Studies in History and Philosophy of Science 88 (2021) 102–109

Contents lists available at ScienceDirect

Studies in History and Philosophy of Science

journal homepage: www.elsevier.com/locate/shpsa

Evaluating community science

Karen Kovaka

Virginia Tech, 229 Major Williams Hall, 220 Stanger St, Blacksburg, VA, 24060, United States

ARTICLE INFO

Keywords:
Community science
Citizen science
Data evaluation
Aims of science

ABSTRACT

Community science—scientific investigation conducted partly or entirely by non-professional scientists—has many advantages. For example, community science mobilizes large numbers of volunteers who can, at low cost, collect more data than traditional teams of professional scientists. Participation in research can also increase volunteers' knowledge about and appreciation of science. At the same time, there are worries about the quality of data that community science projects produce. Can the work of non-professionals really deliver trustworthy results? Attempts to answer this question generally compare data collected by volunteers to data collected by professional scientists. When volunteer data is more variable or less accurate than professionally collected data, then the community science project is judged to be inferior to traditional science. I argue that this is not the right standard to use when evaluating community science, because it relies on a false assumption about the aims of science. I show that if we adopt the view that science has diverse aims which are often in tension with one another, then we cannot justify holding community science data to an expert accuracy standard. Instead, we should evaluate the quality of community science data based on its adequacy-for-purpose.

When I try to learn something new, like sketching, kayaking, or playing the guitar, I judge my progress by comparing what I do to what someone with expertise in the activity does. Odds are, you do too. Of course, my drawings of birds will never be as good as Audobon's, and I'll never be able to paddle a kayak or play a guitar as well as my teachers can. That's fine—I don't want to be a world-class artist, athlete, or musician. What I want is to sketch the birds I see outside, navigate mellow rivers, and entertain friends around a campfire with a song or two. But I can still meaningfully assess the quality of my drawings, paddling, and music-making in comparison to experts. Even though I can meet my personal goals while falling well short of expertise, there is an important sense in which how good I am at making art or running whitewater depends on how similar what I do is to what an expert does.

This basic idea—judging how good a product is by comparing it to what an expert would produce—is an intuitive and often helpful way to assess quality. But not always. Sometimes, being high quality and being expert-like come apart. The Wright brothers' first airplane, for example, had little in common with engineering expert Samuel Langley's aerodrome. But the Wrights, who were bicycle salesmen and had no engineering credentials, launched the first successful flight, while the aerodrome sank in the Potomac River on its failed launch. In such cases, we miss or undervalue excellence if similarity to what an expert would produce is our only resource for assessing quality.

This paper is about one such case, the case of community science. Community science is scientific investigation conducted partly or entirely by non-professional scientists. It's a fast-growing and popular approach to science, but because it involves non-professionals, proponents of community science often find themselves defending the quality of its results. In response to worries about data quality in particular, it has become common to evaluate community science projects by comparing the data they produce to expert or professionally-produced data. The focus on expert accuracy springs from a desire to ensure that community science data are accurately representing the world. Expert data are seen as a useful proxy for assessing the actual accuracy of the data community scientists produce.

For example, thousands of community scientists around the world use a software program called *Nature's Notebook* to document shifts in plant phenology (the timing of life cycle events such as flowering) caused by climate change. Kerissa Fuccillo and collaborators studied the accuracy of non-professionals' phenological observations by sending 28 volunteers and one ecologist to observe and record the phenological phases of 19 different species growing along a park trail in Portland, Oregon. The researchers found that volunteer and expert observations agreed with one another between 70 and 91% of the time, and they concluded that "volunteers can provide reliable observations of plant phenology when collecting explicit, standardized protocols" (Fuccillo et al., 2015, p. 921).

Here I argue that such volunteer-expert comparisons are a bad general model for assessing the success of community science projects or the

E-mail address: kkovaka@vt.edu.

https://doi.org/10.1016/j.shpsa.2021.05.004
Received 18 July 2020; Received in revised form 20 April 2021; Accepted 17 May 2021

quality of the data they produce.[1] Community science projects often have different kinds of goals than traditional or expert-driven science. This is not to say that community science is inferior science. Its goals, even when they differ from those of traditional science, are fully scientific. Still, the difference in goals justifies a standard of evaluation other than expert-likeness. A better standard is known as *adequacy-for-purpose* (Parker, 2020). According to this standard, community science data should be evaluated in terms of whether they are adequate for meeting the goals of the investigation of which they are a part. Sometimes the goals of community science require expert quality data, but not always.

1. Community science

What I am calling community science is defined in different ways in different fields and goes by many names, including citizen science, community-based participatory research, and participatory action research. It also includes a wide range of projects, from crowd-sourced wildlife counts to public health initiatives to partnerships between professional scientists and indigenous peoples. There are many ways of taxonomizing this heterogenous category. For example, some researchers categorize community science projects in terms of their different goals, such as education, conservation, or policy change (e.g Wiggins & Crowston, 2011), while others focus on the level of participant engagement projects achieve (e.g. Shirk et al., 2012). These taxonomies also have their critics, as well (e.g. Kimura & Kinchy, 2016; Ottinger, 2017).

This diversity of labels, projects, and taxonomies reflects the fact that community science is a dynamic and fast-growing area of science. Worldwide financial investment in community science is now in the billions of dollars. There are established institutions dedicated to community science, including professional societies and academic journals. Most importantly, community science promises a range of benefits that many leaders believe make it an especially important area of focus.[2]

Prominent among these advantages is research volume. The best way to answer many urgent scientific questions is by compiling vast datasets. Take the example of climate change-induced shifts in plant phenology. Understanding how the flowering and fruiting times of plants are shifting is vital for conservation planning and adapting agricultural practices to climate change. Scientists have sophisticated tools for studying phenological shifts, including models that predict their future trajectories, but they still need rich datasets from around the world. Finding the resources to pay professionals or their students to make observations and collect this data is a significant problem (McDonough MacKenzie 2020). Community science offers an alternative way to compile these phenological datasets—train willing volunteers to do it.

Sometimes, increasing data volume also increases data quality. For projects like monitoring the invasion front of an invasive species or tracking the distribution of migratory birds at different points in time, more data points often means better science, in the sense that the scientific goals of the investigation are better served. Community science may also increase the quality of scientific research in other ways. A number of researchers have argued that high levels of non-professional engagement, such as consulting community members when developing research questions, designing a study, and interpreting or analyzing data, can reveal insights that professionally trained scientists would miss (Allen, 2017; Irwin 1995; Kimura & Kinchy, 2019; Ottinger, 2017; Whyte & Crease, 2010; Wylie, 2015).

But all of these benefits depend upon community scientists meeting some threshold of competency. For data collection in particular, increasing data volume may *decrease* data quality if community scientists are not skilled enough at identifying samples or measuring quantities. The same is true of the other ways in which community scientists participate in research. Contributions to research question selection, study design, and data analysis may mislead as well as provide insight. According to a survey by Burgess et al. (2017), many professional scientists worry that community science data is of insufficient quality for their purposes, and as a result, they prefer data from other professionals rather than data from community scientists (see also Riesch & Potter, 2013).

Given the potential and rapid growth of community science, as well as the concern that it can decrease the quality of scientific work, researchers involved with community science have worked to develop ways of evaluating the quality of community science data. One way, which I have already mentioned, is to evaluate community science data by comparing it to professionally collected data. I call this approach to evaluating community science the *expert accuracy standard.* According to the expert accuracy standard, a successful community science initiative should produce data that is always comparable to data produced by experts. Most research on community science data quality accepts the expert accuracy standard. For example, Margaret Kosmala and collaborators write, in a review of the literature on community science data quality, that "a reasonable definition of high quality data for citizen science is data of comparable accuracy and bias to that produced by professionals and their trainees" (2016, p. 552). Though some researchers have questioned the expert accuracy standard (e.g. Elliott & Rosenberg, 2019; Ottinger, 2016, 2017), it is the consensus standard in the natural scientific literature on community science.

An admirable feature of the expert accuracy standard is that it treats community science as real science. Community science projects aim to produce knowledge, influence policy, and improve the human condition, just as traditional science does. I will argue later that there are differences between the goals of community science and traditional science, but it is also important to emphasize the overlap. And it is understandable to think that achieving these shared goals requires community science to meet the epistemic standards of traditional science. By holding community and traditional science to the same standard when it comes to data quality, the expert accuracy standard acknowledges that community science has the potential to contribute to the scientific enterprise in genuine, non-trivial ways. This advantage of the expert accuracy standard comes with a corresponding disadvantage, however. The disadvantage is that it is not sensitive to important differences between community science and traditional science. As I will show, these differences mean that the expert accuracy standard is not the appropriate standard for evaluating community science data.

The question of which standard we use to evaluate data quality is an important one for community science. At stake is not only which particular projects and datasets count as high quality, but also which types of community science are deemed capable of producing high quality data. If we use the expert accuracy standard, then only projects where non-professional volunteers can achieve comparable results to experts will count as high quality. These projects will be first in line for funding, and it will be harder to implement "lower quality" projects. Where such projects are implemented, their results will be greeted with skepticism.

So, what types of community science can meet the expert accuracy standard? First, projects that involve simple data collection come closer to meeting the expert accuracy standard than other types of projects. Examples of simple data collection include counting the number of invasive plants one sees (Gallo & Waitt, 2011) or classifying a

[1] Community science also involves much more than data collection, and many scholars have argued that non-professionals can and should be involved in scientific investigation throughout the research process, not merely in the context of data collection. Though I am sympathetic to this claim, I focus on data collection in this paper because of the influence the expert accuracy standard has over the design and implementation over community science projects.

[2] Not only is community science a fast-growing area of science, but interdisciplinary scholarship that raises normative questions about community science is fast-growing as well. In this paper, I am focused on evaluating the quality of community science data, and so I engage most directly with work on data quality from philosophy of science and the natural sciences. For an overview of important and complementary work from other disciplines, see Arnstein, 1969; Greenwood & Levin, 2007; Torre et al., 2012; Kemmis et al., 2014.

photograph of a galaxy according to its shape (Willett et al., 2013). These are projects that only require participants to observe and record, rather than, say, performing an experimental intervention. It is easier to train a non-expert to make a specific kind of observation than it is to train them to set up experimental and control treatments or to work with specialized equipment. Nearly all of the highly publicized community science success stories involve simple data collection, and I suspect this is due to the influence of the expert accuracy standard. Burgess and collaborators, for example, end their paper with the claim that while not all types of biodiversity science can benefit from the participation of community scientists, "evolving technology allows members of the public to participate in monitoring and conservation-oriented data collection without leaving home" (2017, pp. 118–119).

A second and related point is that projects which engage nonprofessionals in shallow ways perform better with respect to the expert accuracy standard than projects that involve deeper engagement. Simple data collection is, to be sure, the easiest way to involve non-professionals in scientific research, but many advocates of community science envision projects where volunteer involvement goes beyond this rote and boring work. Theirs is a vision of community science in which non-professionals are involved in identifying research questions and in data analysis and interpretation (see, e.g. Kemmis et al., 2014, p. 4). More than supplying free labor, community scientists would be genuine partners in knowledge production. But when it comes to these deeper levels of engagement, meeting the expert accuracy standard is even more difficult. In fact, it often does not even apply. We can, of course, formulate versions of the standard that apply to data analysis, etc., but training a non-expert to approximate an expert when it comes to statistical analysis is far more difficult than training a non-expert to recognize the developmental stages of butterfly larvae. As a consequence, projects that strive for deeper engagement count as lower quality science, according to the expert accuracy standard, *in virtue of* their commitment to this deeper engagement.

Finally, if we accept the expert accuracy standard, we will prefer projects that are led and controlled by professional scientists over projects that are grassroots-inspired and led. The more control professional scientists have over a project, the more likely that it will meet or come close enough to meeting the expert accuracy standard. Grassroots projects are often explicitly activist in nature. In southwest Virginia, where I live, groups of citizen activists have been fighting against the construction of a natural gas pipeline through the Appalachian Mountains for years. Part of their work involves documenting harmful effects of the pipeline build, from emerging sinkholes to dried up water supplies to abnormal levels of sedimentation in local waterways.[3] Such projects look unscientific by some lights because they have no veneer of impartiality, whereas an expert-led project would not take an explicit stance about the harmfulness of a pipeline build in advance of data collection. Further, these projects often do not try to meet the expert accuracy standard for data quality, because their purpose is to mobilize community members or get the attention of a governmental agency or non-profit, not to publish results in a scientific journal.

Now, it could turn out to be true that expert-led surveying and monitoring projects with shallow levels of volunteer engagement produce better results than other kinds of community science projects. So far, the expert accuracy standard has run afoul of certain visions of what community science can be, but this doesn't mean it's a bad standard. All I've claimed here is that it matters whether we accept the expert accuracy standard. This standard comes with a particular ideal of what community science ought to be like, one that sidelines some ways of imagining expert and non-expert partnerships.

There is, particularly in the social scientific literature, an alternative approach to evaluating community science and the data it produces. On this approach, some community science projects are different enough from traditional science that they require different standards of evaluation. Gwen Ottinger has done a lot to draw attention to these kinds of projects, which she terms social movement-driven citizen science (2016). In her view, these projects do "not seek merely to reproduce scientific methods in understudied areas; rather, these citizen projects critique, and offer alternatives to, methods and standards accepted by the scientific mainstream" (2017, p. 352). It is the inadequacy of scientific standards for collecting, evaluating, and interpreting data that motivates these community science projects in the first place. As a result, evaluating such projects according to the standards of traditional science misses or discounts their genuine contributions.

The idea that community science and traditional science sometimes require different standards of evaluation has the following advantages: it recognizes genuine and important differences between some types of community science and traditional science, and it does not imply that there is a hierarchy of community science quality with expert-controlled, shallow modes of engagement at the top. But I also worry that this approach concedes too much—that in advocating for different standards for community science, it grants that community science is second-rate science, even if it has other advantages that compensate for this fact. It's a bit like me saying that while my bird sketches aren't great art, this doesn't matter because they bring me pleasure and help sharpen my powers of observation.

At this point it's important to clarify that when critiquing standards for evaluating science, we may be claiming these standards are fundamentally flawed, and thus shouldn't be used at all, or we may be claiming certain standards are appropriate for evaluating science, but not for evaluating other kinds of activities. If, in claiming that community science and traditional science need different standards of evaluation, Ottinger and others mean that existing scientific standards are epistemically flawed, then of course it's wrong to judge community science by these flawed standards. But this is a reason to change the standards we use for evaluating traditional science, not a reason to use different standards for community science and traditional science. My own approach takes standards for evaluating traditional science to be largely appropriate, if imperfect, though I recognize that this is not a commitment shared by all theoretical perspectives.

If, on the other hand, the claim is that even standards appropriate for evaluating traditional science are inappropriate for evaluating community science, then my concern about conceding that community science is second-rate science comes into play. Now, if one really believes that some kinds of community science are a different kind of activity than traditional science, then insisting on different standards make sense. I am going to proceed, however, from the assumption that community science as I have defined it genuine science rather than some other kind of activity. A full defense of this assumption is beyond the scope of this paper.

The adequacy-for-purpose standard I am going to defend is distinct from both the expert accuracy standard and the social scientific alternative I have just described. The adequacy-for-purpose standard acknowledges differences between community science and traditional science that make the expert accuracy standard inappropriate, but also insists that community science really is science—not an extra-scientific form of knowledge production or an activity primarily focused on nonscientific goals. Community science data that is adequate-for-purpose, even if it is not expert-like, is as good, *qua* science, as data produced by experts.

2. Against expert accuracy

My argument against the expert accuracy standard stems from the fact that science has many different aims, some of which are in tension with each other. Traditional science meets some of these aims well, but not others. The same is true of community science. In fact, I'll suggest that community science can meet certain central aims of science better than traditional science can, even when it fails to meet the expert accuracy

[3] To learn about this project, visit the websites for POWHR (Protect Our Water, Heritage, Rights) (https://powhr.org/mvwatch/) and New River Geographics (https://data-nrgeo.opendata.arcgis.com/).

standard. This argument against the expert accuracy standard differs from the social scientific critique, which rests on the idea that community science and traditional science are too different from one another to be evaluated by the same standards.

The expert accuracy standard sets up the professional investigator as the model investigator. In an ideal world—one where money to pay researchers and the time to train them are not scarce—highly trained experts would conduct all scientific investigations. These trained experts would get things right more often than people with less training. They would make observations that those with less-finely-tuned senses would miss. Even though we can't always realize this ideal, the expert accuracy standard still considers it helpful, an ideal that should regulate how science is done.[4]

But this ideal is wrong. And not because it makes the perfect the enemy of the good or because it underestimates how severe the constraints on research actually are. The ideal is wrong because it imagines that squads of highly trained professionals would be the best way to further the aims of science if there were enough time and money to create and pay them.

Here is my argument against the expert accuracy standard and the ideal on which it rests:

1. **Multiple Aims**: Science has multiple aims that are in tension with one another.
2. **Tradeoffs**: Particular projects typically further some aims of science at the expense of others.
3. **Unnecessary Expertise**: Some aims of science can be achieved without expert quality data.
4. **No Hierarchy**: There is no hierarchy of scientific aims such that furthering certain ones is always better than furthering others.
5. No justification for the expert accuracy standard is compatible with premises 1–4.
6. So, no justification for the expert accuracy standard succeeds.

The first premise, Multiple Aims, is the near-consensus view within philosophy of science. The literature on Multiple Aims is extensive, and I can only sketch the reasons philosophers find it persuasive here. If you aren't convinced Multiple Aims is true, you won't accept the rest of the argument. In that case, I'm content to illustrate the logical connection between Multiple Aims and rejecting the expert accuracy standard.

Multiple Aims picks up on an idea with a long history in philosophy of science: that truth is not the primary aim of science. As Philip Kitcher (2001), (California, 1993) has pointed out, the primary aim of science cannot be truth, because there are infinite truths, some of which are uninteresting. Science picks and chooses which truths matter, so in Kitcher's view, science aims to discover *significant* truth. Which truths count as significant depend, in turn, on our values and interests, so in an important sense, all scientific projects are motivated by values.

Other philosophers have focused on the mysterious and important role that falsehoods, rather than truths, play in scientific investigation. Highlights from this literature include Richard Levins' (1966) insight that model-building involves inevitable tradeoffs between realism, precision, and generality, Nancy Cartwright's argument that the laws of physics lie (1983), and Bill Wimsatt's observation that false models are a means to truer theories (1987). For Catherine Elgin (2017) and Angela Potochnik (2017), falsehoods are so central to science that we can't even say the primary epistemic aim of science is truth, significant or otherwise. Instead, the primary epistemic aim of science is understanding. The literature on these ideas is vast, but a common theme emerging from it is that science has diverse aims, and sacrificing truth is the only way to achieve many of them.

There is no fixed list of the aims of science. Widely accepted aims of science include both epistemic aims such as explanation and prediction, as well as pragmatic aims such as informing policy or reducing suffering. There need not be a sharp divide between the epistemic and pragmatic aims, either. Ultimately, humans decide what the aims of science are, and how important different aims are relative to one another. Without taking a position here on how we *should* determine the aims of science and their relative importance, it is easy enough to find examples that demonstrate the existence of different aims and the ways in which they are in tension with one another.

Kevin Elliott and Dan McKaughan (2014) discuss the case of scientific risk assessments of toxic chemicals. One goal of risk assessment research is to accurately understand the health risks toxic chemicals pose to the public. Another goal is to identify risks quickly so that people aren't exposed to toxic chemicals for a long time before researchers establish the need for regulation. Unfortunately, the more accurate one aims to be in estimating the risks, the longer it takes to complete a risk assessment. Researchers and policy-makers must therefore choose which of these aims takes priority as they set the procedure for conducting risk assessments.

Additional support for Multiple Aims comes from Angela Potochnik (2015), who argues that one consequence of the centrality of falsehoods, or idealizations, in science is that, not only does science have many aims, but "it is the norm for the pursuit of one aim to occur at the expense of others" (p. 75). According to Potochnik, this tension among science's different aims is due to two facts: science investigates complex phenomena, and humans have limited powers of cognition and action. What science allows us to do, despite these limitations, is achieve some kind of understanding and control of a complex world. But the way in which science helps us do this is by focusing our efforts on transcending one or a few limitations at a time, not all of them at once. Our limitations also shape the aims of science. The aims we have are responses to our limited abilities to predict, make inferences, build new infrastructure, change behaviors, etc. (Potochnik, 2015, p. 76). This connection between the aims of science and human limitations is a fundamental reason why different aims of science are in tension with one another.

The second premise, Tradeoffs, follows from the same considerations that support Multiple Aims. If the aims of science are in tension with one another, it will be uncommon for particular research projects to further multiple aims at once. Or if they do, we still expect particular projects to further certain aims at the expense of others. What I mean by "at the expense of" is that data or other tools from a project that furthers one set of aims will at best serve different sets of aims less well than alternative data or tools could. For example, tools aimed at providing fast medical diagnostics are not often not able to diagnose with as much accuracy as tools designed with different aims in mind (e.g. Ranya et al., 2020). Or, the data that help us understand the spread of a particular invasive species may not be the ideal data for a meta-analysis aimed at identifying common features of many different biological invasions (see, e.g. the discussion in Thomsen, 2020).

The idea that science has diverse aims is now in place, so we can move to the Unnecessary Expertise premise. This premise is important because even if science has multiple, conflicting aims which cannot all be furthered by any single project, it could be still true that furthering any of these aims requires expert-level data collecting abilities. If so, the expert accuracy standard is in good shape.

The easiest way to see that Unnecessary Expertise is true is by way of an example. One aim of science is to inform policy decisions. In the case of invasive species research, decision-makers need to know if an invasive species is present in an area in before implementing eradication or control techniques. Surveys that determine the presence or absence of the species in question are sufficient to provide this information. Volunteers only need "low" levels of skill or training to carry out such surveys (Kosmala et al., 2016, p. 554), so they are popular community science

[4] Though the issue of who counts as an expert is an important one, I leave that question aside in this paper. Here, I just use the intuitive conception of expertise implicit in the expert accuracy standard—that of a person who is credentialed, institutionally sanctioned, and both knowledgeable and experienced with respect to a given domain.

projects. Since the only information needed to further the relevant aim of science in this case is presence/absence data, community scientists can be less accurate than experts at identifying invasive species, yet still provide the information that policy-makers or land managers need. Community scientists still need to meet some threshold of reliability in order for worries about false positives and false negatives not to swamp the benefits their data provide, but this threshold is lower than expert accuracy. Some scientific research is finicky, with relatively little room for error, but some of it is not, and it is counter-productive to think we need expert quality data in these cases.

I will not take a position here on how often certain aims of science can be achieved without expert-quality data. Obviously, my argument is stronger the more common it is that the aims of science do not require expert-quality data. But how prevalent these situations are is an empirical question, and not one I or anyone else is in a position to answer at present.

Premise 4, No Hierarchy, claims that there is no single aim of science that we should always prefer to further. If there was such an aim, and if expert-quality data was required to further it, then there might be a reason to relegate community science data that fail the expert accuracy standard to second-class status. Useful in their way, but unable to further the "high" aim of science. It is very unlikely that science has such an aim, however. First, if we look at scientific practice, we do not see one aim that emerges as preferred over the others. Second, if we accept Multiple Aims, there is no reason to think there is one fundamental scientific aim that is not in tension with some of science's other aims.

Of course, there are norms all scientific work should follow, such as not fabricating data. Following these norms can be in tension with furthering legitimate aims of science. But the existence of norms that should be followed no matter what is a different matter than the existence of an aim of science that trumps all others.

I should also clarify that No Hierarchy is compatible with the existence of many other kinds of hierarchies among the diverse aims of science. It is compatible, for instance, with the idea that the epistemic aims of science are more important than the pragmatic ones, or vice versa. I will not take a stand here on what, if any, hierarchies there are among the aims of science, except to deny there is a hierarchy in the strong sense required to save the expert accuracy standard.

The argument's final premise states that no justification for the expert accuracy standard (which says that a successful community science initiative should produce data that are always comparable to data produced by experts) is compatible with premises 1–4. To illustrate, let's return to the example of presence/absence data to inform invasive species management. We've established that in this case, expert-quality data isn't necessary to further a legitimate aim of science. So, applying the expert accuracy standard to evaluate a community science project whose volunteers record presence/absence data for an invasive species with less-than-expert accuracy would misclassify what is, in reality, high quality science. It would treat data from this project as lacking to the extent that community scientists are less accurate than experts, when in fact this project is adequate for the purpose it tries to accomplish. More accurate data would be necessary for other scientific purposes, but if the Multiple Aims and Tradeoffs premises are true, then there is not a good reason to expect this project to meet standards appropriate for other scientific purposes. And, given No Hierarchy, it's not true that an aim of science requiring expert accuracy in data collection would always have to trump the aim of the presence/absence study.

There are at least some (just how many is an open question) aims of science that can be furthered without expert quality data, and enough community science projects that contribute to these aims, that the expert accuracy standard fails as a general way of evaluating community science data. No justification for the expert accuracy standard succeeds. There may be particular community science projects where expert-level accuracy is needed, but this is not enough to license adopting it as a general way of evaluating community science projects.

Before abandoning the expert accuracy standard altogether, however, we should consider whether it can be salvaged through revision. After all, the idea that expertise sets the standard for quality dies hard. One thing a defender of the expert accuracy standard can readily concede is that failing to meet the expert accuracy standard does not mean data are useless. It just means that data are not as good as they would be if they did meet the expert accuracy standard. The alternative to less-than-expert quality community science data is likely no data at all, and some data is usually better than no data. It makes sense, then, to acknowledge that sloppy or somewhat inaccurate data may contribute to science. It may even make science better than it otherwise would be.

The expert accuracy standard can accommodate this point by shifting from an all-or-nothing evaluation of the success of a community science project to the idea that the *closer* community science data come to approximating expert data, the better those data are. If we make this revision, we will still look to experts in designing data collection protocols, even when we don't expect community scientists to be as good at data collection as experts are. What the defender of the expert accuracy standard will still insist on, however, is that there's no escaping expert guidance in telling community scientists what to aim for. Sure, we shouldn't dismiss community science that isn't of expert quality, but neither should we pretend that expert accuracy isn't an important regulative ideal and a useful way of assessing the quality of community science.

Depending on how it is developed, I either disagree with this proposal, or it approximates the adequacy-for-purpose standard I will present in the next section. If the core claim is that less-than-expert quality data is worse science than expert quality data, but still useful, then I disagree. On my view, for some data to be worse science than other data, it would have to be worse at furthering the relevant aim of science. We've already seen from the example of presence/absence data for an invasive species that being of less-than-expert quality is an imperfect proxy for how well data further aims of science. I'll provide another example in the next section, along with a principled reason for thinking it's true more generally that the expert accuracy standard doesn't track quality well enough for this revised version to be acceptable.

If, instead, the core claim is that the revised expert accuracy standard is something like, "a successful community science initiative should always produce data that is *as comparable to data produced by experts as it needs to be* in order to further the relevant aims of science," then the view is the same as my own view. What I argue below, however, is that once you go this far, expert accuracy drops out altogether.

One final point before I get to the adequacy-for-purpose standard. Not only *can* non-professionals adequately further some aims of science even when their data collection falls short of expert accuracy (this is the Unnecessary Expertise premise), it may be that some aims of science can *only* be furthered through the involvement of such non-professionals. I'm thinking of an aim like increasing public understanding of science. Research suggests that participating in the process of scientific inquiry is an effective way of increasing someone's understanding of how science works—of the nature of the enterprise itself (Bonney et al., 2016; Weinberg et al., 2018). If this is correct, then community science participation is an excellent tool for increasing the public's understanding of science. And, as survey after survey shows, Americans have a very poor understanding of science, both of particular scientific facts and of its nature (Kennedy & Hefferon, 2019). Poor understanding of the nature of science is a kind of public epistemic crisis, a possible contributor to the rejection of the theory of evolution (Lombrozo et al., 2008), human-caused climate change (Kovaka, 2019), and other scientific consensuses with important implications for public life. Scientific research conducted by professionals has clearly done little to prevent or abate this crisis, but perhaps community science is part of the solution. If this is right, it introduces the possibility of further tensions among the diverse aims of science. Involving non-professionals in research could be important for increasing public understanding of science, even as this involvement makes it difficult or impossible to achieve certain explanatory or predictive aims that do require expert accuracy.

You might object to this point on the grounds that I'm smuggling in an educational aim, when talk about the aims of science is grounded firmly in the sphere of producing new knowledge, not transmitting existing knowledge. What's right about this objection is that not all education counts as scientific inquiry. But this doesn't mean education is not a central aim of science. I am happy to place science education that does not involve genuine inquiry (e.g. teaching plate tectonics to children or doing the classic baking soda volcano "experiment") in the category of not-science, but some kinds of scientific education really do involve scientific investigation and inquiry. And, as any good humanist will tell you, the line between education and research is far from clear. Community science projects are perhaps the pre-eminent example, but even a single person dedicated to nature journaling and observation is participating in a process of inquiry continuous with the most sophisticated laboratory science.

This point about increasing public understanding of science being an aim of science is important because it shows another way in which community science projects don't need to meet the expert accuracy standard to be considered scientific successes. For projects with primarily educational aims, the expert accuracy standard is even more inappropriate than it is in for projects which primarily aim at surveying and monitoring. If educational projects are saddled with the expectation that they need to further both their educational aims and aims that require expert data quality, they are being set up to fail.

I also don't think I've exhausted the aims of science that community science can fulfill better than traditional science can. Perhaps mobilizing large numbers of people to action, or improving health, or a number of other aims also fall into this category. All of this being said, it is still important to evaluate community science. So if we reject the expert accuracy standard, what other options do we have?

3. Adequacy for purpose

As an alternative to the expert accuracy standard, I will borrow a set of ideas from Alisa Bokulich and Wendy Parker's recent work on data evaluation (2021). They present what they call an adequacy-for-purpose standard for evaluating the quality of scientific data. Though they have traditional scientific data in mind, the adequacy-for-purpose standard is flexible enough to apply to community science data as well. In arguing that we use this standard when evaluating community science data, I am developing Kevin Elliott and Jon Rosenberg's recent claim (2019) that when evaluating community science data, we need to consider the aims for which it was produced. At the heart of the adequacy-for-purpose standard is the idea that data quality depends on a number of dimensions, including who uses the data and the purposes for which they are used. The same data can count as high or low quality depending on, among other things, who uses the data and why.

The adequacy-for-purpose standard arises from what Bokulich and Parker call a pragmatic representational view of data. Data, in their view, are records of the results of a process of inquiry, and they represent the world rather than providing an unmediated view of it. Like any representation, data vary in their accuracy, resolution, and precision. Researchers collect data because it serves as evidence—not only for questions the data collector is interested in, but also for other researchers and other purposes. That is, the evidential value of data are not fixed (Leonelli, 2019).

According to Bokulich and Parker, a consequence of this view of data is that "the quality of some data or data model is relative to one or more purposes of interest; the question is not whether some data are 'good' or 'bad' … but rather whether they can be used to achieve the particular epistemic or practical aims that interest their users" (2021, p. 31). In order to determine when data are good enough for particular purposes, or what purposes some data are good enough for, we have to engage in data evaluation. And the appropriate standard for data evaluation is adequacy-for-purpose. Drawing on Parker (2020), Bokulich and Parker present two ways of thinking about adequacy-for-purpose. These aren't the only two possibilities, but they provide a sufficient flavor of the idea. The first sense is adequacy-for-purpose in a given instance:

Data are *adequate*$_I$*-for-P* just in case the use of data *D* in instance *I* would (or would be very likely to) result in the achievement of purpose *P* (ms, p. 11).

The second sense is adequacy-for-purpose given certain resources:

Data are *adequate*$_R$*-for-P* just in case the user U has access to informational, technological, cognitive and practical resources R such that there is some coherent way W that U *could* use data D to achieve purpose P (ms, p. 11).

Adequacy-for-purpose, then, is partly about how data represent the world, but also very much about the complex relationship among data, data users, their purposes, and their resources, including their scientific methods and tools and their background circumstances. These different dimensions create what Parker (2020) calls a problem space. The dimensions in the problem space set what sort of data can fulfill the scientific purpose in question. Depending on the details of the problem space, data will need to have certain properties, including but not limited to particular levels of accuracy, resolution, and precision.

Dimensions of a problem space can vary significantly across contexts. Data users in some community science projects will not be professional scientists but laypeople. This fact expands one dimension of the problem space relative to most traditional science projects. At the same time, a community science project's resources (or the resources of those who will use its data) may be very constrained compared to those of a traditional science project. This combination of a broader set of users but narrower set of resources affects the kind of data that will meet the adequacy-for-purpose standard.

Let's compare how the expert accuracy standard and the adequacy-for-purpose standard would evaluate data from a recent community science project run by Barbara Allen (2017). In 2014, Allen, along with Alison Cohen (an epidemiologist), Yolaine Ferrier and Johanna Lees (both anthropologists) conducted an environmental health study in one of France's largest industrial zones, known as Fos-sur-Mer/Etang de Berre. People who live in town bordering this industrial zone complain of pollution-related health problems, but none of several professionally conducted studies had produced evidence supporting their claims of a systematic problem. Allen and her collaborators set out to study the issue once again, this time with citizen involvement at every stage of the research, from study design, to execution, to interpretation and reporting.

Allen's approach was to ask residents of the towns near the industrial zone about their health, which the previous studies had not done. In consultation with residents and local doctors, the team developed a health survey. Central to the survey were questions about whether residents had ever been diagnosed with a variety of pollution-linked illnesses. Then the team administered the survey to residents by going door-to-door according to a randomly generated pattern. The survey, which had a 45% response rate and yielded responses from 10% of the population, revealed unusually high rates of asthma, cancer, and diabetes, along with skin conditions, nosebleeds, eye irritation, and headaches. Once the data were collected, the research team once again met with the local community to develop a final report, which included recommendations about how to respond to the results. These recommendations ranged from ideas about responding to powerful polluting industries to strategies for harm mitigation.

The French media payed quite a bit of attention to the final report after it was published. Since the results conflicted with the previous studies, Allen and her team had to defend and justify their approach. Ultimately, however, the report has met with success. One town has mobilized around the re-permitting of a large garbage incinerator. Governmental health experts have recommended allocating more resources for combatting chronic disease to the area. Local hospitals are using the report to guide their research agendas.

How would the expert accuracy standard evaluate the data produced by this community science project? If we think that producing data comparable to experts means producing data that is similar in kind or in agreement with the data experts studying the issue had already produced, then this project obviously fails to meet the expert accuracy standard. This project both chose a method that experts had elected not to use, perhaps because they perceived it to be scientifically less rigorous, and generated results that conflicted with expert conclusions.

Perhaps this is not a fair application of the expert accuracy standard. On another interpretation, the data this study produced needs to be comparable to what a team of experts (and not community scientists) would have produced using similar methods. A defender of the expert accuracy standard could say that the contrast class for applying the expert accuracy standard is a study with similar methods that did not solicit public input in survey design, data interpretation, or drafting the final report. But there are two problems here. First, it is quite possible that this hypothetical alternative study would have produced *worse* data than the actual one, in the sense that information centrally relevant to the question of interest would have been missed! According to Allen, the success of the study is due in large part to the fact that it involved dozens of focus groups, interviews, and open meetings with the local community to learn which health issues the survey should include and prioritize (2017). Second, the usefulness of the expert accuracy standard is called into question if we must imagine hypothetical data, the character of which we do not know, in order to evaluate data from actual studies.

The adequacy-for-purpose standard is much better suited to handling this case. To apply it, we must characterize the different dimensions of the problem space. The purpose of the study is to better understand and improve the environmental health of people living in the region. The data the study produced were intended to inform the public, both local and national, to lobby for health resources, and to provide justification for localities to change their policies with respect to polluting industries. This means that the potential users of the data range from any residents of Fos-sur-Mer/Etang de Berre to activists to health professionals to politicians. These users have different sets of resources available to them, so producing data that can be helpful to such different people with different resources is quite a feat. Yet this project seems to have accomplished it.

Even if we did not know that the study has already achieved some of its purposes, we would be able to identify aspects of it that do make it adequate for its purposes in several instances. For example, in designing the study, Allen and her team used random sampling and ensured that they had a sufficient response rate to defend the epidemiological rigor of the results. This technique was important for convincing people outside of the focal zone that the study was credible. Focus groups were another technique that helped make the data actionable for local residents. Simply learning that you live in a town where people get cancer and diabetes more than in other parts of France might not motivate you to action, unless you have the resources and inclination to move. But presenting the data in a context where people can brainstorm solutions like stopping street cleaners from using mechanical blowers, or changing the times of day that schools hold sports practice, achieves adequacy-for-purpose in a problem space with a diverse set of non-expert data users.

The results of this environmental health survey—both the data and the actions stemming from them—speak for themselves. We should want a standard of community science data evaluation that recognizes this study as high-quality science, not merely high-quality activism. The expert accuracy standard does not do this, at least not without a significant departure from the spirit of the standard and a sacrifice in applicability. But the adequacy-for-purpose standard does recognize this study as high-quality science, despite its unusual form, because the data it produced are able to further the relevant scientific aims.

4. Conclusion

I opened this paper by talking about cases where something like the expert accuracy standard is appropriate. When it comes to my personal hobbies, I can evaluate my skills in two different ways. I can ask, how good am I, *objectively*, at drawing or kayaking or playing the guitar? To answer this question, comparing my performance to expert performances is helpful. I can also ask, am I *good enough* at drawing, kayaking, or playing the guitar to accomplish what I want to accomplish with these skills? To answer this question, comparing myself to experts isn't helpful. Instead, something closer to an adequacy-for-purpose standard does the job.

What is distinctive about the adequacy-for-purpose standard for community science is that it doesn't entail admitting that community science is as different from expert science as my attempts at kayaking are from a whitewater champion's. Though there are often differences between the aims of community science and the aims of traditional science, the aims are still scientific aims. I can never take a rightful place among Olympic athletes simply because my athletic pursuits fulfill the aims of being personally rewarding. But my argument here is that community science and traditional science really are peers, and adopting the adequacy-for-purpose standard for evaluating community science data is the proper way to acknowledge this.

Acknowledgements

Thanks to Kevin Elliott, Billy Monks, Wendy Parker, Jessica Taylor, 2 anonymous reviewers, and audiences at Doing Science in a Pluralistic Society and at the Virginia Tech Center for the Humanities, for their help in developing this article.

References

Allen, B. L. (2017). A successful experiment in participatory science for promoting change in a French industrial region. *Engaging Science, Technology, and Society, 3*, 375–381. https://doi.org/10.17351/ests2017.180, 0.
Arnstein, S. R. (1969). A ladder of citizen participation. *Journal of the American Institute of Planners, 35*(4), 216–224. https://doi.org/10.1080/01944366908977225
Bokulich, A., & Parker, W. (2021). Data models, representation, and adequacy-for-purpose. *Euro Jnl Phil Sci, 11*(31). https://doi.org/10.1007/s13194-020-00345-2
Bonney, R., Phillips, T. B., Ballard, H. L., & Enck, J. W. (2016). Can citizen science enhance public understanding of science? *Public Understanding of Science, 25*(1), 2–16. https://doi.org/10.1177/0963662515607406
Burgess, H. K., DeBey, L. B., Froehlich, H. E., Schmidt, N., Theobald, E. J., Ettinger, A. K., HilleRisLambers, J., Tewksbury, J., & Parrish, J. K. (2017). The science of citizen science: Exploring barriers to use as a primary research tool. *Biological Conservation, 208*, 113–120. https://doi.org/10.1016/j.biocon.2016.05.014
S. D. P. K. P. of P. U. of. California. (1993). *The advancement of science: Science without legend, objectivity without illusions: Science without legend, objectivity without illusions*. USA: Oxford University Press
Cartwright, N. (1983). *How the laws of physics lie*. USA: Oxford University Press.
Elgin, C. (2017). *True enough*. USA: MIT Press.
Elliott, K. C., & McKaughan, D. J. (2014). Nonepistemic values and the multiple goals of science. *Philosophy of Science, 81*(1), 1–21. https://doi.org/10.1086/674345. JSTOR.
Elliott, K. C., & Rosenberg, J. (2019). Philosophical foundations for citizen science. *Citizen Science: Theory and Practice, 4*(1), 9. https://doi.org/10.5334/cstp.155
Fuccillo, K. K., Crimmins, T. M., de Rivera, C. E., & Elder, T. S. (2015). Assessing accuracy in citizen science-based plant phenology monitoring. *International Journal of Biometeorology, 59*(7), 917–926. https://doi.org/10.1007/s00484-014-0892-7
Gallo, T., & Waitt, D. (2011). Creating a successful citizen science model to detect and report invasive species. *BioScience, 61*(6), 459–465. https://doi.org/10.1525/bio.2011.61.6.8
Greenwood, D. J., & Levin, M. (2007). *Introduction to action research: Social research for social change*. Thousand Oaks, California: SAGE Publications.
Irwin, A., & Staff. (1995). *Citizen science: A study of people, expertise and sustainable development*. Psychology Press.
Kemmis, S., McTaggart, R., & Nixon, R. (2014). *The action research planner: Doing critical participatory action research*. Singapore: Springer Science & Business Media. https://doi.org/10.1007/978-981-4560-67-2
Kennedy, B., & Hefferon, M. (2019). *What Americans know about science*. Pew Research Center.
Kimura, A. H., & Kinchy, A. (2016). Citizen science: Probing the virtues and contexts of participatory research. *Engaging Science, Technology, and Society, 2*, 331–361. https://doi.org/10.17351/ests2016.99, 0.
Kimura, A. H., & Kinchy, A. (2019). *Science by the people: Participation, power, and the politics of environmental knowledge*. USA: Rutgers University Press.
Kitcher, P. (2001). *Science, truth, and democracy*. USA: Oxford University Press.
Kosmala, M., Wiggins, A., Swanson, A., & Simmons, B. (2016). Assessing data quality in citizen science. *Frontiers in Ecology and the Environment, 14*(10), 551–560. https://doi.org/10.1002/fee.1436
Kovaka, K. (2019). *Climate change denial and beliefs about science*. Synthese. https://doi.org/10.1007/s11229-019-02210-z

Leonelli, S. (2019). What distinguishes data from models? *European Journal for Philosophy of Science, 9*(2), 22. https://doi.org/10.1007/s13194-018-0246-0

Levins, R. (1966). The straegy of model building in population biology. *American Scientist, 54*(4), 421–431 (JSTOR).

Lombrozo, T., Thanukos, A., & Weisberg, M. (2008). The importance of understanding the nature of science for accepting evolution. *Evolution: Education and Outreach, 1*(3), 290–298. https://doi.org/10.1007/s12052-008-0061-8

McDonough MacKenzie, C., Gallinat, A. S., & Zipf, L. (2020). Low-cost observations and experiments return a high value in plant phenology research. *Applications in Plant Sciences, 8*(4). https://doi.org/10.1002/aps3.11338

Ottinger, G. (2016). Social movement-based citizen science. In *The rightful place of science: Citizen science* (pp. 89–103). Arizona State University. Consortium for Science, Policy, and Outcomes.

Ottinger, G. (2017). Reconstructing or Reproducing? : Scientific authority and models of change in two traditions of citizen science. In *The routledge handbook of the political economy of science* (pp. 351–364). Taylor and Francis. https://doi.org/10.4324/9781315685397-31.

Parker, W. S. (2020). Model evaluation: An adequacy-for-purpose view. *Philosophy of Science, 87*(3), 457–477. https://doi.org/10.1086/708691

Potochnik, A. (2015). The diverse aims of science. *Studies in History and Philosophy of Science, 53*, 71–80. https://doi.org/10.1016/j.shpsa.2015.05.008

Potochnik, A. (2017). *Idealization and the aims of science*. University of Chicago Press.

Ranya, M., Jones Hayley, E., Sian, T. P., Justin, S., Keith, P., Shabnam, J., et al. (2020). Accuracy of UK rapid test consortium (UK-RTC) "AbC-19 rapid test" for detection of previous SARS-CoV-2 infection in key workers: Test accuracy study. *BMJ, 371*. https://doi.org/10.37766/inplasy2020.4.0099. m4262.

Riesch, H., & Potter, C. (2013). *Citizen science as seen by scientists: Methodological, epistemological and ethical dimensions: Public Understanding of Science*. https://doi.org/10.1177/0963662513497324

Shirk, J., Ballard, H., Wilderman, C., Phillips, T., Wiggins, A., Jordan, R., McCallie, E., Minarchek, M., Lewenstein, B., Krasny, M., & Bonney, R. (2012). Public participation in scientific research: A framework for deliberate design. *Ecology and Society, 17*(2). https://doi.org/10.5751/ES-04705-170229

Thomsen, M. S. (2020). Indiscriminate data aggregation in ecological meta-analysis underestimates impacts of invasive species. *Nature ecology & evolution, 4*(3), 312–314. https://doi.org/10.1038/s41559-020-1117-6

Torre, M. E., Fine, M., Stoudt, B. G., & Fox, M. (2012). Critical participatory action research as public science. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbooks in psychology®. APA handbook of research methods in psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological* (pp. 171–184). American Psychological Association. https://doi.org/10.1037/13620-011.

Weinberg, A. E., Trott, C. D., & Sample McMeeking, L. B. (2018). Who produces knowledge? Transforming undergraduate students' views of science through participatory action research. *Science Education, 102*, 1155–1175.

Whyte, K. P., & Crease, R. P. (2010). Trust, expertise, and the philosophy of science. *Synthese, 177*(3), 411–425. https://doi.org/10.1007/s11229-010-9786-3

Wiggins, A., & Crowston, K. (2011). From conservation to crowdsourcing: A typology of citizen science. *Proceedings of the 2011 44th Hawaii International Conference on System Sciences, 1–10*. https://doi.org/10.1109/HICSS.2011.207

Willett, K. W., Lintott, C. J., Bamford, S. P., Masters, K. L., Simmons, B. D., Casteels, K. R. V., Edmondson, E. M., Fortson, L. F., Kaviraj, S., Keel, W. C., Melvin, T., Nichol, R. C., Raddick, M. J., Schawinski, K., Simpson, R. J., Skibba, R. A., Smith, A. M., & Thomas, D. (2013). Galaxy zoo 2: Detailed morphological classifications for 304 122 galaxies from the sloan digital sky survey. *Monthly Notices of the Royal Astronomical Society, 435*(4), 2835–2860. https://doi.org/10.1093/mnras/stt1458

Wimsatt, W. (1987). False models as means to truer theories. In *Neutral models in biology* (pp. 23–55). Oxford University Press.

Wylie, A. (2015). A plurality of pluralisms: Collaborative practice in archaeology. In F. Padovani, A. Richardson, & J. Y. Tsou (Eds.), *Objectivity in science: New perspectives from science and technology studies* (pp. 189–210). Springer International Publishing. https://doi.org/10.1007/978-3-319-14349-1_10.