Participation dynamics in crowd-based knowledge production: The scope and sustainability of interest-based motivation

Henry Sauermann
Georgia Institute of Technology, Scheller College of Business
800 W Peachtree St., Atlanta, GA 30308
henry.sauermann@scheller.gatech.edu

Chiara Franzoni
DIG, Politecnico di Milano
P. Leonardo da Vinci 32, Milano, 20133
chiara.franzoni@polimi.it

v. 131128

ABSTRACT

Crowd-based knowledge production is attracting growing attention from scholars and practitioners. One key premise is that participants who have an intrinsic "interest" in a topic or activity are willing to expend effort at lower pay than in traditional employment relationships. However, it is not clear how strong and sustainable interest is as a source of motivation in crowd-based knowledge production. We draw on research in psychology to discuss important static and dynamic features of interest and derive a number of research questions regarding interest-based effort in crowd-based projects. Among others, we consider the specific versus general nature of interest, highlight the potential role of matching between projects and individuals, and distinguish the intensity of interest at a point in time from the development and sustainability of interest over time. We then examine users' participation patterns within and across 7 different crowd science projects that are hosted on a shared platform. Our results provide novel insights into contribution dynamics in crowd science projects. Moreover, given that extrinsic incentives such as pay, status, self-use, or career benefits are largely absent in these particular projects, the data also provide unique insights into the dynamics of interest-based motivation and into its potential as a driver of effort.

PRELIMINARY DRAFT – COMMENTS WELCOME

We gratefully acknowledge support from the Sloan Project "Economics of Knowledge Contribution and Distribution". We thank participants in the Georgia Tech Roundtable for Engineering Entrepreneurship Research (REER) for their feedback. We thank especially Christina Raasch, Cristina Rossi-Lamastra and Uriel Stettner for insightful comments and suggestions. We thank Zooniverse's Chris Lintott, Stuart Lynn, Grant Miller and Arfon Smith for providing access to data and for many valuable discussions.

*"You need to warn people just how addictive this is! Its dangerous! […]. After doing a couple hundred I was starting to burn out … suddenly there was a kelly-green star in the foreground. Whoa! […] being the first to see these things: who *knows* what you might find? Hooked!"*[1]

1 Introduction

An increasing number of organizations reach out to members of the general public – the "crowd" – to help with innovation tasks that have traditionally been performed internally. Among others, this approach is used in innovation contests, crowd sourcing, user innovation, and most recently crowd science (Boudreau et al., 2011; Afuah & Tucci, 2012; Felin & Zenger, 2012; Raasch & Von Hippel, 2012; Boudreau & Lakhani, 2013; Franzoni & Sauermann, 2014). The potential benefits from involving the crowd are twofold. One line of literature emphasizes knowledge related benefits, suggesting that drawing on a large and diverse base of knowledge, skills, and pre-existing solutions can increase the likelihood that a given problem is solved, reduce the time required, and lead to more valuable outcomes (Von Hippel, 2006; Jeppesen & Lakhani, 2010; Afuah & Tucci, 2012). A second line of work suggests that involving the crowd can yield motivational benefits in that its members are often driven by different motivations than workers in traditional employment relationships and may be willing to work for lower pay or even "for free" (Raasch & Von Hippel, 2012; Franzoni & Sauermann, 2014).[2]

One potential reason for contributors to work without financial compensation is "interest", i.e., individuals perform a task because they find it interesting and intellectually stimulating or because they want to satisfy their curiosity. While interest has been suggested as a potentially powerful motivator generally and in the context of crowd-based knowledge production in particular (Silvia, 2006; Shapin, 2008; Raasch & Von Hippel, 2012; Raddick et al., 2013), it is not clear how powerful and sustainable interest is as a source of effort since it is typically difficult to empirically isolate interest from other motivations and incentives. In open source software development, for example, effort may be driven not only be interest but also self-use, career incentives, or social motives (Lerner & Tirole, 2005; Von Krogh et al., 2012). Similarly, participants in innovation contests may be driven by a range of different motivations such as interest, the hope to win a monetary prize, or enjoyment of competition.

In this paper, we provide insights into the potential of interest-based motivation as a source of effort in crowd-based knowledge production. We do so by studying individuals' participation patterns on the largest crowd science platform, Zooniverse, which has thousands of members working on different

[1] Blog entry of a Galaxy Zoo participant. http://chrislintott.net/2007/07/11/galaxy-zoo-press/

[2] In some cases such as innovation contests, participants receive monetary prizes or royalties, yet the expected value of these prizes is typically low, perhaps indicating limited information or overconfidence about one's chance to succeed but also likely reflecting that participants derive nonpecuniary benefits from participation (see Murray et al., 2012).

projects in fields such as astronomy, biology, or climatology. As we will discuss below, Zooniverse has a number of attractive features that allow us to attribute participants' efforts primarily to interest-based motivation, and to observe their efforts within and across projects over time.

To ground our inquiry, we draw on prior work by psychologists who have examined the nature and sources of "interest" for a considerable time (Dewey, 1913; Krapp, 1999; Hidi & Renninger, 2006; Silvia, 2006). Insights from this prior work allow us to formulate a number of specific research questions that guide our analysis of participation patterns in crowd science projects. A first important insight provided by this literature is that interest does not exist in isolation but is the outcome of an interaction between a person and a particular object (e.g., problem, knowledge domain, or task). As such, the same person may have an interest in object A but not in object B. This relational perspective raises questions such as how "specific" interests are, and how willing people are to engage with multiple different objects. Answers to these questions are of great practical relevance for project organizers or platform providers who seek to involve the same crowd in a range of different tasks.

A second insight from prior work is that interest is not static but develops and changes over time (Krapp, 2002). While certain features of a particular object may initially trigger someone's interest, this interest often quickly fades and only few individuals develop a more stable predisposition to repeatedly engage with a particular object. To the extent that this process unfolds also in the context of crowd-based knowledge production, it implies that "free help" is fleeting and that organizers may need to develop tools to repeatedly stimulate interest among existing users, or rapidly replenish their user base with new participants. Yet, some individuals may also develop increasing interest over time, potentially leading them to emerge as key contributors that make significant contributions over an extended period of time. Understanding the role of such individuals and what drives their sustained interest may be critical for project organizers.

After providing a more detailed discussion of interest-based motivation and deriving our research questions, we analyze contribution patterns of over 100,000 Zooniverse members who participated in 7 different projects over a time span of 18 months. We study these data focusing on different levels of analysis, including aggregate participation patterns at the level of projects as well as individual-level patterns of activity within and across projects. Our results show that interest can be a powerful motivator of individuals' contributions to crowd-based knowledge production, as evidenced by thousands of hours of effort invested in the projects we studied. However, both the scope and the sustainability of this interest appear to be rather limited for the large majority of contributors, with many participating only in a single project and only for a few days. At the same time, some individuals show a strong and more enduring interest to participate both within and across projects, and these contributors are ultimately responsible for much of what the project is able to accomplish.

This paper makes several contributions. First, we provide unique descriptive insights into an increasingly important mode of organizing scientific research, and the patterns we observe have direct implications for project leaders who seek to rely to rely on the crowd to provide labor inputs in science or knowledge production more generally. Second, because extrinsic rewards are largely absent and ability matters little, participation data from the Zooniverse platform allow us to gain unique insights into the scope and sustainability of interest-based motivation as a driver of effort. These insights, in turn, may be relevant not only for projects that exclusively rely on contributors' interest in a task, but also for projects that combine a variety of different types of incentives, including more traditional kinds of employment relationships. For example, our results suggest that the compensating wage differentials that need to be paid for jobs offering "interesting" vs. "boring" work will differ both across individuals and over time. Finally, by highlighting interest-based motivation and discussing some of its unique features, this paper suggests a fresh angle for thinking about important questions in the study of science and innovation such as how researchers choose between different research topics, why they may be reluctant to change as new research opportunities emerge, or why research productivity changes over the life cycle.

2 Background

2.1 Interest-based motivation

Scholars in several domains have studied individuals' motivations to exert effort in work and non-work activities. While economists have primarily focused on pecuniary payoffs, psychologists and innovation scholars have made the distinction between extrinsic and intrinsic motivation (Deci & Ryan, 1985; Amabile, 1993; Sauermann & Cohen, 2010). Simply speaking, extrinsic motivation reflects individuals' desire to obtain certain rewards that result from the larger environment and that are external to the individual. Such extrinsic rewards may include pecuniary payoffs such as money, but also social rewards such as recognition from peers. Intrinsic motivation, in contrast, is internal to the individual and does not require rewards provided by the environment. Instead, intrinsic motivation is based on nonpecuniary rewards the individual derives from engaging in the task itself.

While the notion of intrinsic motivation is useful to distinguish sources of motivation that are internal to the individual from those that are external, it captures a range of different motivations that are still quite heterogeneous. For example, intrinsic motivation may result from the enjoyment of solving challenging problems, from curiosity about the object or task, or from anticipated feelings of competence once a problem has been solved (Amabile, 1993; Ryan & Deci, 2000). While all of these sources of motivation are intrinsic in nature, each has different antecedents and may have different consequences.[3]

[3] Similarly, "extrinsic motivation" can reflect a number of different extrinsic rewards such as money, peer recognition, or career advancement. Each of these may operate quite differently depending on the particular context and may lead to different

As such, we will focus specifically on intrinsic motivation based on individuals' "interest" in a particular task or topic. In the following, we draw on prior research in psychology to discuss key features of interest and to derive a number of research questions that will guide our empirical analysis of contribution dynamics in crowd-science projects.

2.2 Interest as the relationship between a person and an object

Interest can be conceptualized as a psychological state that motivates an individual to engage with a particular object, where the object of interest can include not only physical objects (e.g., interest in animals) but also knowledge domains (e.g., interest in chemistry) or activities (e.g., interest in solving problems). Thus, interest is a relational construct that reflects the interaction between the person and a particular object (Krapp, 1999; Silvia, 2006). Assuming for the moment that different people have (potential) interest in different objects (more on individual differences in section 2.4), many person-object pairs may fail to result in interest, i.e., the person and the object are not a match.

Objects of interest can be thought of at different levels of aggregation. For example, individuals may be interested in science at a general level, in particular fields such as biology or astronomy, or even more narrowly in specific topics within a field (Feist, 2006; Krapp & Prenzel, 2011). This multi-level nature of objects suggests a first important research question for our empirical analysis: How specific versus general are the interests of participants on a crowd science platform? If interests tend to be specific to particular topics, we would expect that many participants participate only in single projects and are unlikely to also contribute to other projects in the same field and especially across fields. On the other hand, if contributors' interest tends to be in science at a general level, then we would expect them to participate in multiple projects, potentially spanning different fields of science. A general interest in science would presumably be preferable from the perspective of platform organizers since it may allow them to leverage a given user base for a broader range of different projects.

Research question 1 (generality vs. specificity of interest): What share of participants contributes to just a single project vs. to multiple projects?

The more specific people's interests are (e.g., in certain topics vs. crowd science in general), the more important will be the "matching" between projects and individuals with corresponding interests. In the context of a crowd science platform, individuals may evaluate their match with projects at two stages. First, individuals who receive information about a project may assess it prior to participating, e.g., by reading a project description or looking at examples of tasks. Only those who believe that they may be

---

behaviors. By way of example, a scientist who seeks to maximize income may decide to patent new research findings or keep them secret in order to maximize value appropriation (Cohen et al., 2000). In contrast, a scientist who seeks to maximize peer recognition will publish and diffuse the results widely (Sauermann & Roach, 2014). Thus, while both money and peer recognition are extrinsic rewards, they may lead to very different behaviors.

interested will start working. Second, users may evaluate a match when participating in the project for the first time. Since the projects in our empirical setting are easy to navigate and tasks are very simple, we assume that people will recognize the absence of a match during their first session in the project ("this is not interesting"). As such, we interpret individuals' return to the project for a second time as evidence of an initial match, i.e., that they experienced interest during the first visit. Thus, in order to examine the potential role of matching based on interests, our second research question is:

Research question 2 (matching): What share of individuals who receive information about a project starts to participate? What share of contributors returns to a project a second time?

While discussions of interest typically focus on the intensity of interest with respect to a particular object in an absolute sense, it is useful to also consider the relative strength of interest since increased attention to one object may reduce attention to another. Especially if individuals' total time budget is fixed, we would expect that increased participation in one project comes at the expense of participation in projects in which users have participated previously. However, it is also possible that contributors who join new projects increase the overall amount of time they spend on crowd science activities, e.g., by reducing the amount of time spent on other leisure or work-related activities. To date, there is no systematic evidence on levels of effort devoted to crowd science activities per se, no less on the degree to which participation in new projects results in an expansion of effort vs. crowding-out of effort in existing projects.

Research question 3 (crowding-out vs. expansion of effort): To what extent does participation in new projects reduce effort in old projects vs. increase overall effort devoted to crowd science projects?

2.3 The sustainability of interest

Interest in a given object may change – it may develop and strengthen but it may also decline. Thus, in addition to the existence and intensity of interest at a given point in time (prior section), we now also consider the sustainability of interest over time. Psychologists suggest that when a person first encounters an object, interest may be triggered by certain features of the object that stimulate and catch a person's attention. Attributes that tend to increase the interestingness of an object include its complexity, novelty, uncertainty, and conflict (Berlyne, 1960; Silvia, 2006). In this context, *complexity* refers to the number of elements, the dissimilarity of elements, and the degree to which the whole can be predicted from a part. In experimental studies, for example, people judge complex polygons more interesting than simple polygons and spend more time studying the former. *Novelty* of an object refers to the degree to which it does not fit a person's existing categorizations or is unexpected. Thus, a picture that one has not seen before is typically more interesting than one that has been seen before, and one that includes an unknown or unexpected object (such as a UFO or a green galaxy) will be particularly interesting.

Uncertainty refers to the predictability of events and is highest when there is a large number of alternative events and when the alternatives are likely to occur with similar probabilities. More uncertain events tend to be more interesting, such as a soccer match with two equally good teams rather than one where one team is the clear favorite. Finally, (cognitive) *conflict* is generated if an object entails information that is inconsistent with the information already possessed by the person or that violates certain assumptions the person holds. As such, seeing a human floating in mid-air is interesting because it contradicts our prior knowledge about physics and about which kinds of animals can (and cannot) fly.[4]

Even though complexity, novelty, uncertainty, and conflict tend to raise interest in a particular object, this interest is not necessarily sustained over time. Most obviously, interest should wane as complexity is understood, a novel object becomes familiar, or uncertainty is resolved. As such, the interest in a given object is likely to diminish over time. However, in some circumstances, engaging with an object builds new knowledge, and this knowledge may suggest new questions, may point towards unexplored territory, or may provide raw material for conflict with new incoming information (Prenzel, 1992; Fleming & Sorenson, 2004). To illustrate how prior related knowledge can increase the interest in a particular object, consider that the sighting of a (not particularly attractive) Tasmanian Tiger may not arouse much interest in most people. However, it would be extremely interesting for a biologist who has learned through prior study that this animal was last seen in the 1930's and is believed to be extinct. Despite the possibility that the acquisition of knowledge in the process of engaging with an object may stimulate and renew interest (Silvia, 2006), we expect that important interest-generating attributes such as the object's perceived novelty or complexity will decline for the average person, leading to a reduction in interest over time. The rate of interest decline is an important empirical question, however, because it conditions how long project organizers can rely on a given individual as a source of effort. Assuming that changes in project participation inform us about the dynamics of contributors' underlying interest, we state our fourth research question:

Research question 4 (decline in interest): How fast does the average person's participation in a project decline over time?

[4] Scholars have used an evolutionary perspective to explain why complexity, novelty, uncertainty, and conflict increase interest in an object. In particular, interest in such objects motivates humans to engage with new and complex aspects of the world, which in turn will increase humans' knowledge. Based on this knowledge, humans may be better equipped to understand means-ends relationships, enabling them to navigate and interact with their environment and to shape it to their advantage (Feist, 2006; Silvia, 2006). While this evolutionary perspective provides an explanation for the existence of interest as such, and for why certain objects tend to be more interesting than others, individuals are unlikely to think about the evolutionary benefits of interest when engaging with an object. Rather, interest may involve positively valued emotions that provide more immediate intrinsic rewards (Silvia, 2006). For example, these emotions may come in the form of the satisfaction of curiosity, a continued feeling of enjoyment, involvement, or the absorbing experience of "flow" (see Csikszentmihalyi, 1996; Lakhani & Wolf, 2006).

2.4 Heterogeneity in interest across individuals

Both the intensity of interest in a given object and its development over time may differ significantly across individuals. First, as the example of the Tasmanian Tiger illustrates, individual differences in (static) interest may in part reflect differences in individuals' pre-existing stock of knowledge regarding an object and thus in individuals' evaluations of novelty or cognitive conflict. Second, Silvia (2006) argues that as individuals engage with an interesting object, they realize intrinsic rewards and cognitively link these rewards to engaging with the object. For example, an individual who sees the photograph of a colorful distant galaxy may experience the positive feeling of "wonder" (McDougall, 1960) and may form the cognitive expectation that looking at pictures of galaxies in the future will again provide a pleasurable experience. These cognitive links provide an intertemporal connection and may explain why people return to an interesting object even after a pause or distraction. In particular, even when not engaging with an object (and thus not realizing any intrinsic rewards), people "remember" that the object was interesting in the past and thus return to the object in order to again realize intrinsic benefits.[5] However, these cognitive links between engaging with a particular object and intrinsic rewards may also weaken if subsequent encounters with the object are experienced as "boring", e.g., if a user sees a sequence of images of small and unimpressive galaxies. As such, these cognitive links are constantly evolving and reflect the idiosyncratic experiences different individuals have made with objects over time, possibly leading to diverging levels of interest. Finally, personality psychologists have suggested that heterogeneity in individuals' interest in science generally and even in particular fields may result from differences in individual traits such as dominance, self-confidence, a desire for autonomy, openness to new experiences, or an orientation towards things versus people (Wilson & Jackson, 1994; Holland, 1997; Feist, 1998, 2006).

The discussion of individual differences in prior knowledge, experiences, and personality traits suggests that some individuals may not feel any interest when first encountering a particular crowd science project; these individuals should not start working in the project or should stop participating in the project after the first session – an idea that is already reflected in research question 1 (matching). More interestingly, while we expect a decline in interest for the average person, this discussion also suggests that some individuals may develop strong cognitive links between engaging with an object and intrinsic rewards. If these links result in future interactions that again provide a positive experience, a positive feedback loop may result, potentially leading to a long-term sustained interest in the object.

[5] Similar mechanisms have been used to explain how addiction develops (Silvia, 2006). In a more general sense, this view suggests important path dependencies. As such, early exposure to scientific knowledge or activities provides one potential channel through which the social or educational environment in early life stages may shape individuals' long-term interest in science (Krapp & Prenzel, 2011).

Research question 5 (addiction): What share of individuals shows stable or even increasing activity in a project over time?

The distinction between the intensity of interest and the sustainability of interest, as well as individual differences along both dimensions suggest that the overall distribution of contributions to a project may be highly skewed. That is, most of the contributions may come from a small number of individuals who participate very intensively at a given point in time (e.g., one 20 hour session), who show a continued and sustained activity over time (20 one hour sessions over the course of three months), or who do both. The extent of skew, and the degree to which it results from very intensive activity at a given point in time vs. sustained activity over time, is our final research question.

Research question 6 (top contributors): How concentrated are contributions among top contributors? To what extent does a high number of total contributions to a project reflect intensive activity at a given point in time vs. activity that is sustained over a longer period of time?

We will now examine these research questions empirically. This analysis provides novel insights into participation patterns in crowd-based knowledge production.[6] At the same time, we argue that due to particular features of our empirical setting, the observed patterns of participation provide unique insights into the scope and sustainability of interest as a driver of effort in crowd-based knowledge production.

3 Empirical setting, data, and measures

3.1 Zooniverse crowd science projects as an empirical setting

Our empirical analysis examines participation patterns in projects that involve the crowd in scientific research. Crowd science projects have recently emerged in a variety of fields and have two features that distinguish them from traditional science (Young, 2010; Nielsen, 2011; Franzoni & Sauermann, 2014). First, crowd science projects are open to the participation of any interested individual, including not only professional scientists but also members of the general public that do not know each other in advance. As such, while project leaders are typically professional scientists, contributors in general are not. Second, crowd science projects openly disclose not just final research outputs but also many of the intermediate inputs such as data, problem solving algorithms or logs of discussions among participants. These two features provide many potential benefits including the opportunity to draw on a large and diverse contributor base to increase projects inputs in terms of both knowledge and effort

[6] While patterns of participation in large-scale open projects have been examined in other contexts such as open source software (OSS) development (Lerner & Tirole, 2002; Lakhani & Wolf, 2006), the participants and the projects are qualitatively different in crowd science. For example, while OSS development involves mostly individuals with domain specific skills that are highly valued in the traditional (paid) labor market, crowd science projects rely primarily on members of the general public that have no domain-specific skills and perform relatively simple tasks. Furthermore, while many OSS projects provide contributors with a range of extrinsic rewards (Von Krogh et al., 2012), most crowd science projects do not. There is also prior work exploring the distribution of contributions to crowd-based projects (Wilkinson, 2008; Varshney, 2012) but these analyses are primarily descriptive and provide little discussion of underlying individual-level mechanisms.

(Franzoni & Sauermann, 2014). However, to generate these benefits, projects also have to overcome many challenges including the need to attract and retain a large number of contributors.

While many crowd science projects are stand-alone, an increasing number are hosted on platforms that allow them to share infrastructure and users. Our data come from one of the largest existing crowd science platforms, Zooniverse. Zooniverse began with a single project, Galaxy Zoo, in 2007. Galaxy Zoo was started by astronomers at the University of Oxford who planned to study the formation of galaxies. While the Sloan Digital Sky Survey (SDSS) delivered a massive amount of images that could provide the necessary data, these images could not be reliably analyzed by computers. To overcome this problem, the Oxford scientists established an online interface and asked volunteers to inspect these images and classify astronomical objects based on a number of simple criteria such as their shape. After less than a year, volunteers had classified close to a million images. The project led to a number of scientific publications and the data have been used by the original project team as well as other scientists. Over subsequent years, Galaxy Zoo evolved into a larger online platform – Zooniverse – which currently hosts over 20 projects in fields such as astronomy, archeology, biology, and climatology.[7]

We suggest that crowd science generally, and Zooniverse projects in particular, are not only an increasingly relevant way to produce knowledge, but are also a unique empirical setting that allows us to gain more general insights into interest-based motivation. First, a common problem in seeking to understand sources of motivation is that individuals' effort may reflect a range of different reasons and incentives that are difficult to distinguish based on observed behaviors. In the context of open source software development, for example, behaviors that some observers have interpreted as altruistic or driven by intrinsic motivation have been attributed by others to long-term career concerns (Lerner & Tirole, 2002; Benabou & Tirole, 2003; Von Krogh et al., 2012). Similarly, contributions to innovation contests may reflect a desire for intellectual challenge, the quest for prize money, or a desire to signal skills and ability in the broader labor market (Boudreau et al., 2011). A unique feature of our crowd science projects is that there are virtually no extrinsic incentives available to individual contributors, suggesting that effort is motivated primarily by a narrower set of intrinsic benefits. Supporting this notion, a recent large scale survey asked contributors to Galaxy Zoo about their primary motives for working on this project and shows results that are consistent with the dominance of interest-based motivations (Raddick et al., 2013). By far the most frequently mentioned reasons were an excitement about contributing to scientific research, the opportunity to discover new things, and a general interest in the field of astronomy.

Second, while crowd science projects can ask participants to perform a wide variety of tasks and may require a range of skills (Franzoni & Sauermann, 2014), the Zooniverse projects we study involve

[7] The Zooniverse infrastructure is currently managed by a team located at the Adler Planetarium in Chicago, Illinois. All active projects as well as background information on the platform can be accessed at www.zooniverse.org.

very simple tasks and require only common human skills. As such, individual-level differences in the volume of contributions can be attributed primarily to effort and motivational factors rather than differences in ability.

Third, the "costs" of joining Zooniverse projects are very low and stay relatively constant over time. Entry costs are low because all projects are listed on the same website and use intuitive and similar interfaces, requiring little project-specific investment or knowledge to become a contributor. Moreover, while some crowd-based projects require new entrants to read through or build upon contributions of others – thus potentially changing entry costs over time – Zooniverse participants perform their tasks independently from each other and entry costs remain largely constant throughout the life of a project.[8] As such, the Zooniverse platform allows us to examine individuals' participation patterns across a number of different projects with little interference from entry or switching costs, allowing us to attribute individuals' movements among projects (or lack thereof) more clearly to their interests.

Finally, it is typically difficult to measure and compare effort levels across individuals, especially in settings that involve cognitive effort and where individuals may engage in a range of different types of activities. A unique feature of Zooniverse projects is that the contributions are standardized and well-defined tasks that are measured and recorded at very high levels of resolution. For example, contributors to the project Galaxy Zoo: Hubble inspect images of galaxies taken by the Hubble Telescope and code basic characteristics such as the shape of a galaxy or the direction in which it spins. Each of these contributions is made by the click of a mouse and recorded instantaneously in the Zooniverse system, together with a time stamp. Activity can be tied to individual participants because Zooniverse asks contributors to log in before entering a project, and all projects on the Zooniverse platform use the same login. As such, we can track individuals' participation over time and can compare contributions across individuals using standardized and comparable measures.

3.2 Data

We obtained individual-level participation data directly from the Zooniverse organizers. We use data from all seven projects that were started in 2010: Solar Stormwatch (with project number P04); GalaxyZoo: Supernovae (P05); Galaxy Zoo: Hubble (P06); Moon Zoo (P07); Old Weather (P08); The Milkyway Project (P09) and Planet Hunters (P10).[9] These projects were added sequentially to the platform (see Table 1 for exact dates) and cover the fields of astronomy as well as climatology. Screenshots of the user interfaces of three of the projects are provided in the Appendix for illustration.

[8] In the crowd science projects Foldit or Polymath, for example, participants build on the work of others and the costs of participation may increase over time because newcomers need to process earlier contributions.
[9] Our project numbers start at 04, which reflects that three Zooniverse projects started before our observation window. We do not have detailed participation data for these projects, but we know which individuals joined Zooniverse prior to project 04 (21.17% of all users in our sample). We will account for their "old user" status as appropriate.

Our observation period begins on February 22, 2010, when the project Solar Stormwatch came online. We observe individuals' activity until July 15, 2011.[10] The data include information on the daily activities of 109,253 individuals who participated in at least one of the seven projects during the observation period. Since people joined the platform at different times, the effective observation time differs across individuals; the total number of person-days is roughly 32 million.

3.3 Measures

Our analysis utilizes a set of measures at the level of the individual, the project, and the intersection between the two. The data have a panel structure, i.e., each person can have multiple records in the data set, with each record (row) describing the individual's activity on a given day. In the following description of key measures, ## stands for the project number of the focal project (ranging from 04 to 10).

Classifications in project number ## on a given day (CLASS##, time varying). The basic unit of contributions across all Zooniverse projects is called a "classification". For example, in The Milkyway Project, a classification would be the identification of a hole or a cloud in a picture of galaxies. In the project Old Weather, a classification would be the transcription of a weather report from the scanned logbook of an old ship. The variable CLASS## counts the number of classifications an individual has performed on a given day in project ##.

Participation in a given project on a given day (SESSION##, time varying). This dummy variable takes the value of 1 for each day on which the individual performs at least one classification in project ##. Note that this measure does not count how many times a user visited a project on a given day, but simply indicates whether or not there was a session.

Time spent on project ## on a given day (DURATION##, time varying). This measure captures the time between the first and the last classification in a project on a given day (measured in seconds). If the time between two classifications exceeds 30 minutes, we assume that the session was interrupted and split activity in multiple sub-sessions; DURATION## is then the sum of the time spent in these sub-sessions on a given day. Given the considerable skew in DURATION##, we use the natural log for regression analyses (after adding 1 to the raw score).

Any participation in project ## (ANYCLASS##, time invariant). We code the dummy variable ANYCLASS## as 1 if an individual made at least one classification in project ## during our observation window.

Tenure in project ## (TENURE##, time varying). Tenure in a given project in days, starting with the day on which the individual made the first classification in the project.

[10] We end the analysis in July 2011 because a longer observation period would make the computational requirements too large. The first project in 2011 started on July 25, so we end before this project comes online.

Top contributor status in project ## (TOP##, time invariant). For each project, we sort all contributors by the number of classifications they have made over the whole observation period. We then define as "top contributors" those individuals in the top 10^{th} percentile of contributors. While classifications are comparable within a given project, they are not directly comparable across projects (classifications are more complex and time consuming in some projects than in others). As such, we do not define top contributors at the platform level.

First project in which an individual participated (FIRSTPROJECT##, **time invariant)**. Dummy variable indicating whether project ## was the user's first project on Zooniverse. Individuals who participated in multiple projects on the first day (8.4% of cases), are coded using a separate dummy "**multiple first projects**". Similarly, we code a separate dummy for "**old users**" who joined Zooniverse before project Solar Stormwatch (P04) started.

Participation in the first project in which an individual participated (FIRSTP_SESSION, time varying). This variable records whether the user was active on a given day in the first project in which he or she participated (which could be any one of the seven projects).

Time spent on the first project in which an individual participated (FIRSTP_DURATION, time varying). This variable records the time spent on a given day in the first project in which the user participated (which could be any one of the seven projects). Given the considerable skew, we use the natural log for regression analyses.

Tenure on the platform (TENUREPLAT, time varying). Tenure on the platform in days, starting with the first session in the first project.[11]

User active in second and subsequent projects (SIGNEDUP_SECOND... SIGNEDUP_SEVENTH, time varying). Dummy variable taking on the value of 1 for each day on/after the user started participating in a second or subsequent project.

Number of projects in which the user participated (TOTALPROJECTS, time invariant). This variable captures the number of projects in which the user made at least one classification.

Participation in all projects on a given day (SESSION_ALL, time varying). The sum of all SESSION##.

Participation in at least one project on a given day (SESSION_ALL_D, time varying). Dummy taking the value of 1 if SESSION_ALL>0.

Time spent in all projects on a given day (DURATION_ALL, time varying). The sum of all DURATION##. Given the considerable skew, we use the natural log for regression analyses.

[11] The clock is set to start on February 20, 2010 for old users who joined Zooniverse prior to the start of project 04.

While we do not know the identity of contributors, the data provide information on two individual characteristics that we include as controls in key regressions. First, we distinguish users with and without an academic affiliation based on the domain of the email address they used when signing up on the platform. In particular, we code those participants whose email domain includes ".edu", "ac.uk", or "uni-" as having an academic affiliation (**ACADEMIC**).[12] The share of individuals with an academic affiliation is 5.74% and we control for this variable to account for the possibility that academics may derive certain career related motives from participating in crowd science. Second, we include a dummy variable **ANYCOMMENT** that captures whether a respondent ever posted on one of the Zooniverse discussion boards. This variable is intended to control for potential nonpecuniary benefits users may receive from sharing interesting objects with the broader community. Descriptively, fewer than 5% of users were active on the discussion boards (more details on the discussion boards in section 4.4.).

4 Results

We begin our empirical analysis by providing an impression of the overall level of effort devoted to projects on the platform. We will then examine users' participation in individual projects, speaking to research questions 2, 4, 5, and 6. In a third set of analyses, we will consider individuals' activity across projects, speaking primarily to research questions 1 and 3.

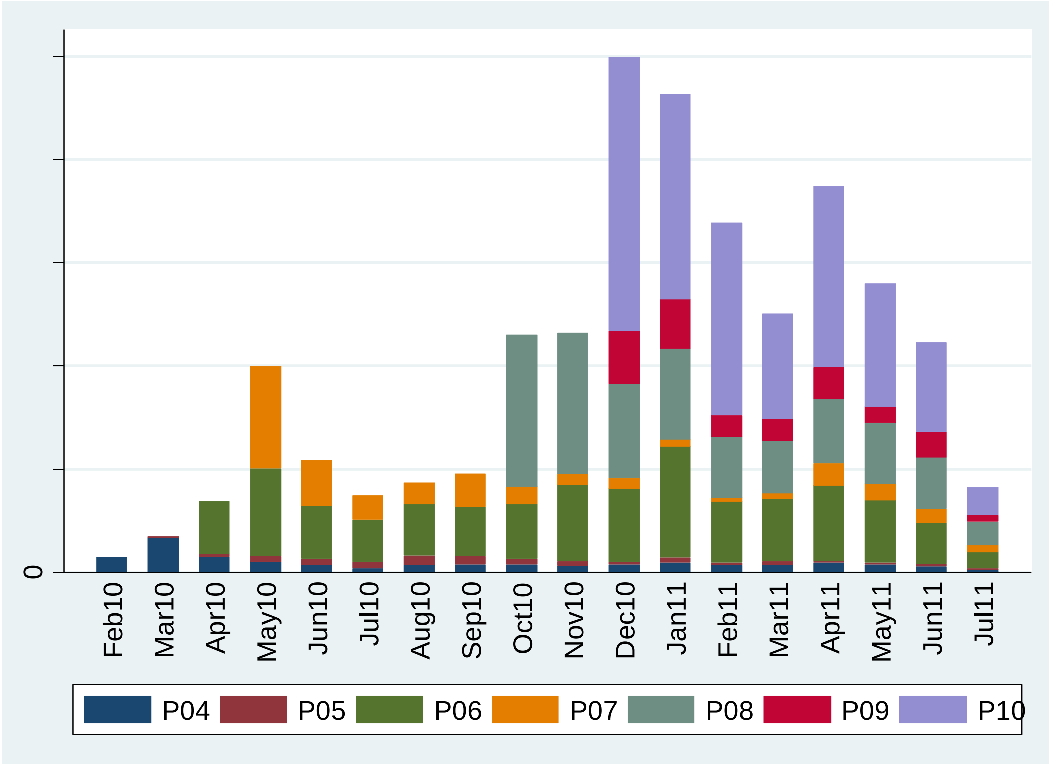
4.1 Project level patterns

To provide a first impression of the effort individuals are willing to commit to crowd science, Figure 1 plots the total amount of time spent by all users on a given project in a given month, showing as much as 25,000 hours spent on the platform in a single month. As expected, the number of hours spent on the platform increases over time as new projects are added. At the same time, we observe that new projects immediately receive a considerable level of contributions when they are started. While one might expect that contributions to a project start low and increase over time as word of mouth makes the project better known in the population of potential contributors, the observed "front loaded" pattern likely reflects that new projects are announced by email to the existing base of Zooniverse members, attracting a considerable number of visitors early on.[13]

[12] A recent survey study suggests that by far the most Zooniverse users come from the U.S. and the UK (Raddick et al., 2013), where the .edu and ac.uk domains should be valid proxies of an academic affiliation. However, users may sign up with non-work email addresses such as gmail.com and we may not identify all individuals with an academic affiliation. Since our data include only email domains (but not the full email), we are unable to obtain more detailed information on users' background.

[13] Figure 1 shows only a relatively small amount of time devoted to project 05. This partly reflects that this project periodically ran out of raw data and did not accept contributions. Results for this particular project should be interpreted accordingly.

Figure 1: Total time spent by month and project (in hours)



Note: Observation period from Feb. 22, 2010 to July 15, 2011

Table 1 provides key statistics for each project and for the overall platform (defined as the seven projects we observe). This table summarizes most of the descriptive results we will discuss in this paper. Among others, Table 1 shows the total number of hours spent on each of the projects during our observation period, as well as the platform total of over 180,000 hours. Even when evaluated at the relatively low U.S. federal minimum wage of $7.25, this effort represents the equivalent of over $1 million in contributions of the "crowd" to scientific work in these 7 projects and over only 18 months alone. The current volume of contributions is likely much larger since Zooniverse has expanded to over 20 projects and has over 800,000 registered users.[14]

--- Table 1 here ---

[14] Similar efforts to quantify the effort invested by non-professional innovators have recently been made in the context of user innovation (Raasch & Von Hippel, 2012; Hienerth et al., 2013). The financial calculations are for illustration only and do not necessarily reflect the "opportunity costs" of effort for Zooniverse contributors. On the one hand, not all contributors are of working age and some are located in countries where wages are lower than in the US. On the other hand, a disproportionate share of Zooniverse contributors have PhDs (Raddick et al., 2013), suggesting that the federal minimum wage underestimates the opportunity cost of their time in the traditional labor market.

4.2 Individual level participation patterns within projects

4.2.1 Matching and decline in interest over time

Our conceptual discussion highlighted the potential importance of the matching between projects with particular topics and individuals with corresponding interests (research question 2). Such a matching may take place in two stages. First, individuals can evaluate projects when receiving information about a new project by email or when exploring the menu of projects available on the Zooniverse website. Only users who think they may be interested will start working on a given project. To get an impression of the share of individuals who match at this stage, we treat individuals who first joined Zooniverse through one of the other six projects as the "installed base" of potential contributors, and examine how many of them start contributing to a particular ("focal") project. While the "installed base" does not encompass all people who receive information about a Zooniverse project, it is a relatively well-defined group that regularly receives project updates from the organizers by email and that can explore additional projects at very low cost through the website of their original project. We find that only a small minority of the users in the "installed base" makes a classification in a focal project (average of 3.49% across all projects, see Table 1).[15] Note that this low percentage may not only reflect that users consciously evaluate a given project and decide not to participate; it may also reflect that users who receive emails about a new project do not read those emails or that users who started in a particular project fail to explore the other projects available on the Zooniverse website.

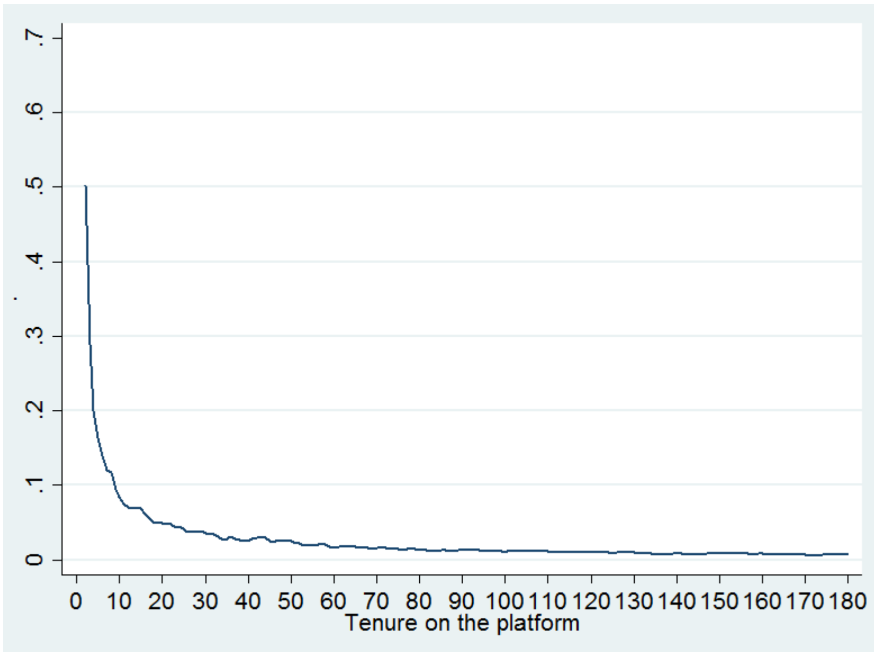
In a second stage, users who started working on a project can experience it during their first session and can decide whether to return a second time. Given that Zooniverse tasks are easy to understand and can largely be experienced during the first session, we suggest that failure to return for a second session indicates that individuals determined that a given project is not a "match". Conversely, returning a second time indicates that a user experienced interest when working on the project for the first time (see Silvia, 2006). Table 1 shows the share of individuals who return for a second session for each of the seven projects, showing that this share averages 27% across projects. Thus, the preponderance of users who try out a given project appears to conclude that it is not a good match. Given our discussion of interest at multiple levels, this result may reflect both a lack of interest in the particular project, but may also reflect a lack of interest in crowd science activities more.

We now turn from the initial matching of projects and users to the dynamics of interest in a project. Hence, we limit our attention to users who return to a project for a second session. Assuming that the lack of a match is discovered during the first session (see above), we interpret declining participation

[15] We observe users who joined Zooniverse prior to the launch of project 04 only if they participate in at least one of the seven projects for which we have data. Since including them in this analysis would overstate the activity of the installed base, we exclude them in this analysis.

in later days as reflecting changes in interest, i.e., that individuals who found the project a good match initially subsequently became less interested. For comparability, we focus on activity in a user's first project (which could be any one of the 7 projects) and we limit the sample to those individuals who are observed for at least 180 days to avoid censoring. Figure 2 shows the average of FIRSTP_SESSION from day 2 to day 180 for the resulting sample of 12,771 individuals.[16]

Figure 2: Likelihood of a session by tenure for a user's first project



Note: Graph shown starting at TENUREPLAT=2, conditional upon at least two sessions in the user's first project over a 180 day observation period (N=12,771).

As expected, Figure 2 shows that the likelihood of participation declines over time, which we interpret as a decline of interest for the average project participant. Complementing Figure 2, Table 1 reports the "half-life" of returning users for each project, i.e., the number of days after which more than 50% of those users who have at least two sessions have dropped out of the project.[17] While there is some variation across projects, this half-life is generally low, with 50% of returning users dropping out after an average of 15 days (research question 4).

We now return to our conjecture that there may be heterogeneity in the dynamics of interest, i.e., that interest in a given project may decline for some people, but may increase for others (research

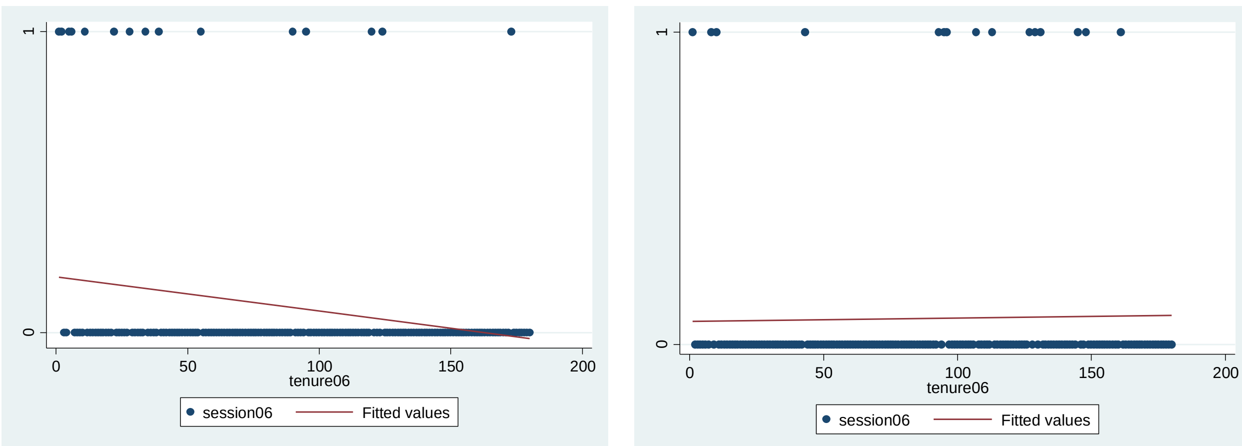
[16] We omit day 1 in this and subsequent figures since FIRSTP_SESSION is by definition equal to 1; excluding day 1 allows us to show subsequent activity at a higher level of detail. Note also that FIRSTP_SESSION is not defined for "old users" who joined Zooniverse before 2010 since we do not observe their first project.

[17] For this analysis, we compute the difference in days between the time a user makes his/her last contribution in a project and the time s/he started. To adjust for right-censoring, we exclude users who made their last classification less than 2 weeks before the end of the observation period (July 15, 2011).

question 5). To examine this possibility, we obtain individual-level estimates of the time trend in participation for each individual with at least 5 sessions in a given project by regressing SESSION## on TENURE##, for a time period of 180 days from the individual's first activity in the project.[18] Figure 3 illustrates this process for two particular participants in Galaxy Zoo: Hubble (P06). Panel A shows 180 daily observations of SESSION06 for contributor 416106, including 15 days with activity in that project (SESSION06=1). For this particular individual, the regression coefficient is negative, indicating that participation tended to become less frequent over time. In contrast, panel B shows the data for individual 54736, who also had 15 days of activity in project 06, but whose activity has become more frequent over time.

Figure 3: Individual-level estimates of time trend in participation

PANEL A (contributor 416106) PANEL B (contributor 54736)



Note: Graphs show SESSION06 for two selected individuals over their first 180 days since starting in project 06. SESSION06=1 indicates activity, SESSION06=0 indicates no activity.

Using this approach, we can examine the share of individuals who have an increasing participation trend in a given project (positive slope) rather than a declining trend in participation (negative slope). Table 1 reports the share of contributors experiencing a positive trend over time, based on all contributors who are observed for at least 180 days and recorded at least five sessions in a given project. Table 1 shows that the share of individuals with a positive slopes ranges from 2.74% to 16.67% across the seven projects. Thus, most individuals participate less over time (likely reflecting a decrease in interest), and only a small minority shows an increase in activity.

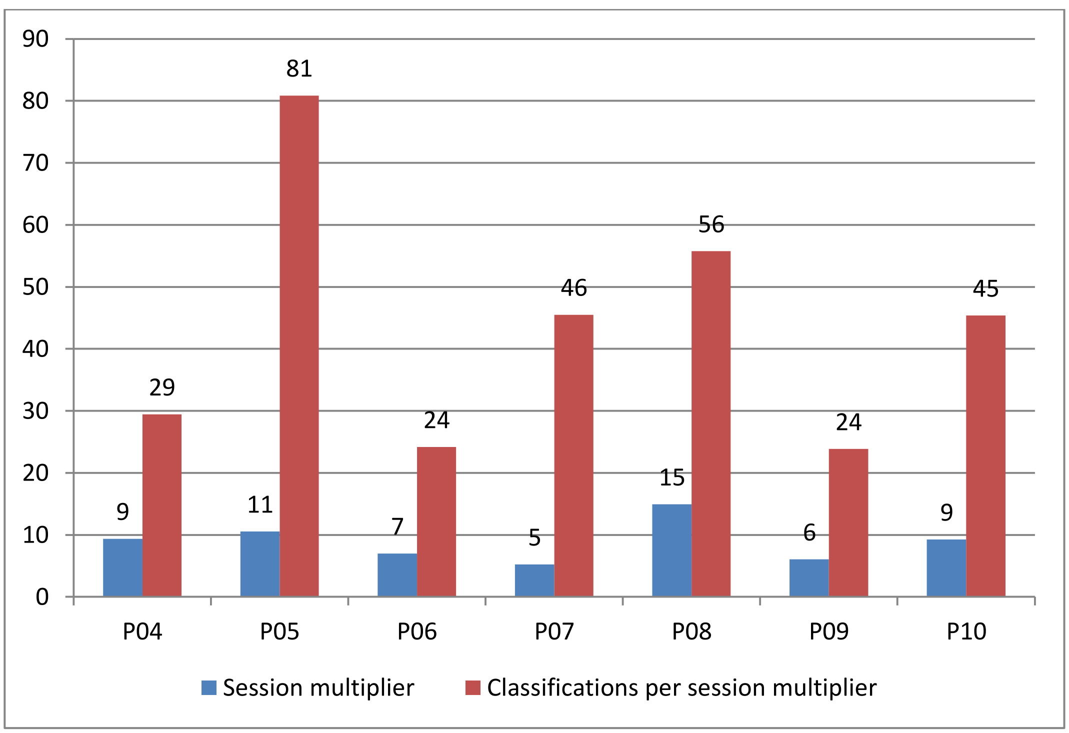
[18] We chose 180 days as the time window for two reasons. First, it allows us to include a larger number of individuals in the estimation, including those who started relatively recently. Second, our descriptive analyses showed that most individuals only participate for a few days, and 180 should be long enough to distinguish those participants with a more sustained interest from those with rapidly declining interest.

4.2.2 Skew in contributions and the role of key contributors

The observation that only a minority of users return to a project and that interest declines for most participants suggests that a small number of individuals will in the end be responsible for much of the output in a crowd science project. To examine the role of top contributors (research question 6), we compute the share of contributions that is made by the top 10% of contributors in terms of their total classifications in a given project. Table 1 shows that across all projects, the contributions of the top 10% of contributors represent over 77% of all classifications. Thus, consistent with prior observations in other settings (Wilkinson, 2008; Varshney, 2012), contributions to crowd science projects are highly skewed.

In principle, top contributor status could reflect a larger number of classifications per session but also a larger number of sessions over time. Figure 4 quantifies both by showing the top contributor "multiplier" with respect to the number of session as well as the number of classifications per session. For example, in project 04, top contributors participated in roughly 9 times as many sessions as non-top contributors, while making roughly 29 times as many classifications per session. Figure 4 shows that the multipliers are generally very large but tend to be higher for the number of contributions than for the number of sessions. Thus, top contributors tend to earn that status not so much because they return more frequently to the project, but primarily because they make more classifications in a given session.

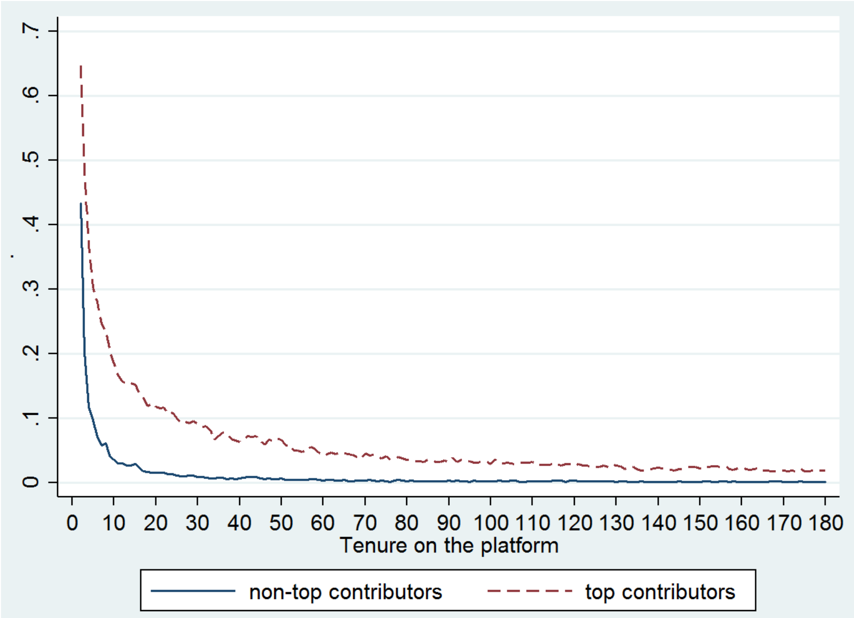
Figure 4: Activity multipliers for top vs. non-top contributors, by project



Note: The session multiplier is the average number of sessions per top contributor divided by the average number of sessions per non-top contributor. The classifications per session multiplier is the ratio of the average number of classifications per session for the two groups.

A natural follow-up question is whether top contributors are those individuals who exhibit increasing interest over time or get “addicted”. If that were the case, we would expect that the individual-level slopes of a regression of SESSION## on TENURE## (see section 4.2.1.) are less negative among top contributors than among other individuals. However, we find just the opposite: the average slope is more negative among top contributors than among non-top contributors. This result reflects that top contributors tend to be very active at the beginning of their tenure in a project, and that this higher starting level of activity provides more opportunity for a steep decline. Figure 5 illustrates this effect by plotting the likelihood of a session in the users’ first project (FIRSTP_SESSION) over the first 180 days (comparable to Figure 2), distinguishing top contributors from non-top contributors.

Figure 5: Likelihood of a session by tenure for a user’s first project, top contributors vs. non-top contributors



Note: Graph shown starting at TENUREPLAT=2, conditional upon at least two sessions in the user’s first project over a 180 day period (N=12,771).

Taken together, these results suggest that top contributors are characterized by particularly heavy activity early in their tenure, both in terms of the frequency of participation (see our analysis of the time trend) and in terms of intensity in a given session (see Figure 4). While they experience a steeper drop in activity than non-top contributors, they continue to have higher absolute levels of activity later in their tenure in a project (Figure 5), but these levels are small compared to their efforts in the beginning.

4.3 Individual level activity across projects

4.3.1 Participation in multiple projects

After having gained insights into contributors' activities within individual projects, we now consider more explicitly activity across projects. This analysis will provide initial insights into the scope of interest, i.e., the degree to which individuals who joined Zooniverse for one project are willing to also work on other projects, and whether movement is more likely within versus across fields (research question 1). At the same time, this analysis provides additional insights into individual heterogeneity in the level of activity across all projects, which we interpret as an interest in crowd science more generally.

Descriptively, we observe that 78.2% of the users who joined Zooniverse during our observation period participated in only one project, 13.4% participated in two projects, and 8.4% participated in more than 2 projects. Thus, a large majority of users appears to be interested in only one particular topic, but some express an interest in crowd science more generally by participating in multiple different projects.[19]

The fact that some users contribute to multiple projects has interesting implications when we switch to the project level of analysis. In particular, even though most users participate in only one project and projects attract only a small share of users from the "installed base" of users who started their activity on the platform in a different project (section 4.2.1.), roughly 40% of contributors to a given project come from the installed base (see Table 1). This observation highlights one potential benefit projects can derive from being hosted on a platform: Drawing on a large installed base of users can yield a substantial number of users whose interests are general enough to lead them to participate in a new project.

4.3.2 Participation and top contributor status as a function of activity in other projects

We now examine participation in multiple projects in more detail. First, we estimate linear probability models predicting whether an individual makes at least one classification to a focal project (ANYCLASS##=1), conditional upon activity in any of the other projects.[20] For example, we take all individuals who started in projects 04-09 and examine which ones are most likely to also contribute to project 10. To account for the fact that some projects came online earlier than others and that individuals joined the platform at different points in time, we include day fixed effects for the date on which an individual joined the platform. The results of this cross-sectional analysis are presented in Table 2.

Models 1-7 use as predictors the dummy variables indicating which project a participant joined first. For ease of interpretation, FIRSTPROJECT08 is omitted. Recall that project 08 (Old Weather) is in

[19] If we limit our attention to users who participate multiple times in their first project (which we define as an initial "match" with that project in terms of interest), we find that 75.9% work on only this first project, 14% contribute to a total of two projects, and 10.1% participate in more than two projects.

[20] We chose linear probability models rather than logit models because LPM estimates have a more intuitive interpretation and LPM is less computationally intensive. The qualitative results are very similar if we use logit estimation.

the field of climatology while all other projects are in astronomy. To the extent that interests are field specific, and assuming that the choice of the first project provides information about a field of primary interest to the user, we would expect that individuals who initially joined an astronomy project are more likely to participate in another astronomy project than individuals who started in the climatology project. Supporting this notion, we find that most of the FIRSTPROJECT## dummies are significant and positive predictors of subsequent participation in another astronomy project. For example, we see that individuals who initially signed up for projects 04, 05, 07, 09, and 10 are more likely to also contribute to project 06, compared to the baseline of the individuals who started with project 08 (model 3). While these results are limited by the fact that the projects span only two different fields and that the data include only one project in climatology, they provide some evidence that interests are somewhat field specific.[21]

Models 8-14 include as predictors a set of dummy variables indicating in which other projects an individual participated (ANYCLASS##) and if s/he was a top contributor (TOP##). With no exception, we find that top contributors in any given astronomy project are more likely than non-top contributors to also participate in another astronomy project. Top contributors in the climatology project are more likely than non-top contributors to join three of the six astronomy projects.[22]

In a second set of regressions (Table 3), we examine which individuals are likely to emerge as top contributors in a given project (TOP##=1), conditional upon participation in that project. We first consider the relationship between top contributor status and the individuals' first project (FIRSTPROJECT##). One conjecture is that individuals may be more likely to emerge as top contributors in their first project on Zooniverse since that project was their "first" (i.e., best) choice at the time of joining the platform and is thus more likely to result in interest. On the other hand, individuals who decide to join a second project (i.e., for whom the focal project is not the first project) are a selected set of individuals who have determined in their first project that they are interested in crowd science activity in general. In other words, if individuals evaluate their interest at multiple levels – that of a particular topic and that of crowd science in general – then individuals who join a second project are likely to be a good match with respect to at least one of these levels and may be more likely to emerge as active contributors. The results in models 1-7 show that only a few of the FIRSTPROJECT## coefficients are significant, and these coefficients are generally negative (with the dummy for the focal project being omitted). As such, users who join a project after having contributed to another project first are not more likely – or even somewhat less likely – to emerge as top contributors in the new project. Thus, the "first choice" effect

[21] Since the data include only one climatology project, we cannot examine whether individuals who signed up for a climatology project are more likely to join another climatology project than an astronomy project. As such, we cannot rule out the alternative explanation that participants in the climatology project are generally less interested in other projects, regardless of the field.
[22] Note that models 8-14 are agnostic as to the timing of participation in different projects, i.e., participation in the project captured in the dependent variable may have occurred after but also before participation captured in the independent variables.

appears to dominate slightly over any existing selection effect associated with the decision to participate in multiple projects.

Models 8-14 of Table 3 include the measures of top contributor status in non-focal projects and show that the associated coefficients are generally positive and significant. For example, a user who is a top contributor in project 06 is 7.6 percentage points more likely than a non-top contributor in project 06 to also emerge as a top contributor in project 09. Thus, top contributors are not only more likely to join additional projects, but they are also likely to be particularly active in those projects, suggesting that their interest is both more general and more intense than that of non-top contributors.

With respect to our two control variables, we see that individuals with an academic affiliation tend to be less likely to participate in an additional project (Table 2). Conditional upon participation, they are more likely to be top contributors in 3 out of 7 projects (Table 3). While one may expect that users with an academic affiliation are more interested in scientific research than members of the larger crowd, one potential explanation for their lower likelihood of participating in additional projects is that interests of "experts" are more specific (i.e., in particular topics) than those of "citizen scientists". Second, we see that people who at some point participated in the discussion board are somewhat more likely to participate in additional projects (Table 2) and are also more likely to be top contributors in about half of the projects (Table 3). Since the forum is primarily used to highlight and comment on particularly interesting objects (see section 4.4.), the positive correlation between forum posts and top contributor status is likely to reflect that individuals who make more classifications are more likely to come across images that are worthy of a post. At the same time, it is possible that posting in the forum increases the intrinsic rewards users gain from discovering interesting objects.

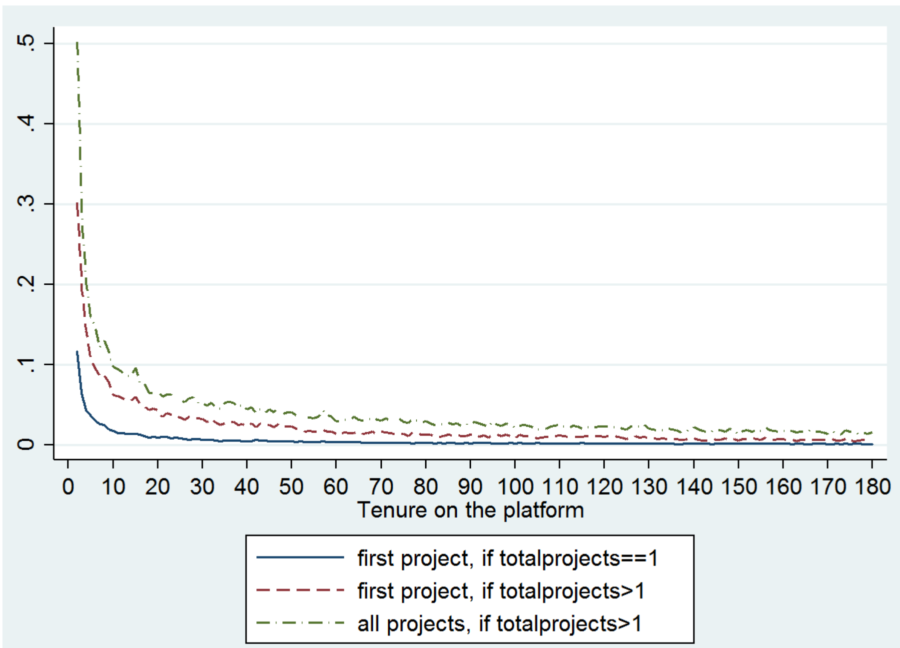
Taken together, our results suggest that users' interests tend to be specific to individual projects. However, there is important heterogeneity in the specificity of interest, with a significant minority of users showing interest in a broader set of projects. Even those who participate in multiple projects, however, appear more likely to expand into related rather than unrelated fields.

--- Tables 2 and 3 here ---

4.3.3 Expansion vs. crowding out of effort

In a final set of analyses, we consider more explicitly the dynamics of effort for those individuals who decide to participate in multiple projects. To establish a baseline, Figure 6 shows participation patterns for individuals who participate in only one project (which is necessarily their first project) and for individuals who participate in more than one project, for a period from day 2 to day 180 since they started on the platform. For those who participated in only one project, we show the average number of sessions in that project on a given day (FIRSTP_SESSION). For those who participate in multiple projects, we show the average number of sessions in the first project (FIRSTP_SESSION) as well as the average number of sessions in all 7 projects combined (SESSION_ALL). Among others, Figure 6 shows that those who decide to participate in multiple projects are overall more active, as evidenced by significantly higher project participation even in the first project.

Figure 6: Average number of sessions per day, by number of total projects



Note: Average FIRSTP_SESSION and SESSION_ALL from TENUREPLAT 2 to 180

Given that some users work on multiple projects, it is useful to consider potential heterogeneity in the time trend in participation not just at the level of individual projects (section 4.2.1) but also at the level of the platform. As such, we estimated the time trend of participation in any of the projects (SESSION_ALL_D) for each individual, conditional upon at least 5 days of activity and for the first 180 days of their tenure on the platform (TENUREPLAT). We find that a significant minority of 14.82% have a positive slope, indicating that their activity became more frequent over time. Since this share of individuals is significantly larger than the share of individuals with positive slopes within single projects (average of 6.25% across projects), some users show declining participation at the level of individual projects but increasing participation at the level of the platform by expanding into new projects.

We now use regression analysis to examine the extent to which an individuals' participation in additional projects is associated with an expansion of overall effort versus a crowding out of effort in the first project. Towards this end, we exploit the panel structure of the data to examine changes in total effort as well as in effort in the first project once a given individual starts to work on subsequent projects (e.g., SIGNEDUP_SECOND, SIGNEDUP_THIRD, etc.). We control for individual-level heterogeneity by employing fixed effects regression.[23] To control for the general decline in activity over time, all regressions include fourth-order polynomials of tenure on the platform.[24] Results are reported in Table 4.

The first model reported in Table 4 uses as the dependent variable the sum of sessions across all projects on a given day (SESSION_ALL). Controlling for a generally declining trend over time, the number of sessions increases by about 0.033 once the individual starts working in a second project, suggesting an expansion of activity.[25] To appreciate the magnitude of this effect, consider that, for those individuals who participate in multiple projects, the median tenure at the time start the second project is 15 days, at which stage the average likelihood of a session in the first project is 0.06 (see Figure 6). Compared to this base level, the increase by 0.033 sessions is quite sizeable. The likelihood of a session on the platform increases further as the participant joins additional projects, although the marginal effect becomes smaller. Complementing this analysis using sessions, model 2 uses as the dependent variable the (log) of the total time spent on all projects in a given day (LNDURATION_ALL) and again shows an increase in time spent once the individual starts his or her second, third and fourth project, although the increase in time levels off faster than the increase in sessions.

To examine potential crowding-out of effort in the user's first project, model 3 in Table 4 uses participation in the first project as the dependent variable (FIRSTP_SESSION). We see that once an individual starts a second project, the likelihood of a session in the first project declines significantly by roughly 0.015. To put this number in perspective, consider again that the likelihood of contributing a session in the first project on day 15 is 0.06. The likelihood of participation in the first project declines further as the user starts to participate in additional projects and we find similar patterns when we use time spent on the first project as the dependent variable (FIRSTP_LNDURATION, model 4).

It is conceivable that effort expansion and crowding out effects depend on whether the user was actually interested in the first project (i.e., a "match"). As such, models 5-8 replicate the analyses for individuals who had at least two sessions in their first project. The qualitative results as well as the effect sizes are very similar to the full sample. Models 9-12 utilize an even narrower sample by focusing on those who are top contributors to their first project. We find that the crowding out in terms of time spent

[23] The analyses reported in this section use data from 8,161 individuals who participate in at least 2 projects and who have a clear first project, i.e., we exclude "old users" who signed up prior to the start of project 04 and users who participated in multiple projects on their first day. While some individuals participate in all 7 projects, the number of cases who participate in more than 4 projects is quite small (N=439) and the respective coefficients should be interpreted with caution.

[24] Figure 6 showed that time trends in participation tend to have a steeper slope for contributors who participate in multiple projects because these individuals start from higher levels of activity. As such, our specification allows for different time trends depending on the total number of projects an individual participates in.

[25] Note that the variables SIGNEDUP_ take on the value of 1 for each day on/after the individual started an additional project. As such, the coefficients do not estimate the change in activity on the particular day on which the user started another project (which would be roughly one additional session), but the average change for all subsequent days.

on the first project is somewhat larger, likely reflecting that top contributors spend more time on the first project and thus have more time to shift away to other projects.

Overall, these results show that individuals who participate in multiple projects increase their overall levels of effort once they start working on additional projects but also reduce their effort in the first project. It should be noted that whether and when an individual joins additional projects is clearly an endogenous choice. As such, our results inform about changes in activity once individuals sign up for additional projects, but they do not necessarily predict what would happen if a random person were to be exogenously made to participate in additional projects. Given the voluntary nature of crowd science activities, insights into the activities of those users who voluntarily decide to participate in multiple projects appear most relevant. At the same time, using fixed effects regression as well as allowing for different time trends for users who participate in multiple projects should largely mitigate concerns about unobserved individual heterogeneity associated with the choice to contribute to multiple projects.

--- Table 4 here ---

4.4 Alternative explanations

The empirical patterns we observed should be of direct interest to scholars and practitioners interested in crowd-based knowledge production. At the same time, we go beyond description by drawing on prior research on the psychology of interest and by interpreting the observed patterns of activity as providing insights into static and dynamic aspects of contributors' interest-based motivation. As such, it is important to consider again how clearly we can attribute observed activity to individuals' motivation generally, and interest-based motivation in particular.

One potential concern is that heterogeneity in contributions reflects heterogeneity in individuals' skills and ability. As noted earlier, all projects require only basic levels of common skills such that differences in individuals' knowledge and ability are unlikely to be the drivers behind the strongly skewed distribution of contributions. Similarly, project-specific skills play only a minor role such that users' tendency to contribute to only one project (or to enter projects within fields rather than across fields) is unlikely to reflect that users have specialized skills and lack the ability to contribute to multiple projects. Finally, heterogeneity in skills and ability provides no explanation for the decline in participation over time – indeed, learning by doing should increase rather than decrease skills and thus the ability to perform the task.

A second concern is that individual differences in participation may be driven by unobserved heterogeneity in opportunity costs. For example, some individuals may contribute actively not because they have a strong interest but simply because they have nothing else to do. While we cannot rule out this possibility, it seems unlikely to explain what we find. First, differences in opportunity cost would not explain why individuals tend to focus on single projects and why participation appears to be somewhat field specific. Second, while it is conceivable that a small number of individuals have extremely *high* opportunity costs, it is less likely that a small number of individuals has extremely *low* opportunity costs, yet the latter would be required to explain why top contributors make 100 or more times the contributions of the average non-top contributor (see Figure 4).[26] Relatedly, it is conceivable that the decline in participation (e.g., Figure 2) reflects an increase in opportunity costs over time. One would expect, however, that an increase in opportunity costs is gradual and linear for the average person, yet the drop in participation is very fast and nonlinear. Similarly, while opportunity costs may increase for some individuals, they should decrease for others, and an opportunity-cost based explanation would suggest a more balanced distribution of increasing and decreasing trends in individuals' participation patterns than what we found in section 4.2.1.

Finally, an important question is whether the observed patterns may reflect motivations other than interest. The particular empirical context of our study allows us to rule out some motivations and suggests that several other candidates are unlikely to play a significant role. For example, the projects provide no financial incentives and, unlike in the context of OSS development, there is little room for "self-use" of the output produced. Similarly, individuals' contributions are not readily visible to the broader community such that peer recognition is unlikely the driver of their efforts. The skills required to contribute to our projects are also very common such that individuals are unlikely to contribute in order to signal skills to potential employers. If signaling or self-use motives were important to any particular group of users, we would expect this to be students and academic scientists, who might think about using project data or results for their own purposes, who may hope to signal skills (or motivation) to other professional scientists, or who may hope to be named on papers for particularly important contributions.[27] Our regressions included a control for users' academic affiliation but this variable had only weak relationships with outcomes of interest and did not impact our key results. One may also wonder about the degree to which working on Zooniverse projects provides "social benefits" and a feeling of community. Zooniverse does not support meaningful individual profiles that could be the basis for more personalized interactions among members. While Zooniverse does have discussion boards for each project, the level of social interaction on these boards is very low. We noted earlier that fewer than 5% of

[26] The distribution of opportunity costs of time is likely to be skewed, but in the opposite direction of what would be needed to explain our results; many people have low opportunity cost while a smaller number of people have very high opportunity cost. Some discussions take the large numbers of hours that people watch TV as a sign that many people face low opportunity costs of time (Cook, 2011). A more economic approach is to look at the distribution of hourly wages, which has a strong positive skew (Congressional Budget Office, 2011).

[27] While contributors are typically not named, in a few cases they appeared as co-authors for unusual contributions. For example, Hanny van Arkel was named in a publication describing a new type of astronomical object she discovered while working on the project Galaxy Zoo (Lintott et al., 2009).

users in our sample ever posted on the discussion boards, and most of the posts lack an interactive component, being limited primarily to highlighting particularly interesting objects (see Figure A4 in the appendix for an example). Our regressions nevertheless control for a user's activity on the discussion boards and including this measure had no impact on the featured coefficients. Finally, some crowd science projects such as Foldit use "gamification" (e.g., leaderboards, competitions, points etc.) in an effort to keep users engaged (Prestopnik et al., 2013). The Zooniverse projects we study use no gamification, except for Old Weather (P08), which allows users to earn special ranks (e.g., Lieutenant or Captain) based on their contributions. Recent qualitative work suggests, however, that the impact of this project feature on participation is ambiguous (Eveleigh et al., 2013) and the founder of Zooniverse has publicly expressed skepticism about gamification for the platform (e.g., in Parr, 2013).

Overall, we believe that the unique empirical setting allows us to attribute the observed patterns of contributions largely to individuals' motivation, especially motivation based on an interest in science or in particular topics and tasks. While some other closely related intrinsic motives such as the desire to contribute to science may also play a role, much of our conceptual discussion as well as our empirical results regarding the scope and sustainability of these motivations would still apply.

5 Summary and discussion

An increasing number of science and innovation projects seek to draw on unpaid labor contributions by the crowd, i.e., members of the general public that are not part of the organizational hierarchy. Especially when pecuniary and other extrinsic incentives are absent, projects' ability to attract and retain such workers can depend critically on the degree to which potential contributors have an intrinsic "interest" in the task. Guided by a number of research questions derived based on prior literature on interest, we provide initial insights into the scope and sustainability of interest-based motivation in the context of crowd science. Our empirical analysis draws on unique individual-level participation data from 7 different crowd science projects hosted on a shared platform, involving a total of over 100,000 individuals.

A first set of analyses examined the scope of interest-based motivation. These analyses built on prior research suggesting that interest should be conceptualized as the relationship between a person and a particular object (e.g., task, project, topic), rather than as a general trait of the person or a general characteristic of the object. Consistent with the notion that interest is quite specific and that many project-person pairs fail to result in a match, we find that most members of the installed base of users on the platform do not sign up for multiple projects, and most of those who try out a project do not return. Even those individuals who participate in multiple projects appear more likely to choose projects in the same scientific field rather than in different fields. Thus, our results suggest that interest-based motivation tends to be quite specific. At the same time, some individuals appear to have an interest that generalizes across topics and fields. Interestingly, controlling for the general time trend, contributors who start with one project and subsequently enter new ones increase their overall level of effort on the platform, although we also observe some crowding-out of effort in the first project.

Building on the notion that a given person's interest in a particular object can develop and change over time, a second set of analyses examined the sustainability of interest. This dynamic analysis shows that interest declines rapidly, with a large majority of the participants who returned to a project (and thus were likely an initial match) dropping out within a few weeks. However, we also observe some contributors whose activity increases over time, especially when we analyze activity at the level of the platform rather than individual projects, thus taking into account switching into additional projects. Individual-level heterogeneity in both initial levels of participation and in the dynamics over time translates into a highly skewed distribution of contributions, with a small share of contributors driving most of the output of projects.

Overall, it appears that interest can be a powerful motivator of individuals' contributions to crowd-based knowledge production, as evidenced by thousands of hours of effort invested in the projects we studied. However, both the scope and the sustainability of this interest appear to be rather limited for the large majority of contributors, with many participating only in a single project and only for a few days. At the same time, some individuals show a strong and more enduring interest to participate both within and across projects, and these contributors are ultimately responsible for much of what crowd science projects are able to accomplish.

These insights have several implications. First, both the low likelihood of a match between projects and people, and the fast decline in interest over time suggest that reaching out to a large number of potential contributors is critical for project success. Indeed, we suggest that "broadcast search" can be powerful not just with respect to identifying individuals who have particular knowledge required for a project (Jeppesen & Lakhani, 2010) but also with respect to identifying individuals that have particular interests.[28] Given the importance of reaching a large number of potential contributors, multi-project platforms may offer several benefits for project organizers: Most obviously, they have a large installed base of users, and a significant number of these users is likely to "check out" newly arriving projects. At the same time, existing projects can benefit from new projects that draw in additional users since some of these users may expand their activity to old projects as well. Such a transfer of users from new to old projects may be particularly important given that a project's existing contributor base "depreciates" quickly. In addition to providing benefits to projects, platforms may also be very attractive to individuals

[28] By way of example, bats are probably not popular animals among most people, but the project Bat Detective has been quite successful in finding those individuals who are interested in this species and are willing to listen to sound recordings and to identify different types of bat calls (Franzoni & Sauermann, 2014).

who are looking for something interesting to do because they provide a range of different projects (increasing the likelihood of a match based on interest) and – due to a shared infrastructure – reduce the cost of exploring and switching between projects. Especially if individuals are unfamiliar with crowd work or are unsure about their interests, the flexibility that platforms are able to offer may give them an important competitive advantage over stand-alone projects in attracting new users.

Second, the low share of matches and the fast decline in interest suggests that developing mechanisms to make projects more interesting and getting people "hooked" should be a central concern of project organizers. Interest theory suggests that project or task characteristics such as novelty, complexity, cognitive conflict, and uncertainty can play an important role, and project organizers should consider how they can establish and maintain these characteristics over time. However, there are likely limits regarding how interesting a task can be made, and even with a wide broadcast search, some topics (e.g., those that are generally perceived to be "boring") may not attract a sufficient number of contributors on the basis on interest alone. As such, a better understanding of the power and limitations of interest-based motivation suggests when there might be a need to consider other types of motives and incentives (see Amabile, 1993). For some projects, e.g., those that are unable to broadcast their needs to a large audience or where only a very small share of the population will have matching interests, the easiest solution may be to use paid crowd labor through platforms such as Amazon's Mechanical Turk (see Kittur et al., 2013). Other projects may be able to increase participation by introducing social and competitive elements, including increasingly popular "gamification" (Prestopnik et al., 2013). Future research is needed, however, on potential interactions between different types of motives and incentives in crowd-based knowledge production. In particular, prior work has suggested that providing extrinsic incentives may undermine intrinsic motivation (Deci et al., 1999; Frey & Jegen, 2001), and there may be similar challenges if projects try to combine interest-based motivation with extrinsic incentives such as competitive rankings or financial rewards (Alexy & Leitner, 2011).

Our insights into interest-based motivation may also be relevant for projects that rely on a broader range of pecuniary and non-pecuniary incentives, including projects using paid crowd labor, innovation contests, or even traditional employment relationships. For example, assuming that effort corresponds to the overall utility individuals are expecting to derive from their work, individuals should be willing to work more or work for lower pay if the project also offers intrinsic benefits such as the opportunity to engage with an interesting object (Rosen, 1986; Stern, 2004). As such, the relational perspective of interest developed in this paper suggests that effort levels and required levels of pay depend on the degree to which there is an interest-based match between a worker and the task and may differ significantly across individuals. One implication is that workers may be willing to work for lower pay on platforms that offer a broader range of tasks and that allow for a better matching of projects and individuals with respect to their interests. Moreover, our findings regarding the decline in interest over time suggests that workers may be willing to work for very low pay or even free in the beginning, but will require additional incentives to remain engaged once their interest in the task has faded. Commonly used static compensation schemes that offer a fixed amount of hourly or performance based pay do not take into account these changes over time, likely "overpaying" individuals early on and "underpaying" them later, potentially explaining why some projects and platforms are able to get a significant number of individuals to join but have a hard time keeping them. Compensation schemes that consider the dynamic nature of interest may be able to address this challenge.

We see several important areas of future research. First, additional work is needed to replicate our analyses in different empirical settings to assess the generalizability of the patterns we observe. We believe that qualitative results such as evidence of the specificity of interest and of a general decline in interest over time will hold also in other empirical settings, but more specific quantitative estimates may differ. To the extent that differences across contexts are observed, follow-on research that theorizes about the sources of such differences and tests them empirically would be particularly valuable.[29] Second, more research is needed to explore the characteristics of those individuals who emerge as the top contributors in crowd-based efforts. While we characterized top contributors with respect to their patterns of activity on the platform (section 4.3.2), additional work drawing on richer data sources including demographic characteristics or survey measures of motivation seem particularly promising. Finally, future work might focus more explicitly on the attributes of projects, including potential levers organizers might use to attract participants and keep them engaged. As noted earlier, prior work suggests that novel, complex, and uncertain objects, as well as objects that generate cognitive conflict tend to be perceived as more interesting. As such, organizers who avoid overly simple tasks, introduce novel tasks or data, allow room for surprises and unexpected discoveries (see our quote at the beginning of the paper), or who provide relevant background knowledge may generate higher levels of interest. Of course, there may be trade-offs. For example, while more complex tasks tend to be more interesting, they are likely to increase coordination challenges and may also preclude users with limited skill levels from participating, potentially reducing the base of potential contributors (Franzoni & Sauermann, 2014). Similarly, introducing novel tasks and data may distract some users and may disrupt established routines. These potential trade-offs point to the need to consider both motivational and organizational aspects when seeking to improve project design.

Finally, our discussion of "interest" also point towards novel directions for research in science and innovation more generally. For example, there is increasing evidence that intrinsic motivation is

[29] For example, prior work has speculated that different rates of decline in user involvement across projects may reflect different costs of entry (Wilkinson, 2008).

particularly beneficial for creativity and innovation (Amabile, 1996; Sauermann & Cohen, 2010). Even though an intrinsic motivational orientation is to some extent a stable individual trait (Amabile et al., 1994), our discussion of the specificity of interest suggests that the strength of a given person's intrinsic motivation may also depend critically on the particular project or task. As such, while studies of pecuniary as well as nonpecuniary incentives typically abstract from the "content" of the particular task (Siemsen et al., 2007; Sauermann & Cohen, 2010; Boudreau et al., 2011; Larkin et al., 2012), it may be useful to characterize task content more explicitly, and to study issues such as the matching between specific tasks and individuals or potential decreases in motivation when workers are moved between tasks. Scholars interested in understanding why individuals (and organizations) choose certain topics or research trajectories may benefit from considering that choices of topics reflect not only expected abstract benefits such as money, recognition or the pleasure from solving puzzles (Stephan, 2012) but also an interest in very "concrete" topics or tasks. Similarly, our discussion of the dynamics of interest suggests that researchers may abandon topics even before their scientific potential has been exploited, and may provide an alternative explanation for declines in research effort over scientists' life cycle (Levin & Stephan, 1991). Of course, studying interest may require a broader range of empirical approaches, including historical and qualitative work exploring what triggered a person's initial interest in a particular topic or problem (for a fascinating early example, see Galton (1874)).[30]

Overall, our work suggests that interest-based motivation may be a powerful source of effort in crowd science projects, but important limitations as well as significant heterogeneity across individuals should not be ignored. We hope that our results provide useful insights for project organizers and stimulates interest among scholars seeking to understand individual motivation within and beyond the context of crowd-based knowledge production.

[30] Research-active readers may enjoy thinking about the role interest has played in shaping their own research trajectories.

REFERENCES

Afuah, A., & Tucci, C. L. 2012. Crowdsourcing as a solution to distant search. ***Academy of Management Review***, 37(3): 355-375.

Alexy, O., & Leitner, M. 2011. A fistful of dollars: Are financial rewards a suitable management practice for distributed models of innovation? ***European Management Review***, 8(3): 165-185.

Amabile, T. 1996. ***Creativity in Context***. Boulder, Colo.: Westview Press.

Amabile, T. M. 1993. Motivational synergy: Toward new conceptualizations of intrinsic and extrinsic motivation in the workplace. ***Human Resource Management Review***, 3: 185-201.

Amabile, T. M., Hill, K. G., Hennessey, B. A., & Tighe, E. M. 1994. The work preference inventory - Assessing intrinsic and extrinsic motivational orientations. ***Journal of Personality and Social Psychology***, 66(5): 950-967.

Benabou, R., & Tirole, J. 2003. Intrinsic and extrinsic motivation. ***Review of Economic Studies***, 70(3): 489-520.

Berlyne, D. E. 1960. ***Conflict, Arousal, and Curiosity***. New York: McGraw Hill.

Boudreau, K. J., Lacetera, N., & Lakhani, K. R. 2011. Incentives and problem uncertainty in innovation contests: An empirical analysis. ***Management Science***, 57(5): 843-863.

Boudreau, K. J., & Lakhani, K. R. 2013. Using the crowd as an innovation partner. ***Harvard Business Review***, 91(4): 60-69, 140.

Cohen, W. M., Nelson, R. R., & Walsh, J. P. 2000. Protecting their intellectual assets: Appropriability conditions and why U.S. manufacturing firms patent (or not), ***NBER Working Paper #7552***.

Congressional Budget Office. 2011. Changes in the distribution of workers' hourly wages between 1979 and 2009: Congressional Budget Office.

Cook, G. 2011. How crowdsourcing is changing science, ***The Boston Globe***.

Csikszentmihalyi, M. 1996. ***Creativity: Flow and the Psychology of Discovery and Invention***: HarperCollins.

Deci, E. L., Koestner, R., & Ryan, R. M. 1999. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. ***Psychological Bulletin***, 125(6): 627-668.

Deci, E. L., & Ryan, R. M. 1985. ***Intrinsic Motivation and Self-determination in Human Behavior***. New York: Plenum.

Dewey, J. 1913. ***Interest and Effort in Education***: Houghton Mifflin Company.

Eveleigh, A., Lynn, S., Cox, A., & Jennett, C. 2013. I want to be a captain! Gamification in the Old Weather citizen science project, ***Working Paper***.

Feist, G. J. 1998. A meta-analysis of personality in scientific and artistic creativity. ***Personality and Social Psychology Review***, 2(4): 290-309.

Feist, G. J. 2006. ***The Psychology of Science and the Origins of the Scientific Mind***. New Haven: Yale University Press.

Felin, T., & Zenger, T. R. 2012. Open innovation, problem-solving and the theory of the (innovative) firm, ***Working Paper***.

Fleming, L., & Sorenson, O. 2004. Science as a map in technological search. ***Strategic Management Journal***, 25(8 9): 909-928.

Franzoni, C., & Sauermann, H. 2014. Crowd Science: The organization of scientific research in open collaborative projects. ***Research Policy***, 43(1): 1-20.

Frey, B. S., & Jegen, R. 2001. Motivation crowding theory. ***Journal of Economic Surveys***, 15(5): 589-611.

Galton, F. 1874. ***English Men of Science: Their Nature and Nurture***. London: Macmillan.

Hidi, S., & Renninger, K. A. 2006. The four-phase model of interest development. ***Educational Psychologist***, 41(2): 111-127.

Hienerth, C., von Hippel, E., & Berg Jensen, M. 2013. User community vs. producer innovation development efficiency: A first empirical study. ***Research Policy***.

Holland, J. L. 1997. ***Making Vocational Choices: A Theory of Vocational Personalities and Work Environments*** (3rd ed.). Odessa, Fla.: Psychological Assessment Resources.

Jeppesen, L. B., & Lakhani, K. 2010. Marginality and problem-solving effectiveness in broadcast search. ***Organization Science***, 21(5): 1016-1033.

Kittur, A., Nickerson, J. V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., Lease, M., & Horton, J. 2013. ***The future of crowd work***. Paper presented at the Proceedings of the 2013 conference on Computer supported cooperative work.

Krapp, A. 1999. Interest, motivation and learning: An educational-psychological perspective. ***European Journal of Psychology of Education***, 14(1): 23-40.

Krapp, A. 2002. Structural and dynamic aspects of interest development: Theoretical considerations from an ontogenetic perspective. ***Learning and Instruction***, 12(4): 383-409.

Krapp, A., & Prenzel, M. 2011. Research on interest in science: Theories, methods, and findings. ***International Journal of Science Education***, 33(1): 27-50.

Lakhani, K., & Wolf, R. 2006. Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects. In J. Feller, B. Fintzgerald, S. Hissam, & K. Lakhani (Eds.), ***Perspectives on Free and Open Source Software***: MIT Press.

Larkin, I., Pierce, L., & Gino, F. 2012. The psychological costs of pay-for-performance: Implications for the strategic compensation of employees. ***Strategic Management Journal***, 33(10): 1194-1214.

Lerner, J., & Tirole, J. 2002. Some simple economics of open source. ***Journal of Industrial Economics***, 50(2): 197-234.

Lerner, J., & Tirole, J. 2005. The economics of technology sharing: Open source and beyond. ***Journal of Economic Perspectives***, 19(2): 99.

Levin, S. G., & Stephan, P. E. 1991. Research productivity over the life cycle: Evidence for academic scientists. ***American Economic Review***, 81(1): 114-132.

Lintott, C. J., Schawinski, K., Keel, W., Van Arkel, H., Bennert, N., Edmondson, E., Thomas, D., Smith, D. J. B., Herbert, P. D., & Jarvis, M. J. 2009. Galaxy Zoo:'Hanny's Voorwerp', a quasar light echo? ***Monthly Notices of the Royal Astronomical Society***, 399(1): 129-140.

McDougall. 1960. ***An Introduction to Social Psychology***. London: Methuan.

Murray, F., Stern, S., Campbell, G., & MacCormack, A. 2012. Grand Innovation Prizes: A theoretical, normative, and empirical evaluation. ***Research Policy***.

Nielsen, M. 2011. ***Reinventing Discovery: The New Era of Networked Science***: Princeton University Press.

Parr, C. 2013. Why citizen scientists help and how to keep them hooked, ***The Times Higher Education***.

Prenzel, M. 1992. The selective persistence of interest. In S. Renninger, S. Hidi, & A. Krapp (Eds.), ***The Role of Interest in Learning and Development***: 71-98: Lawrence Erlbaum.

Prestopnik, N. R., Crowston, K., & Wang, J. 2013. Data quality in purposeful games, ***Working Paper***.

Raasch, C., & Von Hippel, E. 2012. Amplifying user and producer innovation: The power of participation motives, ***Working Paper***.

Raddick, M. J., Bracey, G., Gay, P. L., Lintott, C., Cardamone, C., Murray, P., Schawinski, K., Szalay, A., & Vandenberg, J. 2013. Galaxy Zoo: Motivations of Citizen Scientists, ***Working Paper***.

Rosen, S. 1986. The theory of equalizing differences. In O. Ashenfelter, & R. Layard (Eds.), ***Handbook of Labor Economics***: 641-692: North-Holland.

Ryan, R. M., & Deci, E. L. 2000. Intrinsic and extrinsic motivations: Classic definitions and new directions. ***Contemporary Educational Psychology***, 25(1): 54-67.

Sauermann, H., & Cohen, W. 2010. What makes them tick? Employee motives and firm innovation. ***Management Science***, 56(12): 2134-2153.

Sauermann, H., & Roach, M. 2014. Not all scientists pay to be scientists: Science and engineering PhDs' price of publishing in industrial R&D. ***Research Policy***, 43(1): 32-47.

Shapin, S. 2008. ***The Scientific Life: A Moral History of a Late Modern Vocation***: University of Chicago Press.

Siemsen, E., Balasubramanian, S., & Roth, A. V. 2007. Incentives that induce task-related effort, helping, and knowledge sharing in workgroups. ***Management Science***, 53(10): 1533-1550.

Silvia, P. J. 2006. ***Exploring the Psychology of Interest***: Oxford University Press.

Stephan, P. 2012. ***How Economics Shapes Science***: Harvard University Press.

Stern, S. 2004. Do scientists pay to be scientists? ***Management Science***, 50(6): 835-853.

Varshney, L. R. 2012. ***Participation in crowd systems***. Paper presented at the Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on.

Von Hippel, E. 2006. ***Democratizing Innovation***: MIT Press.

Von Krogh, G., Haefliger, S., Spaeth, S., & Wallin, M. W. 2012. Carrots and rainbows: Motivation and social practice in Open Source Software Development. ***MIS Quarterly***, 36(2): 649-676.

Wilkinson, D. M. 2008. ***Strong regularities in online peer production***. Paper presented at the Proceedings of the 9th ACM conference on Electronic commerce.

Wilson, G. D., & Jackson, C. 1994. The personality of physicists. ***Personality and Individual Differences***, 16(1): 187-189.

Young, J. 2010. Crowd science reaches new heights. ***The Chronicle of Higher Education***, 28.

Table 1: Summary Statistics (from project start to July 15, 2011)

	Solar Stormwatch (# 04)	Galaxy Zoo Supernovae (#05)	Galaxy Zoo Hubble (#06)	Moon Zoo (#07)	Old Weather (#08)	Milkyway Project (#09)	Planet Hunters (#10)	Platform level (defined as all 7 projects) and/or weighted average across projects
Start date	2/22/10	3/26/10	4/19/10	5/11/10	10/12/10	12/7/10	12/16/10	2/22/10 platform
Field	Astronomy	Astronomy	Astronomy	Astronomy	Climatology	Astronomy	Astronomy	n/a
Total users with at least one session	14,917	6,139	52,536	27,089	10,320	13,189	30,518	109,253 platform
Total time spent by all users (in hours)	8,179	3,486	47,134	16,565	38,327	10,872	58,141	182,704 platform
Share of Zooniverse "installed base" who contributes to the project*	2.32%	1.77%	6.43%	2.9%	1.86%	3.69%	5.46%	3.49% average
Share of returning users (i.e., at least two sessions in a given project)	20.5%	17.1%	33.5%	19.9%	39.7%	25.9%	32.3%	26.99% average
Half-life of returning users (time after which >50% of returning users drop out)	10 days	26 days	19 days	17 days	9 days	9 days	15 days	26 days platform 15 days average
Share of individuals with positive slope in regression of SESSION## on TENURE##	2.74%	16.67%	6.28%	5.93%	4.51%	3.86%	3.76%	14.82% platform 6.25% average
Share of classifications made by top 10% of contributors	71.6%	83.8%	71.3%	81.3%	83.7%	70.8%	79.5%	77.43% average
Share of users drawn from "installed base" of users who originally started in another project	40.43%	55.19%	29.98%	30.56%	41.20%	49.12%	33.29%	39.97% average
Share of users who participates in only one project*	49.74%	18.58%	65.45%	63.69%	58.72%	32.88%	59.64%	78.21% platform 49.81% average
Share of returning users who participates in only one project*	42.13%	8.48%	56.98%	41.96%	62.32%	25.36%	52.41%	46.37% platform 41.38% average

* excludes users who joined platform prior to start of observation period.

Table 2: Participation in additional projects

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
VARIABLES	anyclass04	anyclass05	anyclass06	anyclass07	anyclass08	anyclass09	anyclass10	anyclass04	anyclass05	anyclass06	anyclass07	anyclass08	anyclass09	anyclass10
FIRSTPROJECT04		0.013**	0.049**	0	0.011**	0.020**	0.037**							
		[0.003]	[0.008]	[0.004]	[0.004]	[0.005]	[0.006]							
FIRSTPROJECT05	0.005		0.119**	0.006	0.013*	0.036**	0.078**							
	[0.006]		[0.013]	[0.007]	[0.005]	[0.008]	[0.010]							
FIRSTPROJECT06	0.001	0.012**		0.002	0.003	0.020**	0.030**							
	[0.003]	[0.002]		[0.002]	[0.002]	[0.003]	[0.004]							
FIRSTPROJECT07	0.001	0	0.030**			0.012**	0.021**							
	[0.003]	[0.003]	[0.007]			[0.004]	[0.005]							
FIRSTPROJECT08														
FIRSTPROJECT09	0.015**	0.014**	0.053**	0.011**	0.012**		0.032**							
	[0.004]	[0.003]	[0.007]	[0.003]	[0.003]		[0.006]							
FIRSTPROJECT10	0.013**	0.013**	0.043**	0.005	0.007**	0.017**								
	[0.003]	[0.002]	[0.006]	[0.003]	[0.002]	[0.004]								
MULTIPLE FIRST PROJECTS	0.047**	0.043**	0.107**	0.038**	0.038**	0.072**	0.094**							
	[0.004]	[0.003]	[0.008]	[0.004]	[0.003]	[0.005]	[0.006]							
OLDUSER	0.192**	0.076**	0.438**	0.176**	0.111**	0.113**	0.227**							
	[0.004]	[0.020]	[0.052]	[0.042]	[0.020]	[0.033]	[0.038]							
TOP04									0.102**	0.124**	0.127**	0.103**	0.079**	0.096**
									[0.010]	[0.013]	[0.011]	[0.010]	[0.010]	[0.011]
TOP05								0.072**		0.140**	0.060**	0.041*	0.041*	0.136**
								[0.019]		[0.023]	[0.019]	[0.016]	[0.017]	[0.020]
TOP06								0.023**	0.062**		0.047**	0.021**	0.086**	0.061**
								[0.004]	[0.004]		[0.005]	[0.004]	[0.005]	[0.005]
TOP07								0.077**	0.046**	0.140**		0.071**	0.076**	0.076**
								[0.007]	[0.006]	[0.010]		[0.007]	[0.007]	[0.008]
TOP08								0.029**	0.013	0.039**	0.015		0.007	0.033**
								[0.011]	[0.008]	[0.013]	[0.010]		[0.010]	[0.012]
TOP09								0.039**	0.033**	0.167**	0.062**	0.028**		0.100**
								[0.011]	[0.009]	[0.015]	[0.011]	[0.009]		[0.014]
TOP10								0.019**	0.043**	0.094**	0.039**	0.019**	0.059**	
								[0.006]	[0.005]	[0.008]	[0.006]	[0.005]	[0.007]	
ANYCLASS04									0.050**	-0.044**	0.031**	0.042**	0.058**	0.027**
ANYCLASS05								0.091**		0.148**	0.082**	0.043**	0.084**	0.090**
ANYCLASS06								0.001	0.046**		0.020**	0.011**	0.052**	0.007*
ANYCLASS07								0.023**	0.040**	-0.031**		0.037**	0.060**	0.029**
ANYCLASS08								0.053**	0.040**	-0.050**	0.075**		0.066**	0.054**
ANYCLASS09								0.048**	0.046**	0.023**	0.063**	0.039**		0.123**
ANYCLASS10								0.021**	0.043**	-0.059**	0.037**	0.034**	0.095**	
ACADEMIC	-0.019**	-0.004*	0.022**	-0.017**	-0.008**	-0.019**	-0.024**	-0.014**	0	0.007	-0.012**	-0.002	-0.010**	-0.018**
	[0.002]	[0.002]	[0.007]	[0.003]	[0.002]	[0.003]	[0.003]	[0.002]	[0.002]	[0.007]	[0.003]	[0.002]	[0.002]	[0.003]
ANYCOMMENT	-0.012**	0.002	-0.052**	-0.015**	-0.005*	0.022**	0.397**	-0.019**	-0.010**	-0.055**	-0.032**	-0.018**	-0.008	0.377**
	[0.003]	[0.003]	[0.005]	[0.004]	[0.003]	[0.004]	[0.012]	[0.003]	[0.003]	[0.005]	[0.004]	[0.003]	[0.004]	[0.012]
Start plat day fe	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.
Constant	-0.008*	0.007	0.107*	0.099*	0.009	0.04	0.042	-0.020**	-0.055**	0.201**	0.048	-0.038*	-0.03	0.032
	[0.003]	[0.020]	[0.052]	[0.042]	[0.020]	[0.033]	[0.038]	[0.004]	[0.017]	[0.051]	[0.040]	[0.019]	[0.031]	[0.034]
Observations	100,367	106,502	72,469	90,442	103,185	102,543	88,894	100,367	106,502	72,469	90,442	103,185	102,543	88,894
R-squared	0.087	0.032	0.306	0.146	0.052	0.051	0.135	0.115	0.1	0.326	0.179	0.091	0.122	0.177
df	477	509	512	516	512	512	516	487	522	519	522	522	522	522

Note: Sample includes individuals who started in a project other than the focal project. All models include fixed effects for the day of a user's first activity on the platform. *=significant at 5%, **=significant at 1%.

Table 3: Top contributor status

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
VARIABLES	top04	top05	top06	top07	top08	top09	top10	top04	top05	top06	top07	top08	top09	top10
FIRSTPROJECT04		-0.022	-0.050**	-0.009	-0.025	-0.041*	-0.040*							
		[0.033]	[0.014]	[0.034]	[0.024]	[0.020]	[0.016]							
FIRSTPROJECT05	0.064		0.014	-0.024	-0.01	-0.084	-0.04							
	[0.061]		[0.025]	[0.045]	[0.042]	[0.052]	[0.030]							
FIRSTPROJECT06	0.002	-0.025		0.033**	-0.007	0.005	-0.002							
	[0.014]	[0.014]		[0.013]	[0.015]	[0.012]	[0.009]							
FIRSTPROJECT07	-0.024	0.017	-0.011		0.016	0	0.003							
	[0.017]	[0.027]	[0.010]		[0.026]	[0.019]	[0.014]							
FIRSTPROJECT08	-0.024	-0.043	-0.025	-0.023		0.046	0.012							
	[0.033]	[0.035]	[0.021]	[0.051]		[0.028]	[0.025]							
FIRSTPROJECT09	-0.031	-0.035	-0.005	-0.023	-0.051*		-0.036*							
	[0.022]	[0.032]	[0.016]	[0.031]	[0.022]		[0.015]							
FIRSTPROJECT10	-0.017	0.012	-0.063**	-0.018	0.004	0.002								
	[0.014]	[0.023]	[0.009]	[0.017]	[0.018]	[0.013]								
MULTIPLE FIRST PROJECTS	-0.021**	-0.034**	-0.032**	0.003	-0.018*	-0.021**	-0.022**							
	[0.007]	[0.010]	[0.003]	[0.005]	[0.007]	[0.006]	[0.004]							
OLDUSER	0.03	0.592*	-0.164	0.16	0.147**	0.176	0.074							
	[0.029]	[0.301]	[0.108]	[0.107]	[0.041]	[0.169]	[0.094]							
TOP04									0.063**	0.040**	0.025	0.017	-0.008	-0.02
									[0.022]	[0.015]	[0.019]	[0.018]	[0.018]	[0.017]
TOP05								0.079**		0.126**	0.085**	0.047	0.001	0.075**
								[0.030]		[0.021]	[0.027]	[0.034]	[0.027]	[0.023]
TOP06								0.034*	0.056**		0.045**	0.035*	0.076**	0.056**
								[0.013]	[0.012]		[0.013]	[0.015]	[0.012]	[0.011]
TOP07								0.027*	0.040*	0.042**		0.012	0.073**	0.040**
								[0.014]	[0.017]	[0.010]		[0.017]	[0.016]	[0.013]
TOP08								0.036	0.018	0.072**	0.05		0.041	0.031
								[0.027]	[0.033]	[0.021]	[0.028]		[0.028]	[0.022]
TOP09								0.019	0.009	0.133**	0.122**	0.037		0.059**
								[0.019]	[0.021]	[0.016]	[0.023]	[0.023]		[0.015]
TOP10								0.027	0.037*	0.068**	0.077**	0.050**	0.066**	
								[0.016]	[0.018]	[0.012]	[0.018]	[0.019]	[0.014]	
ANYCLASS04									-0.014	-0.022**	0.017*	-0.008	-0.013	-0.011*
ANYCLASS05								0.002		0.022**	0.001	-0.024*	-0.034**	-0.007
ANYCLASS06								-0.016**	-0.008		0.002	-0.005	0.005	-0.009*
ANYCLASS07								0.001	-0.025**	-0.022**		-0.017*	-0.019**	-0.006
ANYCLASS08								0.022**	-0.005	-0.021**	0.017*		-0.014	-0.025**
ANYCLASS09								-0.005	-0.033**	0.009	0	-0.022**		-0.023**
ANYCLASS10								-0.009	0.017*	-0.021**	-0.012	0.006	-0.003	
ACADEMIC	0.067**	-0.001	0.021**	-0.003	-0.018	0.142**	-0.001	0.071**	0.001	0.023**	-0.003	-0.019	0.145**	-0.001
	[0.015]	[0.017]	[0.005]	[0.011]	[0.012]	[0.016]	[0.008]	[0.015]	[0.017]	[0.005]	[0.011]	[0.012]	[0.016]	[0.008]
ANYCOMMENT	0.046**	-0.007	0.076**	0.036*	0.021	0.063**	0.087**	0.037*	-0.018	0.048**	0.015	0.011	0.052**	0.086**
	[0.014]	[0.015]	[0.010]	[0.015]	[0.017]	[0.011]	[0.005]	[0.015]	[0.015]	[0.010]	[0.016]	[0.018]	[0.011]	[0.005]
Start plat day fe	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.
Project total tenure fe	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.
Constant	-0.02	-0.547	0.134	-0.108	-0.141**	-0.166	-0.081	-0.018	-0.54	0.112	-0.14	-0.147**	-0.186	-0.11
	[0.029]	[0.301]	[0.108]	[0.107]	[0.041]	[0.168]	[0.094]	[0.029]	[0.317]	[0.113]	[0.107]	[0.037]	[0.159]	[0.095]
Observations	14,917	6,137	52,536	27,089	10,320	13,189	30,518	14,917	6,137	52,536	27,089	10,320	13,189	30,518
R-squared	0.435	0.514	0.228	0.276	0.393	0.325	0.345	0.438	0.521	0.236	0.283	0.396	0.334	0.349
df	774	581	958	800	673	640	698	798	619	968	820	689	656	709

Note: Sample includes only individuals who participated in a given project. Models include fixed effects for the day of a user's first activity on the platform as well as the total number of days a user is observed in the focal project. *=significant at 5%, **=significant at 1%.

Table 4: Effort expansion vs. crowding-out

	Full Sample				If match in first project				Top 10 in first project			
	Total effort		Effort in first project		Total effort		Effort in first project		Total effort		Effort in first project	
	1	2	3	4	5	6	7	8	9	10	11	12
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
	session_all	lduration_all	firstp_session	firstp_lduration	session_all	lduration_all	firstp_session	firstp_lduration	session_all	lduration_all	firstp_session	firstp_lduration
SIGNEDUP_SECOND	0.033**	0.104**	-0.015**	-0.101**	0.039**	0.080**	-0.016**	-0.122**	0.053**	0.099**	-0.014**	-0.146**
	[0.002]	[0.010]	[0.001]	[0.008]	[0.002]	[0.015]	[0.002]	[0.012]	[0.005]	[0.031]	[0.004]	[0.028]
SIGNEDUP_THIRD	0.050**	0.107**	-0.026**	-0.188**	0.062**	0.119**	-0.027**	-0.209**	0.075**	0.123*	-0.033**	-0.291**
	[0.004]	[0.020]	[0.002]	[0.015]	[0.005]	[0.027]	[0.003]	[0.022]	[0.010]	[0.051]	[0.006]	[0.045]
SIGNEDUP_FOURTH	0.079**	0.154**	-0.036**	-0.263**	0.089**	0.157**	-0.039**	-0.303**	0.106**	0.158*	-0.046**	-0.401**
	[0.008]	[0.039]	[0.004]	[0.028]	[0.009]	[0.047]	[0.005]	[0.036]	[0.015]	[0.080]	[0.010]	[0.067]
SIGNEDUP_FIFTH	0.089**	0.129*	-0.033**	-0.251**	0.104**	0.162*	-0.033**	-0.267**	0.136**	0.198*	-0.034*	-0.335**
	[0.015]	[0.059]	[0.007]	[0.052]	[0.016]	[0.065]	[0.008]	[0.062]	[0.026]	[0.097]	[0.014]	[0.102]
SIGNEDUP_SIXTH	0.094**	0.124	-0.038**	-0.302**	0.101**	0.115	-0.043**	-0.357**	0.138**	0.211	-0.047**	-0.454**
	[0.021]	[0.098]	[0.009]	[0.058]	[0.023]	[0.109]	[0.010]	[0.065]	[0.037]	[0.176]	[0.016]	[0.106]
SIGNEDUP_SEVENTH	0.088*	0.226	-0.046**	-0.338**	0.097*	0.225	-0.051**	-0.387**	0.183**	0.465	-0.047**	-0.458**
	[0.037]	[0.190]	[0.009]	[0.059]	[0.038]	[0.199]	[0.010]	[0.066]	[0.055]	[0.294]	[0.016]	[0.112]
TENUREPLAT (polynom.)	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.
Constant	0.201**	1.142**	0.155**	0.999**	0.264**	1.528**	0.212**	1.404**	0.372**	2.373**	0.311**	2.245**
	[0.002]	[0.013]	[0.002]	[0.012]	[0.003]	[0.019]	[0.002]	[0.018]	[0.006]	[0.040]	[0.005]	[0.038]
Observations	2,077,127	2,077,127	2,077,127	2,077,127	1,287,194	1,287,194	1,287,194	1,287,194	531,333	531,333	531,333	531,333
R-squared	0.053	0.057	0.07	0.066	0.062	0.071	0.086	0.084	0.077	0.097	0.107	0.113
Number of refid	8,161	8,161	8,161	8,161	5,149	5,149	5,149	5,149	2,032	2,032	2,032	2,032
df	30	30	30	30	30	30	30	30	30	30	30	30

Note: Conditional upon participation in at least two projects. OLS, with fixed effects for individual contributors and clustered standard errors. Controls include 6 sets of 4^{th} order polynomials of TENUREPLAT, one set for each group of users who participated in n=2…7 total projects.

Figure A1: Screenshot of Galaxy Zoo: Hubble

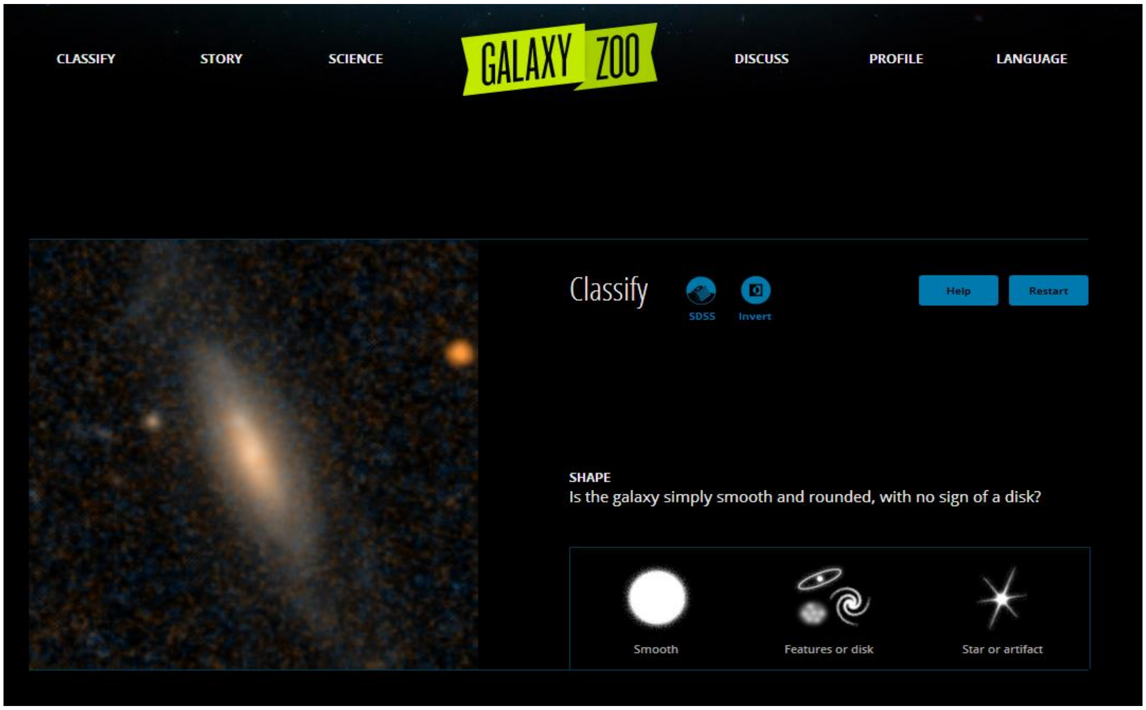


Figure A2: Screenshot of Old Weather

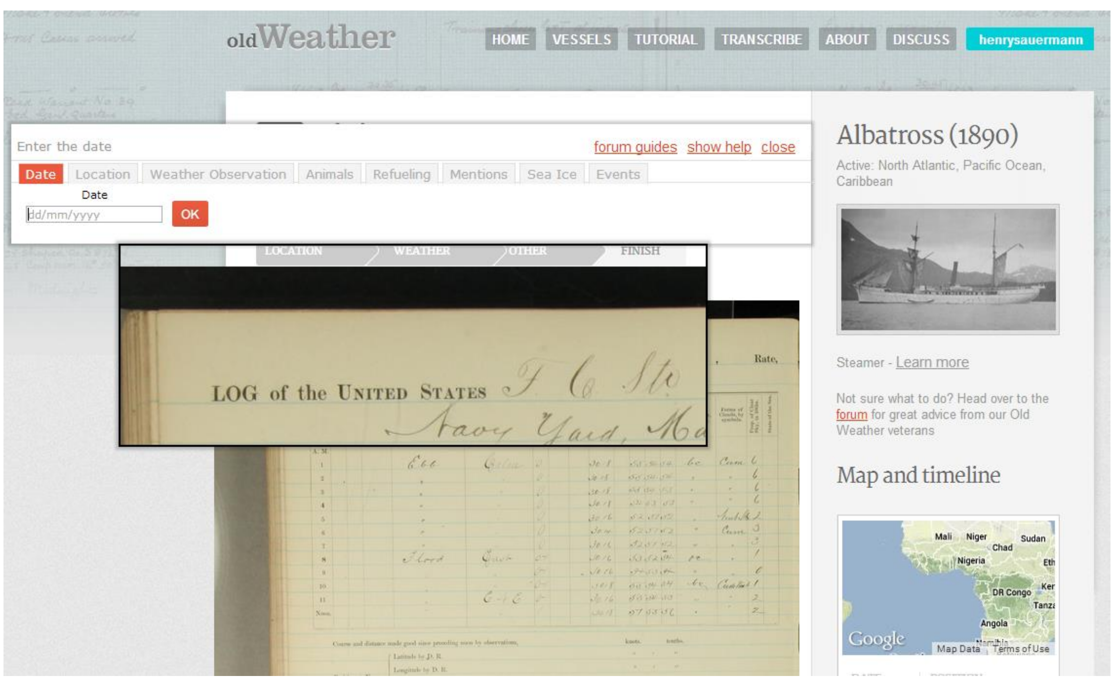


Figure A3: Screenshot of Planet Hunters

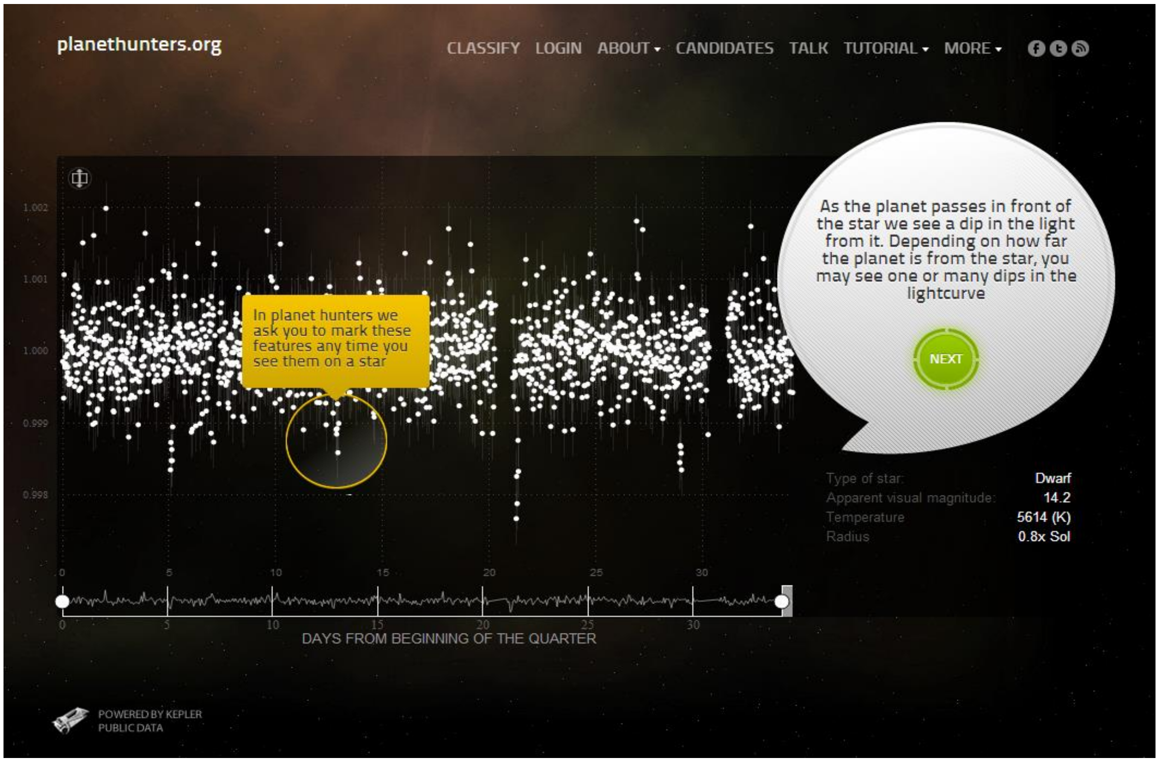
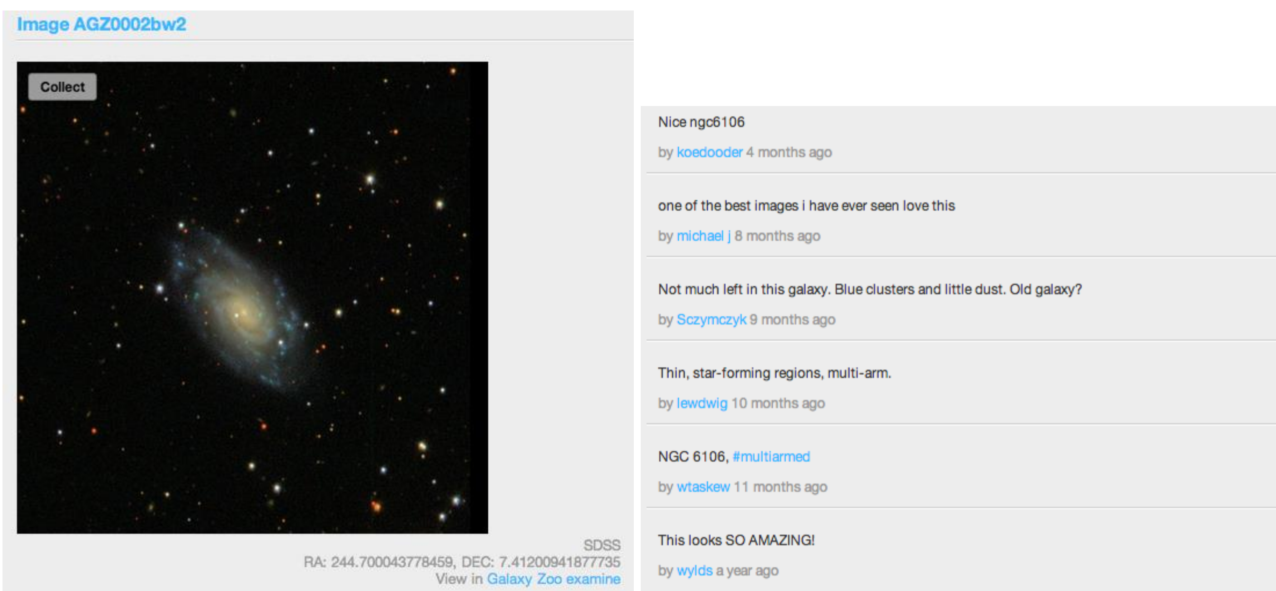


Figure A4: Screenshot of forum entry



Source: http://talk.galaxyzoo.org/#/subjects/AGZ0002bw2, accessed Nov. 22, 2013