Citizen science frontiers: Efficiency, engagement, and serendipitous discovery with human–machine systems

Laura Trouille[a,b,1], Chris J. Lintott[c], and Lucy F. Fortson[d]

[a]Department of Citizen Science, The Adler Planetarium, Chicago, IL 60605; [b]Center for Interdisciplinary Exploration and Research in Astrophysics, Northwestern University, Evanston, IL 60208; [c]Department of Physics, The University of Oxford, Oxford, OX1 3RH, United Kingdom; and [d]Department of Physics and Astronomy, The University of Minnesota-Twin Cities, Minneapolis, MN 55455

Citizen science has proved to be a unique and effective tool in helping science and society cope with the ever-growing data rates and volumes that characterize the modern research landscape. It also serves a critical role in engaging the public with research in a direct, authentic fashion and by doing so promotes a better understanding of the processes of science. To take full advantage of the onslaught of data being experienced across the disciplines, it is essential that citizen science platforms leverage the complementary strengths of humans and machines. This *Perspectives* piece explores the issues encountered in designing human–machine systems optimized for both efficiency and volunteer engagement, while striving to safeguard and encourage opportunities for serendipitous discovery. We discuss case studies from Zooniverse, a large online citizen science platform, and show that combining human and machine classifications can efficiently produce results superior to those of either one alone and how smart task allocation can lead to further efficiencies in the system. While these examples make clear the promise of human–machine integration within an online citizen science system, we then explore in detail how system design choices can inadvertently lower volunteer engagement, create exclusionary practices, and reduce opportunity for serendipitous discovery. Throughout we investigate the tensions that arise when designing a human–machine system serving the dual goals of carrying out research in the most efficient manner possible while empowering a broad community to authentically engage in this research.

citizen science | machine learning | human computing interaction | physical sciences | biological sciences

The 1968 Cybernetic Serendipity exhibition (www.studiointernational.com/index.php/cybernetic-serendipity-50th-anniversary) was an early imagining and exploration of computer-aided creative activity, play, and interplay. The exhibit, curated by Jasia Reichardt at the Institute of Contemporary Arts in London, examined the role of cybernetics in contemporary art and included robots; algorithmically generated movies, poetry, and music; painting machines; and kinetic interactives. Several of the works featured chance as an important ingredient in the creative process, reflected in the priority given in the exhibition's emphasis on machine-enabled serendipity. It is useful to reflect, 50 y later, whether machines have indeed enabled serendipitous discovery, albeit in the realm of science rather than the arts.

We consider this concept in the context of online citizen science projects. These projects, which massively share the task of data analysis among a crowd of volunteers, in many ways exemplify the promise of those early ideas, providing a compelling modern example of the transformative power of the integration of human and machine effort. Online citizen science not only is a powerful tool for efficiently processing our growing data rates and volumes (the "known knowns"), but also can function as a means of enabling serendipitous discovery of the "known unknowns" and the "unknown unknowns"

in large, diverse datasets. In this *Perspectives* piece, we describe our recent efforts and future considerations for designing a human–machine system optimized for "happy chance discovery" (a definition of serendipity that provided guidance for the Cybernetic Serendipity exhibit) that best takes advantage of the efficiencies of the machine while acknowledging the complexity of human motivation and engagement.

What is now called citizen science—the involvement of the general public in research—has a long history. An early example is Edmund Halley's study of timings during the 1715 total solar eclipse, which included observations from a distributed, self-organized group of observers (1). Works by refs. 2 and 3, among others, have linked modern-day efforts to their 19th century antecedents, for example, highlighting the role played by amateur networks of meteorological observers in establishing that field of study (i.e., by 1900, more than 3,400 observers were contributing data to a network organized by George Symons, producing data on a scale that could not be matched by the professional efforts of the time). In recent decades, citizen science has gained renewed prominence, boosted in part by technological advances and digital tools like mobile applications, cloud computing, and wireless and sensor technology which have enabled new modes of public engagement in research (4) and facilitated research projects that investigate questions from data at scales beyond the professional research community's resource capacity (5). Professional citizen science organizations have been created in Europe, Australia, and the United States. In the United States, the Crowdsourcing and Citizen Science Act of 2015 was introduced to encourage the use of citizen science within the federal government and, that same year, the first Citizen Science Association (CSA) conference was held (although some consider the 2012 European Space Agency side event on citizen science the first CSA gathering). CitizenScience.gov (https://www.CitizenScience.gov) currently lists over 400 active citizen science projects. Participation in citizen science today ranges from hands-on data collection, tagging, analysis, and research projects [e.g., iNaturalist.org (https://www.iNaturalist.org) (research grade observations: https://www.gbif.org/dataset/50c9509d-22c7-4a22-a47d-8c48425ef4a7), eBird.org (https://www.eBird.org) (6), and CitSci.org (https://www.CitSci.org) (7)]

to contributing in-person data and participating in hands-on data analysis [e.g., the Denver Museum of Science Genetics of Taste Laboratory (8)] to a growing number of cocreated environmental monitoring projects using low-cost sensors with community members working in collaboration with researchers [e.g., the LA Watershed Project (https://www.epa.gov/urbanwaterspartners/diverse-partners-brownfields-healthfields-la-watershed)] to online data-processing efforts, described in more detail below. There has also been an explosion of citizen science efforts carried out in classroom settings; for example, Sea-Phages (9), Small World Initiative (10), and the Genomic Education Partnership (11) provide standardized curricula for undergraduate students to collect soil and other samples from their local environments, isolate the bacteria in them, annotate the genomes, characterize them, and upload their results into national databases. Over 300 universities participate annually in these recently launched efforts, with dozens of peer-reviewed articles to date and a major impact on these fields of study (e.g., ref. 12).

Online citizen science, which has become a proven method of distributed data analysis, enables research teams from diverse domains to solve problems involving large quantities of data, taking advantage of the inherently human talent for pattern recognition and anomaly detection. For example, the Eterna (13) online gaming environment [the next generation of the FoldIT platform (14)] challenges players to design new ways to fold RNA molecules to find solutions for diseases like tuberculosis. These new molecular structures are then synthesized and tested in Stanford's medical laboratories. Other examples of online citizen science include Eyewire (15) and Cosmoquest (https://cosmoquest.org), among others. There are also a growing number of online crowdsourced transcription efforts in the humanities, including From the Page (https://fromthepage.com), Veridian (16), Smithsonian Transcript Center (https://siarchives.si.edu/collections/siris_sic_14645), and Transcribe Bentham (17). See CitizenScience.gov and SciStarter.org (https://SciStarter.org) for comprehensive listings of citizen science projects and platforms. Technological advances have also enabled a parallel track of "volunteer/distributed computing" efforts, like SETI@Home (https://setiathome.berkeley.edu), which harness computing resources for distributed computing and/or storage.

Numerous studies have outlined the positive impacts of public participation in scientific research, including increases in long-term environmental, civic, and research interests (e.g., ref. 18 and references therein); the empowerment of communities to influence local environmental decision making (18, 19); the increased representation of women and minorities in the scientific process (20); and increases in confidence (21–24), scientific literacy (25–28), domain knowledge (23, 29), and understanding that scientific progress is a collective process (30).

Zooniverse (https://www.zooniverse.org), which we focus on in the remainder of this paper, is the largest platform for online citizen science, host to over 120 projects with 1.7 million registered participants around the world. It is unique among online citizen science platforms as a result of its (*i*) shared open-source software, experience, expertise, and input from users across the disciplines; (*ii*) reliable, flexible, and scalable application programing interface (API), which can be used for a variety of development tasks; (*iii*) free, do-it-yourself (DIY) "Project Builder" (also known as the Project Builder Platform) capabilities as described below; and (*iv*) the scale of its existing audience. Zooniverse partners with hundreds of researchers across many disciplines, from astronomy to zoology, cancer research to climate science, and history to the arts. At a time when citizen science is gaining prominence across the world, Zooniverse has become a core part of the research infrastructure landscape. Since the launch in 2007 of the Galaxy

Zoo project (31, 32), Zooniverse projects have led to over 150 peer-reviewed publications, enabling significant contributions across many disciplines (see zooniverse.org/publications for the full list), e.g., in ecology (33–36), humanities (37, 38), biomedicine (39), physics (40, 41), climate science (42, 43), and astronomy (31, 32, 44–46). These projects have established a track record of online citizen science producing quality data for use by the wider scientific community. This paper provides a compilation of lessons learned and questions raised around the integration of machine learning into online citizen science based on the experiences from myriad projects on the Zooniverse platform. With each Zooniverse project highlighted, we reference the specific project URL and the relevant citations.

The number of projects supported by Zooniverse has recently experienced rapid growth, an acceleration which is a result of the launch in July 2015 of the free Project Builder Platform (https://www.zooniverse.org/lab) which enables anyone to build and deploy an online citizen science project at no cost, within hours, using a web browser-based toolkit. The Project Builder supports the most common types of interaction including classification, multiple-choice questions, comparison tasks, and marking and drawing tools. The Project Builder front end is a series of forms and text boxes a researcher fills out to create the project's classification interface and website. All Project Builder projects come with a landing page; a classification interface; a discussion forum; and "About" pages for content about the research, the research team, and results from the project. Fig. 1 displays screen shots from the Project Builder front end. The Project Builder is transformative; before its development a typical online citizen science project required months to years of professional web development time. Zooniverse went from launching 3–5 projects per year to launching 26 in 2016, 44 in 2017, and over 50 in 2018.

Early Lessons in Design for Cybernetic Serendipity—Finding the Unexpected

From its first project Galaxy Zoo (https://GalaxyZoo.org) (31, 32), the Zooniverse platform has encouraged and enabled serendipitous discovery, taking advantage of having so many eyes on the data and our innate human ability to notice the unusual and unexpected. Just as amateur scientists in the 19th century defined new lines of scientific inquiry based on their observations, volunteer participants on modern citizen science platforms have more often than not adventured far beyond the original framework of a given project to flag an unusual object of interest and/or work on research questions they came up with themselves, typically based on something they noticed while performing the prescribed main classification task (47). The Zooniverse "Talk" discussion forums have been central to these citizen-led investigations. Talk provides a space for a wide range of information sharing and ideas exchange between volunteers and researchers, enabling serendipitous collaboration and facilitating social community building. The interactions in Talk range from the purely social to goal driven and collaborative, benefiting the project both directly, in their question-answering function and as an enabler of novel discoveries, and indirectly, as a means to attract interest and long-term engagement (48).

The serendipitous discoveries made by Zooniverse volunteers over the years have transformed frontier fields of research. For example, the discovery of Galaxy Zoo's "Green Peas" (49) pushed our theoretical understanding of the formation and evolution of galaxies. These compact green galaxies are low mass, low metallicity galaxies with high star formation rates. They were a surprise because despite their low metallicity (which indicates that they are young), they are not very far away. Their discovery highlights the power of Talk (50). A few volunteers
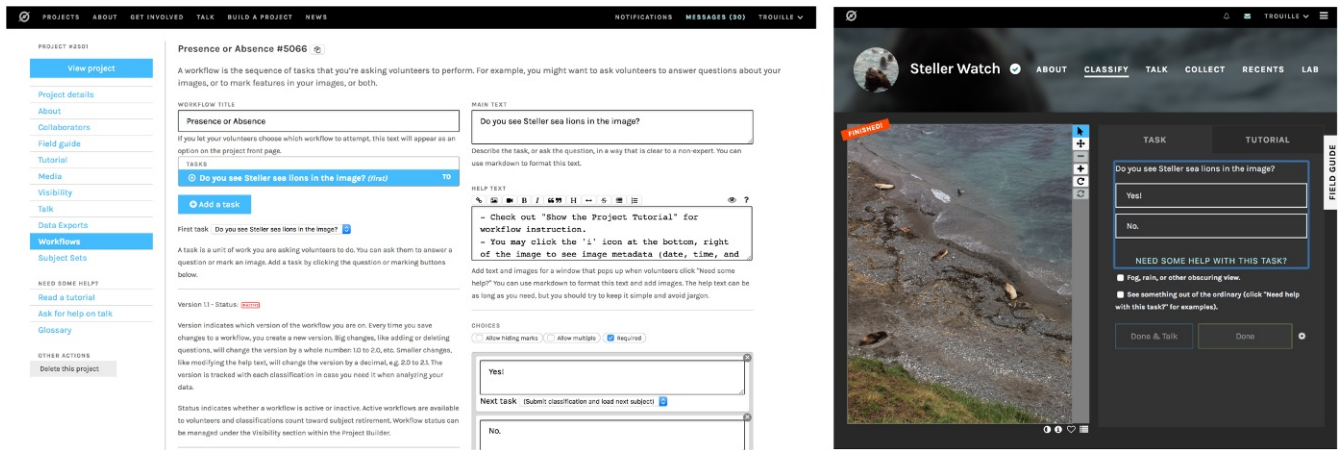
Fig. 1. Screenshots of Zooniverse's free Project Builder Platform (https://www.zooniverse.org/lab). As the user inputs content into the Project Builder interface (*Left*), he or she can immediately view changes in the associated public-facing website (*Right*). Here we have displayed the "Workflow" section of the Project Builder (in which the user sets the classification task the volunteer will carry out) and the associated "Classify" page within the project's actual website that it creates. Through the tabs located along the left-hand side of the Project Builder interface, project leads upload all of the necessary content and data for their crowdsourced research project, from inputting information about their research goals and why they need volunteers' help in the About tab to uploading their subjects that need classifications in the "Subject Sets" tab to exporting the raw classifications provided by the volunteers through the "Data Exports" tab. The Project Builder is democratizing access to online citizen science as a tool for research and has enabled the accelerated expansion of the Zooniverse. Since its launch in July 2015, the Zooniverse has gone from launching 3–5 projects each year to launching over 50 in 2018.

had noticed the strange, compact, green blobs while classifying and had started a Talk discussion forum thread humorously titled "Give peas a chance." The researchers worked alongside the volunteers (who referred to themselves as the "Peas Corps") to refine the collection of objects, ultimately identifying 250 Green Peas in the million-galaxy dataset. Even if the researchers had managed to examine 10,000 of the images, they would have only come across a few 'Green Peas' and would not have recognized them as a unique class of galaxies (49). Numerous other examples of serendipitous discovery pepper Zooniverse's history—from Hanny's Voorwerp, the ghost remnant of a supermassive black hole outflow, offset from its central Galaxy Zoo galaxy and of which only a few dozen other examples have been observed (51), to the discovery of a group of 19th century female scientific illustrators and writers [a volunteer noticed the name "Mabel Rhodes" while annotating 19th century scientific journal pages as part of the Science Gossip project (https://www.ScienceGossip.org) and spurred a cohesive search and collection effort and a new research direction for the project (https://talk.sciencegossip.org/#/boards/BSC0000004/discussions/DSC00004s8)].

Experiments in Machine Integration

As we enter an era of growing data rates and volumes [e.g., the 30 TB of data each night that will be produced by astronomy's Large Synoptic Survey Telescope (52) or the thousands of terabytes produced by ecology projects annually], Zooniverse has been moving toward a more complex system design, one that takes better advantage of the complementary strengths of humans and machines, integrating these efforts to optimize for both efficiency and volunteer engagement, while striving to safeguard and encourage opportunities for serendipitous discovery. Below we provide a few examples of early efforts to integrate humans and machines.

In its simplest form, a number of projects have used volunteer classifications to generate training sets for automated methods to efficiently classify all remaining data. For example, an early project on Zooniverse, Galaxy Zoo: Supernova (53), using data from the Palomar Transient Factory, retired from the system after the volunteer classifications provided a large enough training set for the researchers' machine-learning

algorithms to automatically process the remaining data with confidence.

Combining Human and Machine Classifiers. The current Supernova Hunters project (https://www.zooniverse.org/projects/dwright04/supernova-hunters) provides an illustrative example of the additional power that comes from combining human and machine classifications. Each week, thousands of new Pan-STARRS telescope images are flagged by machine-learning routines as containing potential supernovae candidates. Subjects which the model deems unlikely to be a supernova are automatically rejected and the remaining subjects ($\sim$5,000 each week) are uploaded for our volunteers to classify. Ref. 54 found that the human classifications and machine-learning results in Pan-STARRS were complementary; the human classifications provide a pure but incomplete sample while the machine classifications provide a complete but impure sample. By combining the aggregated volunteer classifications with the machine-learning results, they are able to create the most pure and complete sample of new supernovae candidates.

Transfer Learning, Cascade Filtering, and Early Retirement. The Camera CATalogue project (https://www.zooniverse.org/projects/panthera-research/camera-catalogue) further took advantage of the different strengths in human–machine integration to more efficiently classify new data. Through this project, the Panthera conservation organization harnesses the power of the crowd to tag different species in camera trap images. The researchers first used a transfer learning technique (55) to train a model specific to South Africa based on the much larger Snapshot Serengeti (https://www.SnapshotSerengeti.org) dataset. The images were then passed through the logic of "cascade filtering" wherein the task is broken into a sequence of simple "yes/no" questions (e.g., 'blank/not blank') which volunteers could easily do through the Zooniverse mobile app. Furthermore, instead of requiring a default of five volunteer classifiers to classify each image, the project used new Zooniverse system infrastructure to automatically retire an image if only one to two volunteers agreed with the model prediction. This combination of human–machine classifiers reduced human effort by 43% while maintaining overall accuracy and

Trouille et al.

demonstrated the utility of transfer learning for our smaller camera trap projects.

Near Real-Time Integration. Zooniverse has begun to explore the next level of sophistication in the human–machine system through near real-time integration of the two. For example, ref. 56 ran a series of experiments with the Galaxy Zoo 2 dataset of 200,000 galaxy images. Zooniverse volunteers classified on the Galaxy Zoo 2 project for over 1 y (from 2010 to 2011) to retire the full dataset. In contrast, in the ref. 56 simulations, an off-the-shelf machine-learning algorithm started training in near real time on day 8. The machine was fully trained by day 12 and retired over 70,000 images on its first application. By day 30, through online learning alongside the human classifications, the machine had retired the entire 200,000 galaxy-image dataset. Had this system been deployed on the live project, it would have produced over a factor of 8 decrease in time compared with the 400+ d it took human volunteers to complete the dataset alone.

Intelligent Task Assignment. Additional progress has stemmed from a growing body of theoretical work and practical application, often using data from Zooniverse projects, demonstrating that efficiencies exist through judicious task assignment to volunteers, greatly reducing the total number of classifications needed (e.g., refs. 57–59). Each of these studies used classification data to derive estimates of volunteer performance or ability and used this information to improve allocation. For example, task assignment studies within the Space Warps project (https://SpaceWarps.org) demonstrated that false negatives (images wrongly classified as containing no gravitational lenses) could be eliminated if at least one volunteer with high measured skill reviewed those images (45). For other examples of this type, see refs. 60 and 61.

Design Tensions

Efficiency and Engagement. The above examples outline the promise of human–machine integration within an online citizen science system. In the following, we hope to convey the tensions that exist when designing for efficiency in a human–machine system. For example, ecology camera trap projects like Snapshot Wisconsin (https://SnapshotWisconsin.org) automatically remove a certain fraction of "blank" subjects (those containing no animals) before upload into their Zooniverse project. Naively it would seem best to remove all known "blanks" to increase the efficiency of the system. However, ref. 62 found that removing all blanks decreased volunteer engagement as measured by number of classifications, a phenomenon which we suggested is related to game design and the psychology of slot machines, where effort is invested for intermittent rewards. This simple experiment highlights how optimizing for user experience, not efficiency, is critical if each volunteer's potential for contribution is to be maximized. This is just one experiment, but it is clear that volunteers' behavior in systems which involve them in scientific tasks is likely to be complex and difficult to predict. This is especially so for volunteers who may have a complex relationship with science and who may or may not view their participation as contributing to research (24, 27, 63–66).

As another example, in our discussion above of early machine-learning integration, we cited an example of a Zooniverse project (the first Supernova Hunters project) that was able to retire from the system once humans had provided a large enough training set for the machines to reliably process the remainder of the data. For the Zooniverse web development team and (most of) the volunteers, retiring a project because machines can take over is a victory, allowing us to focus human effort where it is truly needed. An important nuance worth noting, though, is that for those volunteers who become deeply engaged in and attached

to a particular project (e.g., searching for supernovae), their experience of its retirement from the system can be difficult, especially if data are still coming in but are now being processed by machines.

Efficiency and Inclusion. The current Supernova Hunters project (https://www.zooniverse.org/projects/dwright04/supernova-hunters) provides a different lens through which to consider the impact of design choices. Each week the researcher uploads the new subjects on a Tuesday morning and sends out the weekly announcement shortly after. Volunteer classification activity precedes the weekly announcement and the project averages over 20,000 classifications in the first 24 h. Within 48 h enough classifications are submitted to retire that week's data. There is a heightened sense of competition due to the legitimate possibility of being the first person to discover one of the ∼12 supernovae candidates likely present in the small amount of new data released each week, but this opportunity relies on beating other volunteers to the data. Given the extremely high level of dedicated volunteer engagement in this project, it might seem a good idea to recommend that other projects adopt a similar format of incremental data release (even to the point of creating artificial scarcity) to encourage engagement.

Even when data are plentiful, one might imagine operating in a mode where artificial scarcity is used to drive classification activity. However, ref. 67 cautions that perceived scarcity and competition have effects beyond motivating volunteers to interact with a project more frequently. For example, the Supernova Hunters volunteer community has a strong age and gender bias, being most popular among males in the oldest age group (65+). In contrast, Zooniverse projects overall are 45% female with a flat age distribution from 18-y olds to 75-y olds. In addition to the demographics bias, Supernova Hunters is currently the most unequal Zooniverse project in terms of volunteer classification contributions; the majority of the classifications are made by a small cohort of highly dedicated volunteers. We note that it is possible these trends could be alleviated simply by a more judicious timing of the data release (i.e., not in the middle of the week during work hours) when a broader distribution of the population could participate. More importantly, however, the example provides the opportunity to examine the potential for tension when serving the occasionally opposing goals of carrying out authentic research in the most efficient manner possible (scientific efficiency) and empowering a broad community to engage in valued and meaningful ways with real research (social inclusivity).

Ref. 67 explores how the relationship between these two aims, social inclusivity and scientific efficiency, is nuanced. For example, there may be instances where a project is more scientifically efficient if it is more exclusive. Although designing projects for efficiency via exclusivity is not ideal, it is not clear whether the alternative, of reducing scientific efficiency in the name of inclusivity, is preferable. For example, in the Supernova Hunters project, social inclusivity could have been enhanced through providing greater opportunity to classify by artificially increasing the number of classifications required for each image. However, this would not only represent a wasteful use of volunteer time and enthusiasm, but also undermine one of the most common motivations of volunteers—to make an authentic contribution. If volunteers want to complete only tasks that directly contribute to research, but do more work if shown images for which classifications are not needed, the contradiction is apparent.

Strategies for the Future

Machine Integration with "Leveling Up." The Gravity Spy project (https://www.GravitySpy.org) (41, 68–72) suggests one way forward in integrating humans and machines into the system and

optimizing for both efficiency and engagement. Through Gravity Spy, the public helps to categorize images of time-frequency representations of gravitational wave detector glitches from the Laser Interferometer Gravitational Wave Observatory (LIGO) into preidentified morphological classes and to discover new classes that appear as the detectors evolve. This is one of our most popular projects, with over 12,000 registered participants having provided over 3 million classifications, and has led to the accurate classification of tens of thousands of LIGO glitches and the identification of multiple prominent and previously unknown glitch classes (73, 74). This popularity has been a surprise given that the images are not particularly "pretty" and the goal is to categorize noise (not to directly discover a new gravitational wave signal). The project's success is a good reminder to not make assumptions on what will or will not prove engaging to the public.

In parallel with the human effort in Gravity Spy, a deep-learning model with convolutional neural network (CNN) layers is used to categorize images after being trained on human-classified examples of the morphological classes. The system also maintains a model of each volunteer's ability to classify glitches of each class and updates the model after each classification (i.e., increasing its estimate of the volunteer's ability when he or she agrees with an assessment and decreasing it if he or she disagrees). When the volunteer model shows that a volunteer's ability is above a certain threshold, the volunteer advances to the next workflow level, in which he or she is presented with new classes of glitches and/or glitches with lower machine-learning confidence scores. Volunteers progress through five levels within the Gravity Spy system, choosing from 22 different glitch classes in Level 5. In addition, images which neither the volunteers nor the machine confidently classify as a known glitch class are moved from one workflow level to the next through the "None of the Above" category. The volunteers, in concert with machine-learning efforts, are now working to create new glitch categories from these None of the Above images. Fig. 2 shows the Gravity Spy human–machine system architecture and the flow of images through that system.

Gravity Spy thus uses machine learning alongside a leveling-up infrastructure to guide the presentation of tasks to newcomers to help them more quickly learn how to do the image classification task while still contributing to the work of the project.

A recent study by ref. 75 finds that volunteers who experienced the machine-learning scaffolded training performed significantly better on the task (an average accuracy of 90% vs. 54%), contributed more work (an average of 228 classifications vs. 121 classifications), and were retained in the project for a longer period (an average of 2.5 sessions vs. 2 sessions). The project also exemplifies how curious citizen scientists are well situated for serendipitous discovery of unusual objects. Gravity Spy volunteers have identified a number of new glitch categories, including the discovery of the "Paired Doves" class which has proved significant in LIGO detector characterization endeavors, as this glitch resembles signals from compact binary inspirals and is therefore detrimental to the search for such astrophysical signals in LIGO data (73, 74).

Machine Integration with Intelligent Task Assignment. Another example of the potential for thoughtful integration of machines within online citizen science systems comes from the Etch-a-cell project (https://www.zooniverse.org/projects/h-spiers/etch-a-cell). Here volunteers provide detailed tracings of the boundaries of cellular components such as the nucleus, as seen in extremely high-resolution microscopy. Modern instruments can slice a cell to produce a stack of thousands of such images, and the Etch-a-cell project aims to assemble 3D representations from these segmentations. At present, this is achieved by having volunteers work through each image in the stack separately and combining the results. One can easily imagine using machine learning to determine where user intervention is really required; this could be via simple change detection or something as complex as a recurrent neural network (RNN). The net effect would be to save volunteer effort while increasing the variety of the experience. This kind of intervention, which increases both the effectiveness of volunteer contributions and the variety of the task, is, we suggest, more likely to produce successful projects.

Such strategies are likely to be easiest to find for highly ordered datasets. For others, such as those which draw their data from modern astronomical surveys, this may be more problematic. It is also clear that any attempt to direct effort in a classification task more efficiently reduces the possibilities for serendipitous discovery, a key motivation for human classification in many projects. In cases where the discovery is in the background, or otherwise tangential to the main task (as in the
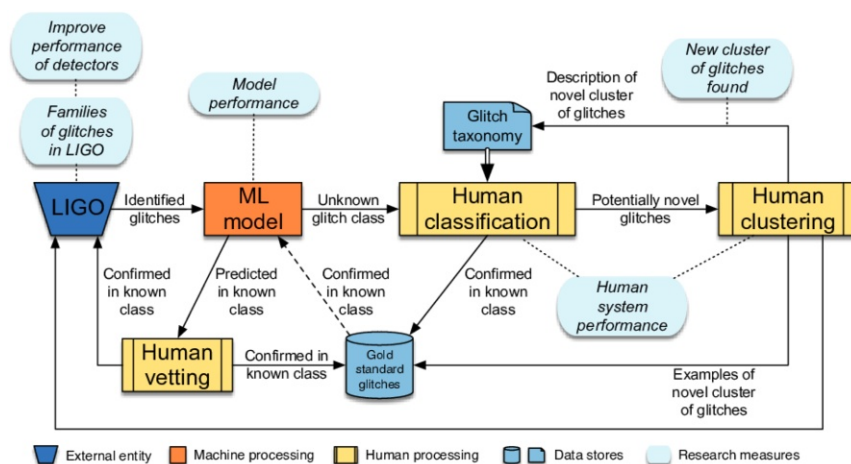

Fig. 2. Gravity Spy system architecture and data flow through the interconnected, interdisciplinary components of the project. Each day, LIGO detects ~2,000 perturbations with a signal-to-noise ratio greater than 7.5 and sends them to the Gravity Spy system. To maximize the gravitational wave detection rate, glitches must be identified and removed from this dataset. The Gravity Spy project couples human classification with machine-learning models in a symbiotic relationship: Volunteers provide large, labeled sets of known glitches to train machine-learning algorithms and identify new glitch categories, while machine-learning algorithms "learn" from the volunteer classifications, rapidly classify new glitch signals, and guide the training provided to the volunteers. In parallel, LIGO engineers work to identify and isolate the physical cause(s) of the identified glitches and, if possible, eliminate them. If they cannot be physically eliminated, the glitches are flagged and removed as part of the LIGO data processing pipeline. See ref. 41 for details.

Trouille et al.

www.manaraa.com

case of the Galaxy Zoo's Green Peas and Hanny's Voorwerp), the odds of such a discovery being made are presumably roughly proportional to the amount of time spent looking at images. There is thus a clear tension between efficiency for classification and designing for serendipity, which has not been evident in projects carried out to date.

Machine Integration with Clustering. A potential solution exists in using machine learning not for classification, but for clustering. Ref. 76 demonstrated that galaxy classification (albeit with color as well as morphology included in the classification) can be approached with a completely unsupervised clustering algorithm, sorting data from the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS) previously classified by Galaxy Zoo (77) into roughly 200 separate categories. The data in Gravity Spy have also been subject to a similar clustering analysis (Fig. 3), with tools being built to allow volunteers to explore these clusters. If clusters produced by such an analysis are uniquely separate groups, then it is easy to see that volunteer classification of just a small number of objects can be leveraged to classify the whole dataset. Even if they are strongly contaminated—for example, blurring several morphological categories in the case of Galaxy Zoo—they can still simplify the task presented to volunteers. In such a hybrid system, it might be expected that truly unusual objects will form their own categories which can then be investigated by volunteers.

Summary

The efforts described above serve as signposts for what a full human–machine system (what would once have been called cybernetic serendipity) might be like. We have shown that efficient combination of human and machine classification can produce results superior to either one on its own, and while the performance of modern deep learning might be expected to continue to improve as further training data become available, the scarcity of many objects of interest makes us believe that human–machine systems will remain competitive for many years to come. However, even the simple experiments carried out so far make clear the tension inherent in such projects, which emerges as complex interventions are considered.

First, we noted that complex task assignment might reduce the willingness of volunteers to participate in tasks rendered increasingly monotonous. Using an automated classifier to identify and remove the animal-free images in the Snapshot Serengeti project seems an obvious way to increase efficiency, yet the resulting loss of variety reduced user engagement as measured by time spent on the site. This tension is especially important when scientific aims—classifying the data—are combined with goals relating to engagement or changing attitudes and which therefore depend on extended time spent in the project.

The loss of variety seen in this intervention is not necessarily a feature of intelligent task assignment. One could artificially alter the frequency of interesting images (selecting the most beautiful or inserting boring images which do not need classification), but this has ethical implications in projects which offer "authentic" engagement with science rather than a manipulated experience. It is better to consider user experience when designing interventions.

We also explored the tensions that arise in serving the dual goals of carrying out authentic research in the most efficient manner possible (scientific efficiency) and empowering a broad community to engage in valued and meaningful ways with real research (social inclusivity). Through the example of the Supernova Hunters project, we noted that incremental (in this case,
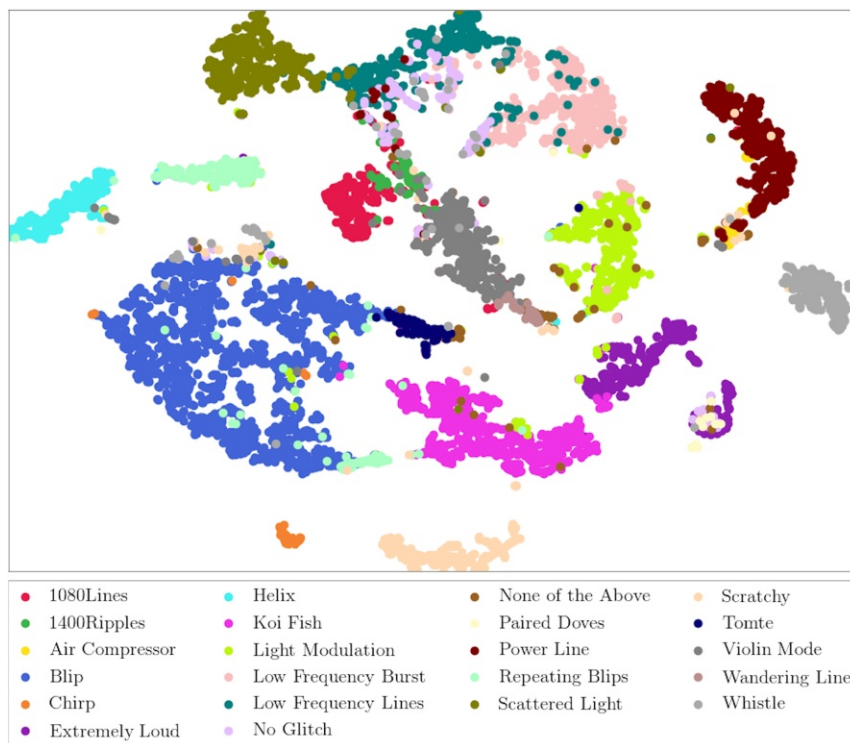

● 1080Lines	● Helix	● None of the Above	● Scratchy
● 1400Ripples	● Koi Fish	○ Paired Doves	● Tomte
● Air Compressor	● Light Modulation	● Power Line	● Violin Mode
● Blip	● Low Frequency Burst	● Repeating Blips	● Wandering Line
● Chirp	● Low Frequency Lines	● Scattered Light	● Whistle
● Extremely Loud	● No Glitch		

Fig. 3. t-stochastic neighbor embedding (t-SNE) mapping of the Gravity Spy training set in the Deep DIscRiminative Embedding for ClusTering (DIRECT) feature space. DIRECT is an unsupervised machine-learning algorithm using deep neural networks developed by ref. 71 for use with Gravity Spy data. The symbols indicate the different known glitch classes, as well as subjects that fall in the None of the Above category. This view shows the clustering along two of many possible dimensions. The Gravity Spy team is building tools that use the results of these clustering analyses to support volunteers in identifying new glitch classes. See ref. 71 for details on Gravity Spy clustering efforts to date and ref. 78 for background on the t-SNE mapping method used.

www.manaraa.com

weekly) data release with the potential for discovery encourages high levels of engagement. Yet the perceived scarcity and competition have also resulted in a skewed volunteer base (in this case, the majority of the classifications are made by a small cohort of highly dedicated volunteers, the vast majority of whom are male and 65+ y in age).

We then discussed strategies for addressing these design tensions and thoughtfully attempting to optimize for both efficiency and engagement, while leaving space for serendipitous discovery. The leveling-up model explored through the Gravity Spy project provides one such opportunity. In this project, machine learning guides the presentation of tasks to newcomers to quickly train them in the image classification task while still contributing work to the project. Volunteers are promoted from one level to the next as they pass certain thresholds of classification counts and quality. In parallel, images that do not conform to an existing class are passed through the levels and eventually retire as None of the Above. The volunteers, in concert with machine-learning efforts, then identify new classes through visual inspection, clustering analyses, and a combination of the two.

There also exist possibilities for more creative uses of machine learning, which become apparent once one's design goal switches from trying to replace humans with machines to one of building complex "social machines" that involve both sorts of worker. Examples of such opportunities detailed above include the ability to use machine effort to direct attention in datasets where classification subjects are highly linked. The example given is in layered microscopy, but this solution holds potential for a range of projects, e.g., planet hunting where the task is change detection. The most extreme machine/human hybrid classifier introduced in the previous section is the use of machine clustering to dramatically reduce the need for human classification; the success of such a scheme is likely, however, to depend on the purity of clusters that can be achieved without significant manual intervention.

Despite the increasing sophistication and complexity of many deployed citizen science systems, therefore, it is clear that there is much more to do in project design. Keeping open the possibility of volunteer serendipitous discovery with the largest of all upcoming datasets will require the development of new and flexible tools for interacting with sorting algorithms. However, the performance of experiments carried out to date gives us confidence that we can succeed, and we should expect citizen scientists to be experiencing cybernetic serendipity for many years to come.

1. Pasachoff JM (1999) Halley and his maps of the Total Eclipses of 1715 and 1724. *J Astron Hist Heritage* 2:39–54.
2. Dawson G, Lintott C, Shuttleworth S (2015) Constructing scientific communities: Citizen science in the nineteenth and twenty-first centuries. *J Vic Cult* 20:246–254.
3. Shuttleworth S (2018) Old weather: Citizen scientists in the 19th and 21st centuries. *Sci Mus Group J* 3:150304.
4. Bonney R, Phillips TB, Ballard HL, Enck JW (2015) Can citizen science enhance public understanding of science? *Publ Understanding Sci* 25:2–16.
5. Miller-Rushing A, Primack R, Bonney R (2012) The history of public participation in ecological research. *Front Ecol Environ* 10:285–290.
6. Sullivan BL, et al. (2009) eBird: A citizen-based bird observation network in the biological sciences. *Biol Conserv* 142:2282–2292.
7. Wang Y, Kaplan N, Newman G, Scarpino R (2015) CitSci.org: A new model for managing, documenting, and sharing citizen science data. *PLoS Biol* 13:e1002280.
8. Boxer EE, Garneau NL (2015) Rare haplotypes of the gene TAS2r38 confer bitter taste sensitivity in humans. *SpringerPlus* 4:505.
9. Caruso SM, Sandoz J, Kelsey J (2009) Non-STEM undergraduates become enthusiastic phage-hunters. *CBE Life Sci Educ* 8:278–282.
10. Caruso JP, Israel N, Rowland K, Lovelace MJ, Saunders MJ (2016) Citizen science: The small world initiative improved lecture grades and California critical thinking skills test scores of nonscience major students at Florida Atlantic University. *J Microbiol Biol Educ* 17:156–162.
11. Shaffer CD, et al. (2014) A course-based research experience: How benefits change with increased investment in instructional time. *CBE Life Sci Educ* 13:111–130.
12. Davis E, et al. (2017) Antibiotic discovery throughout the small world initiative: A molecular strategy for identifying biosynthetic gene clusters involved in antagonistic activity. *MicrobiologyOpen* 6:e00435.
13. Lee J, et al. (2014) RNA design rules from a massive open laboratory. *Proc Natl Acad Sci USA* 111:2122–2127.
14. Cooper S, et al. (2010) Predicting protein structures with a multiplayer online game. *Nature* 466:756–760.
15. Kim JS, et al. (2014) Space–time wiring specificity supports direction selectivity in the retina. *Nature* 509:331–336.
16. Daniels C, Holtze TL, Howard R, Kuehn R (2014) Community as resource: Crowdsourcing transcription of an historic newspaper. *J Electron Resour Librarian* 26:36–48.
17. Causer T, Grint K, Sichani A-M, Terras M (2018) 'Making such bargain': Transcribe Bentham and the quality and cost-effectiveness of crowdsourced transcription. *Digital Scholarship Humanit* 33:467–487.
18. Dicksinson J, Bonney R (2012) *Citizen Science: Public Participation in Environmental Research* (Comstock Pub Assoc, Ithaca, NY). Available at https://www.amazon.com/Citizen-Science-Participation-Environmental-Research-ebook/dp/B007NK5I16?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=B007NK5I16. Accessed July 1, 2018.
19. Newman G, et al. (2012) The future of citizen science: Emerging technologies and shifting paradigms. *Front Ecol Environ* 10:298–304.
20. Groulx M, Lemieux CJ, Lewis JL, Brown S (2016) Understanding consumer behaviour and adaptation planning responses to climate-driven environmental change in Canada's parks and protected areas: A climate futurescapes approach. *J Environ Plann Manag* 60:1016–1035.
21. Raddick MJ, et al. (2010) Galaxy Zoo: Exploring the motivations of citizen science volunteers. arXiv:0909.2925. Preprint, posted March 3, 2016.
22. Raddick MJ, et al. (2013) Galaxy Zoo: Motivations of citizen scientists. arXiv:1303.6886. Preprint, posted March 27, 2013.
23. Masters K, Cox J, Simmons B, Lintott CJ (2016) Science learning via participation in online citizen science. *J Sci Commun*. Available at https://doi.org/10.22323/2.15030207. Accessed January 1, 2017.
24. Greenhill A, et al. (2016) Playing with science. *Aslib J Inf Manage* 68:306–325.
25. Crall AW, et al. (2012) The impacts of an invasive species citizen science training program on participant attitudes, behavior, and science literacy. *Publ Understand Sci* 22:745–764.
26. Cronje R, Rohlinger S, Crall A, Newman G (2011) Does participation in citizen science improve scientific literacy? A study to compare assessment methods. *Appl Environ Educ Commun* 10:135–145.
27. Jones G, Childers G, Andre T, Corin E, Hite R (2018) Citizen scientists and non-citizen scientist hobbyists: Motivation, benefits, and influences. *Int J Sci Educ* 8:287–306.
28. Trumbull DJ, Bonney R, Bascom D, Cabral A (2000) Thinking scientifically during participation in a citizen-science project. *Sci Educ* 84:265–275.
29. Brossard D, Lewenstein B, Bonney R (2005) Scientific knowledge and attitude change: The impact of a citizen science project. *Int J Sci Educ* 27:1099–1121.
30. Ruiz-Mallén I, et al. (2016) Citizen science. *Sci Commun* 38:523–534.
31. Lintott CJ, et al. (2008) Galaxy Zoo: Morphologies derived from visual inspection of galaxies from the Sloan digital sky survey. *Mon Not R Astron Soc* 389:1179–1189.
32. Fortson L, et al. (2012) Galaxy Zoo: Morphological classification and citizen science. *Advances in Machine Learning and Data Mining for Astronomy*, eds Way MJ, Scargle JD, Ali KM, Srivastava AN (CRC Press, Taylor & Francis Group, Boca Raton, FL), pp 213–236.
33. Swanson A, Kosmala M, Lintott C, Packer C (2016) A generalized approach for producing, quantifying, and validating citizen science data from wildlife images. *Conserv Biol* 30:520–531.
34. Matsunaga A, Mast A, Fortes JAB (2016) Workforce-efficient consensus in crowdsourced transcription of biocollections information. *Future Gener Comput Syst* 56:526–536.
35. Arteta A, Lempitsky V, Zisserman A (2016) Counting in the wild. *Computer Vision–ECCV 2016*, eds Leibe B, Matas J, Sebe N, Welling M (Springer International Publishing, Amsterdam), pp 483–498.
36. Anderson TM, et al. (2016) The spatial distribution of african savannah herbivores: Species associations and habitat occupancy in a landscape context. *Philos Trans R Soc B Biol Sci* 371:20150314.
37. Williams AC, et al. (2014) A computational pipeline for crowdsourced transcriptions of ancient greek papyrus fragments. *2014 IEEE International Conference on Big Data (Big Data)*, eds Samad T, Arnold G, Goldgof D, Hossain E, Lanzerotti M (IEEE, Washington, DC).
38. Grayson R (2016) A life in the trenches? The use of operation war diary and crowdsourcing methods to provide an understanding of the british army's day-to-day life on the western front. *Br J Mil Hist*. Available at https://bjmh.org.uk/index.php/bjmh/article/view/96. Accessed February 1, 2018.

www.manaraa.com

39. Candido dos Reis FJ, et al. (2015) Crowdsourcing the general public for large scale molecular pathology studies in cancer. *EBioMedicine* 2:681–689.

40. Hass A (2017) HiggsHunters: A citizen science project for ATLAS publication. *Conference on Computing in High Energy and Nuclear Physics*, eds Mount R, Tull C (IOP Journal of Physics: Conference Series, San Francisco), Vol 898, p 1.0.

41. Zevin M, et al. (2017) Gravity Spy: integrating advanced ligo detector characterization, machine learning, and citizen science. *Classical Quan Gravity* 34:064003.

42. Hennon CC, et al. (2015) Cyclone center: Can citizen scientists improve tropical cyclone intensity records? *Bull Amer Meteorol Soc* 96:591–607.

43. Rosenthal IS, et al. (2018) Floating Forests: Quantitative validation of citizen science data generated from consensus classifications. arXiv:1801.08522. Preprint, posted January 25, 2018.

44. Clifton Johnson L (2015) PHAT stellar cluster survey. II. Andromeda project cluster catalog. *Astrophys J* 802:127.

45. Marshall PJ, et al. (2016) SPACE WARPS - I. Crowdsourcing the discovery of gravitational lenses. *MNRAS* 455:1171–1190.

46. Schwamb ME, et al. (2018) Planet Four: Terrains–Discovery of araneiforms outside of the south polar layered deposits. *Icarus* 308:148–187.

47. Luczak-Rösch M, et al. (2014) Why won't aliens talk to us? Content and community dynamics in online citizen science. *Eighth International AAAI Conference on Weblogs and Social Media*, ed Goel A (AAAI Press, Palo Alto, CA), pp 315–324.

48. Mugar G, Østerlund C, Hassman KD, Crowston K, Jackson CB (2014) Planet hunters and seafloor explorers. *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing–CSCW '14*, eds Fussell S, Lutters W (ACM Press, New York) pp 109–119.

49. Cardamone C, et al. (2009) Galaxy zoo green peas: Discovery of a class of compact extremely star-forming galaxies. *MNRAS* 399:1191–1205.

50. Straub M (2016) Giving citizen scientists a chance: A study of volunteer-led scientific discovery. *Citizen Sci Theor Pract*. Available at http://doi.org/10.5334/cstp.40. Accessed June 1, 2018.

51. Lintot CJ, et al. (2009) Galaxy zoo: 'Hanny's voorwerp', a quasar light echo? *Mon Not R Astron Soc* 399:129–140.

52. Ivezic Z, et al. (2008) *For the LSST Collaboration. LSST: From Science Drivers to Reference Design and Anticipated Data Products.* arXiv e-prints:0805.2366. Preprint, posted May 15, 2008.

53. Smith AM, et al. (2011) Galaxy zoo supernovae. *Mon Not R Astron Soc* 412:1309–1319.

54. Wright DE, et al. (2017) A transient search using combined human and machine classifications. *Mon Not R Astron Soc* 472:1315–1323.

55. Yosinski J, Clune J, Bengio Y, Lipson H (2014) How transferable are features in deep neural networks? *Proceedings of the 27th International Conference on Neural Information Processing Systems, NIPS'14*, eds Ghahramani Z, Welling M, Cortes C, Lawrence ND, Weinberger KQ (MIT Press, Cambridge, MA), Vol 2, pp 3320–3328.

56. Beck MR, et al. (2018) Integrating human and machine intelligence in galaxy morphology classification tasks. *Mon Not R Astron Soc* 476:5516–5534.

57. Simpson E, Roberts S, Psorakis I, Smith A (2013) Dynamic Bayesian combination of multiple imperfect classifiers. *Decision Making and Imperfection*, eds Guy T, Karny M, Wolpert D (Springer, Berlin), pp 1–35.

58. Waterhouse TP (2013) Pay by the bit: An information-theoretic metric for collective human judgment. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, eds Tardos E, McCormick H (ACM, New York), pp 623–638.

59. Kamar E, Horvitz E (2015) Planning for crowdsourcing hierarchical tasks. *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, eds Bordini R, Elkind E, Weiss G, Yolum P (International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC), pp 1191–1199.

60. Russakovsky O, Li L-J, Fei-Fei L (2015) Best of both worlds: Human-machine collaboration for object annotation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, ed Pecht M (IEEE, Boston), pp 2121–2131.

61. Kamar E, Hacker S, Horvitz E (2012) Combining human and machine intelligence in large-scale crowdsourcing. *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*, eds Sierra C, Luck M (International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC), Vol 1, pp 467–474.

62. Bowyer A, Maidel V, Lintott C, Swanson A, Miller G (2015) This image intentionally left blank: Mundane images increase citizen science participation. *Proceedings of the 2015 Conference on Human Computation*, eds Gerber E, Ipeirotis P (AAAI, Palo Alto, CA), pp 209–210.

63. Horn MS, Crouser RJ, Bers MU (2011) Tangible interaction and learning: The case for a hybrid approach. *Pers Ubiquitous Comput* 16:379–389.

64. Kafai Y, Fields D (2012) Connecting play: Understanding multimodal participation in virtual worlds. *Proceedings of the 14th ACM international conference on Multimodal interaction - ICMI '12*, eds Morency L, et al. (ACM Press, New York), pp 265–272.

65. Bandura A (1982) Self-efficacy mechanism in human agency. *Am Psychol* 37:122–147.

66. Williams MM, George-Jackson CE (2014) Using and doing science: Gender, self-efficacy, and science identity of undergraduate students in stem. *J Women Minorities Sci Eng* 20:99–126.

67. Spiers H, et al. (2018) Patterns of volunteer behavior across online citizen science. *WWW '18 Companion Proceedings of The Web Conference 2018* (International World Wide Web Conferences Steering Committee, Geneva), pp 93–94.

68. Crowston K; The Gravity Spy Team (2017) Gravity Spy: Humans, machines and the future of citizen science. *ACM Conference on Computer Supported Cooperative Work and Social Computing*, eds Lee C, Poltrock S (CSCW, New York), pp 163–166.

69. Crowston K, Østerlund C, Lee TK (2017) Blending machine and human learning processes. *Proceedings of the 50th Hawaii International Conference on System Sciences*, ed Bui T (Hawaii International Conference on System Sciences, Honolulu), pp 65–73.

70. Crowston K, Mitchell E, Østerlund C (2018) Coordinating advanced crowd work: Extending citizen science. *Proceedings of the 51st Hawaii International Conference on System Sciences*, ed Bui T (Hawaii International Conference on System Sciences, Honolulu), pp 1681–1690.

71. Bahaadini S, et al. (2018) DIRECT: Deep discriminative embedding for clustering of LIGO data. *IEEE International Conference on Image Processing*, ed Pecht M (IEEE, New York), pp 748–752.

72. Lee TK, Crowston K, Harandi M, Østerlund C, Miller G (2018) Appealing to different motivations in a message to recruit citizen scientists: Results of a field experiment. *J Sci Commun*. Available at https://jcom.sissa.it/archive/17/01/JCOM_1701_2018_A02. Accessed June 1, 2018.

73. Smith J (2015) Glitches from misbehaving PCAL-Y on October 9. Advanced LIGO electronic log 21463 available at https://alog.ligo-la.caltech.edu/aLOG/index.php?callRep=21463. Accessed June 15, 2018.

74. Lundgren A (2016) New glitch class: Paired doves. Advanced LIGO electronic log 27138 available at https://alog.ligo-wa.caltech.edu/aLOG/index.php?callRep=27138. Accessed June 15, 2018.

75. Jackson C, Crowston K, Østerlund C, Harandi M (2018) Folksonomies to support coordination and coordination of folksonomies. *Computer Supported Cooperative Work*, ed Schmidt K (Springer, New York), Vol 27, pp 647–678.

76. Hocking A, Geach JE, Sun Yi, Davey N (2017) An automatic taxonomy of galaxy morphology using unsupervised machine learning. *Mon Not R Astron Soc* 473:1108–1129.

77. Simmons BD (2016) Galaxy Zoo: Quantitative visual morphological classifications for 48 000 galaxies from CANDELS. *Mon Not R Astron Soc* 464:4420–4447.

78. van der Maaten LJP, Hinton GE (2008) Visualizing data using t-SNE. *J Mach Learn Res* 9:2579–2605.

COMPUTER SCIENCES

ECOLOGY

Trouille et al.