Check for updates

Article

S|S|S

Social Studies of Science
1–25
© The Author(s) 2018
BY NC
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0306312718778806
journals.sagepub.com/home/sss
SAGE

The epistemic culture in an online citizen science project: Programs, antiprograms and epistemic subjects

Dick Kasperowski
Department of Philosophy, Linguistics and Theory of Science, University of Gothenburg, Gothenburg, Sweden

Thomas Hillman
Department of Education, Communication and Learning, University of Gothenburg, Gothenburg, Sweden; Oxford Internet Institute, University of Oxford, Oxford, UK

Abstract

In the past decade, some areas of science have begun turning to masses of online volunteers through open calls for generating and classifying very large sets of data. The purpose of this study is to investigate the epistemic culture of a large-scale online citizen science project, the Galaxy Zoo, that turns to volunteers for the classification of images of galaxies. For this task, we chose to apply the concepts of programs and antiprograms to examine the 'essential tensions' that arise in relation to the mobilizing values of a citizen science project and the epistemic subjects and cultures that are enacted by its volunteers. Our premise is that these tensions reveal central features of the epistemic subjects and distributed cognition of epistemic cultures in these large-scale citizen science projects.

Keywords

anti-program, citizen science, epistemic culture, epistemic subject, Galaxy Zoo, program

Correspondence to:
Dick Kasperowski, Department of Philosophy, Linguistics and Theory of Science, University of Gothenburg, Olof Wijksgatan 6, 405 30 Gothenburg, Sweden.
Email: dick.kasperowski@gu.se

Introduction

In the past decade, some areas of science and other research have begun turning to masses of online volunteers, sometimes for finding optimized solutions to well specified problems, but more commonly for generating and classifying very large sets of data. This raises general questions about whether and how the mobilization, through open calls, of online masses into scientific work is distinguishable from the contributions made by professional scientists.

The practice of involving volunteers in scientific work is traceable through scientific publications in currently available databases (Web of Science) back to the mid-1960s, notably with the North American Breeding Bird Survey (Kullenberg and Kasperowski, 2016). According to the standard history offered by proponents of citizen science, the practice has a 120-year history, often typified by the Audubon Christmas bird count starting in 1900 and by various initiatives in astronomy (Kärnfelt, 2015). However, such initiatives decreased as universities and governments took financial stewardship of research in the mid-twentieth century, and the sciences became professionalized (Goodchild, 2007; Star and Griesemer, 1989).

Outsiders as epistemic subjects in scientific work have been both assets and problems for science. The issue of whether non-scientists might violate scientific standards of data collection has been a long-time companion to the inclusion of outsiders. Consequently, outsider co-workers have not always been acknowledged in the production of scientific knowledge, though to what extent is difficult to answer (Cooper et al., 2014). However, it has been argued that 'hundreds of scientific papers' in bird migration studies would not have been possible without unacknowledged contributors (Haklay, 2013: 113). More generally, the visibility of actors in scientific work is not a given, but subject to cultural values and power: Shapin and Schaffer's (1985) notion of 'invisible technicians' denotes vital scientific workers that are part of the invisible infrastructure of science where distributed work and invisibility is combined with very visible epistemic subjects such as Robert Boyle (Shapin and Schaffer, 1985) and Carl von Linné (Fors, 2015). Thus co-workers, citizen scientists, could be thought of as invisible technicians, belatedly made more visible as contributors to science.

Today, citizen science as a scientific method is particularly used in conservation, ecology and other areas of biology, but increasingly also in geography, social epidemiology, medicine and the humanities (Kullenberg and Kasperowski, 2016). With this mobilization of outsiders in scientific research, concerns continue about data quality. Indeed, much of the discussion on the possibility to engage citizens in scientific work has been about volunteers' abilities to collect and classify data properly (Cornwell and Campbell, 2011: 105), and about the potential for the development of skills that ensure the quality of observations or classifications. These issues have overshadowed enquiries into what values volunteers develop as epistemic subjects taking part in such projects, a gap addressed in this article in relation to the epistemic culture of a large on-line citizen science project.

By extending research teams with online volunteers, it becomes possible to pursue time-consuming tasks (such as observation and classification) that cannot be well automated, particularly in fields that need to cover large geographical areas, manage very large datasets or long time-spans. For many research fields, 'big data' carries the promise of new discoveries, if resources can be adequately used to provide for analytic work. In those projects building on pattern recognition and observations, participatory protocols are held to be sustainable and reliable. This potentially includes research initiatives in all disciplines, the humanities, social and natural sciences.

The overarching issue of interest in this paper is the epistemic culture, and more specifically the epistemic subjects and values, created in the mobilization of volunteers for Galaxy Zoo, a large online citizen science project in astronomy that involves the classification of large amounts of data and the possibility of new discoveries.

We chose to examine the 'essential tensions' (Kuhn, 1977) that arise between the mobilizing values of a project and the epistemic cultures and subjects that are enacted by the volunteers. Our premise is that these tensions reveal central features of an epistemic culture resulting from the invitation of outsiders into scientific work. We also view our material through the lens of programs and anti-programs (Akrich and Latour, 1992) as central features of a tension between the individualized and more distributed epistemic subjects in an epistemic culture (Knorr-Cetina, 1999). The following thematic questions guide our investigation:

- Through which values are volunteers mobilized into the epistemic culture of an online citizen science project?
- How do tensions in the epistemic culture arise as epistemic subjects enact these mobilization values?
- What epistemic subjects are present in an online citizen science project?

The empirical material consists of interviews with scientists, online discussion forum moderators and programmers, and participant observation at project meetings, but first and foremost a corpus of extensive digital traces from discussion forums connected to the Galaxy Zoo, the first established project on the now multidisciplinary online citizen science platform, Zooniverse.

Citizen science, participation and epistemic representation

It would be a mistake to see citizen science as homogenous, as there are different intertwined forms of relations between institutionalized science and the outsiders.

There are a number of terms used by Science and Technology Studies (STS) scholars to describe social movement-based initiatives that seek epistemic representation for and from local communities: 'street science', 'popular epidemiology', 'community based participatory research', 'civic science' and 'citizen science', among others (see Brown, 1992; Corburn, 2005; Kimura, 2016; Ottinger, 2010; Wylie et al., 2017). These initiatives involve local communities battling environmental or health issues by collecting data or disrupting protocols and research designs (Epstein, 1996). This is done in response to scientific knowledge that cannot represent the particular needs and problems of local communities, which tend not to be visible in conventional scientific approaches and modes of representation (Kimura, 2016; Ottinger, 2010). Even though these initiatives emerge from outside of the institutions of science, they often rely on scientific standards – and in some cases scientific laboratories – for validating data (see Ottinger, 2010). The use of data becomes a form of civic empowerment in relation to corporate or government interests (Kullenberg, 2015; Orta-Martínez and Finer, 2010; Ottinger, 2010; Wylie et al., 2017).

Some seminal STS works on citizen science by Wynne (1992), Irwin (1995, 2001) and others attend to different actors' possibilities of opening up scientific knowledge and science policy to relevant stakeholders, often with reference to deliberative and democratic processes. Here, the knowledge produced may be used in public consultations and participatory exercises, to try to establish legitimate governance of decision-making involving science and technology, and understanding the role of the citizen as a representative stakeholder in science and policy.

These citizen sciences stand in stark contrast to the model of citizen science, crowd science or crowdsourced science (Franzoni and Sauermann, 2014) developed in the natural and physical sciences. Here, citizen science is mainly understood and practiced as a method of including contributors in clearly defined and restricted aspects of the research process, often to collect observations and make classifications. This citizen science is a method of including contributors in top down projects initiated by professional researchers. This is the 'form' of citizen science in focus for this study.

Conceptual perspective

Epistemic cultures, subjects and mobilization values in citizen science

The concept of epistemic culture was introduced in STS as way of understanding the 'arrangements and mechanisms … which, in a given field make up how we know what we know' (Knorr-Cetina, 1999: 1). Through her extensive fieldwork, Knorr-Cetina shows that high-energy physics and molecular biology either distribute or individualize the epistemic subject in their respective epistemic cultures. In high-energy physics, the individual person is minimized as an epistemic subject, responsibility and authority is not centralized, rewards for scientific discovery are shared (as in authorship) and a high level of trust is developed and valued. Thus, this field's epistemic culture displays a strong displacement of the knowing subject in favour of the experiment as a collective endeavour. Central values in this epistemic culture are not associated with the individual and no single person can be identified as producing knowledge. In contrast stands molecular biology, where the individual epistemic subject 'structures' research, both in projects and in managing the laboratory as a workplace, and is associated with high cultural value.

Giere (2002) and Giere and Moffat (2003) argues that, irrespective of research field, distributed cognition is central to scientific practice and epistemic gains, and yet that some individual human epistemic agency is still unavoidable. Magnus (2007) asserts that 'humans with normal human cognitive capacities' are able to do research, since they have constructed systems of distributed cognition (p. 297). The emphasis on 'normal humans' is strikingly close to the values by which volunteer outsiders are mobilized to perform scientific work as citizen scientists. As our account of how the work produced by volunteers in Galaxy Zoo unfolds, we will return to the issues of the relation between human epistemic agency and distributed cognition.

The possible tension between distribution and individualization in epistemic cultures are empirical questions and are, we argue, of importance and interest for understanding the epistemic cultures developed when ‘outsiders’ are mobilized into scientific work. In its original instantiation, the concepts of epistemic culture and distributed cognition concerned only communities of professional scientists and technicians. However, Knorr-Cetina (2007) acknowledges that not all contexts of knowledge production are ‘bounded spaces’, suggesting the need for studies of more distributed locations and networks of different size and scale, where ‘[s]uch networks are made possible by electronic connections, and … have global reach’ (p. 367).

In this study, we make restricted use of the epistemic culture concept, as an analytic entry point for exploring tensions between individual and distributed characteristics of the epistemic subject in scientific work. We illustrate how distributed and epistemic subjects are configured in performing scientific tasks, and display tensions when outsiders are mobilized – specifically in projects using large-scale online platform technology and volunteers for classificatory work. We hypothesize that an important aspect of widening participation in this form of citizen science is the mobilization of volunteers on the basis of constructions of intrinsic human perceptual abilities, thus reinforcing the agency and value of the individual as an epistemic subject. However, in the case of citizen science, the task most often asked for in projects cannot be performed by an individual subject (c.f. Magnus, 2007: 298). Instead, the aggregated cognitive ability of the crowd is constructed as comparable with the individual cognitive ability of a scientist, which in turn promises the quality of data in observations and classifications. This construction of the citizen scientist results in a volunteer contributor without a pre-conceived perspective. In fact, researchers do not want a conceptual perspective in a mobilized citizen, because their valid input relies on following standardized protocols (or programs). The cognitive threshold for the participation of the non-scientists in scientist-initiated projects is therefore usually designed to be low, minimizing the need for instruction and learning. However, projects may also rely on existing domain expertise in the mobilized citizens. This way, the scientist and the citizen are understood to be on par with each other as epistemic subjects, securing the validity of the ‘method’ and of the data created by citizens.

The qualities of citizen scientists must be standardized and distributed to mobilize them *en masse*. Tasks need to be simple, yet relevant and valid, as Riesch and Potter (2014) note in outlining strategies for ensuring data quality in citizen science (p. 112). Simplicity makes participation ‘accessible to anyone’. Yet Danielsen et al. (2005) argue that ‘locally-based methods are generally more vulnerable than professional techniques to various sources of bias’ (p. 2524). One solution could be to provide extensive and ‘thorough training’. However, extensive training is both expensive and demands infrastructural solutions, so the answer often found in citizen science is to create stable protocols that put citizens on par with professional scientists with regard to their tasks, typically at the level of program and distributed cognition. If such stabilization cannot be attained, ‘professional scientists will remain sceptical about the results of local monitoring schemes’ (Danielsen et al., 2005: 2527). Protocols are constructed to realize citizens as the eyes of professional scientists, i.e. enabling the citizen to ‘see’ what the scientist sees, because the level of standardization ensures the validity and accuracy of the observations and classifications performed by outsiders (Cohn, 2008: 194).

Our focus is on the ways distributed and/or more individualistic epistemic abilities are configured in online citizen science through, for instance, protocols for classification or through serendipitous discovery. For the purposes of this article, we bracket issues about the cognition citizen science can be considered to occasion. However, the focus of our study on the development of antiprograms in citizen science could serve to illustrate how cognition can develop in unforeseen ways.

Protocols, programs and anti-programs

A protocol is not just a means of producing scientifically valid data, but is also a specification of the imagined epistemic subject that will perform it. In citizen science, this subject is often conceived of in terms of perception, though a significant part of the work is to reduce the scope of subjects' perception through a program or protocol that renders them able to see on the behalf of professional scientists. This strategy relies on the power to ensure alignment among observations and classifications. There are interesting questions about where and how perception can be related to this power, where the power lies and how it is extended or compromised over time. We propose that power in citizen science projects is largely understood as residing within the protocol that constructs volunteer contributors as epistemic subjects and renders their involvement in the observation and classification of data possible. Following our argument for seeing power in protocols, we empirically explore the paradox of power introduced by Latour (1992), when he argues that

> When you simply have power – in potentia – nothing happens and you are powerless; when you exert power – in actu – others are performing the action and not you. (p. 265)

At the heart of many citizen science-projects is the need to persuade people other than professional scientists to perform observations and classifications. But how are volunteer contributors constructed as epistemic subjects through the protocols of a citizen science project to perform this? To distribute and scale up the possibility of classifying objects involves a risk, since, according to Callon and Latour (1981) power is basically equal for all actors, scientists and volunteers, and

> the spread in time and space of anything – claims, orders, artifacts, goods – is in the hands of people; each of these people may act in many different ways, letting the token drop, or modifying it, or deflecting it, or betraying it, or adding to it, or appropriating it. (Latour, 1986: 267)

An inherent possibility of a protocol, then, is that epistemic subjects have the power to contest it or deviate from it and create an alternative program – or anti-program (Akrich and Latour, 1992; Latour, 1986). The practices that can be described as anti- or alternative to the program do not imply a rejection of the program itself (Akrich and Latour, 1992). However, the relationship between the citizen scientist as instantiated epistemic subject in the protocol and their development as perceptual actor implies some degree of contestation or tension. It is on this tension in particular that we focus here. We do not focus on it as an essential characteristic of the relationship between project members in citizen science projects, but instead we use the distinction to orient our analysis of the performance of boundaries and tensions in the large corpus of interactional data that forms the empirical basis for the study.

Epistemic tension and the case of Galaxy Zoo

For this study, we will turn to empirical data from one of the larger platforms for 'people powered research', Zooniverse (July 2007), and the oldest project in the Zooniverse catalogue, the Galaxy Zoo. This project was launched in 2007 to let volunteers contribute to the classification of images of galaxies, and possibly 'find' or 'witness what few have done before' (Galaxy Zoo, 2017a). That is, a feature of the epistemic culture of Galaxy Zoo is the possibility of becoming a discoverer, being recognized as the individual who was the first to see something new. In Knorr-Cetina's study of high-energy physics, the epistemic culture displayed was one of distribution and displacement, where the knowing subject was displaced by the experiment as a collective endeavour. In the epistemic culture of Galaxy Zoo, the experiment as the vehicle for epistemic progress is replaced by the protocol for classification, either in classification of galaxies as form of distributed cognition without a recognized epistemic subject, or through the anomaly that does not fit in the classificatory protocol, rendering a possible discovery by an individual epistemic subject. As we will see, discovery in a large on-line top down citizen science project can be both the result of a strong displacement of the knowing subject as well as intimately bound to it. In fact, the cultural hero of the epistemic culture of Galaxy Zoo, is an outsider discoverer.

Galaxy Zoo is a 'digital astronomical practice' (Hoeppe, 2014) in that researchers and volunteer contributors are 'removed' from the instruments that have generated the data in need of classification. In that sense, Galaxy Zoo is different from the many citizen science projects in biodiversity and conservation research. Images have always been a central part of optical astronomy, but what is different here is the remarkable number of those images, generated through innovations in digital technology. Development in electronic image making (CCD chips) means that images of natural objects in the sky are now collected in large repositories of data and astronomy has become an image-processing science (Knorr-Cetina, 1999: 28).

The limits of automation present themselves in Galaxy Zoo. The history of mobilizing citizen scientists in the digital realm often returns to the limitations of algorithmic automation. According to standard tales in citizen science, automation is limited by the hope for serendipitous discoveries, as humans have an 'eye out for the weird and the odd, even while sorting most objects into more mundane categories' (Citizen Science Alliance, 2017; see also Cooper et al., 2007). According to scientists involved in Galaxy Zoo, the project's millions of images of galaxies cannot be meaningfully processed by computers alone, but can be processed by volunteers (Lintott et al., 2008).

As images in need of classification grow in number, classifiers become a scarce resource and citizen scientists become important assets. However, the perceptual qualities of citizen scientists must be harnessed in a standardized way. To extend the chains of reference to every image, observations must be reported using standardized protocols (Latour, 1999). The tasks need to be simple, yet relevant, valid and 'accessible to anyone'. The deployment of methodologies and 'forms' must be simple enough to 'sustain' the participation of volunteers: 'as simple and locally appropriate as possible' (Danielsen et al., 2005: 2516, 2522). Yet protocols must be immutable to the degree that they adhere to scientific standards. The standardized protocol is taken to structure and stabilize the perceptual qualities of the citizen scientist, successfully negotiating the long-standing issues of data quality in volunteer observations.

Structured and stabilized protocols are expected to assure that citizen scientists can become an epistemic subject for *all scientific fields* in need of classification and observation of data. In 'citizen humanities', large repositories of text are being digitized and transcribed by volunteer contributors. Currently, the Zooniverse platform offers more than 100 projects turning to volunteer contributors for help, ranging from the humanities and social sciences to physics and astronomy (Zooniverse, 2017). Volunteers are invited to accelerate the research process by processing large amounts of data and making 'real discoveries together' with professional researchers. As they become manifest in large citizen science platforms like Zooniverse, new digital technologies are reconfiguring epistemic relationships between scientists and outsiders .

Methods

Trace ethnography

As a general approach to examining tensions that arise between the mobilizing values and epistemic cultures enacted in a citizen science project, we drew on trace ethnography (Geiger and Ribes, 2011; Hassman et al., 2013; O'Keeffe, 2016). This method is an extension of the established practices of documentary ethnography that account for and take advantage of the particular conditions of digital and distributed practices. It seeks to exploit the richness of traces left as people interact with and through digital technologies that can be 'assembled into rich narratives of interaction, allowing researchers to carefully follow coordination practices, information flows, situated routines, and other social and organizational phenomena across a variety of scales' (Geiger and Ribes, 2011: 1). While these traces or documents can often be relatively thin, '[b]y knowing the specificities of the sociotechnical landscape in which these documents are produced, a skilled observer can examine them to quickly trace the history of the document' (Geiger and Ribes, 2011: 5).

Identifying mobilizing values

To identify the values through which volunteers are mobilized by a citizen science project, we analysed material available on the Zooniverse and Galaxy Zoo websites, conducted interviews and engaged in participant observation at the project meetings of Zooniverse team members. The Zooniverse website serves as a portal for all the available citizen science projects on the platform, while the Galaxy Zoo site presents information specific to that project. Together, these sites act as the primary points of contact through which volunteers enter the project. For our analysis, we focused on the presentation of information about the character, aims and expectations of Zooniverse and Galaxy Zoo. This information could be found on 'about us' and 'frequently asked questions' pages on the Zooniverse site, and the 'story so far' page on the Galaxy Zoo site. These pages were collected and their content thematically analysed (Braun and Clarke, 2006) for patterns of description of how volunteers are mobilized. We then examined our earlier interviews with members of the scientific and development teams for accounts that provide richer understanding of the themes identified from the material presented on the Zooniverse websites.

Identifying tensions in Galaxy Zoo

To identify tensions in volunteer engagement in anti-programs and project staff responses, we collected and analysed the very large public discussion forum that is associated with Galaxy Zoo. We focused our attention on the discussion forum and not on the classification task that is at the core of the Galaxy Zoo project because, as mentioned earlier, this core task is deliberately limited in scope and volunteers leave relatively limited traces of their activity.

The Galaxy Zoo discussion forum has had two major technical generations over the lifespan of the project. It began as a separate website from the classification activity site and was later moved to become an integrated feature of the main project site. When we collected the corpus of threaded posts examined in this study, the project was in its eighth year and nearly 675,000 posts had been contributed to the two generations of the forum. Since the technologies of the two generations are different and the posts are organized in different ways, several tools were used to collect the data, including the WebScraper system and the Scrapy Python library. This process resulted in a corpus where the content of forum posts including attached images is displayed with metadata such as thread, poster and date. The data in the corpus was then processed so that the organization of the data was uniform for both generations of the forum and a continuous timeline of activity could be established.

Given the very large size of the Galaxy Zoo forum, it became important to find ways to both summarize broad trends and discern particularly relevant areas of the corpus for analysis. We worked with an iterative combination of sustained ethnographic engagement (Hine, 2000) and exploratory data analysis (Morgenthaler, 2009; Tukey, 1977) to identify recurring patterns and significant features of relevance for our aim of unpacking aspects of the tensions that arise between the mobilizing values of a project and the epistemic cultures that are enacted by volunteers. This led to the identification of a set of practices related to an anti-program involving the examination and discussion of imaging artefacts in galaxy images by volunteers. This is a practice notably similar to what has been observed in studies of the digitization of astronomy as searching astronomical images with unexplained artefacts for traces of the scientific instruments that produced them (Hoeppe, 2014; Mulkay and Edge, 1976; c.f. Knorr-Cetina, 1999). This choice was informed by sustained ethnographic engagement with forum content that indicated that the topic of artefacts and the instruments that produce them emerged relatively often, and that discussions of the topic had the interesting feature that they often began with misconceptions on the part of volunteers. Rather than discovering a new astronomical phenomenon, volunteers were often met with the news from moderators, project staff or sometimes other volunteers that they had discovered something extra-astronomical, outside the bounds of the empirical data the project had asked them to classify. Our observations indicated that sometimes these revelations led to detailed discussions of the properties of the scientific instruments and methods that had led to an artefact rather than the astronomical content of an image itself.

With a chosen focus on discussion of imaging artefacts, we decided to further constrain the scope of inquiry to the science forums. These forums are dedicated to discussion of scientific topics raised by citizen-scientists in relation to their participation in Galaxy Zoo. Ethnographic engagement indicated that threads on the science forums tended to consist of more interaction among volunteers and scientific staff than those on other forums. This observation was confirmed by examination of the corpus data that shows the level of scientific staff posts in the science forums at eight percent to be notably higher than the overall proportion of posts by scientific staff of just over two percent across the entire corpus.

Our observations also pointed toward posts on imaging artefacts, that sometimes developed into extensive and detailed discussions of the science and technology behind those artefacts. However, despite the observation that artefacts are a significant topic of discussion on the forums, corpus data shows that they are only relatively infrequently topicalized through the use of the term ‘artefact’ or ‘artifact’, with only five percent of posts containing the term. The mean length of threads in which the term artefact is used on the science forums, at just under 14 posts, is notably longer than the forum-wide mean of six posts, suggesting that artefacts are discussed they receive significantly more attention on the forum than do other topics. Combined with the relatively high number of posts made by moderators and scientific staff members on these forums, the length of threads including the term ‘artefact’ confirmed our observation that the science forums are a particularly relevant part of the corpus to focus on for this study.

While the terms ‘artefact’ and ‘artifact’ are relatively uncommon in the corpus (see Figure 1), our ethnographic engagement indicated that these few instances sometimes marked larger discussions where only a few posts included the term, but artefacts were a sustained topic of interest.

From Figure 1, it is possible to discern time periods in the life of the science forums where artefacts are topics. Immediately obvious is the transition from the first to the second technical generation of the forum in 2011 and 2012. In the second generation, categorizing galaxy images by adding hashtags was made available and we see a large increase in the incidence of the term. It is also possible to identify those periods where project staff and moderators became involved in the discussions. We used this temporal data to identify critical instances in the forums where artefacts were a sustained topic of discussion. For each date for which ‘artefact’ was mentioned in a post, we examined the length of time to the next mention. To identify periods in which the term was used repeatedly within short timeframes, we selected those periods where the number of days between mentions was less than the forum mean of 4.16 days. Given our interest in the way that projects respond to volunteer production of unexpected knowledge, we then added the criteria that at least one post containing ‘artefact’ during a selected period was

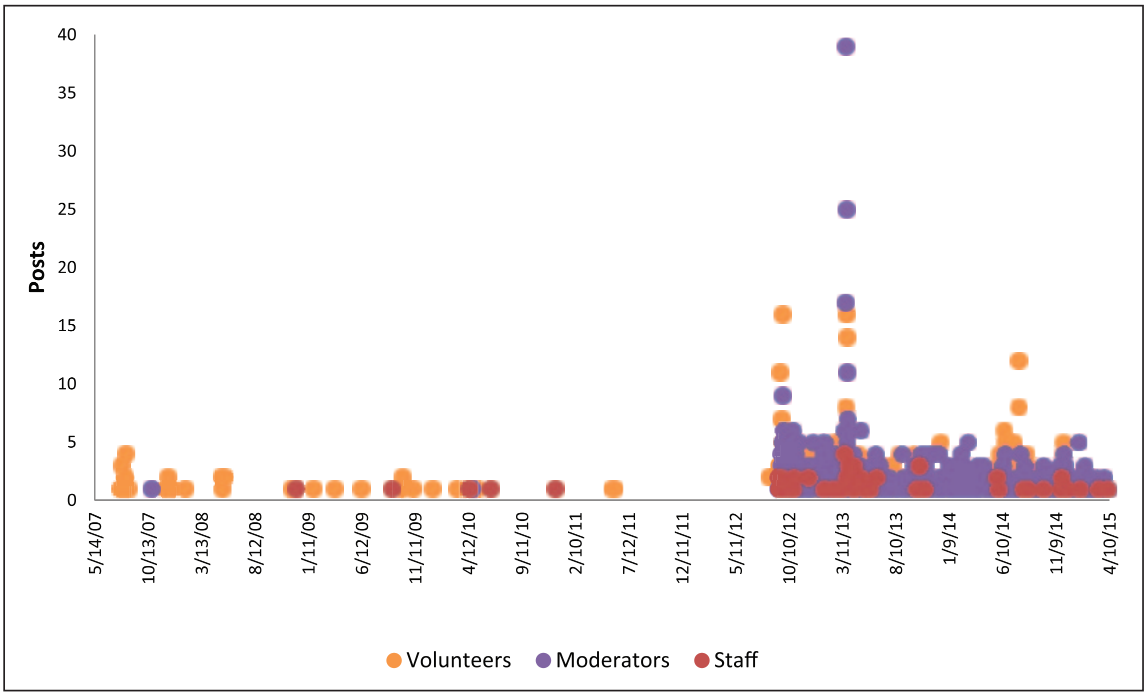

Figure 1. Use of term 'artefact' on the Galaxy Zoo science forums.

made by a moderator, a scientific staff member and a volunteer. This identified 17 periods that were then grouped to produce 15 timeframes during which all threads where the term artefact was included in one or more posts were selected for further examination. This yielded a collection of 64 threads for manual content analysis. The content of these threads was then manually coded through a process of thematic coding (Braun and Clarke, 2006), in relation to the different groups and for the presence of engagement with or resistance to the anti-program.

Findings: Mobilizing values

Based on our analysis of the Zooniverse.org and Galaxyzoo.org websites, we identified four key recurring mobilizing themes in the statements made. These themes identified on these webpages consist of *individual contributor, individual discoverer, collective contributor*, and *collective discoverer*. Although there are relatively few mobilizing value statements on the websites, they have prominent positions. Taken together, the four mobilizing values identified from these statements indicate possible tension between individual and collective and between the positioning of volunteers as contributors and discovers. In the following, these positions are examined as tensions between individual epistemic subjects and distributed collectives, both as discoverers and contributors.

The contributor as individual and collective

When volunteers are positioned as contributors to science in value statements, those statements speak slightly more often to the volunteer as an individual epistemic subject capable of scientific discovery. The emphasis of the statements is on the incremental contributions to science that volunteers can make no matter their skill level or background:

> At the Zooniverse, anyone can be a researcher. You don't need any specialised background, training, or expertise to participate in any Zooniverse projects. We make it easy for anyone to contribute to real academic research, on their own computer, at their own convenience. (Zooniverse, 2017)

At the same time, in projects in citizen science the research process is designed so that outsiders can be considered reliable, not compromising data and results. This reliability is most often not constructed based on the characteristics of volunteers as individual contributors, but through their mass aggregation as collectives of distributed cognition. While somewhat contrary to the positioning of volunteers as individual contributors and recognized epistemic subjects, the contribution volunteers make as collectives, relying on a displacement of the individual for algorithmic discovery, is recurrently stated on the Zooniverse and Galaxy Zoo websites:

> Our projects combine contributions from many individual volunteers, relying on a version of the 'wisdom of crowds' to produce reliable and accurate data. (Zooniverse, 2017)

The statements suggest an algorithmic epistemic equality established between the collective of volunteers and scientists. As one member of the Zooniverse development team described in an interview,

> What we normally do is we set up a beta project and give that to a select group of users that would go through and use the interface and provide answers and then we would aggregate over those answers and look at where those thresholds lie. And normally we try and over-sample in this so we don't come under the threshold and we'd be like, 'Oh, we only had five people look at this, we need twenty.' So we normally try and get twenty, twenty five, thirty to look at this thing and then we say, 'OK, we only need seventeen to do this one with that much confidence.' So normally we just iterate over the project design, over the interface and then over the aggregation as well during our beta runs which gives us an idea about how well it's going to work in a production system … we used to do a majority coding algorithms, now it's mostly just taking like proportional representation of them. If 97% of people say it's that and we hand that back to the scientists and we say 97%, an error bar, and say, 'Off you go,' and it's up to them to interpret that how they want to. Because they'll all have different science cases which will have different thresholds at where they can use that data. (Interview, Zooniverse developer, March 15, 2015)

An object for classification, for instance a galaxy, is taken out of 'circulation' once volunteers have classified it according to the set threshold value of an algorithm. This varies between different projects on the Zooniverse platform but all projects have in common an algorithmic representation of the volunteer contribution to the scientific task of classification. In interviews and during participant observations, the default retirement limit for Galaxy Zoo was described as 40 volunteer classifications, but this could be overridden for individual data sets (Interview, Zooniverse developer, June 19, 2017). While this positioning of volunteers as both epistemic subjects and as distributed and displaced in the program algorithm may seem contradictory, earlier STS studies of wildlife monitoring citizen science projects show that volunteers primarily regard themselves as data collectors or classifiers, disassociating themselves from analytical tasks, thus not claiming epistemic equality with scientists (Cornwell and Campbell, 2011: 107 115). In these cases, there may be little tension between individual and collective contribution mobilizing values, but this is not the case for Galaxy Zoo.

Discovery as individual and collective

The simultaneous positioning of volunteers as both individual and collective epistemic subjects in mobilizing value statements is also associated with the value of discovery in the epistemic culture of Galaxy Zoo. Statements positioning the volunteer as an individual discoverer build on the value of the cultural hero and the assertion that the project grants access to data in a way that means anyone can find something new:

> In some cases, Zooniverse volunteers have even made completely unexpected and scientifically significant discoveries. (Zooniverse, 2017)

For Galaxy Zoo, the sheer volumes of astronomical imagery collected by large scale sky surveys and then shown to volunteers means that it is possible to claim that a volunteer may be the first person to ever see a particular image or even galaxy:

> To understand how galaxies formed we need your help to classify them according to their shapes. If you're quick, you may even be the first person to see the galaxies you're asked to classify. (Galaxyzoo.org)

This kind of rhetoric that frames participation in Galaxy Zoo as the possibility for volunteers to discover something new and scientifically relevant is significantly strengthened by an early and well publicized example of individual discovery in the project, Hanny's Voorwerp (Sartori et al., 2016). In this archetypal example of a volunteer becoming the discoverer of something significant, a Dutch school teacher, Hanny van Arkel, found an unusual green cloud-like object in an image during the first few months of the project in 2007. This identification led to the tentative classification and discussion of an entirely new astronomical phenomenon and to a number of academic publications, including some with van Arkel as co-author (e.g. Józsa et al., 2009). Van Arkel is thus recognized as an epistemic subject and raised to the status of cultural hero. The story is engaging and gives credence to the notion in Galaxy Zoo that any volunteer could make an individual discovery. However, while positing volunteers as potential individual discoverers makes for a potent mobilizing value, statements are also made on the websites about the value of volunteers as collective discoverers:

> Volunteers and professionals make real discoveries together. (Zooniverse, 2017)

This collective and distributed form of discovery is perhaps a more plausible experience for the vast majority of volunteers offering a kind of downgraded middle ground in the epistemic culture. It sits between being the next one in a million-individual discoverer and being one of a million collective contributors.

Programs and antiprograms

Topicalizations of artefact

The thematic analysis of threads in which artefacts are discussed resulted in some more common patterns of interaction among volunteers, moderators and scientists as epistemic subjects. The coding resulted in the identification of three ways in which artefacts are topicalized by participants: through a *question, statement of fact* or *explanation*.

Perhaps unsurprisingly, 21 of 23 instances where artefacts are topicalized through *questions* are performed by volunteers. Some of these topicalizations are short formulations, such as '#artefact?', that make use of the hashtag functionality of the Talk platform. More often, however, the formulations used are longer and call more explicitly for explanations of the phenomenon leading to the artefact, rather than just for confirmation of the presence of the artefact itself. These more elaborated questions by volunteers tend to take the form of asking, for example, 'what is going on in this image?'.

In response to the question of whether or not there is an artefact in an image, volunteers and moderators most often provide *statements* of fact such as simply stating 'it's an artefact'. These 59 statements generally offer short assessments of the presence of an artefact, but volunteers often hedge their statements with formulations such as 'I think it is an artefact' or 'it looks like an artefact to me'. By contrast, the majority of statement of fact responses from moderators are formulated in an authoritative tone without hedges, such as:

> Moderator: Hi [volunteer] and welcome to the Zoo
> The dark spots are optical artifacts. Sorry !
> Happy hunting!

As this example illustrates, while assessing the presence of an artefact in authoritative terms, moderators often also orientate to the potentially sensitive nature of telling a volunteer that something unusual or interesting that they have discovered is not an astronomical phenomenon. This kind of remediation work where moderators express compassion for the volunteer who may have believed that they had found something special, while encouraging further attempts at discovery, is rarely performed by volunteers themselves. We can see it as illustrating how moderators take responsibility for balancing the mobilizing values of the project.

Beyond assessing the presence of artefacts through statement of fact, in 28 instances volunteers, moderators and scientific staff choose to go further and to offer explanations of the artefacts themselves and how they had been produced. In the following example a volunteer posts an image with a bright object in the centre with colourful spikes emanating from it. The volunteer asks whether or not the object is a star, to which a moderator responds that it is a star, but that many of the colours in the image are artefacts. Soon after, another volunteer joins the thread and challenges the moderator's assessment. In the post, the second volunteer signals some knowledge of the ways that imaging artefacts are produced in telescopes and offers an alternative assessment that the image does not contain artefacts but is another type of astronomical phenomenon. This post calls for a more detailed assessment of the phenomenon in the image and a scientific staff member contributes with an assessment that supports the original moderator response:

> Scientist: [Moderator] is right – they are artifacts. You can verify it by going into SkyServer and zooming out: everything nearby looks like this. Not every star shows up with diffraction spikes. In fact, most don't: you can only see them (in both SDSS and HST) for the bright stars.

The scientific staff member orients to both the original volunteer who started the thread by simply confirming that the moderator is right, and to the second volunteer who challenged the moderator's assessment. The mentions of the Sloan Digital Sky Survey (SDSS), Hubble Space Telescope (HST) and to the SkyServer website are all references that require experience with images from Galaxy Zoo or elsewhere to understand. These mentions respond to the knowledge that the second volunteer provides by offering an explanation of why the phenomenon is an artefact, why it is possible to mistake it for something else, and how the assessment can be verified. This type of formulation is typical for the few times that scientific staff members choose to respond to particularly elaborate discussion of artefacts.

Engagement and resistance to antiprograms in artefact analysis

We coded the threads for the presence of engagement with or resistance to an artefact analysis antiprogram. Across volunteers, moderators and scientific staff more posts were made indicating *engagement* with an artefact analysis antiprogram than resistance to one with 32 of the 43 identified instances indicating engagement. Often, posts with indications of engagement from moderators and scientific staff members consist of encouraging statements and offers of help, such as providing a description or link to a resource that explains the phenomenon or acknowledging how neat or beautiful an image is. As volunteers discuss artefacts amongst themselves, their conversations often draw out detailed knowledge about artefact production and about the analysis of astronomical phenomenon.

In the following example, a volunteer has started a thread in which they are collecting images that they think might include an 'inner ring', an astronomical phenomenon where a bright ring appears towards the centre of a galaxy. Having posted several possible examples, the volunteer questions whether some of the bright rings they have identified are inner rings or imaging artefacts. Rather than dismissing the call to explain which of the phenomena are artefacts with simple assessments, several other volunteers begin to offer advice on techniques for investigating the images, introducing the possibility of becoming an epistemic subject in realizing the antiprogram. First, a volunteer suggests using the 'examine' feature in Galaxy Zoo to look at images of different light wavelengths, noting that it is common to find differences between the three wavelengths available. This suggestion speaks to a feature of Galaxy Zoo, that the images presented for classification by volunteers are composites of several images taken at different light wavelengths. The second volunteer suggests that a more conclusive analysis of whether or not an artefact is present can be done by examining each of the available wavelengths. The first volunteer responds to this post thanking the second volunteer for their advice and noting that they think they have found a case where an inner ring appears to be visible in the composite image but is missing from the red wavelengths version. This indicates a continued interest in determining whether or not a ring is an imaging artefact; a third volunteer responds by suggesting that there are tools available outside Galaxy Zoo that would make such a determination possible. Starting with a question about the resources of the first volunteer is aware, this suggestion offers both scientific resources external to Galaxy Zoo that could be used for further examination of the galaxies and models an even more detailed analysis comparing two of the galaxies posted by the first volunteer. The analysis demonstrates that it is possible to identify the New General Catalogue (NGC) number for an object depicted in Galaxy Zoo, find its Atlas of Peculiar Galaxies (Arp) number, and then use those numbers to find data about the object from the United Kingdom Infrared Telescope Infrared Deep Sky Survey (UKIDSS) and the Spitzer Space Telescope. It also provides an example of one ring that appears to be an imaging artefact and another that appears to be an astronomical phenomenon, along with the wavelength (8μ) at which that assessment could be made. In response to this elaborate suggestion, the first volunteer answers the initial question about the resources of which they are aware and thanks the other volunteers for their help noting that they are ‘just a regular joe’ and indicating that they are not familiar with resources outside of Galaxy Zoo, but that they are interested to gain access. The thread illustrates a relatively common occurrence in which examination and discussion of imaging artefacts leads to detailed analyses of astronomical phenomena and to the sharing of knowledge and resources amongst volunteers. However, not all attempts to engage in examination and discussion of artefacts is met with encouragement.

Roughly a quarter of the incidences of engagement or resistance to artefact analysis antiprograms were coded as *resistance*. Notably, nine of the eleven instances were found in posts made by moderators, with volunteers and scientific staff members only contributing one such post each. In the case of the staff member post, the thread begins with a volunteer who posts a picture with what they describe as a ‘glitch’ in it. They ask what the artefact is and how it is produced. The first to respond is a moderator who notes that the artefact seen is produced by a foreground star in the image that has caused a series of spikes and a blue hue. Shortly afterwards, the scientific staff member responds:

> Scientist: Yep! The main goal is to classify the galaxy at the center of the image (here and in all the others), and mostly to ignore the optical artifacts or foreground stars in your classifications.

The response clearly states that the goal of Galaxy Zoo, the program, is to perform classifications of galaxies, and other aspects of images such as optical artefacts and foreground stars should be ignored. While this post is short, it motivates the suggestion to ignore artefacts, as not part of the main goal of Galaxy Zoo.

By contrast, the posts indicating resistance made by moderators do not motivate the suggestion to focus on the elements of images required for Galaxy Zoo classification. In one example, a volunteer posts an image with a bright object at the centre, asking what aspects of it can be classified. The volunteer calls for a description of how images are selected for classification in Galaxy Zoo and indicates that they have found many similar images. A moderator replies, noting that the bright object is a star and that software involved in the Galaxy Zoo analysis process has made a mistake:

> Moderator: There is a star in the center of the image and the software has mistakenly labelled it as a galaxy. Just click star/artifact.

The moderator acknowledges the issue raised by the volunteer but does not reply to the call for information about how the Galaxy Zoo analysis process works instead instructing them to simply classify the image as 'star/artefact' and move on. As a response, the volunteer again calls for more discussion of the artefact production process and provides a display of their own knowledge by referring to Skyserver, a resource outside Galaxy Zoo that includes a database of the images and other data available from the Sloan Digital Sky Survey. In the response, the volunteer indicates that they know that they should classify the image as 'star/artifact', but that they have an analytic interest that goes beyond what is required to perform the Galaxy Zoo classification task. They indicate that they have pursued their analysis further with a resource outside Galaxy Zoo and have found more unexplained data. Rather than orient to the antiprogram with which the volunteer is engaged, the moderator continues to frame their response within the confines of the Galaxy Zoo classification task:

> Moderator: Then the software really made a mistake!

This short exchange illustrates the different orientations that volunteers and moderators can indicate in their posts. Whilst many volunteers orient to an interest in discovery and the possibility of becoming an epistemic subject by analysing and understanding what they see in the images they are presented with on Galaxy Zoo, moderators often orient to the value of contribution in the sense of distributed cognition to the scientific project encouraging volunteers to focus on the task of classification.

Distributed and individual epistemic subjects in Galaxy Zoo

The task of classifying galaxies as a perceptual citizen scientist relies on following strict protocols or programs, and is an example of the distribution of the epistemic subject in a scientific project. From its inception, leaders of the project have acknowledged the distributed and collective effort of contributors in Galaxy Zoo:

> This achievement was made possible by inviting the general public to visually inspect and classify these galaxies via the Internet. The project has obtained more than 4 × 107 individual classifications made by 105 participants. (Lintott et al., 2008: 1)

> The data in this paper are the result of the efforts of the Galaxy Zoo 2 volunteers, without whom none of this work would be possible. Their efforts are individually acknowledged at http://authors.galaxyzoo.org. (Hart et al., 2016: 3681)

The making algorithmic of the classification task by which volunteers are mobilized is reiterated in scientific publications from the project with formulations emphasizing the collective and distributed epistemic subject embodied by the volunteer contributor:

> Images from the main AEGIS, GEMS, and GOODS data sets had a median of 122 independent classi-fications per image'. (Willett et al., 2013: 10)

> This publication has been made possible by the participation of more than 95,000 volunteers in the Galaxy Zoo project. The contributions of the more than 40,000 of those who registered a username with Galaxy Zoo are individually acknowledged at http://authors.galaxyzoo.org/'. … We combine, on average, 43 independent classifications of each galaxy to produce detailed, quantitative morphological descriptions of these distant galaxies along many physical axes of interest'. (Simmons et al., 2016: 2–3)

The rewards for scientific results and discoveries are shared, both in collective but named authorship for the relevant professional researchers and in collective acknowledgment of volunteer contributors in all publications. The individual contributor is minimized as an epistemic subject for scientific publications, but some are individually acknowledged by name on the Galaxy Zoo website that at the time of access (29 May, 2017) included over 180,000 names (Galaxy Zoo, 2017b). The epistemic culture of Galaxy Zoo thus seems to exhibit several of the features that have come to be associated with the distributed character of scientific research, where the individual person is minimized as an epistemic subject and rewards for scientific discovery are shared. Thus, the epistemic culture displays a strong distribution and displacement of the knowing subject in favour of classification according to protocols that enable citizens to act as the eyes of professional scientists. This displacement works because the level of standardization and algorithmization ensures the validity and accuracy of observations and classifications performed. Central values in the resulting epistemic culture are not associated with the individual and no single person can be identified as producing the knowledge. For Knorr-Cetina (1999), such traits of a distributed epistemic subject are associated with a distinct epistemic culture. However, we suggest that an online citizen science epistemic culture might contain elements of tension in which distributed subjects are realized through the practice of classification, and yet the individual epistemic subject, the cultural hero personified by Henny van Arkel, is a highly valued and essential part of the epistemic culture.

Discovery in an online citizen science project

Discovery can be attained as a collective distributed endeavour though classification, identifying new features on the aggregated level. However, the epistemic culture of Galaxy Zoo is also highly individualized. As volunteers classify galaxies, they may come across phenomena and anomalies that can't be accommodated by the protocol at hand. Sometimes this leads to extended discussions, as we have shown in our data. The question for the classifier in these cases is whether or not a new phenomenon has been discovered or if it is just an imaging artefact.

In 2009, a paper was published in the *Monthly Notices of the Royal Astronomical Society* with the title 'Galaxy Zoo: 'Hanny's Voorwerp', a quasar light echo?' (Lintott et al., 2009). In the paper, an unusual stellar structure, 'Hanny's Voorwerp', was discussed. The structure was named after its discoverer, the Dutch schoolteacher Hanny van Arkel, who, together with nineteen professional scientists, is also an author of the paper. This instance of discovery raises several issues concerning the epistemic culture of Galaxy Zoo in relation to epistemic subjects and distribution. First, an individual epistemic subject is recognized as an individual amongst the authors – indeed the phenomenon is named after the discoverer. Galaxy Zoo thus includes values connected to both distribution and displacement of the knowing subject as well as recognizing it. Discoveries associated with classification are distributed, while the discoveries of 'anomalies' (not artefacts) are attributed to an individual epistemic subject, as are also the discovery of artefacts mistaken for phenomena. In the former case, however, the epistemic subject is associated with cultural value as it is still understood to be within the program.

In a sense, when it concerns the professional scientists in the project, discoveries are always distributed, however in the case of Hanny's Vorweerp, the volunteer contributor (Hanny van Arkel) was recognized as an epistemic subject, while at the same time also being part of the distributed epistemic culture. Volunteers are mobilized as a distributed collective of cognition with the promise of a possibility for individuality as epistemic subjects. However, as our data show, the practices of individualized epistemic subjects acting as discoverers within the distributed collective may turn into anti-programs, as volunteers start to search and discuss the instruments producing galaxy images. This produces a tension in the epistemic culture of Galaxy Zoo: Volunteers find features in the images that do not comply with the protocol, providing them a fleeting chance of acquiring the status of discoverer and cultural hero, but more often identifying them as someone in danger of developing an antiprogram and devoting time and effort to their own centres of calculation.

This central tension, or paradox, in the epistemic culture of Galaxy Zoo creates value and promise for volunteer contributors, as evident in the central message meeting a volunteer about to start classifying galaxies: 'Few have witnessed what you're about to see' (Galaxy Zoo, 2017a). However, this also prompts volunteers to investigate and ask questions about images on the discussion forums or in emails directly to scientists:

> Sometimes. Yes, sometimes. There are people who are just clearly really motivated by the idea of being the famous person. We just try to … not encourage it. We're not trying to discourage it, we're just trying to say everybody is making a contribution, even the people who don't get famous. (Interview, Zooniverse scientist, March 15, 2015)

This potential tension between distributed and individualized epistemic subjects in Galaxy Zoo is intimately intertwined with the programs and anti-programs enacted. The possibility of becoming a recognized epistemic subject can result in anti-programs and volunteers developing their own centres of calculation where they generate advanced knowledge of imaging artefacts in a search for the instruments that have produced the images.

The perceptual qualities of the volunteer contributor rely on strict protocol following to accomplish large numbers of classification tasks. However, the epistemic culture of mobilizing the masses into programs also relies on the highly valued possibility of serendipitous discovery and the recognition as individual epistemic subject. The chance to see something never seen before motivates participation and human calculation. The call for serendipity thus invites the possibility of anti-programs when volunteers investigate whether or not artefacts are new stellar phenomena, resulting in the development of elaborate knowledge of how artefacts are produced by telescope technology. Over time, new centres of calculation can form within these anti-programs, with volunteer contributors spending time discussing the origin of images containing artefacts relying on resources outside the project. However, researchers and moderators may try to manage the central tension in the epistemic culture, trying to re-mobilize amateurs into classification programs by performing boundary work.

The main goal of many online citizen science projects is not to contribute to citizens' education. In fact, online citizen science projects seldom include or offer extensive tutorials or instructional materials for volunteer contributors, with notable exceptions such as Foldit (2018). Projects tend to rely on protocols for observation and classification that enable mobilization of volunteers at large scales. In the highly distributed algorithmic epistemic participatory culture of Galaxy Zoo this approach puts volunteers on par with professional scientists in certain limited stages of scientific work, creating an epistemic equality of intrinsic perceptual qualities that is taken to transcend or precede learning and expertise. Simultaneously, there are cognitive inequalities between professional scientists and volunteer contributors. Research design, theoretical development and hypothesis formulation are all stages of the scientific process where volunteers may be present as epistemic subjects, but where professional scientists have decision power, as it is generally their interpretation and analysis of data that leads to publishable results.

In our study of Galaxy Zoo, the path to learning to identify and understand artefacts in the imaging process generally begins with a moment of discovering something unique. This moment is configured through the way the project is described to potential contributors – as an opportunity to discover something no one has seen before. Volunteers often go to the discussion forum to check if they have discovered something new and are generally met with the response that they have in fact discovered an imaging artefact. This presents an opportunity for an epistemic journey of unpacking different imaging artefacts, sometimes searching for the instruments that produced them and creating a classification anti-program.

An earlier STS study of citizen science in this journal came to the conclusion that volunteers do not directly contribute to 'institutionalized' knowledge, but may influence the practices of science in the field (Cornwell and Campbell, 2011). However, this finding may be biased on the fact that similar to many citizen science projects, the specific project in focus concerned ecological field-work. Galaxy Zoo, by contrast, is an example where volunteers contribute to the 'institutional' aspects of science, if these aspects can be understood in terms of scientific publications. Similarly, questions of learning on the part of volunteers in Cornwell and Campbell's study are also affected by the work in ecology and conservation. Volunteers were found to learn and appreciate the role of science in influencing policy (Cornwell and Campbell, 2011: 115). These findings stand somewhat in contrast to our case study, where some volunteers displayed sophisticated knowledge about the production of artefacts in scientific imaging. However, as volunteers are involved in scientific practice there is always the possibility for epistemic subjects to contest that practice or deviate from it and create an anti-program. For example, in their study of citizen-based sea turtle monitoring, Cornwell and Campbell found that while volunteers appreciated the

> role of science in conservation, this does not translate into an unconditional acceptance. They used their 'improved understanding' of science to challenge it. Specifically, they exploited scientific uncertainty and the competing theories on the cost and benefits of nest relocation. (Cornwell and Campbell, 2011: 115)

Whether understanding artefacts in astronomical imaging processes or moving turtle nests in a conservation project are expressions of the possible creation of anti-programs is an empirical question that could shed some light on the features of the epistemic cultures of citizen science.

Conclusion

The purpose of this study has been to investigate the tensions in the epistemic culture of a large online citizen science project. We explored the values with which volunteers were mobilized into the epistemic culture of an online citizen science project. We identified the strong value of inclusion, where 'anyone' is eligible for the particular epistemic culture, and also the central cultural importance of discovery and discoverer. These cultural values were also associated with tensions as epistemic subjects enact them. We could observe that the participatory epistemic culture of Galaxy Zoo encompasses both distributed as well as more individual aspects of the epistemic subject. Volunteer contributors are valued and mobilized as a collective of classifiers who each have the possibility of individual discovery. An essential feature of this mobilization takes place as participants are configured as an algorithmic collective to perform classifications of galaxies, but with the possibility of individuality through the discovery of hitherto unknown stellar phenomena.

Thus, the epistemic culture of Galaxy Zoo values at least two forms of scientific discovery. First, discovery can be a distributed collective endeavour, where the individual person is minimized as an epistemic subject and rewards for scientific discovery are shared, as in collective and anonymous authorship on behalf of the volunteer contributors. Here, the epistemic culture displays a strong distribution and displacement of the outsider as a knowing subject in favour of the collective. Second, discovery can be attributed to an individual having witnessed something no one else has. When volunteers classify an image, there is always the possibility for aspects that cannot be fit into the protocol. Such instances spur questions taken to the forum for evaluation and discussion, thus enacting the mobilization value of possible discovery. This is most vividly exemplified by Hanny van Arkels discovery of a new stellar phenomenon and also by her named authorship on the resulting publications. Thus, the outsider can be on par with the scientist both in the form of a collective and as an individual depending on the nature of discovery.

The epistemic subjects present in an online citizen science project thus encompass several kinds. Volunteers perform the possibility of discovery often seeking individual recognition as epistemic subjects. Both moderators and scientists perform boundary work to maintain focus, offering explanations of artefact production that often occasion volunteers to pursue anti-programs. Anti-programs deviate from the central classification program or protocol of the project to develop expertise in the technical infrastructure. However, moderators take on responsibility for guiding volunteers back to the program of distributed collective endeavour of the project.

Online citizen science projects are often designed to minimize learning as a necessity for contribution. However, learning might be said to be unavoidable and outside the control of project creators. If projects are aiming for scientific output (peer-reviewed publications) there must be instances in the scientific process were citizens are constructed as on par with scientists. Over the history of the development of top-down citizen science, this relation has most often been associated with the quality of data. These instances are usually situated in participatory protocols harnessing some kind of volunteer ability, in this study perception as distributed cognition, that makes their participation and contributions valid for scientific work. At the same time, most citizen science projects must also clearly uphold a boundary between citizens and scientists where they are not on par with each other. Intuitively, this might not seem necessary, as scientists by their professional training have abilities beyond what is generally possible for volunteer contributors. In practice, however, boundaries between volunteers and scientists and between programs and antiprograms are not often clear. As shown in this study, epistemic cultures can form where outsiders can be seen to be on par with professional scientists, both as part of an algorithmic distributed collective, but also as an individual epistemic subject.

Acknowledgements

We would like to thank the members of the research project Taking science to the crowd: Researchers, programmers and volunteer contributors transforming science online at the University of Gothenburg, professor Chris Lintott, University of Oxford, and project members at Zooniverse.org for generously offering their time and knowledge. We are grateful for the comments of two anonymous reviewers and the help given during the editorial process by Sergio Sismondo whose insights significantly improved this paper. In conducting this study, the ethical guidelines developed by the ethical committee of the Association of Internet Researchers (AoIR) (2012) along with those of the Ethical Guidelines for Internet Research of the Norwegian National Committee for Research Ethics in the Social Sciences and the Humanities (NESH) (2014) were applied. In addition, all direct quotes from participants used in this paper have been done so with informed consent.

Funding

This research was supported by the Marianne and Marcus Wallenberg Foundation (MMW 2013.0020)

ORCID iDs

Dick Kasperowski https://orcid.org/0000-0003-4539-2055
Thomas Hillman https://orcid.org/0000-0003-3325-1204

References

Akrich M and Latour B (1992) A summary of a convenient vocabulary for the semiotics of human and nonhuman assemblies. In: Bijker W and Law J (eds) *Shaping Technology/Building Society: Studies in Sociotechnical Change*. Cambridge, MA: MIT Press, 259–264.

Association of Internet Researchers (AoIR) (2012) Ethical decision-making and Internet research: Recommendations from the AoIR ethics working committee (Version 2.0). Available at: http://aoir.org/reports/ethics2.pdf (accessed 23 April 2018).

Braun V and Clarke V (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology* 3(2): 77–101.

Brown P (1992) Popular epidemiology and toxic waste contamination: Lay and professional ways of knowing. *Journal of Health and Social Behavior* 33(3): 267–281.

Callon M and Latour B (1981) Unscrewing the big Leviathan: How actors macro-structure reality and how sociologists help them to do so. In: Knorr-Cetina K and Cicourel AV (eds) *Advances in Social Theory and Methodology: Towards an Integration of Micro- and Macro-sociologies*. Boston: Routledge & Kegan Paul, 277–303.

Citizen Science Alliance (2017) Why citizen science? Available at: https://www.citizensciencealliance.org/philosophy.html (accessed 18 September 2017).

Cohn JP (2008) Citizen science: Can volunteers do real research? *BioScience* 58(3): 192–197.

Cooper CB, Dickinson J, Phillips T, et al. (2007) Citizen science as a tool for conservation in residential ecosystems. *Ecology and Society* 12(2): 11–21.

Cooper CB, Shirk J and Zuckerberg B (2014) The invisible prevalence of citizen science in global research: Migratory birds and climate change (ed R Guralnick). *PLoS ONE* 9(9): e106508.

Corburn J (2005) *Street Science – Community Knowledge and Environmental Health Justice*. Cambridge, MA: MIT Press.

Cornwell ML and Campbell LM (2011) Co-producing conservation and knowledge: Citizen-based sea turtle monitoring in North Carolina, USA. *Social Studies of Science* 44(1): 101–120.

Danielsen F, Burgess ND and Balmford A (2005) Monitoring matters: Examining the potential of locally-based approaches. *Biodiversity and Conservation* 14(11): 2507–2542.

Epstein S (1996) *Impure Science. AIDS, Activism, and the Politics of Knowledge*. Berkeley, CA: University of California Press.

Foldit (2018) Foldit: Solve puzzles for science. Available at: https://fold.it/portal/ (accessed 2 May 2018).

Fors H (2015) *The Limits of Matter: Chemistry, Mining and Enlightenment*. Chicago, IL: University of Chicago Press.

Franzoni C and Sauermann H (2014) Crowd science: The organization of scientific research in open collaborative projects. *Research Policy* 43(1): 1–20.

Galaxy Zoo (2017a) Few have witnessed what you're about to see. Available at: https://www.galaxyzoo.org/?_ga=2.155407803.1651389702.1495806158-639543924.1495803420 (accessed 26 May 2017).

Galaxy Zoo (2017b) Project volunteers. Available at: https://authors.galaxyzoo.org./authors.html#galaxyzoo (accessed 29 May 2017).

Geiger RS and Ribes D (2011) Trace ethnography: Following coordination through documentary practices. In: *Proceedings of the 44th Annual Hawaii International Conference on Systems Sciences.* Available at: http://www.stuartgeiger.com/trace-ethnography-hicss-geiger-ribes.pdf (accessed 14 May 2018).

Giere RN (2002) Discussion note: Distributed cognition in epistemic cultures. *Philosophy of Science* 69(4): 637–644.

Giere RN and Moffat B (2003) Distributed cognition: Where the cognitive and the social merge. *Social Studies of Science* 33(2): 1–10.

Goodchild MF (2007) Citizens as sensors: The world of volunteered geography. *GeoJournal* 69(4): 211–221.

Haklay M (2013) Citizen science and volunteered geographic information: Overview and typology of participation. In: Sui D, Elwood S and Goodchild M (eds) *Crowdsourcing Geographic*

Knowledge. Dordrecht: Springer, 105–122. Available at: http://www.springerlink.com/index/10.1007/978-94-007-4587-2_7 (accessed 4 May 2015).
Hart RE, Bamford SP, Willett KW, et al. (2016) Galaxy Zoo: Comparing the demographics of spiral arm number and a new method for correcting redshift bias. *Monthly Notices of the Royal Astronomical Society* 461: 3663–3682.
Hine C (2000) *Virtual Ethnography*. London: SAGE.
Hoeppe G (2014) Working data together: The accountability and reflexivity of digital astroniomical practice. *Social Studies of Science* 44(2): 243–270.
Irwin A (1995) *Citizen Science: A Study of People, Expertise and Sustainable Development*. London: Routledge.
Irwin A (2001) Constructing the scientific citizen: Science and democracy in the biosciences. *Public Understanding of Science* 10(1): 1–18.
Józsa GIG, Garrett MA, Oosterloo TA, et al. (2009) Revealing Hanny's Voorwerp: Radio observations of IC 2497. *Astronomy & Astrophysics* 500(2): L33–L36.
Kärnfelt J (2015) Follow the information: Comets, communicative practices and Swedish amateur astronomers in the Twentieth Century. *Journal of Astronomical History and Heritage* 18(2): 161–176.
Kimura AH (2016) *Radiation Brain Moms and Citizen Scientists: The Gender Politics of Food Contamination after Fukushima*. Durham, NC: Duke University Press.
Knorr-Cetina K (1999) *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, MA: Harvard University Press.
Knorr-Cetina K (2007) Culture in global knowledge societies: Knowledge cultures and epistemic cultures. *Interdisciplinary Science Reviews* 32(4): 361–375.
Kuhn TS (1977) *The Essential Tension*. Chicago: University of Chicago Press.
Kullenberg C (2015) Citizen science as resistance: Crossing the boundary between reference and representation. *Journal of Resistance Studies* 1(1): 50–76.
Kullenberg C and Kasperowski D (2016) What is citizen science? A scientometric meta-analysis. *PLoS ONE* 11(1): e0147152.
Latour B (1986) The power of associations. In: Law J (ed.) *Power, Action, and Belief: A New Sociology of Knowledge?* Keele: Routledge & Kegan Paul, 261–277.
Latour B (1992) Where are the missing masses? The sociology of a few mundane artifacts. In: Bijker W and Law J (eds) *Shaping Technology/Building Society: Studies in Sociotechnical Change*. Cambridge, MA: MIT Press, 225–259.
Latour B (1999) *Pandora's Hope: Essays on the Reality of Science Studies*. Cambridge, MA: Harvard University Press.
Lintott C, Schawinski K, Keel W, et al. (2009) Galaxy Zoo: 'Hanny's Voorwerp', a quasar light echo? *Monthly Notices of the Royal Astronomical Society* 399: 129–140.
Lintott C, Schawinski K, Slosar A, et al. (2008) Galaxy Zoo: Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society* 389(3): 1179–1189.
Magnus PD (2007) Distributed cognition and task of science. *Social Studies of Science* 37(3): 297–310.
Morgenthaler S (2009) Exploratory data analysis. *Wiley Interdisciplinary Reviews: Computational Statistics* 1: 33–44.
Mulkay M and Edge DO (1976) *Astronomy Transformed: The Emergence of Radio Astronomy in Britain*. New York: Wiley.
Norwegian National Committee for Research Ethics in the Social Sciences and the Humanities (NESH) (2014) *Ethical Guidelines for Internet Research*. Available at: https://www.etikkom.no/en/ethical-guidelines-for-research/ethical-guidelines-for-internet-research/ (accessed 23 April 2018).

O'Keeffe C (2016) Producing data through e-assessment: A trace ethnographic investigation into e-assessment events. *European Educational Research Journal* 15(1): 99–116.

Orta-Martínez M and Finer M (2010) Oil frontiers and indigenous resistance in Peruvian Amazon. *Ecological Economics* 70(2): 207–218.

Ottinger G (2010) Buckets of resistance: Standards and the effectiveness of citizen science. *Science, Technology, & Human Values* 35(2): 244–270.

Riesch H and Potter C (2014) Citizen science as seen by scientists: Methodological, epistemological and ethical dimensions. *Public Understanding of Science* 23(1): 107–120.

Sartori LF, Schawinski K, Koss M, et al. (2016) Extended X-ray emission in the IC 2497 – Hanny's Voorwerp system: Energy injection in the gas around a fading AGN. *Monthly Notices of the Royal Astronomical Society* 457(4): 3629–3636.

Shapin S and Schaffer S (1985) *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life*. Princeton, NJ: Princeton University Press.

Simmons BD, Lintott C, Willett KW, et al. (2016) Galaxy Zoo: Quantitative visual morphological classifications for 48,000 galaxies from CANDELS. *Monthly Notices of the Royal Astronomical Society* 464(4): 4420–4447.

Star SL and Griesemer JR (1989) Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in Berkeley's museum of vertebrate zoology, 1907–39. *Social Studies of Science* 19(3): 387–420.

Tukey JW (1977) *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.

Willett KW, Lintott CJ, Bamford SP, et al. (2013) Galaxy Zoo 2: Detailed morphological classifications for 304,122 galaxies from the Sloan Digital Sky Survey. *arXiv*. DOI: 10.1093/mnras/stt1458

Wylie S, Wilder E, Lourdes V, et al. (2017) Materializing exposure: Developing an indexical method to visualize health hazards related to fossil fuel extraction. *Engaging Science, Technology, and Society* 3: 426–463.

Wynne B (1992) Public understanding of science research: New horizons or hall of mirrors? *Public Understanding of Science* 1(1): 37–43.

Zooniverse (2017) About Zooniverse. Available at: https://www.zooniverse.org/about (accessed 18 September 2017).

Author biographies

Dick Kasperowski is Associate Professor of Theory of Science at the University of Gothenburg. His interests include citizen science, governance of science, participatory practices in science and the humanities and open collaborative projects in scientific work. The analytical focus of his research concerns how new technologies configure relations and the development of epistemic cultures between actors claiming different experiences and knowledge; how political and scientific representations are related to each other and how (scientific) citizenship is connected to research policy and scientific practices. Lately this interest has translated into studies of the role of large scale digital platforms in citizen science.

Thomas Hillman is Associate Professor of Information Technology and Learning at the University of Gothenburg. With a background in the design of products and environments for learning, his research investigates the ongoing reconfiguration of technology for learning in both formal and informal settings with a focus on the mutually constitutive relationship between the development of technologies and the transformation of epistemic practices. In recent years, his work has focused on the blurring boundaries between online and offline activities in many aspects of contemporary life including citizen science.