Research as Leisure: Experimental Evidence on Voluntary Contributions to Science *

Working Draft

Please do not cite without permission

Elizabeth Lyons[†] Laurina Zhang[‡]

February 16, 2018

Abstract

Organizations that depend on voluntary contributions of time and money face unique managerial challenges. In this paper, we investigate the impact of two distinct approaches organizations can use to attract and motivate impure altruistic volunteer labor using both a field and survey experiment on voluntary contributions to science. In both experiments, we examine whether the salience of project outputs (i.e., project outcome) or project inputs (i.e., research hours) affect the quantity and quality of contributions. We find that increasing the salience of both input and output value decreases voluntary participation, but increases the match quality between the task and the volunteer. Furthermore, we find that individuals that select out of volunteering in response to the type of information provided substitute volunteering time by donating money from wage work.

Key words: Non-pecuniary Incentives; Voluntary Contributions; Experiment; Crowd Science; Crowdsourcing

*The authors thank Grant Miller, and the rest of the Zooniverse team for assistance in implementing the study. The authors gratefully acknowledge funding and research support from the NBER's Innovation Policy Grant and the Policy Design and Evaluation Lab at UC San Diego. We are also grateful for helpful feedback from Shane Greenstein, Milan Miric, Laura Rosendahl-Huber, Henry Sauermann, and conference participants and at REER, the IGL Winter 2017 Meeting, and the NBER Productivity Lunch. Eddy Chebelyon provided excellent research assistance. This research is approved by UCSD's Research Ethics Board, project number 171711.

†School of Global Policy and Strategy, University of California, San Diego, 9500 Gilman Drive, MC 0519 La Jolla, CA 92093-0519. lizlyons@ucsd.edu

‡Scheller College of Business, Georgia Institute of Technology, 800 W Peachtree St NW, Atlanta GA 30308. laurina.zhang@scheller.gatech.edu

1 Introduction

Organizations that depend on voluntary contributions of time and money face unique managerial challenges. Unlike incentive designs for paid labor, these organizations have to attract and manage volunteers exclusively through non-pecuniary incentives. Given that many contributors are not purely altruistic (Andreoni, 1990), organizations may be able to alter their behavior by emphasizing the association between their motivations for contributing and the organizations' desired contribution (e.g., Lacetera and Macis, 2010). However, evidence on the relationship between non-pecuniary incentives and contribution efficiency demonstrates important trade-offs between the quantity and quality of contributions across incentive schemes (Grant, 2008; Lilley and Slonim, 2014).

In this paper, we investigate the performance of two distinct approaches organizations can use to attract and motivate impure altruistic volunteer labor. In particular, we study the quantity and quality of volunteer hours donated when a task is advertised as having a high output value (i.e., completing the task can lead to important social welfare gains), and when it is advertised as having a high input value (i.e. completing the task will save the organization labor costs) relative to a neutral task advertisement. We study this research question by running a randomized control trial (RCT) on the largest crowd science platform in the world, Zooniverse, which is an increasingly important field setting for voluntary contributions. An attractive feature of this platform is that other forms of extrinsic and reputational motivations are muted.

Our experiment design is motivated by the theory proposed in Andreoni et al. (1996) in which the utility people get from voluntary activities depends both on the act of giving (i.e., time or money) and on how much the recipient values what is given. In particular, individuals may contribute voluntarily to scientific discovery if the output of the task is framed more meaningfully (Ariely et al., 2008; Chandler and Kapelner, 2013; Kosfeld and Neckermann, 2011)[1] or if their labor supply is more valuable than the money they could donate. In other words, volunteers contribute because of their beliefs about the expected value of the output or input of their contributions.

[1] For instance, Chandler and Kapelner (2013) hired M-Turk workers to label tumor cells in order to assist medical researchers. Some workers were explicitly told the purpose of their task was to help researchers identify tumor cells, while other workers were not given any reason for their work. The authors found that when the task was framed more meaningfully, workers were more likely to participate and, conditionally on participating, they labelled a higher quantity of images.

We build on this theory by combining it with intuition proposed in Cassar and Meier (2016) to better fit our research setting. In our framework, we show that providing information about the value of contributor output or the value of contributor input have theoretically ambiguous impacts on effort, as the effort exerted will depend on whether the additional information leads to upward or downward changes in beliefs about the value of their output and input.

In our RCT, we examine whether emphasizing the role of contributions in affecting a scientific outcome (i.e., output) and emphasizing the role of contributions in terms of giving time (i.e., input) impacts the quantity and quality of crowd science contributions on Zooniverse. Zooniverse has three features that makes it an attractive setting for our study. First, it is economically important. While the concept of crowd science is not new[2], the number and scale of crowd science projects have sharply increased in the past decade primarily because the Internet has lowered the cost of aggregating and comparing results from different participants. These activities have contributed to a number of high-profile publications in scientific outlets such as Nature and Proceedings of the National Academy of Sciences (PNAS), and have been shown to significantly reduce the cost of doing research (Sauermann and Franzoni, 2014).[3] To further capitalize on these potential gains, policy makers are seeking ways to expand participation in crowd science and to better understand its implications (Haklay, 2015).[4] Second, each contributor is uniquely identified on Zooniverse and each contribution is time-stamped, which allows us to accurately trace contributions by each contributor. Third, and most importantly, contributors are anonymous and do not see each other's contributions (unlike Wikipedia). Thus, there is little reason that contributors volunteer for reputation or social reasons. This allows us to isolate two potential reasons contributors receive 'warm glow' returns from volunteering.

[2]The first recorded example of the use of the term is from 1989, describing how 225 volunteers across the US collected rain samples to assist the Audubon Society in an acid-rain awareness raising campaign. The volunteers collected samples, checked for acidity, and reported back to the organization (Haklay, 2015).

[3]These projects range from classifying images of galaxies (Galaxy Zoo) to modifying the shape of a visual 3D model of protein to optimize its shape (Foldit). Sauermann and Franzoni (2014) tracked seven crowd science projects over a 180-day study period and found that 100,386 participants contributed a total of 129,500 hours of unpaid labor, which amounted to more than $1.5 million of help assuming the rate normally paid to undergraduate students ($12/h).

[4]Eye on Earth Summit, which was part of the United Nations Environmental Programme (UNEP) explicitly called for "establishing citizen science as an important source of knowledge within the diversity of knowledge communities" for environmental policy. The UN Education, Scientific, and Cultural Organization (UNESCO) identified it as an important area within the agenda of Information and Communication Technologies (ICT) use and called on UN members to "encourage the use of ICTs, including the Internet and mobile technologies, to facilitate greater participation in the entire scientific process including public participation in scientific research (citizen science) activities." Similar recognition can be seen in a white paper on citizen science in Europe that was supported by the digital science unit of the European Commission. This is also echoed in the US Open Government National Action Plan which commits to extend the use of citizen science.

To eliminate concerns about participant selection into a particular project based on an experimental treatment, we created three similar project pages on the Zooniverse platform – one control page and two treatment pages that differ only in their emphasis on the value of the individual's output and input respectively. Each version is emailed to a randomly selected subset of Zooniverse participants who are invited to contribute to the projects. These participants have all indicated their willingness to contribute to new pages and, therefore, are likely more willing to contribute to crowd science then the general population. In order to ensure participants are not aware that there are several versions of the same project (i.e., different experimental treatments), only those who are invited to contribute are able to view the respective pages. This design allows us to test how framing the scientific task based on input and output value of contributions affects selection into the task and contribution outcomes.

Mean outcome comparisons between treatment and control pages demonstrate that first, consistent with our theoretical predictions, we find that the treatments impact the types of people that contribute to the project. The average contributor on the project page that emphasizes the value of contributors' inputs are less active on Zooniverse, less likely to indicate their abilities motivate them to contribute to Zooniverse, and are less likely to have science related work experience compared to those who contribute to the output and control pages. These patterns suggest that individuals who contribute to the input page perceive the value of their labor supply as relatively low and are induced to engage when they learn that their labor supply is valuable relative to the organization's outside option. Second, the treatments also affect the quantity and quality of work completed. Both treatments lead to a reduction in the number of classifications each contributor makes relative to the control page. However, once we weight classifications for quality (measured by the accuracy of classifications against "external" classifications), the differences between number of contributions across pages become insignificant. In other words, while increasing the amount of information about input and output value decreases the intensity of participation, information about input and output value increases the accuracy of classifications.

While our experiment design allows us to test how framing the scientific task based on input and output value of contributions affects selection into the task and contribution outcomes, it is unclear whether voluntary contributions are impacted by individuals' outside options as predicted by Andreoni et al. (1996). Moreover, we are not able to link individual demographic data with individual contributors, nor are we able to observe how the population of those who select into crowd science platforms differs from those who do not. To address these limitations, we complement our RCT with a survey experiment in which respondents are randomly assigned high or low outside options and an output or input-based volunteer task motivation. Respondents are asked whether they prefer to work for their outside option and donate one hour of their earnings to the science project, volunteer an hour of their time to the science project, or engage in leisure activities.

Consistent with our theoretical prediction, we find that survey respondents whose prior beliefs about the value of their scientific contribution were low are more likely to volunteer given an output-motivated task. However, input-motivated tasks do not appear to disproportionately increase the likelihood of volunteering from respondents who had low initial beliefs about their ability. Instead, we find evidence that those who select out of volunteering due to the type of task information provided choose to donate money from wage work. Specifically, individuals with high perceived ability and high outside options are disproportionately more likely to work and donate money rather than volunteer their time. Taken together, our findings from both experiments suggest that voluntary contributions, whether time or money, can be affected by non-pecuniary motivations and also depend on individuals' beliefs about the project outcome, the value of their labor, and their outside options.

Our study makes several distinct contributions. First, we provide the first experimental study on the differential effects of motivating a voluntary task through the value of contributors' inputs or outputs on the intensity and quality of participation. While the theoretical effect of these types of information provision is ambiguous ex-ante, we show that crowd science outcomes can be impacted based on whether the information changes beliefs about the value of the individual's output or the value of the individual's labor to the organization. While the fact that certain types of people may be willing to give up economic rewards for science is not a new insight (Stern, 2004), it is still puzzling that individuals are investing time that could otherwise be spent on paid work or leisure into voluntary scientific research. Recent studies have documented that image motivation (Ariely et al., 2009; Gallus et al., 2016), social effects (e.g., Zhang and Zhu, 2011), competition (e.g., Boudreau et al., 2011, 2016), and the composition of tasks (e.g., Lakhani and Von Hippel, 2003) all impact voluntary contributions. In addition, Bénabou and Tirole (2006) show that extrinsic incentives can crowd out pro-social behavior. Our setting allows us to test the effect of two distinct intrinsic motivations in a setting where other forms of extrinsic and reputational motivations (e.g., competition, social, image) are absent. Moreover, our survey findings add nuance to the relationship between non-pecuniary motivations and voluntary activity by showing that outside options can impact not only the type of individuals that engage in voluntary activity, but also whether they donate time or money.

Our results yield novel insights for organizations that use voluntary labor. For instance, while some routine tasks simply require a large number of participants to generate a solution due to the sample size requirements of some algorithms (e.g., Kelling et al., 2015), more complex tasks require knowledge diversity and specialized skills (Terwiesch and Xu, 2008; Boudreau et al., 2011)[5]. Our results suggest that there are trade offs associated with providing information about the expected value of the output or input of the task. While additional information on the value of the contributor output or input can lead to improvements in the quality of the outcomes, the improved matching between the volunteer and the task may come at the expense of a reduced number of contributions. Thus, organizations that value contributions from individuals with specific backgrounds and skills may benefit from providing additional information on the value of the task input or output. Alternatively, if the organization's main objective is to generate a large number of contributions and quality considerations are minimal, it may be optimal to avoid placing emphasis on the expected value of contributor output or input. However, our survey evidence suggests those who select out due to the mismatch between expectations and reality may compensate through monetary donations, potentially improving the overall allocation of labor and resources.

The paper proceeds as follows. The next section provides a theoretical framework to motivate the experimental design. Section 3 describes Zooniverse and the experimental setup. Sections 4 and 5 present the data and results, respectively. Section 6 presents the survey experiment design

[5]The tasks the crowd performs vary from routine, well-understood tasks that can be broken into a series of linear steps that include a defined range of acceptable solutions (e.g., tagging images, improving search results), to non-routine, complex tasks (e.g., generating product ideas, solving complex problems) that can be approached in different ways (Perrow, 1967; Howe, 2008; Lindley, 2009; Jeppesen and Lakhani, 2010).

and findings. Section 7 offers concluding remarks and next steps.

2 Motivating Framework

This section provides a framework that motivates the experimental design of our study. It integrates existing theoretical and empirical evidence on non-pecuniary motivations for supplying volunteer work in order to distinguish between the theoretical impacts of increasing the salience of contributors' expected value of output and increasing the salience of contributors' expected value of input. Our framework is largely based on the theories developed in Cassar and Meier (2016) and Andreoni et al. (1996). The utility function used in the framework is consistent with our experiment design, but intended to be general enough to be useful in other settings.

2.1 Framework Setup

Suppose each potential contributor receives some benefit $U(\mathbf{x},e)$ from volunteering her effort, where $\mathbf{x} = (x_o, x_c, x_u)$, x_o is the contributor's expected value of the output she generates by her effort in the task, x_c is the contributor's perceived value of the effort she supplies to the organization relative to the organization's outside option, and x_u is any additional value the contributor gets from supplying labor to the organization, including learning and social benefits. Suppose the contributor's only outside option is wage work where wages are equal to w. Consistent with Andreoni et al. (1996), we assume that x_c is equal to the contributor's expectation of the minimum amount the organization would have to pay a worker to exert the equivalent effort on the task. Therefore, if contributor utility was only a function of x_c and effort, contributors would invest effort up to the point that the net utility from contributing is equal to her net utility from wage work, such that if the cost of exerting effort into the volunteer task and wage work is the same, contributors only volunteer effort if $x_c \geq w$. Otherwise, they prefer to earn wages and donate money to the organization to volunteering. This definition of x_c is consistent with the concept of competence described in Cassar and Meier (2016) in which workers have a preference for performing tasks they are able to do better or at lower cost than alternative workers.

In this model, we allow the contributor to care about the organization's mission by getting utility from contributing to its output directly, and from saving it money and, thus, potentially allowing it to increase or maintain output. The idea that contributors get utility from participating in an organization's output independently of the amount they are paid is consistent with existing evidence (e.g., Akerlof and Kranton, 2005; Besley and Ghatak, 2005). We assume that utility sources are independent of one another, and that utility is weakly increasing in x_i, $i\ \varepsilon\{o,c,u\}$ for all levels of effort.[6]

To allow for the possibility that ability may affect the expectations people have about the value of their contributions to the task,[7] and that preferences for contributing to science may affect expectations about the output value of scientific contributions, we allow x_c to vary with ability and x_o to vary with scientific contribution preferences. For simplicity, we allow people to have either low or high ability, and either low or high preference for science.

Therefore, we write contributor utility as:

$$U_{p,a} = R(\lambda_{op}x_{op}, \lambda_{ca}x_{ca}, x_u, e) - C(e) \tag{1}$$

where $C(e)$ is a convex cost of effort, $\lambda_{op}\ \varepsilon[0,1]$ is a preference weighting that differs by contributor motivations for contributing and $p\ \varepsilon\{low, high\}$. $\lambda_{ca}\ \varepsilon[0,1]$ is a preference weighting that differs by contributor ability and $a\ \varepsilon\{low, high\}$, and $R(\lambda_{op}x_{op}, \lambda_{ca}x_{ca}, x_u, e)$ is the return workers receive from investing effort e into the task. In our setting effort includes the quantity and quality of work performed.

[6] It is likely that increasing the expected value of output of a contributor's contribution also increases the value of their labor supply, and vice versa. We do not consider the interaction between the expected value of output and input in our experiment as we only increase the salience of one of the two at a time.

[7] Specifically, contributors with high outside options may be more likely to donate w instead of time because x_{cnew} has to be sufficiently large to be greater than x_c. Contributors with high outside options likely have a higher wage compared to contributors with low outside options. This impacts our predictions because it changes the likelihood that the new information provided in our experiment will improve or reduce a contributor's utility from effort.

2.2 Framework Predictions

2.2.1 Change in Expected Value of Output

Suppose that contributors receive an increase in the amount of information they are given about the objective of the organization. In our setting, this would be an increase in information about how or why the science produced from the task that contributors can work on will contribute to project outcomes. Given that the cost of effort is not impacted by this new information, this change will impact x_{op} and the first order condition is $U_{x_{op}} = \lambda_{op} R_{x_{op}}$. Suppose $x_{op,new} \neq x_{op,old}$, for instance because people over- or underestimated the value of science being undertaken. Given the initial belief heterogeneity, contributor expectations may be adjusted downwards or upwards with the introduction of new information. In particular, because $x_{ohigh} \geq x_{olow}$, it is possible that $x_{ohigh} \geq x_{op,new} \geq x_{olow}$, such that the new information reduces the effort of high types and increases the effort of low types. Moreover, the reduction in effort among high types will be larger than the increase in effort among low types because $\lambda_{ohigh} \geq \lambda_{olow}$. Therefore,

Proposition 1 *If increasing the amount of information about the goal of the contributor's task leads to an increase (decrease) in the contributor's expected value of her output, then this information provision will lead to an increase (decrease) in the amount of effort invested in task. This increase in the amount of information about the goal of the contributor's task is more likely to lead to an increase in effort among contributors whose prior beliefs about the importance of the task goal were lower.*

2.2.2 Change in Expected Value of Input

Suppose that contributors receive an increase in the amount of information they are given about the value of their labor supply to the organization. In our setting, this could be information on the type and amount of labor researchers would have to hire in order to substitute for the contributor's labor. This change will impact x_{ca} and the first order condition is $U_{x_{ca}} = \lambda_{ca} R_{xca}$. Suppose $x_{ca,new} \neq x_{ca,old}$, for instance, because contributors had previously misinterpreted the type of skill the organization would have to hire to replace contributor work. With initial belief heterogeneity, contributor expectations may be adjusted upwards or downwards with the introduction of new information. As in the case of information about output value, it is possible that $x_{chigh} \geq x_{ca,new} \geq x_{clow}$, such that the new information reduces the effort of high types and increases the effort of low types. Moreover, the reduction in effort among high types will be larger than the increase in effort among low types because $\lambda_{chigh} \geq \lambda_{clow}$. Therefore,

Proposition 2 *If increasing the amount of information about the goal of the contributor's task leads to an increase (decrease) in the contributor's expected value of her output, then this information provision will lead to an increase (decrease) in the amount of effort invested in task. This increase in the amount of information about the value of the contributor's labor is more likely to lead to an increase in effort among contributors whose initial perceived value of their labor supply relative to the organization's outside option were lower.*

2.3 Framework Discussion

This framework demonstrates that providing contributors with more information about the value of their output or the value of their input have theoretically ambiguous impacts on effort. This is because it is not clear ex-ante what contributors' beliefs are in the absence of this information. In particular, if the information improves their perceived value of their contribution, either by making their output or input seem more valuable, then it will increase the contributor's effort in the task. On the other hand, if it reduces their perceived value of their contribution, then it will decrease the effort they invest in the task.

Importantly, these patterns will be impacted by contributor characteristics. In this framework, we focus on the impacts of contributor heterogeneity in skills and motivations on the effects of information provision.[8] Given our assumptions about the source of utility that are impacted by contributor characteristics,[9] the framework predicts that ability will impact the magnitude of the effect of information about the input value on effort, and that preference for science will impact the magnitude of the effect of information about the output value on effort. Moreover, given our

[8] Other contributor characteristics may also impact this relationship. For instance, age may affect contributors' outside options which may alter how they interpret information about the value of their input.

[9] These assumptions are based on prior theoretical and empirical evidence (e.g. Andreoni et al., 1996; Cassar and Meier, 2016).

assumption that these characteristics impact ex-ante expectations, the framework predicts that they will also impact the direction of the relationship between new information and effort.

We test these predictions in our data by analyzing whether increasing information on output or input value changes the quantity and quality of contributor effort and the types of people who contribute to the project. Based on our predictions, we expect that those with lower perceived skills may be more likely to contribute to the project when the value of volunteer labor supply is clarified. Similarly, we expect that those who have lower intrinsic motivations for contributing to science will be more likely to contribute to the project when the value of contributor's output is clarified.

3 Experimental Design

3.1 Research Setting

In order to test whether the desire to contribute to scientific discovery can be increased through task framing, we ran a natural field experiment on Zooniverse, the largest voluntary crowd science platform in the world. Scientists post projects on Zooniverse that require contributors to answer questions about images the scientists would like to classify. For instance, Galaxy Zoo which was the project that motivated the creation of Zooniverse, asks contributors to classify galaxies in images taken by telescopes according to their shapes. Citizen science contributions to this project have been used in 55 publications between 2007 and 2017.[10] Over 75 projects have been posted on Zooniverse since its launch in 2009. Zooniverse contributors assist in these projects with the understanding that they will not be paid or formally acknowledged for their contributions.

Responses to a survey administered by Zooniverse demonstrate that the majority of contributors have at least a Bachelor's degree and over 25% have a graduate degree. In addition, the majority are employed in some capacity in a wide range of professions including engineering, education, archaeology, software development, entrepreneurship, and finance. The majority of contributors reside in North America and Europe.

[10] Galaxy Zoo pre-dates Zooniverse, and was using its own crowd science platform before joining Zooniverse.

One notable feature of Zooniverse is that contributors are anonymous and do not see each other's contributions (unlike Wikipedia, for instance). As a result, reputation and social reasons are unlikely motivations for contributions. While many project pages have a message board ("Talk" tab) that allows contributors to post messages, the extent to which it is used varies widely across pages. In our setting, only two messages were posted on the three project page message boards combined and their purpose was to get researcher clarification on how to classify specific images. Consistent with our experience, most posts on the Talk tabs of other projects are questions or difficulties that contributors have encountered during classifications (e.g., "There are two animals in the background that I can't identify."). In cases where the messages are not classification questions, they are usually remarks or observations about an image (e.g., "Lovely #bird_other in the foreground here with some #waterbuck") and not directed toward any one contributor. In other words, the contributors do not appear to interact with each other which again suggests that there is limited social and reputation motivations for contributing in this setting.

3.2 Experimental Treatments

The project we posted on Zooniverse is part of a larger East African rangeland crowdsourcing project that seeks to improve interventions aimed at assisting pastoralists in the face of increasing drought risks. During their routine herding, participating pastoralists took pictures of rangelands in Northern Kenya and completed a survey in which they classified the types of vegetation in the photos.[11]

To differentiate between changes in selection and effort in response to changes in expectations about the value of scientific contribution output and input, we generated three separate project pages on Zooniverse. All three pages have the same scientific content and include the same background information on the project including its overall objective and the why Zooniverse contributors were being asked to classify the images. We ask contributors to classify a total of 1,061 images on six dimensions: 1) Is there any green grass?, 2) Are there trees?, 3) Are there shrubs?, 4) Is this a picture of a rangeland?, 5) Is this picture poor quality (i.e., blurry, poor lighting or angle)?, 6)

[11] For more information about this project, see https://www.udiscover.it/applications/pastoralism/.

Are there wild/domesticated animals?

The three pages differ in the extent to which they highlight the value of contributor output and input. Specifically, the output treatment page emphasized how contributors' efforts can help advance science. This page included the statement "Your contributions will help the advancement of crowd science and machine-learning algorithms!" on the landing tab and project description tab. In addition, the project description tab included a more detailed description of how contributions to the project can advance machine-learning and crowd science.

The input treatment page emphasized the value of contributors' labor supply by describing the type of labor their effort replaces. This page included the statement "Your contributions will help us replace the hiring of university student research assistants and save hours of our research time!" on the landing tab and project description tab. In addition, the project description tab included a more detailed description of the value of contributors' labor supply.

The control page do not include separate statements about the value of contributors' output or input. We tried to ensure that the additional text included in the output and input treatment pages was not lengthy enough to deter contributions simply because of the higher effort required to read the page content relative to the control page. In particular, the additional text in the input page relative to the control page amounts to a total of 21 words in two of the five page tabs. The full text displayed on the project pages are provided in Appendix B.

3.3 Intervention Implementation

In order to test whether the treatments have an effect on the types of people who contribute to the pages, we need to ensure that all three pages are viewed by comparable populations of contributors. Ideally, we would show the three pages to a set of contributors and have them choose which page to contribute to based on their motivations and preferences. However, this was not possible because it would have alerted contributors to the experiment and led to potential bias in our estimates if people responded differently than they would have if they were not aware that their selection was being observed (e.g., Levitt and List, 2011). To address these concerns, we worked with the Zooniverse team to keep the project pages private from the general population of contributors and allow only a randomly selected subset of contributors to access each page through a link provided in an email.

Each page was sent to around 10,000 randomly selected regular Zooniverse contributors. The email text inviting people to make contributions to the pages was identical in all three cases expect for the inclusion of the statement “Your contributions will help the advancement of crowd science and machine-learning algorithms!” in the output page treatment email, and the inclusion of the statement “Your contributions will help us replace the hiring of university student research assistants and save hours of our research time!” in the input page treatment email. In addition, the three pages had slightly different titles because the platform does not allow separate pages to have the same title.[12] Including statements that were consistent with the treatment framings in the emails allowed us to test whether click rates differed across the pages. The full email text is provided in Appendix B.

Each page also included an invitation for contributors to complete a short, anonymous survey about themselves and their motivations for contributing. This invitation was provided in a banner displayed on every tab and stated “In an effort to better understand who our contributors are and how we might be able to attract more, please fill out a short, anonymous questionnaire at the top right corner of our project page.” Although project pages generally do not include a contributor questionnaire, after consulting with the Zooniverse team on the issue, we concluded that including one was unlikely to be considered inappropriate or disconcerting by contributors. Moreover, it was Zooniverse’s preferred method for us to collect survey responses.

4 Data and Analysis

Data on contributor characteristics, including education, income, employment status, and Zooniverse activity was collected from the subset of contributors who responded to the anonymous contributor questionnaire posted on the project pages. The full set of survey questions is provided in Appendix B. The survey response rate was 48% for the control page, 44% for the output page, and 39% for the input page. In addition, some respondents did not answer all the questions. While

[12]The title of the control page project was “Crowdsourcing for Rangeland Vegetation", the title of the output page project was “Digital Crowdsourcing for Rangeland Vegetation", and the title of the input page project was “Digital Crowdsourcing for Vegetation".

this response rate is not ideal, we think it is reasonable to believe that those who did fill in the survey are those contributors most engaged by the project and, therefore, may be the most relevant population for understanding differences in motivation and ability across the pages.

We examine contributor effort both by the number of classifications they complete, and by the quality of their completed classifications, which is measured by the accuracy of their classifications. To generate a measure of classification accuracy on which to compare contributor classifications, all 1,061 were separately classified by both authors. In addition, we used data from the pastoralists' classifications on grass, shrub, and trees.[13] Both of these baseline accuracy measures are useful for different reasons. Pastoralists have expert knowledge of the rangelands in the project images and, therefore, their classifications are based on better knowledge of the rangelands than those completed by the authors. The authors' classification data also includes information on image quality, the presence of animals, and whether or not the image is of a rangeland. Our preferred measure of question accuracy restricts the sample of image-question observations to those for which both authors and the pastoralist agree in the case of grass, trees, and shrubs, and to those for which both authors agree for the remaining questions, and drops image-questions for which there does not appear to be a clear answer. However, our findings are robust to less stringent measures of accuracy, such as comparing classifications only against those of pastoralists.

With this measure of image-question accuracy, we generate contributor accuracy measures for each question of the image that equals one if the contributors' answers match the accurate answer (i.e., the answer given by all external classifiers), and zero otherwise. We also generate a combined accuracy measure that equals one if each question of the image was answered accurately, and zero otherwise. In addition to generating quality measures at the question and image level of observation, we also generate a count of the number of high quality classifications made by each contributor, which is the sum of images that are classified accurately on all dimensions weighted by the proportion of contributor classifications that are included in the combined image accuracy measure.

[13] Pastoralists did not classify whether there were animals in the photos. Moreover, they were required to submit high quality photos of rangelands so they were not asked whether their photos were of rangelands are of poor quality.

4.1 Summary Statistics

Figure 1 plots the number of clicks to the email link that invites individuals to contribute and the number of contributors per page. The graph indicates that both treatment emails, particularly the output treatment, generated the highest interest, as more people clicked on the output and input email links compared to the control. However, a greater share of people who clicked on the link became contributors on the control page compared to the output and input pages. This suggests that while the treatment emails generated higher initial interest, it did not convert to a higher level of participation compared to the control page. This is likely because individuals on the treatment page were not motivated by the task involved in generating their expected value of output or input. Thus, the clickers that became users are ones that were motivated by the project and the task.

Sample averages for survey responses, and quality and quantity outcomes are presented in Table 1. Panel A of the Table presents summary statistics for the survey responses. The average contributor spends around 7-10 hours on Zooniverse per month, and they almost all indicate that at least one of the reasons they contribute to Zooniverse projects is because they like to know that they are contributing to science. Almost 60% indicated that they contribute because of their skill.[14] In addition, about 67% have some science related education or work experience and over a quarter are employed full-time. Over a third have a bachelor's degree while a quarter have a graduate degree. The average age is between 35-49. The average income is between \$40,000 - \$74,999. Almost 60% of contributors are female.

Panel B of Table 1 reports average quantity and quality outcomes at the contributor level. On average, contributors make about 13 classifications and 6 are accurate on every dimension. Panel C of Table 1 reports average quality outcomes at the image level. On most dimensions, images are classified accurately in the majority of cases. However, only about 50% of images are accurately classified across all dimensions.

[14] "Contribute because of Skill" is equal to one if respondents indicate they contribute to projects on Zooniverse because they're good at classifying things on Zooniverse, or because they have a science background and want to put it to good use and zero otherwise.

5 Treatment Effect Estimation

In order to estimate the effects of increasing the salience of the value of inputs and outputs, we focus primarily on mean comparisons across the two treatment and the control pages. Given that our pages are randomly assigned to a large population, the sample of potential contributors across each page is likely the same.[15] Moreover, the project pages are identical except for the treatment statements. Therefore, we can infer a causal estimate of the treatment based on differences in means across outcomes.

We begin our analysis by comparing mean contributor characteristics across the three pages. As our motivating framework demonstrates, we expect that the primary mechanisms through which our treatments will function is selection into contributions. Thus, we expect that contributor characteristics will differ across the three pages. We then compare average contributor quantity and quality, and average classification quality across the three pages. We perform these mean comparisons across all classification dimensions. However, our preferred measure of quality is the accuracy measure that combines all dimensions because it is the most practically relevant, and it overcomes concerns associated with testing treatment effects on multiple outcome measures (List et al., 2016).

Panel A of Table 2 examines differences in contributor characteristics across treatment and control pages. Several interesting results stand out. First, control participants spend the most time per month contributing to Zooniverse projects (7-10 hours), and it is significantly more than the time spent by input participants (3-6 hours). While most participants across all three pages indicate they contribute for science, both output and control participants are significantly more likely to indicate that they contribute because of skill compared to input participants. This is interesting given the three groups do not differ significantly in the likelihood of having science-related education. However, both output and control participants are more likely to have science related work experience. All three groups do not differ in education level, employment status, age, or income. Interestingly, participants on the output page are significantly more likely to be female compared to the input and control pages. This is consistent with Tonin and Vlassopoulos (2010)

[15] We do not have data on the characteristics of contributors who received the invitation to contribute but opted not to, so we cannot test this assumption by comparing characteristics across the three groups.

who find that warm glow altruism only increases women's effort in a field setting. Figure 2 plots the average Zooniverse contributor characteristics across control and treatment pages along with standard errors.

Panel B presents the differences in classification quantity and quality at the contributor level. Control participants classify significantly more images than output and input participants. In fact, control participants classify more than double the number of images than input participants.[16] However, once we weight the number of classifications by quality, the number of classifications are not statistically different across pages. Figure 3 plots the average quantity and quality of image classifications on Zooniverse across control and treatment pages along with standard errors.

Panel C examines differences in means across pages at the image level. On many dimensions, input and output treatments are associated with more accuracy on classification questions. In particular, classifications on the output page have significantly higher accuracy across all dimension when compared to the control classifications. 55% of output classifications are accurate across all dimensions compared to 45% of classifications on the control page. The results are even stronger if we use less stringent measures of accuracy, where the classification is evaluated only against the pastoralist classifications. Also noteworthy is that output classifications are also more accurate for the image quality question, which is the question that produces the most variation in accuracy. This suggest that output classifications are higher quality compared to the control even for the most difficult classification question.

Although images are randomly assigned to contributors on Zooniverse and all of our project pages included the same set of images, one concern with our mean comparison estimates is that they could be affected by the images that were classified across pages. For instance, if contributors on one project page systematically received images that were "easier" to classify than those on the other pages, our results could suffer from bias. To address this concern, we estimate the effect of our treatments on image level outcomes using regressions that include image fixed effects. The estimates from Table A1 allow us to compare how the treatments affect the accuracy of classifications

[16] One may be concerned that individuals on the input page are sympathetic to replacing university research assistant hours and thus are deterred from contributing. While we do not believe this to be the case, the likelihood of being deterred from contributing due to sympathy for replacing RA hours is likely also correlated with contributor characteristics. Individuals with ability and outside options that exceed what is conveyed on the page are more likely to select out of volunteering due to sympathy for university students.

within the same image. Consistent with our mean comparisons, the output treatment is associated with a 11 percentage point increase in accuracy compared to the control. The input treatment does not have a significant impact on accuracy outcomes.

6 Survey Experiment

The results from the field experiment on Zooniverse suggest that more information about the expected value of input and output leads to fewer contributions but higher quality contributions, measured by agreement with the expert classifications. Consistent with our theoretical framework, these results suggest that the treatments lead people who had previously underestimated their ability ($x_{ca} \leq x_{ca,new}$) and the contribution of their output ($x_{op} \leq x_{op,new}$) to contribute. Conversely, the treatments also lead to people who had overvalued the value of their input (x_{ca}>$x_{ca,new}$) and output (x_{op}>$x_{op,new}$) to not contribute. This explains why individuals on control pages contribute more but the quality of classifications is lower than those of output contributors. In addition, this also explains why the images classified on the input pages is lower quality than those classified on output pages, as only individuals with skill level lower or equal to what is conveyed on the input treatment page will contribute. Individuals with skill level higher than what is conveyed on the input treatment page will not contribute due to poor skill match.

While our field experiment allows us to test our research questions in an important, real-world crowd science setting and allows us to causally compare rates of selection and contributions into the project pages by treatment, there are several shortcomings associated with the Zooniverse setting. First, we are unable to observe how regular Zooniverse contributors differ from non-volunteers. Second, although we have average contributor characteristics across our three treatments, we cannot link individual characteristics to their contributions which limits our ability to speculate on how outside options impact volunteering decisions across the pages. Third, we are unable to test whether those who select out of volunteering due to the treatments substitute by donating something other than time or select out of contributing altogether. To address these limitations, we implement a survey experiment that allows us to examines whether people with different outside options are differentially motivated by the input and output treatment, whether individuals who are never willing to volunteer are systematically different from contributors, and whether organizations may be able to induce high wage earners to contribute money rather than time.

6.1 Survey Experiment Description

The survey experiment is a 2x2x2 design that varies the input/output framing, input/output task, and outside option across participants. Participants are given either an input or output framing, where we emphasize the labor and time required to a complete a scientific project, and the expected output of the project, respectively. We then ask participants a series of demographic questions and questions to measure their belief in the importance of science and belief in their own abilities to contribute to science. In the last question participants are asked to allocate one hour of their time to either working one hour at a wage rate and donating the wage to the scientific project, volunteering for the science project, or pursuing leisure activity. In the input-motivated task, we state that volunteering one hour saves researchers one hour of research time. In the output-motivated task, we state that they advance the accuracy of the algorithm before it is launched for public use. We also vary the wage rate to convey a high (40% above mean wage rate) versus low (40% below mean wage rate) outside option.[17] The full survey text is provided in Appendix B.

We built the survey using *Qualtrics* which allows for automated question and answer randomization across respondents. The survey was distributed by *Survey Sampling International* to adults over the age of 18 living in the US. In total, we collected 4,189 responses over a 1 week period in October, 2017. This gives us approximately 500 respondents per survey type.

6.2 Survey Experiment Findings

Table 3 provides average survey responses across all survey types and demonstrates that our average survey taker looks quite similar to our Zooniverse experiment participants. In particular, around 30% of respondents have a bachelor's degree, the average income is between $40,000-$74,999, over 60% have some science-related education or work experience, over half are female,

[17] The mean wage rate is taken from the 2016 Bureau of Labor Statistics report on National Occupational Employment and Wage Estimates. https://www.bls.gov/oes/current/oes_nat.htm#00-0000 (accessed September 18, 2017)

and the average age is between 35-49. The main difference between the survey respondents and the Zooniverse contributors are in the employment status and graduate degree. Almost half of the survey respondents are employed full-time compared to just over a quarter for Zooniverse respondents; About 17% of survey respondents reported to having a graduate degree, compared to a quarter for Zooniverse respondents. In addition, slightly less than half of the respondents volunteer at least once per month and about 40% believe they have the skills to contribute to science. Interestingly, while three quarters believe scientific discovery is important for society, only one third would donate money to science over other non-profit causes, such as supporting the environment or the homeless.

We start by comparing survey respondent characteristics by their time use preference response within the the input and output-based task motivation (Table 4). Respondents that choose to volunteer their time or donate money are different from those that pursue leisure in meaningful ways. For both the input and output-based task motivations, those that pursue leisure are on average less likely to believe in the importance of science and less likely to indicate they have the ability to contribute to science compared to those that give money or time. They tend to be less educated, have less science-related education or work experience, less likely to be employed full time, and have lower income. They also tend to be older and female. Most of these characteristics are also similar across the input and output-based task motivations. The only difference is that people who pursue leisure on the output motivated page are more likely to indicate that they have abilities to contribute to science compared to those who pursue leisure on the input motivated page.

Unlike in our field experiment, respondents did not select into surveys based on the task framing because they were not provided the framing until the last survey question and it is unlikely respondents dropped out of the survey to avoid answering the last question. However, in order to verify that people did not differently select into the survey, and to verify that survey treatments are randomly assigned, we compare respondent characteristics across outside options and task motivations in Table 5.[18] This table shows that average characteristics of respondents look very similar across treatment groups and they are all statistically the same with the exception of their belief

[18]We do not present these comparisons by survey framing because as Table A2 demonstrates, this framing had no impact on our time use preference question. Nevertheless, mean respondent characteristics are similar across survey framings.

on the importance of scientific discovery, which is significantly higher in the low wage offer-input task motivation group than in the high wage offer-input task motivation group. In addition, the preference for donating to science over other causes is statistically higher in the high wage offer-output task motivation group than in the low wage offer-input task motivation group. Given the number of respondent characteristics and the number of treatment groups, it is not surprising that we see some economically small but statistically significant differences across some characteristics and treatment groups. We control for these differences in our regressions in Table A4 and find that our results are qualitatively unchanged.

The most notable findings presented in Table 5 and summarized in Figure 4 are the differences in responses to our time use preference question across treatments. Specifically, we find that respondents are more likely to switch from leisure to work and donate in the presence of high outside options. Second, we find that within task motivations, volunteering does not respond to outside options. In contrast to our theoretical predictions, this suggests that people do not appear to be adjusting their volunteer labor supply to account for differences in outside options. Third, individuals are more likely to volunteer when given the input-based task motivation compared to the output-based task motivation, whereas individuals are more likely to work and donate money given the output-based task motivation. This suggests that respondents may have had higher ex-ante beliefs about the output of their volunteer contribution and/or that respondents undervalued their ability to contribute to the task in the absence of information about the type of labor required to complete the task. Moreover, it suggests that people who do not volunteer due to the types of information provided about the task may instead donate money. We verify that these patterns hold in a multinomial logit regression that allows us to specify that respondents were choosing between three separate options (see Table A3).

Next, we explore these patterns further and examine whether the impact of task motivation varies depending on people's ex-ante views about the importance of scientific discovery (expected value of the output) and their ability to contribute to science (expected value of the input). Columns 1-3 in Table 6 present results of OLS regressions estimating the effects of treatment on time use preference responses including interactions between the output-based task motivation and whether respondents agreed that scientific discovery is important for improving quality of life. Not surprisingly, having a belief that scientific discovery is important significantly increases both the likelihood of volunteering and donating money, and decreases the likelihood of engaging in leisure. However, consistent with the role of prior beliefs in determining how additional information affects behavior, the coefficients on the interactions term demonstrates that output-based task motivation is more likely to increase the willingness to volunteer from individuals whose prior beliefs on the importance of science was lower. Instead, individuals with higher beliefs on the importance of science increased their willingness to work and donate money when presented with the output-based task motivation. This further suggests that the reduced volunteer labor supply caused by information provision on the output of the task may be compensated by monetary donations.

Columns 4-6 in Table 6 examine the heterogeneous effects of task motivations based on different levels of perceived ability. Again, not surprisingly, respondents who believe they can contribute to science are significantly more likely to volunteer and donate money at the expense of engaging in leisure. Interestingly, this belief does not appear to be impacted by the input-based task motivation which clarifies the type of skill required to perform the task. Although the coefficient is insignificant, the direction of the interaction term suggests that information may increase the likelihood of volunteering among individuals with high perceived ability. This is likely because the respondents in the survey sample are different from Zooniverse contributors in terms of their expectations about contributing to science.

Finally, we explore how the relationship between respondents' ability and survey treatments is impacted by outside options. Table 7 suggests that respondents with high ability and high outside option (measured by whether they have high wage) are disproportionately less likely to give time and more likely to give money compared to those with high perceived ability in science and low outside option and also compared to those with low ability and high outside option. Overall, these finding further suggest that high ability individuals with high earning are much less likely to volunteer but they may be quite willing to donate money, particularly in the presence of output-based task motivation.

7 Conclusion

Many organizations increasingly rely on voluntary contributions of time or money. Organizations that depend on voluntary contributions face unique challenges of soliciting volunteers and matching them to appropriate tasks. In this paper, we examine the non-pecuniary motivations for contributing to digital public goods in an increasingly prominent setting, crowd science.

We begin with a conceptual framework that defines the utility of contributors as a function of the expected value of the output they generate and the expected value of their input to the organization. We predict that how contributors' change their effort in response to additional information about the value of their output and input will depend on whether the additional information leads to upward or downward changes in their beliefs about these values. Using an RCT among contributors to crowd science, we find general support for these predictions. In particular, contributors given additional information about the value of their inputs are more likely to perceive the value of their labor supply (i.e., input) as relatively low ex-ante and are induced to contribute when they learn that their labor supply is valuable relative to the organization's outside option. While providing additional information about both output and input value reduces the number of contributors relative to the control, the difference becomes insignificant once we account for the accuracy of the classifications. In other words, while increasing information about input and output value decreases participation, it increases the accuracy of contributions.

We further investigate how outside options and individual-level perceptions about the value of science and abilities impact voluntary crowd science contributions. We find support for the findings in our field experiment that people's ex-ante expectations about science projects impact how they respond to information about the value of the scientific project. Moreover, while people do not appear to adjust their volunteer labor supply when their outside options change, we find evidence that some people who select out of volunteering due to the type of task information provided choose to donate money from wage work instead.

Combined, our results provide novel evidence that different types of science contributors are motivated by different non-pecuniary rewards and the information provided about the task can lead to improvements in matching between the volunteer and the task. Our results have distinct implications for organizations employing crowd science, and volunteer labor more generally, and for policy. Our findings suggest that organizations can affect the quantity and quality of their volunteer labor supply and the quantity of their monetary donations by adjusting the extent and type of information they provide volunteers. More specifically to our setting, as the availability of data continues to grow and machine learning algorithms become more efficient at processing information, the intensity and type of participation required by human volunteers will change. Given this, we believe that the incentives that impact crowd science participation and how to best match tasks to the appropriate skills and knowledge of crowds will continue to be a prominent line of inquiry in the years to come.

8 References

Akerlof, George A and Rachel E Kranton, "Identity and the Economics of Organizations," *The Journal of Economic Perspectives*, 2005, *19* (1), 9–32.

Andreoni, James, "Impure altruism and donations to public goods: A theory of warm-glow giving," *The Economic Journal*, 1990, *100* (401), 464–477.

_ , William G Gale, and John Karl Scholz, "Charitable contributions of time and money," 1996.

Ariely, D., A. Bracha, and S. Meier, "Doing good or doing well? image motivation and monetary incentives in behaving prosocially," *The American Economic Review*, 2009, *99* (1), 544–55.

Ariely, Dan, Emir Kamenica, and Dražen Prelec, "Man's search for meaning: The case of Legos," *Journal of Economic Behavior & Organization*, 2008, *67* (3), 671–677.

Besley, Timothy and Maitreesh Ghatak, "Competition and incentives with motivated agents," *The American economic review*, 2005, *95* (3), 616–636.

Bénabou, Roland and Jean Tirole, "Incentives and Prosocial Behavior," *The American Economic Review*, 2006, *96* (5), 1652–1678.

Boudreau, K. J., K. R. Lakhani, and M. Menietti, "Performance responses to competition acrosss kill levels in rank-order tournaments: Field evidence and implications for tournament design," *RAND Journal of Economics*, 2016, *47* (1), 140–165.

_ , N. Lacetera, and K. R. Lakhani, "Incentives and Problem Uncertainty in Innovation Contests: An Empirical Analysis," *Management Science*, 2011, *57* (5), 843–863.

Cassar, Lea and Stephan Meier, "Non-Monetary Incentives and the Quest for Work Meaning," Technical Report, Working paper 2016.

Chandler, Dana and Adam Kapelner, "Breaking monotony with meaning: Motivation in crowdsourcing markets," *Journal of Economic Behavior & Organization*, 2013, *90*, 123–133.

Gallus, Jana et al., "Fostering voluntary contributions to a public good: a large-scale natural field experiment at Wikipedia," Technical Report, The Field Experiments Website 2016.

Grant, Adam M, "Does intrinsic motivation fuel the prosocial fire? Motivational synergy in predicting persistence, performance, and productivity.," *Journal of applied psychology*, 2008, *93* (1), 48.

Haklay, Muki, "Citizen science and policy: a European perspective," *The Wodrow Wilson Center, Commons Lab: Washington, WA, USA*, 2015.

Howe, J., "Crowdsourcing: Why the power of the crowd is driving the future of business," in "in," New York: Crown Business, 2008.

Jeppesen, L. B. and K. R. Lakhani, "Marginality and Problem-Solving Effectiveness in Broadcast Search," *Organization Science*, 2010, *21* (5), 1016–1033.

Kelling, Steve, Daniel Fink, Frank A La Sorte, Alison Johnston, Nicholas E Bruns, and Wesley M Hochachka, "Taking a 'Big Data'approach to data quality in a citizen science project," *Ambio*, 2015, *44* (4), 601–611.

Kosfeld, Michael and Susanne Neckermann, "Getting more work for nothing? Symbolic awards and worker performance," *American Economic Journal: Microeconomics*, 2011, *3* (3), 86–99.

Lacetera, Nicola and Mario Macis, "Social image concerns and prosocial behavior: Field evidence from a nonlinear incentive scheme," *Journal of Economic Behavior & Organization*, 2010, *76* (2), 225–237.

Lakhani, Karim R and Eric Von Hippel, "How open source software works:"free" user-to-user assistance," *Research policy*, 2003, *32* (6), 923–943.

Levitt, Steven D and John A List, "Was there really a Hawthorne effect at the Hawthorne plant? An analysis of the original illumination experiments," *American Economic Journal: Applied Economics*, 2011, *3* (1), 224–238.

Lilley, Andrew and Robert Slonim, "The price of warm glow," *Journal of Public Economics*, 2014, *114*, 58–74.

Lindley, D., "Managing data," *Communications of the ACM*, 2009, *52* (10), 11–13.

List, John A, Azeem M Shaikh, and Yang Xu, "Multiple hypothesis testing in experimental economics," Technical Report, National Bureau of Economic Research 2016.

Perrow, C., "A Framework for the Comparative Analysis of Organizations.," *American Sociological Review*, 1967, *32* (2), 194–208.

Sauermann, Henry. and Chiara. Franzoni, "Crowd science user contribution patterns and their implications," *PNAS*, 2014, *112* (3), 679–684.

Stern, Scott, "Do scientists pay to be scientists?," *Management Science*, 2004, *50* (6), 835–853.

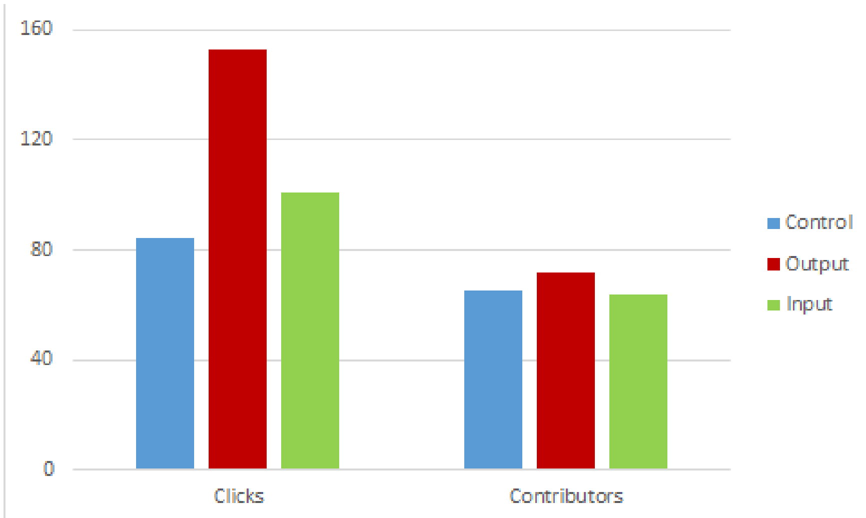
Terwiesch, C. and Y. Xu, "Innovation Contests, Open Innovation, and Multiagent Problem Solving," *Management Science*, 2008, *54* (9), 1529–1543.

Tonin, M and M. Vlassopoulos, "Disentangling the Sources of Pro-socially Motivated Effort: A Field Experiment," *Journal of Public Economics*, 2010, *94* (11-12), 1086–1092.

Zhang, Xiaoquan and Feng Zhu, "Group size and incentives to contribute: A natural experiment at Chinese Wikipedia," *American Economic Review*, 2011, *101* (4), 1601–1615.

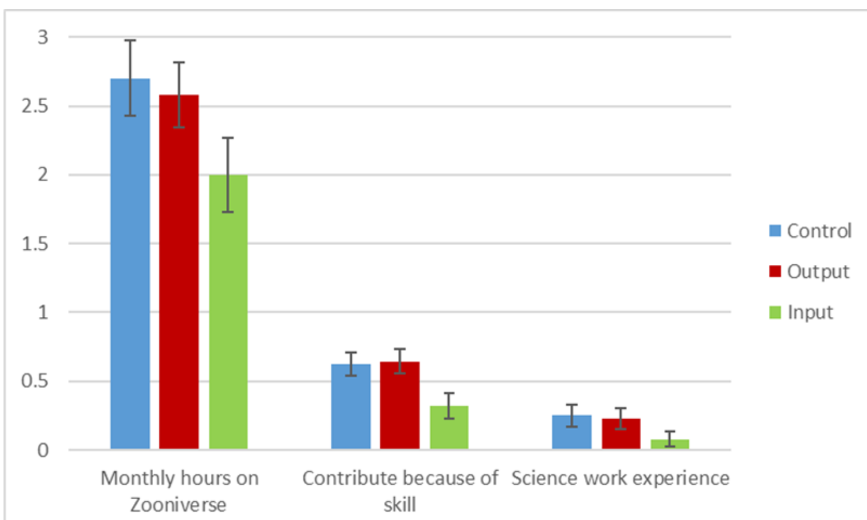
9 Tables and Figures

Figure 1: Clicks and Users by Treatment Groups



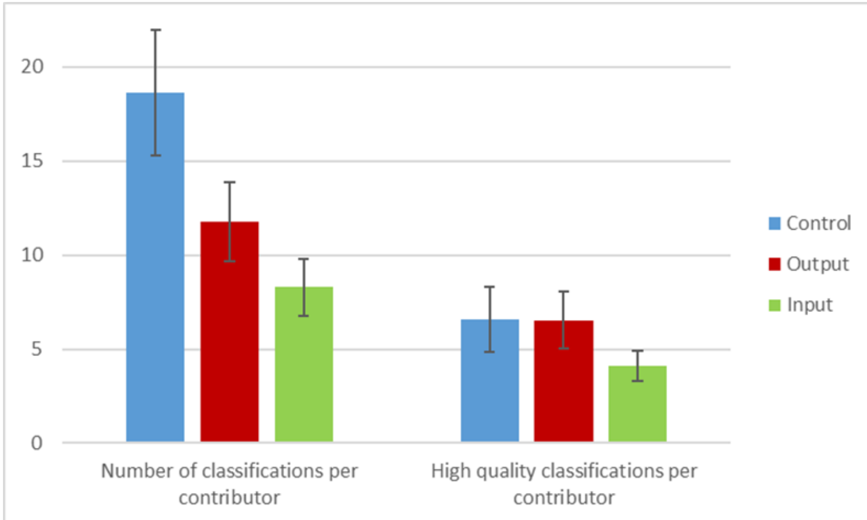
Notes: Figure plots the number of clicks on the email link that invites individuals to contribute and the actual number of contributors across control and treatment pages.

Figure 2: Contributor Characteristics by Treatment Groups



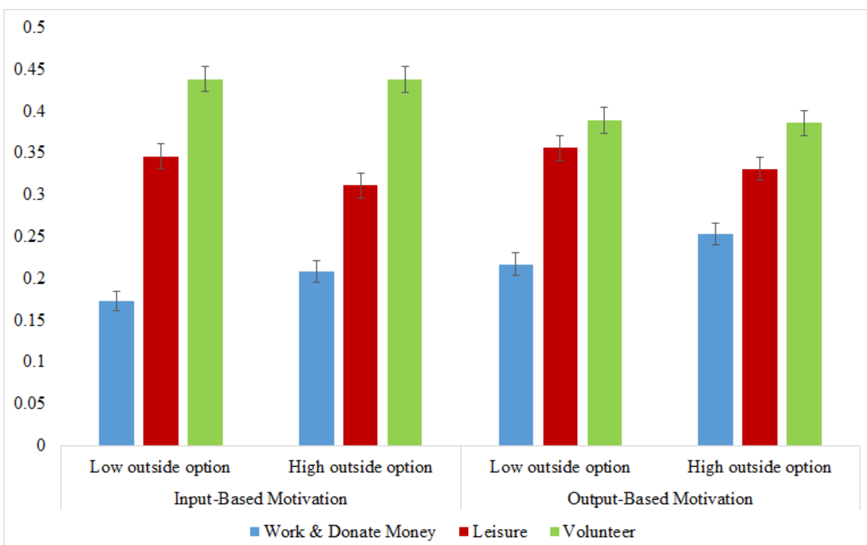
Notes: Figure plots the average Zooniverse contributor characteristics across control and treatment pages along with standard errors.

Figure 3: Contribution Quantity and Quality by Treatment Groups



Notes: Figure plots the average quantity and quality of image classifications on Zooniverse across control and treatment pages along with standard errors.

Figure 4: The Effect of Survey Treatments on Time Use Preferences



Notes: Figure plots the mean response selection to the question of how respondents would prefer to spend an hour of their time by survey treatments along with standard errors.

Table 1: Summary Statistics

Variable	Mean	(Std. Dev.)	N
Panel A: Survey Responses			
Monthly Hours on Zooniverse (Scale ranges from 1-5)	2.476	(1.401)	84
Contribute because of Skill	0.545	(0.500)	88
Contribute because of Scientific Contribution	0.909	(0.289)	88
Science Education/Experience	0.670	(0.473)	88
Bachelor Degree	0.307	(0.464)	88
Graduate Degree	0.25	(0.435)	88
Employed Full-time	0.261	(0.442)	88
Income (Scale ranges from 1-4)	2	(1.048)	61
Female	0.588	(0.495)	85
Age (Scale ranges from 1-5)	3.235	(1.306)	85
Panel B: Quality and Effort, Contributor Level of Observation			
Number of Classifications per User	12.756	(20.407)	197
High Quality Classifications per Contributor	5.76	(11.359)	197
Panel C: Quality and Effort, Image Level of Observation			
Tree Classification Accuracy	0.976	(0.154)	1771
Shrub Classification Accuracy	0.925	(0.263)	1568
Grass Classification Accuracy	0.805	(0.396)	1272
Animal Classification Accuracy	0.981	(0.138)	1916
Rangeland Classification Accuracy	0.888	(0.315)	1918
Picture Quality Classification Accuracy	0.738	(0.44)	1661
Accuracy Across all Dimensions	0.493	(0.500)	805
Accuracy against Pastoralist	0.698	(0.459)	824

Notes: Monthly Hours on Zooniverse is measured on a scale from 1-5 along the following bins: 0-2 hours, 3-6 hours, 7-10 hours, 11-20 hours, and 21+ hours. Age is measured on a scale from 1-5 along the following bins: less than 18, 18-34, 35-49, 50-64, and 65+. Income is measured on a scale from 1-4 along the following bins: Less than $40,0000, $40,000-$74,999, $75,000-$119,999, and $120,000+. Accuracy is evaluated against classifications by the pastoralist and authors. High Quality Classifications per Contributor is a count of all images that are accurate on all dimensions and is weighted by the proportion of contributor classifications included in image accuracy measure.

Table 2: Summary Statistics By Treatment and Control Groups

Variables	**Control**	**Output Page**	**Input Page**	Significant Differences
Panel A: Survey Responses				
Monthly Hours on Zooniverse (Scale ranges from 1-5)	2.710	2.581	2.000	C-I*
	(0.275)	(0.235)	(0.271)	
Contribute because of Skill	0.625	0.645	0.32	C-I**, O-I**
	(0.087)	(0.087)	(0.095)	
Contribute because of Scientific Contribution	0.875	0.935	0.92	
	(0.059)	(0.045)	(0.055)	
Science Education	0.594	0.452	0.52	
	(0.088)	(0.091)	(0.101)	
Science Work Experience	0.25	0.226	0.08	C-I*
	(0.078)	(0.076)	(0.055)	
Bachelor Degree	0.281	0.355	0.28	
	(0.081)	(0.087)	(0.092)	
Graduate Degree	0.344	0.226	0.16	
	(0.085)	(0.076)	(0.075)	
Employed Full-time	0.344	0.194	0.240	
	(0.085)	(0.072)	(0.087)	
Age	3.266	3.333	3.080	
	(0.230)	(0.241)	(0.276)	
Income	1.95	2.167	1.824	
	(0.198)	(0.231)	(0.274)	
Female	0.467	0.767	0.52	O-C**, O-I*
	(0.093)	(0.078)	(0.102)	
Panel B: Quality and Effort, Contributor Level of Observation				
Number of Classifications per User	18.62	11.750	8.297	C-O*, C-I**
	(3.356)	(2.094)	(1.520)	
High Quality Classifications per Contributor	6.582	6.541	4.099	
	(1.739)	(1.494)	(0.794)	
Panel C: Quality and Effort, Image Level of Observation				
Tree Classification Accuracy	0.965	0.982	0.982	O-C*, I-C*
	(0.007)	(0.005)	(0.006)	
Shrub Classification Accuracy	0.936	0.914	0.931	
	(0.010)	(0.011)	(0.010)	
Grass Classification Accuracy	0.797	0.786	0.846	I-C*
	(0.019)	(0.019)	(0.020)	
Animal Classification Accuracy	0.983	0.986	0.969	
	(0.004)	(0.005)	(0.008)	
Rangeland Classification Accuracy	0.900	0.903	0.850	C-I***
	(0.011)	(0.011)	(0.016)	
Quality Classification Accuracy	0.730	0.774	0.703	O-C*
	(0.018)	(0.017)	(0.022)	
Accuracy Across all Dimensions	0.455	0.550	0.465	O-C**
	(0.029)	(0.029)	(0.035)	
Accuracy against Pastoralist	0.671	0.752	0.654	O-C***, O-I**
	(0.027)	(0.024)	(0.033)	

Notes: Standard errors are in parentheses. High Quality Classifications per Contributor weighted by proportion of classifications included in image accuracy measure. Accuracy measures are restricted to images for which all external classifiers agree with each other on the given classification category. Difference * significant at 10%; ** significant at 5%; *** significant at 1%

Table 3: Survey Respondent Characteristics & Responses

Variable	Mean	(Std. Dev.)	N
Experiment Question Response			
Prefers to Work & Donate Earnings	0.214	(0.41)	4189
Prefers to Volunteer	0.412	(0.492)	4189
Prefers Leisure	0.336	(0.473)	4189
Treatments			
Volunteer Task Motivated by Output	0.508	(0.5)	4189
High Wage Offered in Work Option	0.500	(0.5)	4189
Survey Framed with Value of Scientific Output	0.489	(0.5)	4189
Characteristics			
Regularly Volunteers	0.482	(0.5)	4189
Scientific Discovery is Important	0.776	(0.417)	4189
Has Skills for Scientific Contribution	0.411	(0.492)	4189
Prefers to Donate to Science than Other Causes	0.306	(0.461)	4189
Science Education/Experience	0.644	(0.479)	4189
Bachelor Degree	0.287	(0.452)	4189
Graduate Degree	0.172	(0.377)	4189
Employed Full Time	0.490	(0.5)	4189
Income (1-4 scale)	2.011	(1.043)	4188
Female	0.515	(0.5)	4189
Age (1-4 scale)	1.914	(0.830)	4188

Notes: This table presents average responses of the time use preference question and average respondent characteristics. *Regularly Volunteers* is an indicator variable that equals 1 if the individual volunteers at least once a month, and 0 otherwise. *Scientific Discovery is Important* and *Has Skills for Scientific Contribution* are indicator variables that equal 1 if the individual indicates that they agree or strongly agree to the statement that "Scientific discovery is critical for improving people's well-being and quality of life" and "I believe I have the abilities required to contribute to science", respectively. Income is coded on a scale from 1-4 along the following bins: less than $40,000, $40,000 - $74,999, o $75,000 - $119,999, over $120,000. Age is measured on a scale from 1-4 along the following bins: 18-34, 35-49, 50-64, and 65+. See Appendix B for details on the survey experiment.

Table 4: Survey Respondent Characteristics by Time-Use Preference and Treatment

	Input Task Motivation			Output Task Motivation		
	Leisure	Give Money or Time	p-value	Leisure	Give Money or Time	p-value
Regularly Volunteers	0.256	0.593	0.000***	0.263	0.596	0.000***
	(0.017)	(0.013)		(0.016)	(0.013)	
Believes Scientific Discovery	0.586	0.862	0.000***	0.614	0.867	0.000***
is Important	(0.019)	(0.009)		(0.018)	(0.009)	
Believes has Skills for	0.180	0.505	0.000***	0.216	0.532	0.000***
Scientific Contribution	(0.015)	(0.013)		(0.)	(0.)	
Prefers to Donate to Science	0.103	0.392	0.000***	0.122	0.415	0.000***
than Other Causes	(0.012)	(0.013)		(0.012)	(0.013)	
Bachelor Degree	0.247	0.314	0.002**	0.257	0.294	0.078*
	(0.017)	(0.012)		(0.016)	(0.012)	
Graduate Degree	0.102	0.202	0.000***	0.101	0.212	0.000***
	(0.012)	(0.011)		(0.011)	(0.011)	
Some Science Experience/Education	0.443	0.726	0.000***	0.477	0.748	0.000***
	(0.019)	(0.012)		(0.018)	(0.012)	
Employed Full Time	0.402	0.538	0.000***	0.408	0.528	0.000***
	(0.019)	(0.013)		(0.018)	(0.013)	
Annual Income (1-4 scale)	1.791	2.090	0.000***	1.811	2.144	0.000***
	(0.037)	(0.028)		(0.037)	(0.028)	
Female	0.577	0.494	0.000***	0.595	0.464	0.000***
	(0.019)	(0.013)		(0.018)	(0.013)	
Age (1-4 scale)	2.150	1.805	0.000***	2.145	1.787	0.000***
	(0.031)	(0.022)		(0.030)	(0.022)	

Notes: This table presents the mean differences in respondent characteristics by those that choose leisure and those that choose to donate money or time for input and output task motivation.

Table 5: Survey Responses by Treatment

	Input Task Motivation			Output Task Motivation		
	Low Wage Offer	High Wage Offer	p-value	Low Wage Offer	High Wage Offer	p-value
Experiment Question Response						
Work & Donate Earnings	0.174	0.208	0.049**	0.217	0.254	0.048**
	(0.012)	(0.013)		(0.013)	(0.013)	
Volunteer	0.437	0.438	0.993	0.389	0.386	0.869
	(0.015)	(0.016)		(0.015)	(0.015)	
Leisure	0.346	0.311	0.096*	0.356	0.331	0.214
	(0.015)	(0.016)		(0.015)	(0.014)	
Characteristics						
Regularly Volunteers	0.489	0.476	0.554	0.474	0.488	0.514
	(0.015)	(0.016)		(0.016)	(0.015)	
Believes Scientific Discovery is Important	0.791	0.750	0.027**	0.782	0.778	0.848
	(0.013)	(0.014)		(0.013)	(0.013)	
Believes has Skills for Scientific Contribution	0.414	0.381	0.127	0.417	0.429	0.590
	(0.015)	(0.015)		(0.015)	(0.015)	
Prefers to Donate to Science than Other Causes	0.286	0.307	0.295	0.299	0.329	0.130
	(0.014)	(0.015)		(0.014)	(0.014)	
Bachelor Degree	0.293	0.292	0.942	0.276	0.286	0.621
	(0.014)	(0.014)		(0.014)	(0.014)	
Graduate Degree	0.167	0.171	0.816	0.175	0.173	0.921
	(0.011)	(0.012)		(0.012)	(0.011)	
Some Science Experience/Education	0.637	0.629	0.700	0.660	0.650	0.627
	(0.015)	(0.015)		(0.015)	(0.014)	
Employed Full Time	0.483	0.503	0.352	0.484	0.489	0.803
	(0.015)	(0.016)		(0.016)	(0.015)	
Annual Income (1-4 scale)	1.973	2.010	0.425	2.014	2.045	0.490
	(0.032)	(0.033)		(0.032)	(0.032)	
Female	0.532	0.510	0.325	0.515	0.503	0.588
	(0.015)	(0.016)		(0.016)	(0.015)	
Age (1-4 scale)	1.944	1.892	0.148	1.911	1.908	0.939
	(0.026)	(0.026)		(0.026)	(0.025)	

Notes: This table presents the mean differences in treatments, and respondent characteristics by outside options within input and output task motivation.

Table 6: Effect of Survey Treatments on Responses by Initial Motivation and Perceived Ability

	Belief in Importance of Science			Belief in Scientific Abilities		
	(1)	(2)	(3)	(4)	(5)	(6)
	Work & Donate	Volunteer	Leisure	Work & Donate	Volunteer	Leisure
High Wage	0.036***	0.004	-0.038***	0.036***	-0.000	-0.033**
	(0.013)	(0.015)	(0.014)	(0.013)	(0.015)	(0.014)
Output task motivation	-0.013	0.006	0.007	0.020	-0.039**	0.018
	(0.023)	(0.028)	(0.032)	(0.015)	(0.019)	(0.020)
High perceived scientific importance	0.046**	0.271***	-0.347***			
	(0.020)	(0.023)	(0.025)			
High perceived scientific ability				0.104***	0.131***	-0.301***
				(0.018)	(0.022)	(0.020)
Output task motivation*	0.074***	-0.075**	0.015			
High perceived scientific importance	(0.028)	(0.033)	(0.035)			
Output task motivation*				0.051*	-0.034	0.010
High perceived scientific ability				(0.026)	(0.031)	(0.027)
Observations	4,189	4,189	4,189	4,189	4,189	4,189
R-squared	0.013	0.043	0.091	0.030	0.016	0.096
Mean dep var	0.214	0.412	0.336	0.214	0.412	0.336

Notes: This table presents the effect of survey treatments on the time-use preference question by respondents' initial belief in the importance of science (expected output) and by respondents' initial belief in their ability to contribute to science (expected input). Robust standard errors are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 7: Effect of Survey Treatments on Responses by Outside Option and Perceived Ability

	(1) Work & Donate	(2) Volunteer	(3) Leisure
Output task motivation	0.041***	-0.052***	0.023
	(0.0125)	(0.015)	(0.014)
High wage	0.003	0.039	-0.029
	(0.016)	(0.020)	(0.018)
High perceived scientific ability	0.089***	0.161***	-0.291***
	(0.018)	(0.022)	(0.020)
High perceived scientific ability*	0.083***	-0.097**	-0.011
High wage	(0.025)	(0.031)	(0.028)
Observations	4,189	4,189	4,189
R-squared	0.032	0.018	0.096
Mean dep var	0.214	0.412	0.336

Notes: This table presents the effect of survey treatments on the time-use preference question by respondents' perceived ability and outside options. Respondents' outside options are determined by whether they were assigned a high or low wage in the survey treatment. Robust standard errors are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Appendix A Robustness

Table A1: Effect of Output and Input Treatments on Accuracy

Accuracy	(1) All Dimensions
Output Treatment	0.113**
	(0.055)
Input Treatment	0.018
	(0.063)
Image FE	Yes
Observations	805
R-squared	0.482
Mean dep var	0.493

Notes: Robust standard errors are in parentheses. All columns include image fixed effects. * significant at 10%; ** significant at 5%; *** significant at 1%

Table A2: Effect of Survey Treatments on Responses

	(1) Work & Donate	(2) Volunteer	(3) Leisure
High Wage	0.035***	-0.002	-0.030**
	(0.013)	(0.015)	(0.015)
Output Framing	-0.011	0.005	0.007
	(0.013)	(0.015)	(0.015)
Output task motivation	0.045***	-0.050***	0.015
	(0.013)	(0.015)	(0.015)
Constant	0.179***	0.436***	0.340***
	(0.012)	(0.015)	(0.014)
Observations	4,189	4,189	4,189
R-squared	0.005	0.003	0.001
Mean dep var	0.214	0.412	0.336

Notes: This table presents the effect of the survey treatments on the time-use preference question. Robust standard errors are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table A3: Effect of Survey Treatments on Responses, Multinomial Logit

	(1) Work & Donate	(2) Volunteer	(3) Leisure
High Wage	0036***	-0.003	-0.032**
	(0.013)	(0.016)	(0.015)
Output framing	0.007	0.005	-0.012
	(0.015)	(0.016)	(0.013)
Output task motivation	0.044***	-0.056***	0.012
	(0.013)	(0.015)	(0.015)
Observations	4,189	4,189	4,189
Mean dep var	0.214	0.412	0.336

Notes: This table presented the average marginal effects of survey treatments on reported time us preferences from a multinomial logit regression. * significant at 10%; ** significant at 5%; *** significant at 1%

Table A4: Effect of Survey Treatments on Responses with Controls for Variables that Differ by Treatment Group

	(1) Work & Donate	(2) Volunteer	(3) Leisure
High Wage	0.037***	0.004	-0.038***
	(0.013)	(0.015)	(0.014)
Output framing	-0.012	0.002	0.010
	(0.013)	(0.015)	(0.014)
Output task motivation	0.042***	-0.054***	0.022*
	(0.013)	(0.015)	(0.014)
Believes Scientific Discovery is Important	0.062***	0.212***	-0.290***
	(0.014)	(0.017)	(0.018)
Some Science Experience/Education	0.092***	0.095***	-0.220***
	(0.013)	(0.016)	(0.015)
Constant	0.072***	0.212***	0.706***
	(0.016)	(0.019)	(0.020)
Observations	4,189	4,189	4,189
R-squared	0.023	0.050	0.139
Mean dep var	0.214	0.412	0.336

Notes: This table presents the effect of the survey treatments while controlling for variables that different by treatment group. Robust standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Appendix B Zooniverse Project Pages

B.1 Text on Zooniverse Pages

The text on the output treatment, input treatment, and control pages are below. The text displayed in italics is only included on the input page, and text underlined is only included on the output page.

*****

COVER PAGE

Help our team of computer scientists, social scientists, and Kenyan pastoralists classify vegetation in remote East African Rangelands! Your contributions will help the advancement of crowd science and machine-learning algorithms! *Your contributions will help us replace the hiring of university student research assistants and save hours of our research time!*

ABOUT DIGITAL CROWDSOURCING FOR RANGELAND VEGETATION

The Digital Crowdsourcing for Rangeland Vegetation project aims to improve our knowledge and understanding of the rangeland conditions facing pastoralists, or livestock herders, in East Africa. Current policy and program interventions designed to insure against livestock and other drought related losses in these regions are primarily based on data collected through satellite imagery which is limited in its breadth of information. To improve on this data, researchers have developed a crowdsourcing system in which pastoralists contribute photos of rangelands that can complement satellite data. The machine-learning algorithm then uses these data to improve satellite collected information in order to better target program intervention.
In order to test the accuracy of this method, we need to compare these classifications against the ones made by Zooniverse contributors. By contributing to this project, you will help scientists learn about rangelands, advance crowdsourcing science, and improve machine learning algorithms. *By contributing to this project, you will help us replace the hiring of university student research assistants and save hours of our research time!*

Project Objective

The primary objective of the Digital Crowdsourcing for Vegetation project is to determine whether harnessing local pastoral knowledge through crowdsourced data can be an effective means for measuring and understanding of pastoral rangeland conditions.

Why Do We Need Your Help?

With over 100,000 photos submitted, we are unable to classify them all on our own. We need to match the pastoralists' classifications of the photos with your classifications to determine any discrepancies. We are asking you to help us classify several aspects of the photos. For example, we would like to know if the photos have grass, trees, and shrubs. We would also like you to tell us about the quality of the photos, as some pictures may be blurry, taken in poor lighting (too light or dark), or a poor angle (tilted or not taken at the ground level). <u>While there have been significant advances made in machine learning over the last decade (TechRepublic, 2016), its application to developing economies are still in early stages. Your classifications can contribute to the fine-tuning of machine-learning algorithms that will eventually help target agricultural issues facing many countries. Specifically, machine-learning algorithms can analyze satellite and crowd sourced data like ours to recognize plant diseases, pinpoint areas that require immediate resource provision (e.g., water), and areas in which vegetation is usable and not in need of intervention (Innovation Enterprise, 2016).</u> *We realize that your time is valuable and classifying pictures is a time-intensive activity. By contributing to this project, you will help us replace the hiring of university student research assistants and save hours of our research time!*

Project Background

The northern rangelands of Kenya are inhabited by nearly 3 million pastoralists, who rely on their livestock for food and cash. They search the rangelands for pasture and water; key resources required for the survival of their livestock. Drought, which leads to exhaustion of these resources, is perhaps the most pervasive risk encountered by this population. Development agencies have used remotely sensed (e.g., satellite) data to target interventions aimed at assisting pastoralists make better migration decisions in the face of increasing drought risks. Famine Early Warning Systems Network (FEWSNET), for instance, also uses satellite data to help government decision makers and relief agencies plan for and respond to humanitarian crises. Remotely sensed data for these purposes assesses whether an area being observed has green vegetation. Whether this correlates with available pasture for livestock or not is not clear. This lack of clarity could lead to less-optimal interventions that are costly and ineffective.

The research team first ran exploratory studies to determine how pastoralists classified rangelands, and how much they differed from scientific classifications. These explorations led to a classification system based on the pastoralists' considerations, and to the subsequent development of an android application that was used to crowdsource the rangeland condition information. During their routine herding, pastoralists used the app to take pictures and complete a short survey indicating whether the picture they took had any trees, bushes, or grass. They further indicated whether the trees and bushes had any leaves, and, if so, whether they were green or brown. In addition, they indicated whether each type of vegetation was palatable to goats and sheep, cows, and camels. Lastly, they were required to estimate the carrying capacity of the area in which they took the photo, and how far a water point was from the point the photo was taken. The herders were intensively trained on how to use smartphones, how to take clear pictures with their phones, and how use the application developed for this purpose. They were paid to submit up to ten photos and surveys per day.

We will use your classifications to improve the accuracy of the crowdsourcing algorithm. We also expect to improve our understanding of the types of vegetation that contribute to the survival of livestock and pastoralists and the conditions under which remotely sensed data is informative.

B.2 Email Text Inviting People to Contribute to the Pages

Dear Zooniverse Community,
A new project has just been posted on Zooniverse - Digital Crowdsourcing for Rangeland Vegetation.

This project aims to improve our knowledge and understanding of the rangeland conditions facing pastoralists, or livestock herders, in East Africa. To improve on the satellite data currently being used to inform policy and program interventions, researchers have developed a crowdsourcing system in which pastoralists contribute photos of rangelands that can complement the satellite data. A machine-learning algorithm can then use these data to improve satellite collected information in order to better target program intervention. In order to test the accuracy of this method, we need to compare these classifications against the ones made by Zooniverse contributors. By contributing to this project, you will help scientists learn about rangelands, advance crowdsourcing science, and improve machine-learning algorithms. *By contributing to this project, you will help us replace the hiring of university student research assistants and save hours of our research time!*

To begin contributing to the Digital Crowdsourcing for Rangeland Vegetation, follow this link: (url).

Sincerely,

The Zooniverse Team

B.3 Copy of Contributor Survey

Thank you for your interest in our project on Rangeland Conditions in Northern Kenya. In order to develop a better understanding of why people are contributing to our project, we are asking all our contributors to provide some information about themselves by filling out this 8 question survey. We expect that completing the questions will take you no more than 5 minutes. Thank you!

* Required

1. **1. On average, how many hours per month do you spend contributing to Zooniverse projects?** *

   Mark only one oval.

   - ◯ 0-2
   - ◯ 3-6
   - ◯ 7-10
   - ◯ 11-20
   - ◯ 21 or more
   - ◯ Prefer not to answer

2. **2. Why do you contribute to projects on Zooniverse? Check all that apply.** *

   Check all that apply.

   - [ ] The projects are fun
   - [ ] It's a good way to pass the time
   - [ ] I like that I am contributing to science
   - [ ] I have a background in science and want to make use of it
   - [ ] I am good at classifying items on Zooniverse
   - [ ] Prefer not to answer
   - [ ] Other: ____________________

3. **3. Have you studied or worked in a social or natural science-related field? Check all that apply.** *

   Check all that apply.

   - [ ] I am taking or have taken a course in science
   - [ ] I am majoring or have majored in a science-related discipline in school
   - [ ] I am working or have worked in a science-related field
   - [ ] I have experience with science-related fields in other ways
   - [ ] Prefer not to answer
   - [ ] Other: ____________________

4. **4. What is the highest degree or level of school you have completed? If currently enrolled, highest degree received.** *
Mark only one oval.

- Some high school, no diploma
- High school graduate, diploma or equivalent (e.g. GED)
- Some college, no degree
- Trade/technical/vocational training
- Bachelor's Degree
- Master's Degree
- Doctorate, law, or medical degree
- Prefer not to answer

5. **5. What is your current employment status?** *
Mark only one oval.

- Employed for income full-time
- Employed for income part-time (less than 25 hours/week)
- Self-employed
- Unemployed
- Homemaker
- Student
- Retired
- Prefer not to answer

6. **6. What was your total household income before taxes during the past 12 months?** *
Mark only one oval.

- Less than $40,000
- $40,000 - $74,999
- $75,000 - $119,999
- $120,000 or more
- Prefer not to answer

7. **7. What is your age?** *
Mark only one oval.

- Below 18
- 18-34
- 35-49
- 50-64
- 65 or older
- Prefer not to answer

8. **8. What is your gender?** *

Mark only one oval.

- Female
- Male
- Transgender Female
- Transgender Male
- Gender Variant/Non-Conforming
- Other
- Prefer not to answer

B.4 Copy of Survey Experiment

Understanding Scientific Research Motivation

Disclosure statement:

Elizabeth Lyons, who is a Professor at UCSD, and Laurina Zhang, a professor at Georgia Tech University, are conducting a research study to better understand incentives for participating in scientific research. If you agree to be in this study, the following will happen to you: You have the option to fill in an 11 question multiple-choice survey, which will take 10-15 minutes to complete. At no point in the survey are you asked for your name, or any other information that could allow the researchers to identify you. Research records will be kept confidential to the extent allowed by law. Your responses will be anonymous to the researchers. The survey answers will be stored on a password-protected computer in the PI's office, maintained and updated by University of California staff, and connected to the University of California network. These data security measures meet the University of California Network Security Minimum Standards. Participation in research is entirely voluntary. You may refuse to participate or withdraw at any time. If you want additional information or have questions or research-related problems, you may reach Elizabeth Lyons at lizlyons@ucsd.edu. Please read all instructions and questions carefully.

Start of Block: Motivation framing

Input Framing:

We are trying to learn more about people's willingness to volunteer for scientific research projects, and in particular, for research projects on machine learning, which is a subfield of computer science that gives computers the ability to learn from and make predictions on data. These projects often have significant time lags between when an initial investigation of a possible machine learning application begins and when their applications can be made available for wide use. For instance, a large group of researchers at Microsoft spent more than two years developing the algorithm behind Kinect, the motion-sensing input device used with Xbox.

Output Framing:

We are trying to learn more about people's willingness to volunteer for scientific research projects, and in particular, for research projects on machine learning, which is a subfield of computer science that gives computers the ability to learn from and make predictions on data. These projects can have significant impacts on scientific progress with applications in many fields. For instance, Microsoft's development of Kinect – a motion sensing input device – has significantly advanced the fields of surveillance and robotics, for instance, by increasing the precision by which robots can navigate and avoid obstacles.

Start of Block: Demographics & Background

Q1 What is your highest level of education?

1. Some high school, no diploma
2. High school graduate, diploma or equivalent (e.g. GED)
3. Some college, no degree
4. Trade/technical/vocational training
5. Bachelor's Degree
6. Master's Degree
7. Doctorate, law, or medical degree

Q2 Have you studied or worked in a social or natural science-related field? Check all that apply.

1. I am taking or have taken a course in science after high school
2. I am majoring or have majored in a science-related discipline in school
3. I am working or have worked in a science-related field
4. I have experience with science-related fields in other ways
5. Other ____

Q3 What is your current employment status?

1. Employed for income full-time
2. Employed for income part-time (less than 25 hours/week)
3. Self-employed
4. Unemployed
5. Homemaker
6. Student
7. Retired

Q4 How many hours of volunteer work do you do on average each month?

1. 0
2. 1-4
3. 5-10
4. 11-15
5. 16 or more

Q5 What is your annual income?

1. Less than $40,000
2. $40,000 - $74,999
3. $75,000 - $119,999
4. $120,000 or more

Q6 What is your gender?

1. Female
2. Male
3. Transgender Female
4. Transgender Male
5. Gender Variant/Non-Conforming
6. Other

Q7 What is your age?

1. 18-34
2. 35-49
3. 50-64
4. 65 or older

Start of Block: Scientific Discovery

Q8 For each statement below, please select whether you strongly agree, agree, somewhat agree, neither agree nor disagree, somewhat disagree, disagree, or strongly disagree.

	Strongly Agree	Agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Disagree	Strongly disagree
Scientific discovery is critical for improving people's well-being and quality of life.							
I believe I have the abilities required to contribute to science.							
If I had $100 to donate to a non-profit cause, I would donate it to increasing scientific discovery rather than to other causes such as, supporting the homeless, improving educational opportunities for low income children, or environmental preservation.							

Start of Block: Donation

Q9 Suppose you have the following options on how you would spend one hour of your time. Which option would you prefer?

High Outside Option:

1. Work 1 hour at a rate of $33/hour and donate the $33 you earned to a scientific research study on machine learning.

Low Outside Option:

1. Work 1 hour at a rate of $17/hour and donate the $17 you earned to a scientific research study on machine learning.

Input Motivation:

2. Volunteer 1 hour to a scientific research study on machine learning (e.g., classifying images) and save researchers 1 hour of research assistance time.

Output Motivation:

2. Volunteer 1 hour to a scientific research study on machine learning (e.g., classifying images) and advance the accuracy of the algorithm before it is launched for public use.

Leisure:

3. Do neither and use the time to engage in your favorite leisure activity.