level differences in support. Moreover, evaluations may differ depending on the particular day on which a participant saw the project (e.g., because the amount of funding raised had increased). Given our focus on individual-level differences in evaluations and project support, we control for differences between projects using project fixed effects and for timing using day fixed effects. The projects ran on different days such that a series of *project-day fixed effects* captures both aspects.

--- Table 1 about here ---

4 Results

4.1 Income and education as predictors of criteria evaluations and project support

Models 1-3 in Table 2 use as dependent variables the standardized ratings of *social impact*, *scientific merit*, and *team capabilities*. We find no significant coefficients for either income or education. Although not hypothesized, we find that women tend to rate projects' social impact higher than men; ratings of scientific merit or team capabilities do not differ by gender. This result is consistent with prior work suggesting that women may have a greater inclination to see social impact potential in entrepreneurial ventures or may be more responsive to social impact messages (Bloodhart & Swim, 2020; Guzman et al., 2020; Harrison & Mason, 2007; Nyborg, 2000).

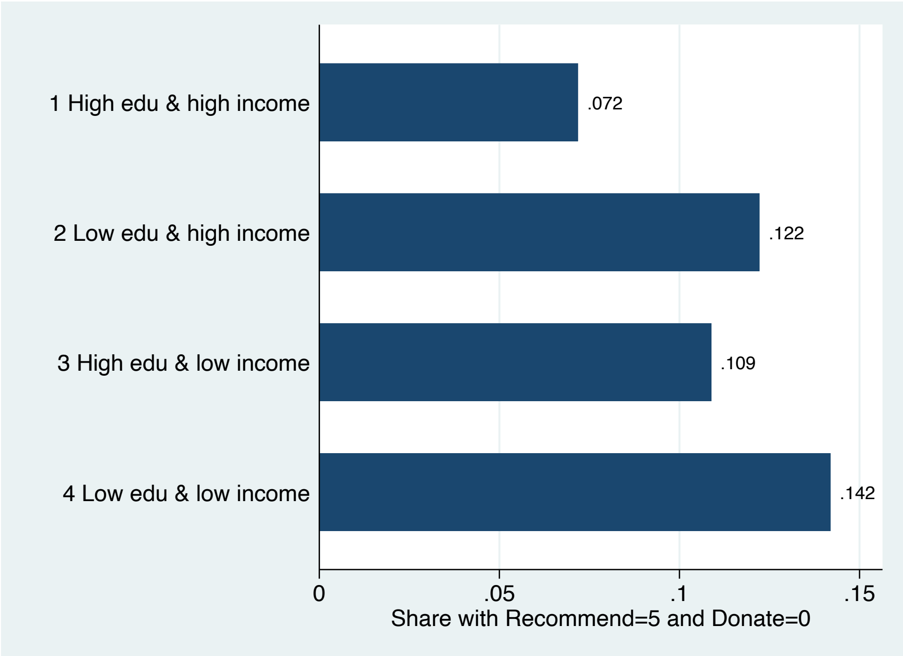
Hypothesis 1 predicted that support for a project depends more strongly on income and education if such support is more costly for the evaluator. To examine this prediction, we assume that donating own funds is more costly than recommending funding (Norwood et al., 2019). Model 4 in Table 2 uses as dependent variable *recommend*, estimated using ordered logistic regressions. Model 5 regresses *donate* using a linear probability model (logistic regressions yield the same result; see Appendix). We find that income or level of education do not predict an individual's recommendation to fund the project. Consistent with H1, however, both income and level of education predict an individual's decision to support the project by donating own funds. As noted in Section 2, this result may reflect that individuals with higher income and education have a greater ability to bear the personal costs of donations but also that they have greater trust in science or perceive greater benefits from scientific research that justify incurring such costs.[12]

An interpretation of these patterns is that asking for funding recommendations yields insights into the preferences of a broad cross section of the population, while using more costly mechanisms such as crowdfunding biases outcomes to reflect primarily the preferences of individuals with higher income and education. To visualize these selection effects, we split the sample by high vs. low education and high vs. low income, and then computed what share of evaluators in each of the four groups strongly recommended funding (recommend=5) but did not donate their own money (donate=0). Low-income participants with low education were nearly twice as likely to recommend but not fund (14.2%) compared to high-income participants with high education (7.2%). The over-

[12] Interestingly, the relationship between income and donations is not linear, with a sharp break between the lowest and higher income brackets and less of a difference between higher income brackets (Model 5). This result is intuitive – given that the required donation was only $1, income constraints should be most binding for low income citizens. Future research could examine whether higher income brackets predict project support if such support is even more costly.

representation of people with high income and education among supporters may be even more pronounced if participation in agenda setting involves greater personal costs than in this study.

Figure 1: Share of respondents who strongly recommend funding but did not donate; by education and income



Note: Differences between group 1 and group 2 as well as 4 are statistically signficant at 5%.

--- Table 2 about here ---

4.2 Personal experience with the problem

Hypothesis 2a predicted that citizens who have experience with a problem are more likely to support a research project addressing that problem. Models 4 and 5 in Table 2 are consistent with this prediction: evaluators with *prior experience* recommend a project more strongly for funding and are also significantly more likely to support the project with their own money.

Our conceptual discussion considered two possible mechanisms: One was that personal experience increases citizens' expected private benefits from research, e.g., a person suffering from Alzheimer's may hope that Alzheimer's research will yield improvements for her own treatment. The second mechanism was that people who have prior experience with a problem expect a greater social impact of a project addressing this problem, because the problem is highly salient to them, or they over-estimate its societal importance, or they inflate the chances of project success due to *wishful thinking* (H2b). Contrary to H2b, Model 1 in Table 2 shows no significant differences in the ratings of social impact between evaluators with and without prior experience with the problem. Although this suggests expected private benefits as a likely mechanism for the greater level of support among those with prior experience, we do not have measures to probe this possibility further.

4.3 Criteria evaluations and project support

In a final set of analyses, we probe the relationships between evaluations of social impact, scientific merit and team capabilities on the one hand, and project support on the other. As expected, Models 6 and 7 in Table 2 show that citizens who evaluate the three criteria more favorably also recommend the project more strongly for funding and are more likely to donate their private funds.[13]

Although we consider recommendations and donations to be two separate outcomes of interest, we also explore whether evaluations of project attributes predict citizens' willingness to *donate* even when controlling for *recommend* (Model 8). We find that the scores for scientific merit and team capabilities lose their significance, while social impact continues to have a significant positive association with *donate*. Our interpretation is that social impact stands out among the three project attributes in that social impact considerations shape citizens' willingness to bear a personal cost of supporting a project even holding constant their general assessment of the projects' "fundability".[14]

In Section 2, we suggested that evaluations of the three criteria may predict project support not only individually but also jointly. In particular, we expected that social impact ratings predict project support less strongly if the evaluator believes that the project has low scientific merit and team capabilities (Hypothesis 3). We test this prediction in two ways. First, we include in the regressions the interaction terms between social impact and both scientific merit and team capabilities (Models 9 and 10). The interaction terms have no significant coefficients, suggesting that the relationship between evaluations of social impact and project support is largely independent of evaluations of other project criteria. As an alternative approach, we create two dummy variables: One takes the value of 1 when scientific merit and team capabilities were rated high (score >3; *z impact if merit and team high*) and the other takes the value of 1 when scientific merit or team were rated low (score<4; *z impact if merit or team low)*. We then estimate two interaction models where we separate the effect of social impact when merit or team are *high* from the effect of social impact when merit and team are *low*. The estimated coefficients (*impact if merit and team high: impact if merit or team low*) are always positive and significant, consistent to prior estimates, but they are not statistically different from one another at conventional significance levels, suggesting no difference in the role of social impact ratings in the two conditions (Models 11 and 12). Taken together, the estimates suggests that citizens are more likely to support a project if they see a greater social impact, irrespective of whether or not they

---

[13] In the regression of *recommend* (Model 6), the coefficient of scientific merit is significantly larger than that of social impact (p=0.015) and that of team capabilities (p=0.009). For *donate* (Model 7), the coefficients of social impact and scientific merit are not significantly different, while the coefficient of team capabilities is significantly smaller than that of scientific merit (p=0.024) and of social impact (p=0.068).

[14] Models 1-3 showed that women rate social impact higher than men (Section 4.1). In supplementary analyses, we explored whether women also give different weights to the three criteria when deciding to support a project. However, regressions using interaction terms show no significant gender differences (see Appendix). This result is particularly interesting given that prior research showing gender differences did not disentangle the potential role of gender in estimates of social impact versus weights given to social impact evaluations. For example, Ganguli et al. (2021) as well as Guzman et al. (2020) found gender differences in the response to social vs. profit oriented framing of opportunities but it is not clear whether those results reflected that women saw greater impact in the same message or gave greater weight to social impact when making participation decisions.

perceive scientific merit and team capabilities to be high or low. Thus, the estimates do not support the predictions of Hypothesis 3.

Even though there is no interaction between social impact and the other two criteria, it could be that evaluators already incorporated assessments of scientific merit and team capabilities when evaluating the social impact of a project. In particular, evaluators may realize that for a project to have social impact, it needs to address an important problem but also have the capacity to come up with an effective solution. Consistent with that idea, evaluations of social impact have a positive correlation with evaluations of scientific merit (r=0.51) as well as team capabilities (r=0.32). However, these correlations do not imply that the evaluations are causally connected – their correlation may reflect a common underlying overall project quality but also common methods bias (Edwards & Bagozzi, 2000). To explore citizens' rationales for evaluations more directly, we coded open-ended explanations that evaluators provided after scoring a project's social impact in response to the question "Please briefly explain why you gave this particular rating". We coded answers in a number of non-mutually exclusive rationales and grouped these rationales according to the related criteria (Table 3). We find that approximately 19% of the answers mention one or more rationales that are related to scientific merit and only about 1% mention rationales that are related to the project team. Indeed, the majority of evaluators (75%) clearly focused on rationales that relate to the importance of the problem that the project was trying to address. Consider the following illustrative examples:

- Well, it deals with a specific area of Florida and so it will provide benefit to the people and otters of that area, but perhaps not a more general or geographically larger benefit. [Project 2 on human-otter conflict]
- COVID-19 has ravaged the world, causing widespread death, suffering, and also economic implosion. To do anything that would lessen this virus's threat would be, perhaps, greater than Jonas Salk's work on the polio vaccine. [Project 3 on Covid drug development]
- I see this as a high impact since many people experience Alzheimer's disease at an older age and it affects families greatly in some cases who care a lot for their family and their health and well-being. [Project 4 on Alzheimer's]

Even though most evaluators focused on problem importance, some also considered the project itself and distinguished scientific and social outcomes:

- We have found that studies do no good in overall social impact. Look at how blacks are still making less and have more incarcerations, even though we KNOW and have studied this. It makes no difference to people at all. [Project 1 on sexual orientation, preferences, and pay gaps]
- This project could have a major social impact but only if it finds scaffolding for a useful drug. I think there are already several other drugs much farther into development. This project might take too long. [Project 3 on Covid drug development]
- The impact is potentially large but it's also down the road. This is basic research, without an immediate clinical application. [Project 4 on Alzheimer's]

Taken together, our analysis of open ended explanations suggests that citizens' evaluations of project social impact tend to focus on the importance of the problem that the project seeks to address. Although some citizens consider both the likely success of the project and the importance of the problem, very few explicitly considered scientific merit and team capabilities.

5 Discussion

Lay evaluators or non-professional "citizens" play an increasing role in shaping the direction of scientific research. Advocates argue that citizen involvement can yield a better picture of the public's needs and preferences, while citizens' experiential knowledge may also help in assessing the importance of research problems or the promise of potential solutions. On the other hand, there are concerns that those who participate are not representative of the broader population, and that prior experience with a topic may lead evaluators to prioritize their personal interests or hold inflated expectations of projects' social impact. We examine these concerns using data from over 2,300 lay people recruited from Amazon Mechanical Turk, who evaluated research proposals asking for funding on the platform Experiment.com.

A first set of analyses explores the role of evaluators' education and income. Consistent with our expectations, we find that higher levels of education and income are strong predictors of decisions to support a project with private funding, while they do not predict whether an evaluator recommends a project for funding. The important implication is that the particular approach used to involve citizens in agenda setting matters: Citizens appear to participate on an equal basis in shaping research agendas when preference expression does not require out-of-pocket costs (e.g., via unexpensive recommendations). Conversely, when expressing one's preferences requires financial means, albeit little, individuals with higher income and education are more likely to shape the direction of research. Thus, if policy makers and funding agencies seek to learn about the preferences of the general public, they should carefully consider what costs are imposed upon citizens in the expression of preferences. Reducing those costs, or even offsetting them for citizens who may otherwise be deterred from participation is likely to increase the diversity and representativeness of citizen evaluations. Note that we do not argue against costly mechanisms, such as crowdfunding. Indeed, crowdfunding has provided a first and important boost to democratize the discussion on the direction of research, compared to the traditional approaches based on expert-only committees. It also increases the resources that are available to science. However, funding agencies such as the Fathom Fund as well as policy makers should be aware of the possible implications of costly approaches. If crowdfunding outcomes are to be interpreted as indicative of the general public's preferences or of the potential social impact of different research projects, special care should be taken to the rules of participation. This is especially important if the approach will be expanded and made more pervasive in the future, as it may leave the door open to misuses from wealthy individuals, interests groups or corporations. Our results suggest that policy makers and funding agencies may partly mitigate such effects by eliciting social impact assessments directly rather than interpreting citizens' overall recommendations for funding or commitment of own resources.

A second set of analyses examined how citizens' prior experience with a problem is related to assessments of social impact and with support for a project addressing that problem. We find that prior experience is associated with stronger recommendations for funding as well as greater willingness to donate private funds. However, this relationship does not seem to be driven by higher – and potentially inflated – assessments of problem importance and social impact. Nevertheless, the observed strong association between prior experience and project support suggests self-selection into the pool of evaluators. Greater involvement of people who are personally affected by certain problems could be positive, as these people may contribute important insights from lay expertise (Epstein, 1995).

However, the resulting recommendations do not necessarily inform about societally optimal levels of investments (Bergstrom, 2006; Nyborg, 2000).

Finally, we examined what role social impact evaluations play in the context of other project characteristics – including scientific merit and the team's capabilities to perform the proposed research. Our regressions suggest that evaluations of scientific merit and team capabilities predict project support independently, but they do not moderate the role of social impact assessments. In other words, evaluators who see a greater social impact in a project are more likely to support it even if scientific merit and team capabilities are judged to be low.[15] Similarly, qualitative analyses suggest that evaluations of a project's social impact primarily reflect citizens' assessments of problem importance, while citizens pay little attention to the project's ability to deliver an effective solution. On the one hand, these results could be seen as problematic in that citizens may allocate resources to projects that seek to solve important problems but have little chance to succeed. On the other hand, these results may be interpreted as positive by observers who feel that the traditional system of peer-evaluations is too risk averse or that scientists shy away from addressing important societal problems that are difficult to solve (Franzoni et al., 2021; Irwin, 1995). Future research is needed to compare directly how citizens and professional scientists differ in their evaluations of scientific projects. Future work should also explore how evaluation mechanisms can be designed that best leverage the expertise of both experts and citizens to advance research for the benefit of science and society.

[15] Of course, it is not clear how accurate evaluators' assessments of scientific merit and team capabilities are, given the general uncertainty in science around those issues but also given citizens' lack of expert knowledge that is typically used to make such assessments (Franzoni & Stephan, 2021; Lamont, 2009). The question of accuracy is even more difficult to address for social impact given that this construct is inherently subjective and it is not clear what "ground truth" could be used to evaluate accuracy of social impact judgments (see section 2.1).

Table 1: Descriptive Statistics

Variable	N	Mean	SD	Min	Max
education: up to HS	2,337	0.11	0.31	0	1
education: up to Bachelors	2,337	0.61	0.49	0	1
education: graduate degree	2,337	0.28	0.45	0	1
income: up to 25k	2,288	0.14	0.35	0	1
income: up to 50k	2,288	0.28	0.45	0	1
income: up to 75k	2,288	0.25	0.43	0	1
income: up to 100k	2,288	0.16	0.36	0	1
income: more than 100k	2,288	0.17	0.38	0	1
personal experience	2,350	0.43	0.50	0	1
social impact	2,350	3.42	1.15	1	5
scientific merit	2,350	3.51	0.96	1	5
team capabilities	2,350	4.04	0.75	1	5
z_social impact	2,350	0.00	1.00	-3.22	2.78
z_scientific merit	2,350	0.00	1.00	-3.86	2.28
z_team capabilities	2,350	0.00	1.00	-4.40	1.66
recommend	2,350	3.88	1.03	1	5
donate	2,350	0.39	0.49	0	1
female	2,326	0.48	0.50	0	1
age class	2,342	2.46	1.23	1	5
donated before	2,350	0.60	0.49	0	1
project clarity	2,350	3.57	0.98	1	5
reward	2,350	178.01	11.35	150	200
projectday	2,350	255.98	107.94	101	403

Table 2: Main results

	1	2	3	4	5	6	7	8	9	10	11	12
	OLS	OLS	OLS	ologit	OLS	ologit	OLS	OLS	ologit	OLS	ologit	OLS
	z_p_impact	z_p_merit	z_p_team	recommend	donate	recommend	donate	donate	recommend	donate	recommend	donate
education: bachelors	0.045	0.014	-0.002	-0.064	0.074*	-0.181	0.070**	0.078**	-0.179	0.071**	-0.181	0.070**
	[0.064]	[0.078]	[0.097]	[0.108]	[0.027]	[0.141]	[0.022]	[0.021]	[0.140]	[0.022]	[0.140]	[0.022]
education: grad degree	0.045	0.047	0.079	-0.069	0.116**	-0.295	0.107**	0.120**	-0.295	0.107**	-0.295	0.107**
	[0.085]	[0.111]	[0.117]	[0.185]	[0.028]	[0.183]	[0.027]	[0.024]	[0.180]	[0.026]	[0.183]	[0.027]
income: 25-50k	0.040	0.096	-0.023	-0.092	0.080*	-0.265	0.072*	0.087**	-0.269	0.072*	-0.265	0.072*
	[0.068]	[0.053]	[0.067]	[0.143]	[0.031]	[0.198]	[0.028]	[0.025]	[0.202]	[0.028]	[0.198]	[0.028]
income: 51-75k	0.005	0.086	-0.014	-0.114	0.092**	-0.257	0.086**	0.100**	-0.257	0.086**	-0.257	0.086**
	[0.105]	[0.080]	[0.088]	[0.120]	[0.018]	[0.150]	[0.015]	[0.019]	[0.159]	[0.014]	[0.150]	[0.014]
income: 76-100k	0.103	0.128	0.068	0.077	0.104**	-0.082	0.086**	0.091**	-0.078	0.087**	-0.082	0.086**
	[0.078]	[0.070]	[0.055]	[0.125]	[0.025]	[0.162]	[0.024]	[0.023]	[0.171]	[0.024]	[0.161]	[0.023]
income: over 100k	-0.036	-0.061	-0.050	-0.357*	0.070*	-0.354*	0.078**	0.101**	-0.358*	0.078**	-0.354*	0.078**
	[0.124]	[0.088]	[0.054]	[0.179]	[0.029]	[0.142]	[0.021]	[0.021]	[0.150]	[0.021]	[0.143]	[0.021]
personal experience	0.088	0.016	0.082	0.257**	0.045**	0.237**	0.036	0.023	0.235**	0.035	0.237**	0.036
	[0.053]	[0.049]	[0.057]	[0.096]	[0.014]	[0.090]	[0.019]	[0.018]	[0.088]	[0.019]	[0.090]	[0.019]
z social impact						0.742**	0.070**	0.021*	0.737**	0.070**		
						[0.096]	[0.014]	[0.010]	[0.094]	[0.014]		
z scientific merit						0.965**	0.067**	0.005	0.949**	0.067**	0.965**	0.067**
						[0.072]	[0.011]	[0.008]	[0.077]	[0.012]	[0.072]	[0.011]
z team capabilities						0.599**	0.029*	-0.010	0.604**	0.029*	0.599**	0.029*
						[0.081]	[0.010]	[0.012]	[0.093]	[0.011]	[0.081]	[0.010]
recommend								0.191**				
								[0.011]				
z social impact X z scientific merit									-0.085	-0.007		
									[0.062]	[0.011]		
z social impact X z team capabilities									0.036	0.007		
									[0.068]	[0.009]		
z social impact if merit or team low											0.742**	0.069**
											[0.118]	[0.020]
z social impact if merit and team high											0.743**	0.072**
											[0.079]	[0.018]
female	0.154**	0.066	0.081	0.138	0.057*	-0.015	0.039	0.043	-0.010	0.040	-0.015	0.039
	[0.025]	[0.040]	[0.044]	[0.071]	[0.026]	[0.053]	[0.023]	[0.021]	[0.051]	[0.023]	[0.052]	[0.023]
age	-0.009	-0.041*	-0.019	-0.001	0.033*	0.040	0.037**	0.036**	0.045	0.037**	0.040	0.037**
	[0.023]	[0.016]	[0.014]	[0.030]	[0.011]	[0.043]	[0.012]	[0.010]	[0.043]	[0.011]	[0.044]	[0.012]
donated before	0.106*	0.009	-0.004	0.084	-0.003	0.086	-0.011	-0.016	0.089	-0.011	0.086	-0.011
	[0.037]	[0.051]	[0.060]	[0.147]	[0.028]	[0.129]	[0.023]	[0.016]	[0.128]	[0.023]	[0.129]	[0.023]
project clarity	0.240**	0.312**	0.295**	0.700**	0.105**	0.333**	0.059**	0.038**	0.335**	0.059**	0.333**	0.059**
	[0.028]	[0.030]	[0.022]	[0.065]	[0.007]	[0.046]	[0.007]	[0.007]	[0.046]	[0.007]	[0.046]	[0.007]
reward	0.001	-0.001	-0.002	0.001	0.002**	0.003	0.002**	0.002*	0.003	0.002**	0.003	0.002**
	[0.004]	[0.003]	[0.003]	[0.007]	[0.000]	[0.005]	[0.001]	[0.001]	[0.005]	[0.001]	[0.005]	[0.001]
Project-day fixed effects	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.
Constant	-1.439*	-1.199*	-0.814		-0.758**		-0.553**	-1.109**		-0.552**		-0.554**
	[0.538]	[0.472]	[0.408]		[0.072]		[0.090]	[0.132]		[0.095]		[0.093]
Observations	2,269	2,269	2,269	2,269	2,269	2,269	2,269	2,269	2,269	2,269	2,269	2,269
R-squared	0.069	0.100	0.086		0.098		0.164	0.239		0.164		0.164
r2_p				0.0937		0.288			0.288		0.288	

*=5%, **=1%. Standard errors clustered at level of project-day.

Table 3: Rationales provided to explain the score of impact

RATIONALES[+]	DESCRIPTION	PERCENT RESPONSES MENTIONING
IMPACT		74.79
scale	Extent to which the problem affects many people	31.64
expected results	Perceived relevance of expected results	14.21
importance	Refers generically to importance	13.51
success	Extent to which the research can succeed	13.36
severity	Extent to which the problem causes severe consequences	8.54
generalizable	Extent to which results could be applied to other areas	5.07
frequency	Extent to which the problem is rare	2.81
practical application	Extent to which the results could be directly applicable	2.11
need	Extent to which there is a need to solve a problem	1.56
time	Perceived time needed to have results	1.31
implementation	Extent to which the research can make a practical difference	1.21
awareness	Perceived importance of raising awareness about a problem/topic	1.05
SCIENTIFIC MERIT		18.58
knowledge	Extent to which the results could expand current knowledge	13.76
quality	Perceived quality/rigor of the project	4.47
novelty	Extent to which the topic is new	1.10
TEAM		1.36
competition	Extent to which other researchers are engaged in similar projects	1.36
OTHER		16.12
gut feeling	Reference to own perceptions	1.10
hype	Extent to which the topic is covered in press/media	1.05
don't know	Does not know/ say	0.70
personal experience	Reference to own experience/knowledge	0.65
other/ various	Reference to other reasons, different from the above	13.31

+ Coded from open-ended responses. Each response was coded into up to 4 non-mutually exclusive rationales.

References

Aguinis, H., Villamor, I., & Ramani, R. S. 2020. MTurk Research: Review and Recommendations: SAGE Publications Sage CA: Los Angeles, CA.

Arechar, A. A., Kraft-Todd, G. T., & Rand, D. G. 2017. Turking overtime: how participant characteristics and behavior vary over time and day on Amazon Mechanical Turk. ***Journal of the Economic Science Association***, 3(1): 1-11.

Beck, S., Brasseur, T.-M., Poetz, M. K., & Sauermann, H. 2019. What's the problem? Crowdsourcing research questions in science, ***Working Paper***.

Bergstrom, T. C. 2006. Benefit-cost in a benevolent society. ***American Economic Review***, 96(1): 339-351.

Bloodhart, B., & Swim, J. K. 2020. Sustainability and consumption: What's gender got to do with it? ***Journal of Social Issues***, 76(1): 101-113.

Bonney, R., Shirk, J. L., Phillips, T. B., Wiggins, A., Ballard, H. L., Miller-Rushing, A. J., & Parrish, J. K. 2014. Next steps for Citizen Science. ***Science***, 343(6178): 1436-1437.

Briley, A., & Singh, C. 2017. Patient involvement in medical research. ***Introduction to Research Methodology for Specialists and Trainees***: 150.

Buhrmester, M., Kwang, T., & Gosling, S. D. 2016. Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data?

Buhrmester, M. D., Talaifar, S., & Gosling, S. D. 2018. An evaluation of Amazon's Mechanical Turk, its rapid rise, and its effective use. ***Perspectives on Psychological Science***, 13(2): 149-154.

Butticè, V., Franzoni, C., Rossi-Lamastra, C., & Rovelli, P. 2017. The road to crowdfunding success: A review of extant literature. In A. Afuah, C. L. Tucci, & G. Viscusi (Eds.), ***Creating and Capturing Value Through Crowdsourcing***: Oxford University Press.

Callon, M., & Rabeharisoa, V. 2008. The growing engagement of emergent concerned groups in political and economic life: Lessons from the French association of neuromuscular disease patients. ***Science, Technology, & Human Values***, 33(2): 230-261.

Caron-Flinterman, J. F., Broerse, J. E., & Bunders, J. F. 2005. The experiential knowledge of patients: a new resource for biomedical research? ***Social science & medicine***, 60(11): 2575-2584.

Cohen, W. M., Sauermann, H., & Stephan, P. 2020. Not in the job description: The commercial activities of academic scientists and engineers. ***Management Science***, 66(9): 4108-4117.

Collyar, D. 2005. How have patient advocates in the United States benefited cancer research? ***Nature Reviews Cancer***, 5(1): 73-78.

Davis, M., & Laas, K. 2014. "Broader impacts" or "responsible research and innovation"? A comparison of two criteria for funding research in science and engineering. ***Science and engineering ethics***, 20(4): 963-983.

de Graaf, B., Rinnooy Kan, A., & Molenaar, H. 2017. ***The Dutch National research agenda in perspective: A reflection on research and science policy in practice***: Amsterdam University Press.

Den Oudendammer, W. M., Noordhoek, J., Abma-Schouten, R. Y., Van Houtum, L., Broerse, J. E., & Dedding, C. W. 2019. Patient participation in research funding: An overview of when, why and how amongst Dutch health funds. ***Research Involvement and Engagement***, 5(1): 1-10.

Dresser, R. 1992. Wanted single, white male for medical research. ***The Hastings Center Report***, 22(1): 24-29.

ECSA. 2020. ECSA's characteristics of citizen science.

Edlin, A., Gelman, A., & Kaplan, N. 2007. Voting as a rational choice: Why and how people vote to improve the well-being of others. ***Rationality and Society***, 19(3): 293-314.

Edwards, J. R., & Bagozzi, R. P. 2000. On the nature and direction of relationships between constructs and measures. ***Psychological Methods***, 5(2): 155-174.

Epstein, S. 1995. The construction of lay expertise: AIDS activism and the forging of credibility in the reform of clinical trials. ***Science, Technology, & Human Values***, 20(4): 408-437.

Experiment.com. 2018. Experiment.com researcher guide.

Eyal, G. 2019. ***The crisis of expertise***: John Wiley & Sons.

Franzoni, C., & Stephan, P. 2021. Uncertainty and Risk-Taking in Science: Meaning, Measurement and Management, ***NBER Working Paper #28562***.

Franzoni, C., Stephan, P., & Veugelers, R. 2021. Funding Risky Research, ***NBER Working Paper***

#28905.
Franzoni, C., & Sauermann, H. 2014. Crowd Science: The organization of scientific research in open collaborative projects. ***Research Policy***, 43(1): 1-20.
Füchslin, T., Schäfer, M. S., & Metag, J. 2019. Who wants to be a citizen scientist? Identifying the potential of citizen science and target segments in Switzerland. ***Public Understanding of Science***, 28(6): 652-668.
Gallo, S. A., Sullivan, J. H., & Glisson, S. R. 2016. The influence of peer reviewer expertise on the evaluation of research funding applications. ***PLoS ONE***, 11(10): e0165147.
Ganguli, I., Huysentruyt, M., & Le Coq, C. 2021. How Do Nascent Social Entrepreneurs Respond to Rewards? A Field Experiment on Motivations in a Grant Competition. ***Management Science***.
Geoghegan, H., Dyke, A., Pateman, R., West, S., & Everett, G. 2016. Understanding motivations for citizen science: UK Environmental Observation Framework.
Golumbic, Y. N., Orr, D., Baram-Tsabari, A., & Fishbain, B. 2017. Between Vision and Reality: A Study of Scientists' Views on Citizen Science. ***Citizen Science: Theory and Practice***, 2(1).
Greenberg, J., & Mollick, E. 2017. Activist choice homophily and the crowdfunding of female founders. ***Administrative Science Quarterly***, 62(3): 341–374.
Guzman, J., Oh, J. J., & Sen, A. 2020. What Motivates Innovative Entrepreneurs? Evidence from a Global Field Experiment. ***Management Science***, 66(10): 4808-4819.
Haklay, M. 2015. Citizen science and policy: A European perspective. ***The Wodrow Wilson Center, Commons Lab***.
Haklay, M. M., Dörler, D., Heigl, F., Manzoni, M., Hecker, S., & Vohland, K. 2021. What Is Citizen Science? The challenges of definition. ***The Science of Citizen Science***: 13.
Harrison, R. T., & Mason, C. M. 2007. Does gender matter? Women business angels and the supply of entrepreneurial finance. ***Entrepreneurship Theory and Practice***, 31(3): 445-472.
Hogg, C., & Williamson, C. 2001. Whose interests do lay people represent? Towards an understanding of the role of lay people as members of committees. ***Health Expectations***, 4(1): 2-9.
Hug, S. E., & Aeschbach, M. 2020. Criteria for assessing grant applications: A systematic review. ***Palgrave Communications***, 6(1): 1-15.
INVOLVE. 2006. Peer reviewing research proposals: Guidelines for members of the public.
Irwin, A. 1995. ***Citizen Science: A Study of People, Expertise and Sustainable Development***. London and New York: Routledge.
Irwin, A. 2001. Constructing the scientific citizen: science and democracy in the biosciences. ***Public understanding of science***, 10(1): 1-18.
Jasanoff, S. 2003. Technologies of humility: Citizen participation in governing science. ***Minerva***, 41(3): 223-244.
Kahne, J., & Bowyer, B. 2017. Educating for democracy in a partisan age: Confronting the challenges of motivated reasoning and misinformation. ***American Educational Research Journal***, 54(1): 3-34.
Koning, R., Samila, S., & Ferguson, J.-P. 2020. ***Inventor Gender and the Direction of Invention***. Paper presented at the AEA Papers and Proceedings.
Lakhani, K., & Wolf, R. 2006. Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects. In J. Feller, B. Fintzgerald, S. Hissam, & K. Lakhani (Eds.), ***Perspectives on Free and Open Source Software***: MIT Press.
Lamont, M. 2009. ***How Professors Think***: Harvard University Press.
Landemore, H. 2012. ***Democratic reason***: Princeton University Press.
Lehner, O. M. 2013. Crowdfunding social ventures: a model and research agenda. ***Venture Capital***, 15(4): 289-311.
Li, D. 2017. Expertise versus Bias in Evaluation: Evidence from the NIH. ***American Economic Journal: Applied Economics***, 9(2): 60-92.
Lloyd, K., & White, J. 2011. Democratizing clinical research. ***Nature***, 474(7351): 277-278.
Mollick, E., & Nanda, R. 2016. Wisdom or madness? Comparing crowds with expert evaluation in funding the arts. ***Management Science***, 62(6): 1533-1553.
Murray, C. J., & Lopez, A. D. 2013. Measuring the global burden of disease. ***New England Journal of Medicine***, 369(5): 448-457.
Norwood, F. B., Tonsor, G., & Lusk, J. L. 2019. I will give you my vote but not my money:

Preferences for public versus private action in addressing social issues. ***Applied Economic Perspectives and Policy***, 41(1): 96-132.
Nyborg, K. 2000. Homo Economicus and Homo Politicus: interpretation and aggregation of environmental values. ***Journal of Economic Behavior & Organization***, 42(3): 305-322.
Paolacci, G., & Chandler, J. 2014. Inside the Turk: Understanding Mechanical Turk as a participant pool. ***Current Directions in Psychological Science***, 23(3): 184-188.
Payne, J. W., Bettman, J. R., & Johnson, E. J. 1993. ***The adaptive decision maker***. Cambridge ; New York, NY, USA: Cambridge University Press.
Pidgeon, N., Demski, C., Butler, C., Parkhill, K., & Spence, A. 2014. Creating a national citizen engagement process for energy policy. ***Proceedings of the National Academy of Sciences***, 111(Supplement 4): 13606-13613.
Raddick, M. J., Bracey, G., Gay, P. L., Lintott, C., Cardamone, C., Murray, P., Schawinski, K., Szalay, A., & Vandenberg, J. 2013. Galaxy Zoo: Motivations of Citizen Scientists. ***Astronomy Education Review***, 12(1).
Robinson, J., Rosenzweig, C., Moss, A. J., & Litman, L. 2019. Tapped out or barely tapped? Recommendations for how to harness the vast and largely unused potential of the Mechanical Turk participant pool. ***PloS one***, 14(12): e0226394.
Rothwell, P. 2001. The high cost of not funding stroke research: a comparison with heart disease and cancer. ***The Lancet***, 357(9268): 1612-1616.
Sauermann, H., Shafi, K., & Franzoni, C. 2019. Crowdfunding scientific research: Descriptive insights and correlates of funding success. ***PLOS ONE***.
Sauermann, H., Vohland, K., Antoniou, V., Balaz, B., Goebel, C., Karatzas, K., Mooney, P., Perello, J., Ponti, M., Samson, R., & Winter, S. 2020. Citizen science and sustainability transitions. ***Research Policy***, 49(5): 103978.
Scheufele, D. A., & Krause, N. M. 2019. Science audiences, misinformation, and fake news. ***Proceedings of the National Academy of Sciences***, 116(16): 7662-7669.
Shapiro, R. Y., & Bloch-Elkon, Y. 2008. Do the facts speak for themselves? Partisan disagreement as a challenge to democratic competence. ***Critical Review***, 20(1-2): 115-139.
United Nations. 2018. Sustainable development knowledge platform.
US Congress. 2016. Crowdsourcing and Citizen Science Act.
Van Brussel, S., & Huyse, H. 2018. Citizen science on speed? Realising the triple objective of scientific rigour, policy influence and deep citizen engagement in a large-scale citizen science project on ambient air quality in Antwerp. ***Journal of Environmental Planning and Management***: 1-18.
Von Hippel, E., & Von Krogh, G. 2015. Identifying viable "need–solution pairs": Problem solving without problem formulation. ***Organization Science***, 27(1): 207-221.
Von Krogh, G., Haefliger, S., Spaeth, S., & Wallin, M. W. 2012. Carrots and rainbows: Motivation and social practice in Open Source Software Development. ***MIS Quarterly***, 36(2): 649-676.
Weinberg, A.M. 1972. Science and trans-science. ***Science***, 177(4045):211.
West, S., & Pateman, R. 2016. Recruiting and Retaining Participants in Citizen Science: What Can Be Learned from the Volunteering Literature? ***Citizen Science: Theory and Practice***, 1(2).

APPENDIX

Social impact in the evaluation criteria of major funding agencies

Prior research has investigated the criteria used by funding agencies and peer-reviewers to evaluate grant applications (Davis & Laas, 2014; Gallo et al., 2016; Hug & Aeschbach, 2020; Li, 2017). To gain additional direct insights especially on the role of social impact expectations, we collected information from the websites and review guidelines of major funding agencies in Asia, Europe, and the North America (see Table A1). Almost all agencies include aspects related to social impact, although such aspects tend to be secondary to criteria related to scientific merit as well as the capabilities and resources of principal investigators. We also note that agencies use different terminology and definitions, which may partly reflect their different disciplinary foci or organizational goals. We built on these insights when conceptualizing "social impact" in the present article and when designing the evaluation tool used by our study participants.

Experiment.com projects

The project proposals used in this study are real projects from the platform Experiment.com. This platform allows researchers to raise crowdfunding for projects across all fields, although the environmental and life sciences are particularly well represented. Recent descriptive research shows that Principal Investigators on the platform are primarily academics but tend to be less experienced than those applying to traditional funding agencies such as NIH or ERC (Sauermann et al., 2019). Moreover, projects tend to be smaller in size, with funds raised primarily for equipment, research materials, and travel rather than salary. Most proposals have a brief introductory video and succinctly describe the project under the headings such as "About this project"; "What is the significance of this project?", and "What are the goals of this project?". In addition, proposals include a section explaining the requested budget, an anticipated timeline for the research, as well as a brief bio of the PIs. Projects also often feature "endorsements" by other researchers.

Prior to going live, projects are vetted by Experiment.com staff. This screening focuses on clarity of the exposition, scientific accuracy, as well as feasibility (see https://experiment.com/start; confirmed in personal interviews with the platform leaders). The thresholds for acceptability are quite low, such that projects can span a wide range of quality.

Table A2 provides summary information on the four projects used in this study. Activity on Experiment.com was lower than usual during 2020, likely reflecting the ongoing Covid-19 situation. Since our empirical strategy required projects to be "live" during the data collection, we collaborated with the leaders of Experiment.com to learn about upcoming projects. We chose projects that were clearly research-oriented (vs. development), were proposed by professional scientists (vs. students), and had non-trivial fundraising goals.

Coding of open-ended responses

After rating project social impact on the 5-point scale, respondents also provided their reasons in response to the question "Please briefly explain why you gave this particular rating." In a first round of coding, two evaluators coded common themes that emerged from the responses. The project team then considered these themes in light of the conceptual discussion and the quantitative results and decided on a set of smaller themes as the basis for a second round of coding. These themes and the respective frequencies will be reported in Table A4 and discussed in the main text (Section 4.3).

Additional analyses and robustness checks

Our main analysis uses linear probability models (LPM) to analyze evaluators' decisions to contribute their own funds to a project (*donate*). We prefer LPM over logistic regression because they easier to interpret and more reliable when estimating interaction terms. LPM is also appropriate in our particular context because the mean of the variable (0.39) is in the middle of the range (Angrist & Pischke, 2008). To check robustness, we estimate key models using logistic regression (Table A5, models 1-2). The results are very similar to our featured analysis.

Our main analysis showed that women rated social impact more highly than men (Table 2). To investigate potential gender differences in the weights attached to the three project criteria, we estimated regressions of *recommend* and *donate* while also including interactions between project criteria and *female* (Table A5, models 3-4). None of the interactions is significant.

A challenge with Mturk studies is that the quality of data can be low (Buhrmester et al., 2016; Paolacci & Chandler, 2014). Following best practices, we used attention check questions as well as manual review of open-ended answers to eliminate cases from respondents who clearly did not pay attention (e.g., failed all attention checks) or who are likely to have been bots rather than humans (e.g., filled open entry fields by copying long text from the internet). However, our sample may still include respondents who spent little time reviewing the research proposal and thinking about their answers to our questions. To eliminate such cases, we estimated key regressions dropping those respondents who completed the study in less than 10 minutes (15.20% of the main sample; the average time taken by respondents in this study was 18.2 minutes). The results are virtually unchanged (Table A5, models 5-9).

Table A1: Overview of evaluation criteria used by major funding agencies

Country	Agency	Key criteria	Subcriteria
USA	National Institutes of Health (NIH)	• Overall impact score arising from scored review criteria and additional review criteria	Scored review criteria (scored individually) include: • significance, (assuming that will succeed) • investigator(s), • innovation, • approach • environment Additional review criteria (not scored individually but considered in overall impact score) depend on the project type. overall impact is some combination with free weights
USA	National Science Foundation (NSF)	• Intellectual merit and broader impacts	• What is the potential for the proposed activity to: a. Advance knowledge and understanding within its own field or across different fields (Intellectual Merit); and b. Benefit society or advance desired societal outcomes (Broader Impacts)? WITHIN both of the main ones, consider • To what extent do the proposed activities suggest and explore creative, original, or potentially transformative concepts? • Is the plan for carrying out the proposed activities well-reasoned, well-organized, and based on a sound rationale? Does the plan incorporate a mechanism to assess success? • How well qualified is the individual, team, or organization to conduct the proposed activities? • Are there adequate resources available to the PI (either at the home organization or through collaborations) to carry out the proposed activities?
Canada	Social Sciences and Humanities Research Council	• Challenge—The aim and importance of the endeavour (40%) • Feasibility—The plan to achieve excellence (20%) • Capability—The expertise to succeed (40%)	CHALLENGE • originality, significance and expected contribution to knowledge; • appropriateness of the literature review; • appropriateness of the theoretical approach or framework; • appropriateness of the methods/approach; • quality of training and mentoring to be provided to students, emerging scholars and other highly qualified personnel, and opportunities for them to contribute; and • potential for the project results to have influence and impact within and/or beyond the social sciences and humanities research community. FEASIBILITY • appropriateness of the proposed timeline, and probability that the objectives will be met; • expertise of the applicant or team in relation to the proposed research; • appropriateness of the requested budget, justification of proposed costs, and, where applicable, other cash and/or in-kind contributions; and • quality and appropriateness of knowledge mobilization plans, including effective dissemination, exchange and engagement with stakeholders within and/or beyond the research community, where applicable. CAPABILITY • quality, quantity and significance of past experience and published and/or creative outputs of the applicant and any co-applicants, relative to their roles in the project and to the stage of their career; • evidence of past knowledge mobilization activities (e.g., films, performances, commissioned reports, knowledge syntheses, experience in collaboration /
EU	European Research Council	• Scientific excellence	Excellence of the research project (see definitions via guiding questions) • Ground breaking nature • Potential impact • Scientific approach Excellence of the principal investigator • Intellectual capacity • Creativity • Commitment
UK	Engineering and Physical Sciences Research Council	• Quality (primary criterion) • Importance (secondary major criterion) • Applicant and partnerships (secondary criterion) • Resource and management (secondary criterion)	QUALITY • The novelty, relationship to the context, timeliness and relevance to identified stakeholders. • The ambition, adventure, transformative aspects or potential outcomes. • The suitability of the proposed methodology and the appropriateness of the approach to achieving impact. IMPORTANCE • Contributes to, or helps maintain the health of other disciplines contributes to addressing key UK societal challenges and/or contributes to future UK economic success and development of emerging industry(s). • Meets national needs by establishing/maintaining a unique world leading activity. • Complements other UK research funded in the area, including any relationship to the EPSRC portfolio. APPLICANT AND PARTNERSHIPS
Japan	Japanese Society for the Promotion of Science	• Academic importance of research project. • Validity of research method. • Appropriateness of ability to conduct research and research environment	Subcriteria depend on the category of the research. ACADEMIC IMPORTANCE OF RESEARCH PROJECT • Is it an important research project to be promoted from the academic point of view? •Is the "key scientific question" comprising the core of the research plan clear, and scientific significance, and originality recognized? •Is it clear that the history leading to the conception of the research plan and domestic and overseas trends related to the proposed research and the positioning of this research in the relevant field? • Can we expect an effect the wave to a wider academic, scientific, technological or society by conducting this research project? VALIDITY OF RESEARCH METHOD • Is the research method concrete and appropriate in order to achieve its research objective? Also, do the research expenditure ensure consistency with the research plan? • Is the preparation status appropriate in order to achieve its research objective? APPROPRIATENESS OF ABILITY TO CONDUCT RESEARCH (AND RESEACH ENVIRONMENT) • Does it possess sufficient ability to conduct the research plan based on research activity over the past years? • (Have the research environment been arranged by the research facilities, equipment, research materials, etc.
China	National Natural Science Foundation of China		Article 15. With respect to an application for funded projects, an evaluation expert shall render an independent judgment and evaluation in terms of scientific value, innovation, social influences as well as the feasibility of research schemes, and bring forward the evaluation opinions thereof. When giving the evaluation opinions about the applications for funded projects, an evaluation expert shall also consider the research experiences of the applicants and participants, the rationality of the plan for using the funds, the information on other grants for the research topics, the information on the implementation of the funded projects by the applicants as well as the necessity for continuous subsidies. The evaluation opinions as brought forward through the meeting-based evaluation shall be determined by voting.

Table A2: Overview of research projects evaluated by study participants

Proposal Title & Project Page	Abstract	Principal Investigators	Funding Goal	Data Collected
Does sexual orientation matter as much as sex for economic preferences? https://experiment.com/projects/does-sexual-orientation-matter-as-much-as-a-sex-for-economic-preferences	An influential economic hypothesis links men's higher compensation to a willingness to compete higher than for women. We propose to run a series of behavioral experiments to test whether sexual orientation exerts a distinct role from sex and gender on the development of economic preferences, possibly contributing to explanation of the gay-penalty and the lesbian-premium in labor markets.	Ryan McWay and Dr. Alessandra Cassar (University of San Francisco)	$4,000	April 24-April 28, 2020
The influence of rabies and otter behavior on otter-human conflict https://experiment.com/projects/otter-human-conflict-the-influence-of-rabies-and-otter-behavior	North American river otters are found throughout Florida but few studies have focused on the health of this species in FL. While most people consider them playful animals, there are occasional conflicts with humans including bites on people and pets that are often attributed to a rabies infection or defensive behavior. Our project will look at the incidence of otter-human conflict in Florida and what factors may be important. We will also investigate reports of rabies infections in otters.	Megan Stolen, MSc. (Hubbs-SeaWorld Research Institute)	$4,300	Sept. 27-Sept.28, 2020
Target-based drug discovery for coronavirus disease 2019 https://experiment.com/projects/target-based-drug-discovery-for-coronavirus-disease-2019	Therapeutics in any modality to combat Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) infections are in urgent demand across the entire world as the Coronavirus Disease 2019 (COVID-19) pandemic continues to interrupt the well-being of human life. Without clinical therapeutics available for any of the human coronaviruses, a key starting point is to focus on target-based high-throughput screening (HTS) to reveal potent anti-SARS-CoV-2 hit-to-leads that are safe to human cells.	Dr. Edward D'Antonio (University of South Carolina) and Dr. Gustavo Fernando Mercaldi (Brazilian Biosciences National Laboratory)	$25,000	Oct. 20-Nov.1, 2020

Could Alzheimer's disease be triggered by bacterial DNA? https://experiment.com/projects/could-alzheimer-s-disease-be-triggered-by-bacterial-dna	Alzheimer's disease (AD) is associated with prion-like aggregation of the β-amyloid and Tau proteins, that form neurotoxic aggregates in the brain. Studies have pointed out the association between brain-localized bacteria and AD; however. Recently, we identified that bacterial DNA can aggregate Tau protein (Tetz et al, 2020). We will explore whether DNA from brain-localized bacteria triggers β-amyloid misfolding; thus, being, a previously overlooked cause of AD.	Dr. George Tetz (Human Microbiology Institute)	$4,900	Dec. 9-Dec.11, 2020

Table A3: Correlations between key variables

	recommend	donate	z soc impact	z sci merit	z team cap	personal exp	female	age	education	income
recommend	1									
donate	0.4350*	1								
z social impact	0.5009*	0.2683*	1							
z scientific merit	0.5547*	0.2701*	0.5147*	1						
z team capabilities	0.4286*	0.1929*	0.3230*	0.3915*	1					
personal experience	0.1410*	0.0758*	0.0508*	0.0199	0.0438*	1				
female	0.0181	0.0605*	0.0660*	0.0174	0.0247	0.0483*	1			
age	-0.0172	0.1051*	-0.0067	-0.0427*	-0.0155	0.0386	0.0766*	1		
education	-0.0301	0.0966*	0.03	0.0236	0.0367	0.0072	0.0326	0.1008*	1	
income	-0.0410*	0.0535*	0.0012	-0.0086	0.0099	-0.0188	-0.0434*	0.0531*	0.3342*	1

Table A4: Coded reasons for social impact ratings

To be completed.

Table A5: Additional analyses and robustness checks

	Full sample		Full sample		drop if duration <10				
	1	2	3	4	5	6	7	8	9
	logit	logit	ologit	LPM	ologit	ologit	LPM	LPM	OLS
VARIABLES	donate	donate	recommend	donate	recommend	recommend	donate	donate	z soc impact
education: bachelors	0.382**	0.382**	-0.189	0.069**	-0.102	-0.222	0.078*	0.072**	0.043
	[0.137]	[0.118]	[0.143]	[0.022]	[0.117]	[0.145]	[0.029]	[0.023]	[0.078]
education: grad degree	0.568**	0.562**	-0.307	0.106**	-0.099	-0.343	0.126**	0.115**	0.015
	[0.136]	[0.132]	[0.189]	[0.027]	[0.177]	[0.184]	[0.034]	[0.031]	[0.094]
income: 25-50k	0.386*	0.414**	-0.259	0.072*	-0.147	-0.328	0.076	0.066	0.043
	[0.153]	[0.159]	[0.204]	[0.029]	[0.177]	[0.234]	[0.043]	[0.038]	[0.075]
income: 51-75k	0.433**	0.482**	-0.248	0.086**	-0.109	-0.197	0.091**	0.088**	-0.012
	[0.098]	[0.066]	[0.153]	[0.015]	[0.120]	[0.150]	[0.024]	[0.019]	[0.099]
income: 76-100k	0.493**	0.467**	-0.068	0.087**	0.029	-0.136	0.103**	0.085**	0.110
	[0.129]	[0.118]	[0.164]	[0.024]	[0.135]	[0.194]	[0.028]	[0.026]	[0.072]
income: over 100k	0.336*	0.434**	-0.344*	0.078**	-0.336	-0.330*	0.076	0.082**	-0.020
	[0.138]	[0.107]	[0.148]	[0.021]	[0.202]	[0.159]	[0.037]	[0.025]	[0.128]
personal experience	0.212**	0.171	0.239**	0.036	0.242**	0.234*	0.041*	0.032	0.080
	[0.065]	[0.091]	[0.090]	[0.018]	[0.093]	[0.104]	[0.016]	[0.019]	[0.041]
z social impact		0.360**	0.671**	0.070**		0.765**		0.071**	
		[0.072]	[0.133]	[0.017]		[0.098]		[0.016]	
z scientific merit		0.337**	0.984**	0.061**		0.944**		0.074**	
		[0.060]	[0.070]	[0.018]		[0.078]		[0.011]	
z team capabilities		0.161**	0.638**	0.018		0.599**		0.031**	
		[0.051]	[0.097]	[0.012]		[0.085]		[0.010]	
z social impact X female			0.158	-0.000					
			[0.095]	[0.026]					
z scientific merit X female			-0.048	0.015					
			[0.076]	[0.021]					
z team capabilities X female			-0.082	0.022					
			[0.119]	[0.018]					
female	0.260*	0.193	-0.012	0.039	0.146	0.008	0.051	0.034	0.133**
	[0.120]	[0.118]	[0.051]	[0.023]	[0.084]	[0.062]	[0.029]	[0.027]	[0.030]
age	0.149**	0.176**	0.039	0.037**	-0.020	0.048	0.027*	0.034*	-0.023
	[0.053]	[0.062]	[0.045]	[0.012]	[0.034]	[0.049]	[0.012]	[0.012]	[0.025]
donated before	-0.015	-0.059	0.086	-0.011	0.073	0.059	0.003	-0.006	0.099*
	[0.131]	[0.111]	[0.128]	[0.022]	[0.143]	[0.125]	[0.030]	[0.026]	[0.042]
project clarity	0.494**	0.291**	0.338**	0.058**	0.708**	0.338**	0.109**	0.059**	0.245**
	[0.038]	[0.032]	[0.047]	[0.007]	[0.071]	[0.056]	[0.008]	[0.008]	[0.028]
reward	0.010**	0.012**	0.003	0.002**	0.002	0.003	0.002**	0.002*	0.002
	[0.002]	[0.002]	[0.005]	[0.001]	[0.006]	[0.005]	[0.001]	[0.001]	[0.003]
project-day FE	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.
Constant	-5.973**	-5.574**		-0.561**			-0.704**	-0.478**	-1.445**
	[0.358]	[0.397]		[0.094]			[0.117]	[0.114]	[0.431]
Observations	2,269	2,269	2,269	2,269	1,929	1,929	1,929	1,929	1,929
R-squared				0.165			0.098	0.170	0.070
r2_p	0.0774	0.135	0.288		0.0913	0.283			

*=5%, **=1%. Standard errors clustered at level of project-day

Appendix references

Aguinis, H., Villamor, I., & Ramani, R. S. 2020. MTurk Research: Review and Recommendations: SAGE Publications Sage CA: Los Angeles, CA.

Angrist, J. D., & Pischke, J.-S. 2008. ***Mostly harmless econometrics: An empiricist's companion***: Princeton University Press.

Arechar, A. A., Kraft-Todd, G. T., & Rand, D. G. 2017. Turking overtime: how participant characteristics and behavior vary over time and day on Amazon Mechanical Turk. ***Journal of the Economic Science Association***, 3(1): 1-11.

Beck, S., Brasseur, T.-M., Poetz, M. K., & Sauermann, H. 2020. What is the problem? Crowdsourcing research questions in science, ***Working Paper***.

Bergstrom, T. C. 2006. Benefit-cost in a benevolent society. ***American Economic Review***, 96(1): 339-351.

Blind, K., Cuhls, K., & Grupp, H. 2001. Personal attitudes in the assessment of the future of science and technology: A factor analysis approach. ***Technological Forecasting and Social Change***, 68(2): 131-149.

Bloodhart, B., & Swim, J. K. 2020. Sustainability and consumption: What's gender got to do with it? ***Journal of Social Issues***, 76(1): 101-113.

Bonney, R., Shirk, J. L., Phillips, T. B., Wiggins, A., Ballard, H. L., Miller-Rushing, A. J., & Parrish, J. K. 2014. Next steps for Citizen Science. ***Science***, 343(6178): 1436-1437.

Briley, A., & Singh, C. 2017. Patient involvement in medical research. ***Introduction to Research Methodology for Specialists and Trainees***: 150.

Buhrmester, M., Kwang, T., & Gosling, S. D. 2016. Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data?

Buhrmester, M. D., Talaifar, S., & Gosling, S. D. 2018. An evaluation of Amazon's Mechanical Turk, its rapid rise, and its effective use. ***Perspectives on Psychological Science***, 13(2): 149-154.

Callon, M., & Rabeharisoa, V. 2008. The growing engagement of emergent concerned groups in political and economic life: Lessons from the French association of neuromuscular disease patients. ***Science, Technology, & Human Values***, 33(2): 230-261.

Caron-Flinterman, J. F., Broerse, J. E., & Bunders, J. F. 2005. The experiential knowledge of patients: a new resource for biomedical research? ***Social science & medicine***, 60(11): 2575-2584.

Cohen, W. M., Sauermann, H., & Stephan, P. 2020. Not in the job description: The commercial activities of academic scientists and engineers. ***Management Science***, 66(9): 4108-4117.

Collyar, D. 2005. How have patient advocates in the United States benefited cancer research? ***Nature Reviews Cancer***, 5(1): 73-78.

Davis, M., & Laas, K. 2014. "Broader impacts" or "responsible research and innovation"? A comparison of two criteria for funding research in science and engineering. ***Science and engineering ethics***, 20(4): 963-983.

de Graaf, B., Rinnooy Kan, A., & Molenaar, H. 2017. ***The Dutch National research agenda in perspective: A reflection on research and science policy in practice***: Amsterdam University Press.

Den Oudendammer, W. M., Noordhoek, J., Abma-Schouten, R. Y., Van Houtum, L., Broerse, J. E., & Dedding, C. W. 2019. Patient participation in research funding: An overview of when, why and how amongst Dutch health funds. ***Research Involvement and Engagement***, 5(1): 1-10.

Dresser, R. 1992. Wanted single, white male for medical research. ***The Hastings Center Report***, 22(1): 24-29.

ECSA. 2020. ECSA's characteristics of citizen science.

Edlin, A., Gelman, A., & Kaplan, N. 2007. Voting as a rational choice: Why and how people vote to improve the well-being of others. ***Rationality and Society***, 19(3): 293-314.

Edwards, J. R., & Bagozzi, R. P. 2000. On the nature and direction of relationships between constructs and measures. ***Psychological Methods***, 5(2): 155-174.

Epstein, S. 1995. The construction of lay expertise: AIDS activism and the forging of credibility in the reform of clinical trials. ***Science, Technology, & Human Values***, 20(4): 408-437.
Experiment.com. 2018. Experiment.com researcher guide.
Franzoni, C., & Sauermann, H. 2014. Crowd Science: The organization of scientific research in open collaborative projects. ***Research Policy***, 43(1): 1-20.
Franzoni, C., & Stephan, P. 2021. Uncertainty and Risk-Taking in Science: Meaning, Measurement and Management, ***NBER Working Paper #28562***.
Franzoni, C., Stephan, P., & Veugelers, R. 2021. Funding Risky Research, ***NBER Working Paper #28905***.
Füchslin, T., Schäfer, M. S., & Metag, J. 2019. Who wants to be a citizen scientist? Identifying the potential of citizen science and target segments in Switzerland. ***Public Understanding of Science***, 28(6): 652-668.
Gallo, S. A., Sullivan, J. H., & Glisson, S. R. 2016. The influence of peer reviewer expertise on the evaluation of research funding applications. ***PLoS ONE***, 11(10): e0165147.
Ganguli, I., Huysentruyt, M., & Le Coq, C. 2021. How Do Nascent Social Entrepreneurs Respond to Rewards? A Field Experiment on Motivations in a Grant Competition. ***Management Science***.
Geoghegan, H., Dyke, A., Pateman, R., West, S., & Everett, G. 2016. Understanding motivations for citizen science: UK Environmental Observation Framework.
Golumbic, Y. N., Orr, D., Baram-Tsabari, A., & Fishbain, B. 2017. Between Vision and Reality: A Study of Scientists' Views on Citizen Science. ***Citizen Science: Theory and Practice***, 2(1).
Guzman, J., Oh, J. J., & Sen, A. 2020. What Motivates Innovative Entrepreneurs? Evidence from a Global Field Experiment. ***Management Science***, 66(10): 4808-4819.
Haklay, M. 2015. Citizen science and policy: A European perspective. ***The Wodrow Wilson Center, Commons Lab***.
Haklay, M. M., Dörler, D., Heigl, F., Manzoni, M., Hecker, S., & Vohland, K. 2021. What Is Citizen Science? The challenges of definition. ***The Science of Citizen Science***: 13.
Harrison, R. T., & Mason, C. M. 2007. Does gender matter? Women business angels and the supply of entrepreneurial finance. ***Entrepreneurship Theory and Practice***, 31(3): 445-472.
Hogg, C., & Williamson, C. 2001. Whose interests do lay people represent? Towards an understanding of the role of lay people as members of committees. ***Health Expectations***, 4(1): 2-9.
Hug, S. E., & Aeschbach, M. 2020. Criteria for assessing grant applications: A systematic review. ***Palgrave Communications***, 6(1): 1-15.
INVOLVE. 2006. Peer reviewing research proposals: Guidelines for members of the public.
Irwin, A. 1995. ***Citizen Science: A Study of People, Expertise and Sustainable Development***. London and New York: Routledge.
Irwin, A. 2001. Constructing the scientific citizen: science and democracy in the biosciences. ***Public understanding of science***, 10(1): 1-18.
Jasanoff, S. 2003. Technologies of humility: Citizen participation in governing science. ***Minerva***, 41(3): 223-244.
Kahne, J., & Bowyer, B. 2017. Educating for democracy in a partisan age: Confronting the challenges of motivated reasoning and misinformation. ***American Educational Research Journal***, 54(1): 3-34.
Koning, R., Samila, S., & Ferguson, J.-P. 2020. ***Inventor Gender and the Direction of Invention***. Paper presented at the AEA Papers and Proceedings.
Lakhani, K., & Wolf, R. 2006. Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects. In J. Feller, B. Fintzgerald, S. Hissam, & K. Lakhani (Eds.), ***Perspectives on Free and Open Source Software***: MIT Press.
Lamont, M. 2009. ***How Professors Think***: Harvard University Press.
Landemore, H. 2012. ***Democratic reason***: Princeton University Press.
Lehner, O. M. 2013. Crowdfunding social ventures: a model and research agenda. ***Venture Capital***, 15(4): 289-311.

Lench, H. C., Smallman, R., & Berg, L. A. 2016. Moving toward a brighter future: The effects of desire on judgments about the likelihood of future events. ***Motivation Science***, 2(1): 33.
Li, D. 2017. Expertise versus Bias in Evaluation: Evidence from the NIH. ***American Economic Journal: Applied Economics***, 9(2): 60-92.
Lloyd, K., & White, J. 2011. Democratizing clinical research. ***Nature***, 474(7351): 277-278.
Norwood, F. B., Tonsor, G., & Lusk, J. L. 2019. I will give you my vote but not my money: Preferences for public versus private action in addressing social issues. ***Applied Economic Perspectives and Policy***, 41(1): 96-132.
Nyborg, K. 2000. Homo Economicus and Homo Politicus: interpretation and aggregation of environmental values. ***Journal of Economic Behavior & Organization***, 42(3): 305-322.
Paolacci, G., & Chandler, J. 2014. Inside the Turk: Understanding Mechanical Turk as a participant pool. ***Current Directions in Psychological Science***, 23(3): 184-188.
Payne, J. W., Bettman, J. R., & Johnson, E. J. 1993. ***The adaptive decision maker***. Cambridge ; New York, NY, USA: Cambridge University Press.
Pidgeon, N., Demski, C., Butler, C., Parkhill, K., & Spence, A. 2014. Creating a national citizen engagement process for energy policy. ***Proceedings of the National Academy of Sciences***, 111(Supplement 4): 13606-13613.
Raddick, M. J., Bracey, G., Gay, P. L., Lintott, C., Cardamone, C., Murray, P., Schawinski, K., Szalay, A., & Vandenberg, J. 2013. Galaxy Zoo: Motivations of Citizen Scientists. ***Astronomy Education Review***, 12(1).
Robinson, J., Rosenzweig, C., Moss, A. J., & Litman, L. 2019. Tapped out or barely tapped? Recommendations for how to harness the vast and largely unused potential of the Mechanical Turk participant pool. ***PloS one***, 14(12): e0226394.
Rothwell, P. 2001. The high cost of not funding stroke research: a comparison with heart disease and cancer. ***The Lancet***, 357(9268): 1612-1616.
Sauermann, H., Shafi, K., & Franzoni, C. 2019. Crowdfunding scientific research: Descriptive insights and correlates of funding success. ***PLOS ONE***.
Sauermann, H., Vohland, K., Antoniou, V., Balaz, B., Goebel, C., Karatzas, K., Mooney, P., Perello, J., Ponti, M., Samson, R., & Winter, S. 2020. Citizen science and sustainability transitions. ***Research Policy***, 49(5): 103978.
Scheufele, D. A., & Krause, N. M. 2019. Science audiences, misinformation, and fake news. ***Proceedings of the National Academy of Sciences***, 116(16): 7662-7669.
Shapiro, R. Y., & Bloch-Elkon, Y. 2008. Do the facts speak for themselves? Partisan disagreement as a challenge to democratic competence. ***Critical Review***, 20(1-2): 115-139.
United Nations. 2018. Sustainable development knowledge platform.
US Congress. 2016. Crowdsourcing and Citizen Science Act.
Van Brussel, S., & Huyse, H. 2018. Citizen science on speed? Realising the triple objective of scientific rigour, policy influence and deep citizen engagement in a large-scale citizen science project on ambient air quality in Antwerp. ***Journal of Environmental Planning and Management***: 1-18.
Von Hippel, E., & Von Krogh, G. 2015. Identifying viable "need–solution pairs": Problem solving without problem formulation. ***Organization Science***, 27(1): 207-221.
Von Krogh, G., Haefliger, S., Spaeth, S., & Wallin, M. W. 2012. Carrots and rainbows: Motivation and social practice in Open Source Software Development. ***MIS Quarterly***, 36(2): 649-676.
West, S., & Pateman, R. 2016. Recruiting and Retaining Participants in Citizen Science: What Can Be Learned from the Volunteering Literature? ***Citizen Science: Theory and Practice***, 1(2).