Ref. Ares(2021)3964279 - 17/06/2021

Crowd4SDG
Citizen Science for the Sustainable Development Goals

Deliverable 4.1

Report on an epistemological analysis of metrics/descriptors for citizen science

Deliverable identifier: D4.1

Due date: 30/04/2021

Justification for delay: due to exceptional activity as agreed with Project Officer

Document release date: 18/06/2021

Nature: Report

Dissemination level: Public

Work Package: WP4

Lead beneficiary: UP

Contributing beneficiaries: UNIGE, CSIC

Document status: Final

For more information on Crowd4SDG, please check: http://www.crowd4sdg.eu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 872944.

Abstract:

Humanity is currently confronted with a series of profound existential crises. Beneath it all lies a crisis of meaning, of being able to make sense of the world. If we are to overcome this situation, we need robust scientific knowledge of the world and our place within it. Unfortunately, the authority of science itself is under attack from many sectors of society. One reason for this is that science often fails to fulfil the unrealistic expectations it has set itself to meet. Such expectations are generated by an outdated view of knowledge production, which persists among many members of the public and scientists alike. This has led to an ultra-competitive system of academic research, which sacrifices long-term productivity through an excessive obsession with short-term efficiency. Efforts to diversify this system come from a movement called democratic citizen science. Here, we argue that this kind of citizen science can serve as a model for scientific inquiry in general, and highlight corresponding opportunities in the context of the projects generated during the GEAR cycles of the Crowd4SDG project. It requires an alternative theory of knowledge with a focus on the role that diversity plays in the process of discovery. Here, we present such an epistemology, which is based on three central philosophical pillars: perspectival realism, a naturalistic process-based epistemology, and deliberative social practices. Together, these three pillars broaden our focus from immediate research outcomes towards the cognitive and social processes which shape research strategies that facilitate sustainable long-term productivity and scientific innovation. They provide a general paradigm and template for scientific inquiry in the 21st century, which marks a shift from an industrial to an ecological vision of how scientific research should be done, and how it should be assessed. At its core are research communities that are diverse, representative, and democratic. This leads to a new, processual, paradigm of scientific project management, monitoring, and assessment, which is outlined in the last section of our report.

Document History

	Name	Partner	Date
Authored by	Johannes Jaeger, Camille Masselot, Bastian Greshake Tzovaras, Enric Senabre Hidalgo, Marc Santolini	UP	11/04/21
Revised by	Barbara Pernici Francois Grey	POLIMI UNIGE	17/04/21 22/04/21
Edited by	Johannes Jaeger, Camille Masselot, Marc Santolini	UP	02/05/21
Reviewed by	Francois Grey	UNIGE	10/05/21
Edited by	Johannes Jaeger, Camille Masselot, Marc Santolini	UP	18/05/21
Approved by	Francois Grey	UNIGE	24/05/21

Table of Contents

Project Partners

	Partner name	Acronym	Country
1 (COO)	Université de Genève	UNIGE	CH
2	European Organization for Nuclear Research	CERN	CH
3	Agencia Estatal Consejo Superior de Investigaciones Científicas	CSIC	ES
4	Politecnico di Milano	POLIMI	IT
5	United Nations Institute for Training and Research	UNITAR	CH
6	Université de Paris	UP	FR

Crowd4SDG in Brief

The 17 Sustainable Development Goals (SDGs), launched by the UN in 2015, are underpinned by over 160 concrete targets and over 230 measurable indicators. Some of these indicators initially had no established measurement methodology. For others, many countries do not have the data collection capacity. Measuring progress towards the SDGs is thus a challenge for most national statistical offices.

The goal of the Crowd4SDG project is to research the extent to which Citizen Science (CS) can provide an essential source of non-traditional data for tracking progress towards the SDGs, as well as the ability of CS to generate social innovations that enable such progress. Based on shared expertise in crowdsourcing for disaster response, the transdisciplinary Crowd4SDG consortium of six partners is focusing on SDG 13, Climate Action, to explore new ways of applying CS for monitoring the impacts of extreme climate events and strengthening the resilience of communities to climate related disasters.

To achieve this goal, Crowd4SDG is initiating research on the applications of artificial intelligence and machine learning to enhance CS and explore the use of social media and other non-traditional data sources for more effective monitoring of SDGs by citizens. Crowd4SDG is using direct channels through consortium partner UNITAR to provide National Statistical Offices (NSOs) with recommendations on best practices for generating and exploiting CS data for tracking the SDGs.

To this end, Crowd4SDG rigorously assesses the quality of the scientific knowledge and usefulness of practical innovations occurring when teams develop new CS projects focusing on climate action. This occurs through three annual challenge based innovation events, involving online and in-person coaching. A wide range of stakeholders, from the UN, governments, the private sector, NGOs, academia, innovation incubators and maker spaces are involved in advising the project and exploiting the scientific knowledge and technical innovations that it generates.

Crowd4SDG has six work packages. Besides Project Management (UNIGE) and Dissemination & Outreach (CERN), the project features work packages on: Enhancing CS Tools (CSIC, POLIMI) with AI and social media analysis features, to improve data quality and deliberation processes in CS; New Metrics for CS (UP), to track and improve innovation in CS project coaching events; Impact Assessment of CS (UNITAR) with a focus on the requirements of NSOs as end-users of CS data for SDG monitoring. At the core of the project is Project Deployment (UNIGE) based on a novel innovation cycle called GEAR (Gather, Evaluate, Accelerate, Refine), which runs once a year.

The GEAR cycles involve online selection and coaching of citizen-generated ideas for climate action, using the UNIGE Open Seventeen Challenge (O17). The most promising projects are accelerated during a two-week in-person Challenge-Based Innovation (CBI) course. Top projects receive further support at annual SDG conferences hosted at partner sites. GEAR cycles focus on specific aspects of Climate Action connected with other SDGs like Gender Equality.

Grant Agreement description of the deliverable

T4.1: Conceptual/epistemological foundations for new metrics and descriptors for citizen science (UPD, UNIGE)

In order to develop new metrics that are more appropriate to capture the diversified incentives and aims of Citizen Science, a careful analysis will be performed of the epistemological foundations and the societal conditions on which Citizen Science projects can flourish. In particular, the emphasis will move from the 'output of science' to a 'process of open inquiry'. Metrics for the success of Citizen Science projects should not only monitor scientific output, but also the quality and diversity of interactions within a group of citizens between them and professional researchers. In addition, metrics that measure productivity must be supplemented by descriptors for the originality, relevance, robustness, adaptiveness of the resulting data and insights, as these are core aspects of research which Citizen Science aims to achieve. Lastly, an increase of the number of valid perspectives on a specific phenomenon is a desirable attribute in research on complex topics like climate resilience. Therefore, we will explore metrics and descriptors that measure the validity and diversity of the multiple perspectives produced by Citizen Science projects. The results from this task will form the foundation for the metrics and descriptors to be developed in Task 4.2. These new metrics will be used to evaluate and improve the CS projects created through the GEAR cycles, and the GEAR methodology itself, in T3.4 and T3.3.

D4.1: Report on an epistemological analysis of metrics/descriptors for citizen science. (M12)

Report: an epistemology for democratic Citizen Science

Authors: Johannes Jaeger (UP), Camille Masselot (UP), Bastian Greshake Tzovaras (UP), Enric Senabre Hidalgo (UP), Marc Santolini (UP)

1. Introduction

The way we do science and the role of science in society are rapidly changing. Since its earliest beginnings, scientific inquiry has been inspired by the ideal of research as a leisurely and disinterested pursuit of knowledge for knowledge's sake. However, this ideal now seems hopelessly outdated, since science is no longer carried out by a small elite of independently wealthy gentlemen scientists. Much has changed for the better: a steady and massive increase in global science funding has enabled the establishment, professionalization, internationalization, and diversification of many flourishing research fields. At the same time, however, the increased flow of money through the system has also resulted in increased fragmentation, specialization and the commodification of science, creating a departure from its initial ideals. Doing science has become much like a factory business—a production line based on fine-tuned division of labor—geared and optimized towards the efficient generation of research output. Efficiency is certainly not a bad thing per se, especially considering the many urgent problems facing humanity in the 21st century. And yet, an obsessive focus on efficiency can hamper long-term productivity when the free exploration of ideas is pitted against the expectation of short-term local returns, preferably in terms of monetary gains and technological applications (see, for example, Melo-Martín & Intemann, 2018; Stephan, 2015). Besides, scientific explorers are more and more often accused of academic elitism, being pampered and detached from the harsh realities outside the ivory tower. This goes hand-in-hand with a growing suspicion towards scientific expertise in general.

In light of this complicated situation, it is timely for us to reconsider the way we do science in our present socio-political context. Humanity is currently living through a period of profound crisis, affecting our ecological sustainability, the stability of our socio-economic and political systems, and—at the bottom of it all—our ability to make sense of the world. We need trustworthy knowledge more than ever before. We need science for the public good, thriving beyond the reach of shortsighted business and political lobbies—a science which serves the interests of all. But how can we ensure the free pursuit of knowledge, without reverting to the elitist gentlemen-science of yore, without neglecting the fact that we urgently need actionable solutions for real-world problems? How can we restore the public's trust in science without losing academic freedom, without rendering it vulnerable to populist political whims? In essence, how can we achieve the democratization of science—a democratization of the process of inquiry itself, but also of the relationship between professional scientists and those who are affected by their work (see Dewey, 1927, for an early exposition of this problem; reviewed in Mirowski, 2018)? More specifically, how can we achieve it without jeopardizing the independence, authority, and long-term productivity of science? This may be the most important question we face as a scientific community today. It may also be one of the most difficult questions we have ever had to tackle. In the context of the Crowed4SDG project, this raises the question of how we can monitor whether such a democratic process is at play, and how we can generate procedures that both measure and ensure that the corresponding incentives are met?

Here, we look at this fundamental question from a philosophical (but also practical) angle, using the kind of citizen science that is promoted and studied by Crowd4SDG as a case study. Citizen science, broadly defined, includes any scientific project in which individuals or communities who are not professional researchers (and may not even have formal scientific training) participate in research design, data collection, and/or data analysis (Eitzel et al., 2017; Mac Domhnaill et al., 2020). Citizen science denotes a way of organising scientific inquiry, rather than a specific type of research project (Franzoni & Sauermann, 2014). In fact, citizen science projects are extremely diverse. On one end of the spectrum, there are those that are still driven by traditional scientific authorities and follow a model of traditional research, but collect data through gamified crowdsourcing (as in the FoldIt project on protein folding: http://fold.it), or community-wide observations and annotation efforts (e.g. Galaxy Zoo, analysing images from different space surveys: https://www.zooniverse.org/projects/zookeeper/galaxy-zoo; these and other examples are reviewed in Franzoni & Sauermann, 2014; Nielsen, 2011). In this type of citizen science, research design, data analysis, and quality control remain exclusively in the hands of professional scientists, and although participants may learn something new through their taking part in a study, their scientific education is not a major focus. On the other end of the spectrum, there are community-based projects that do not necessarily aim to generate new conceptual breakthroughs or theoretical frameworks. Instead, they have more practical scientific knowledge as their primary aim: monitoring and improving water or air quality in a community, for example. Such projects are often driven by nonscientists.

In the context of our discussion here, we are most interested in a kind of citizen science that lies between these two extremes. In particular, we are interested in projects that actively involve a broad range of participants in project design, data analysis, and quality monitoring, with the triple aim of generating new scientific knowledge, of teaching participants about science, and of improving understanding and relations between scientists and nonscientists in a mutually beneficial way. Such projects are often as much concerned with the process of scientific inquiry itself as with generating new knowledge about the world. They study the way by which such knowledge is gained, how individual participants learn in the process, and how teams are best managed to tap into the collective intelligence emerging from the collaboration. We take this definition of a democratic, participatory, social-movement-based or socially-engaged citizen science as an ideal worth aspiring to (as do others; see Haklay, 2018; Mahr & Dickel, 2019; Ottinger, 2017; Strasser et al., 2019, but also Mirowski, 2018, for a more critical assessment). More generally, we believe that it serves as a good model for the kind of reforms we need for the democratization of scientific research in general, beyond the domain of citizen science.

Much has been written about the historical, political, and sociological aspects of democratic citizen science (see, for example, Mahr & Dickel, 2019; Smart et al., 2019; Strasser et al., 2019). It differs significantly from traditional academic research in its goals, values, attitudes, practices, and methodologies. Apart from its focus on the process of inquiry, democratic citizen science has a number of obvious advantages when considered from a political or ethical point of view. It not only taps into a large base of potential contributors, generally incurring a relatively low amount of costs per participant, but also fosters inclusion and diversity in scientific communities, opens a channel of communication between scientists and nonscientists, and provides hands-on science education to interested citizens. Democratic citizen science can help to address the problems of undone science—important areas of inquiry which are neglected due to competing political agendas (Frickel et al., 2010)—and of epistemic injustice—inequalities in the accessibility and distribution of scientific knowledge (Bai, 2020). It aims to bring scientific knowledge to those who most urgently need it, rather than those who provide the bulk of the funding. Its open design is intended to increase the reproducibility and robustness of its scientific results, and to promote collaboration over competition in the process of inquiry.

All these benefits, of course, rely on the strict and proper implementation and monitoring of procedures and protocols that ensure good scientific practice, management, and data quality control. Other challenging aspects of democratic citizen science are its relatively low per-person productivity (compared to that of full-time professional researchers who generally require less instruction and supervision), and an increased complexity in project management—especially if citizen scientists are not merely employed for data collection, but are also involved in project design, quality monitoring as well as the analysis and interpretation of results.

Beyond these practical considerations, there is a more philosophical dimension to democratic citizen science that has received surprisingly little attention so far (see Kasperowski & Hillman, 2018; Ottinger, 2017; Strasser et al., 2019, for a number of notable exceptions). It concerns the theory of knowledge, the kind of epistemology able to describe, analyse, and support the efforts of democratic citizen science. In other words, to assess the practicality, usefulness, ethics, and overall success of democratic citizen science, we need to take seriously the kind of knowledge it produces, and the way by which it produces it. It is this largely unexamined epistemological aspect of citizen science that we want to analyse in this report.

To precisely pinpoint and highlight the differences between knowledge production in democratic citizen science and in traditional academic research, we make use of an argumentative device: we present an epistemology ideally suitable for citizen-science projects of the democratic kind described above by contrasting it with a very traditional view of scientific epistemology. Our intention is not to build a straw man argument, or to paint an oversimplified black-and-white picture of the world of (citizen) science. We are very well aware that the epistemic stances of many scientists and citizens are much more sophisticated, nuanced, and diverse than those that will be depicted here (see, for example, Yucel, 2018). However, even though the philosophy of science may have moved on, many practicing scientists and stakeholders of science still do retain remnants of a decidedly old-fashioned view of science, which we will call naïve realism (ibid.). In most cases, this view is not explicitly formulated in the minds of those who hold it and its assumptions and implications remain unexamined. Nor does this view amount to a consistent or systematic philosophical doctrine. Instead, naïve realism consists of a set of more or less vaguely held convictions, which often clash in contradictions, and leave many problems concerning the scientific method and the knowledge it produces unresolved. And yet, somehow, these ideas tenaciously persist and hold a firm grip on what we—as communities of scientists, stakeholders, and citizens—consider to be the epistemic goal and the societal role of scientific research.

It should be quite clear that the persistence of naïve realism is not a purely theoretical or philosophical problem. One of its major practical implications concerns the way we assess the success of research projects (see Leydesdorff, 2005, for an historical overview). What we value crucially depends on how we define the epistemic (and non-epistemic) goals of science, and what we consider high-quality scientific knowledge. We will argue below that naïve realism leads to a system of incentives which is excessively focussed on a misguided notion of accountability and short-term productivity—in particular, the efficient generation of measurable research output (Muller, 2018). We could call this the industrial model of doing science, since it treats research as a system of mechanical production, which must be put under tight, top-down control.

In such an industrial system, projects of democratic citizen science are at a fundamental disadvantage. Standard assessment practices do not do justice to the diversified ways by which such projects generate knowledge and other benefits for the participants and stakeholders involved (Kieslinger et al., 2018; Schaefer et al., 2021). Even more importantly, democratic citizen science cannot compete with traditional academic science in terms of production efficiency, mainly due to its large organizational overhead, but also because the efficient production of knowledge is often not its primary goal. All of this implies that merely encouraging (or even enforcing) inclusive and open practices, while generating technological platforms and tools to implement them, will not be sufficient to propel citizen science beyond its current status as a specialized niche product—often criticised, belittled, or ignored by commentators and academic researchers for its lack of rigour and efficiency. This is a serious problem, which is philosophical down to its core, and therefore calls for a philosophical solution. In order for citizen science to succeed beyond its current limitations, we need a fundamental reexamination of the nature and purpose of scientific knowledge, and how it is produced. In particular, we need to move beyond our increasing obsession with productivity metrics in science. Simply put, we require a new model for doing research, with corresponding descriptors and procedures for quality control, that is more tolerant and conducive to diversity and inclusive participation (see also Couch et al., 2019; Mahr & Dickel, 2019).

In what follows, we outline an epistemology of science, which is formulated explicitly with our discussion of democratic citizen science in mind. It is centered around three main philosophical pillars (Fig. 1). The first is perspectival realism (also called scientific perspectivism), providing an alternative to naïve realism which is appropriate for the 21st century (Giere, 2006; Wimsatt, 2007). The second is process philosophy, in the form of naturalistic epistemology, which focuses our attention away from knowledge as the product, or final outcome, of scientific research, towards the cognitive processes underlying knowledge production (Kitcher, 1992; Rescher, 1996; Seibt, 2020). The third and final pillar is deliberative practice, with its focus on social interactions among researchers, which yields the surprising insight that we should not always reach for consensus in science (Beatty & Alfred, 2010). These three pillars tightly intertwine and combine into a new model, which we could call the ecological model of doing science, because just like an ecosystem, it is centered around diversity, inclusion, interaction, self-organization, and robustness, in addition to long-term productivity. This model is based on a completely different notion of accountability, leading to process-oriented, participatory, and integrated assessment strategies for scientific projects that go far beyond any narrow set of metrics to measure research output.

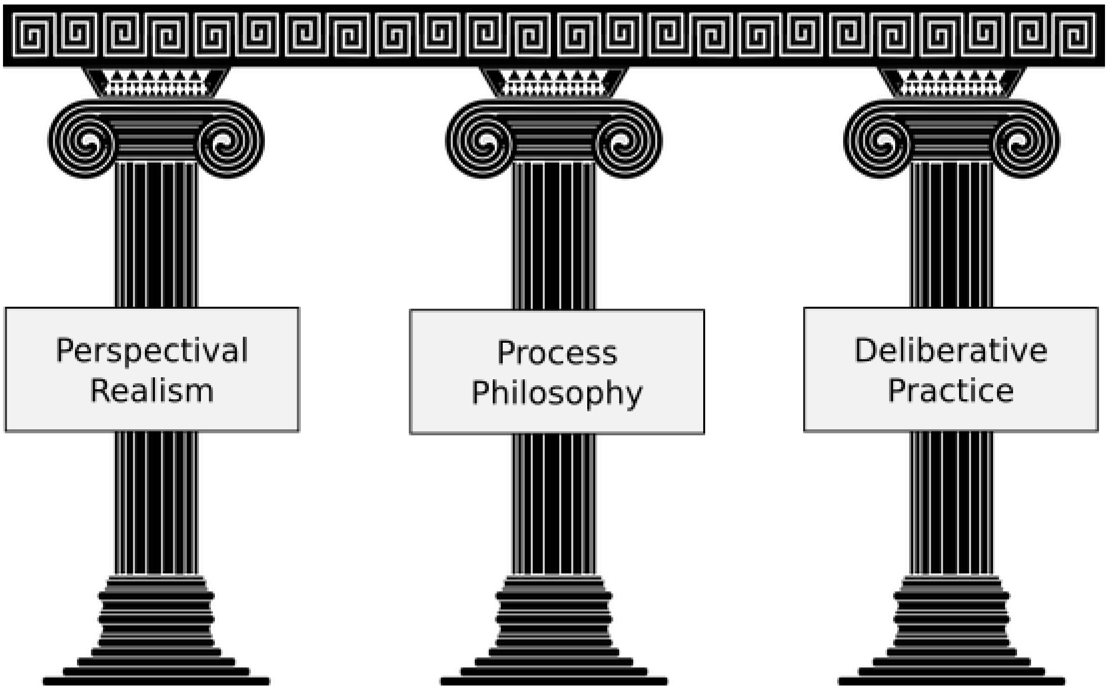

Figure 1: The three pillars of an ecological model for scientific research developed in this report: perspectival realism as an alternative to naïve realism which is appropriate for the 21st century; process philosophy that places cognitive processes underlying knowledge production above knowledge as a product; deliberative practice that aims to go beyond consensus as a goal of social interactions in science.

Box 1: Connection with Crowd4SDG

In this and the following boxes, we provide short summaries of how the epistemological analysis in this report translates into practical advice for the metrics and descriptors that will be used in the Crowd4SDG project. This advice is then combined into a general summary in Section 8: "Implications for Crowd4SDG".

The Crowd4SDG project has set itself the challenge of developing "metrics and descriptors that measure the validity and diversity of the multiple perspectives produced by citizen science projects." In the rest of this report, we use approaches of epistemology to "provide a solid philosophical and conceptual foundation for such metrics and descriptors."

When we refer to **metrics** in this report, we mean not only traditional measures of the success of Citizen Science projects that monitor scientific output (amount of useful data generated, number of scientific articles published) but also the quality and diversity of interactions within a group of citizens, and between them and professional researchers and domain experts.

When we refer to **descriptors** in this report, we are concerned with ways of describing the originality, relevance, robustness and adaptiveness of the resulting data and insights, on the basis that these are core aspects of research which Citizen Science aims to achieve.

Given the abundance of data on interactions between teams and citizens that we expect to obtain within the Crowd4SDG project, we will highlight how we propose to define and refine metrics and descriptors that enable us to monitor the epistemic pillars presented in this report. These metrics and descriptors will in turn be used to evaluate and improve the CS projects created through the so-called GEAR cycles, which are cycles of innovation of citizen science projects running annually during the course of the Crowd4SDG project. The same metrics and descriptors will be applied to improve the GEAR methodology itself.

2. Naïve realism and the cult of measurable productivity

What we mean here by naïve realism is a form of *objectivist realism* that consists of a loose and varied assortment of philosophical preconceptions that, although mostly outdated, continue to shape our view of science and its role in society. Its central tenet is that the main (and only) epistemic goal of science is to find objective and universal Truth. The ideas behind this popular notion are drawn from three main historical sources: the logical positivism of the Vienna Circle, Popper's falsificationism, and Merton's sociology of science.

Positivism in general, and *empirical* or *logical positivism* in particular, hold that information derived from sensory experience, interpreted through reason and logic, forms the source of all certain knowledge (Creath, 2021; Grayling, 2019; Sigmund, 2017). The logical positivists asserted that meaningful discourse is either purely analytic (in logic and mathematics) or empirically testable (in science). Everything else is cognitively meaningless, in particular what became labeled as "metaphysics:" abstract philosophical theory that has no basis in reality. This is still reflected in the "I have facts, and therefore do not need any philosophy" attitude of many current-day researchers.

At the heart of positivism lies *the principle of verification*: scientific hypotheses are positively confirmed by empirical evidence, which comes in the form of condensed summaries of direct observations, where all terms are defined ostensively, *i.e.* in an obvious and unambiguous manner. This firmly anchors scientific knowledge in objective reality, but it demands an exceptional degree of clarity, detachment, and objectivity on the part of the observer. The fact that human beings may not be able to achieve such detached, objective clarity was acknowledged by several logical empiricists themselves. Even our most basic observations are coloured by mood and emotions, biased assumptions, and the things we already know.

In the meantime, Karl Popper—probably the world's best-known philosopher of science—revealed an even more serious and fundamental problem with verification: he showed that it is impossible, amounting to a logical fallacy (an affirmation of the consequent) (Creath, 2021; Grayling, 2019; Sigmund, 2017). In contrast, Popper argued that it *is* possible to *falsify* hypotheses by empirical evidence. Therefore, the only way to empirically test a scientific conjecture is to try to refute it. In fact, if it is not refutable, it is not scientific. This part of Popper's argument still stands strong today, and, because of it, (logical) positivism has become completely untenable among philosophers of science.

The doctrine of *falsificationism* may be the most widely held view of science among practicing researchers and members of the wider public today. However, unknown to most, it has a number of problems and some very counterintuitive implications. First of all, falsificationism is completely incompatible with positivism, even though both views often co-exist in the minds of naïve realists. In fact, falsificationism is incompatible with any kind of realism. In Popper's view, scientific hypotheses stand as long as they have not yet been falsified, but they are never confirmed to be true in the sense of accurately reflecting some specific aspect of reality. Popper called this *verisimilitude*, which literally translates as the appearance of being true. Furthermore, falsificationism provides a rather simplistic account of how science actually works. In practice, scientific theories are rarely discarded, especially not if viable alternatives are lacking. Instead of refuting them, theories are often amended or extended to accommodate an incompatible observation. Quite often, scientists do not even bother to adjust their theories at all: paradoxical results are simply ignored and classified as outliers. Finally, falsificationism has nothing to say about how hypotheses are generated in the first place. It turns a blind eye to the sources of scientific ideas, which remain a mystery, beyond philosophical investigation. Seen from this angle, the creative aspects of science seem rather irrational, and the scientific method acts as a selection mechanism to objectively filter out yet another silly idea.

On top of a fluctuating mix of positivist and Popperian ideas, naïve realism often incorporates a simple *ethos of science* that goes back to the work of sociologist Robert Merton (1973). This ethos is based on four basic principles: (1) *universalism*—criteria to evaluate scientific claims must not depend on the person making the claim; (2) *communism* (or *communality,* for our American readers)—scientific knowledge must be commonly owned once it is published; (3) *disinterestedness*—scientists must disengage their interests from their judgments and actions; and (4) *organised skepticism*—scientific communities must disbelieve, criticise, and challenge new views until they are firmly established. According to Merton, scientists who conform to his ethos should be rewarded, while those that violate it should be punished. In this way, the ethos ensures that science can fulfil its primary societal role: to provide a source of certified, trustworthy knowledge.

It should be evident—even from such a brief and cursory overview—that the ideas underlying naïve realism do not form a coherent doctrine, even though they generally adhere to a vaguely defined objectivist realism. Nor do they paint a very accurate picture of actual science, performed by actual human beings. In fact, the naïve realist view is highly idealized: more about what science *should be like* in our imagination than about what it actually is. It provides a deceptively simple epistemological framework for an ideal science whose progress is predictable, under our control. This is probably why it is still so attractive and influential today. Everybody can understand it, and it makes a lot of intuitive sense, even though it may not hold up to closer scrutiny. Its axiomatic nature provides an enticing vision of a simpler and better world than the complicated and imperfect one we actually live in. Because of its (somewhat ironic) detachment from reality, there will likely be unintended consequences and a lack of adaptability if we allow such an overly simplistic vision to govern our way of measuring the success of science. Let us highlight some of the specific features of naïve realism that lead to unforeseen negative consequences in science today.

First of all, naïve realism suggests that there is a single *universal scientific method*—based on logical reasoning and empirical investigation—which is shared by researchers across the natural and social sciences. This method allows us to verify, or at least falsify, scientific hypotheses in light of empirical evidence, independent of the aim or object of study. Considered this way, the application of the scientific method turns scientific inquiry into a formal activity. It works like an algorithm. If applied properly, scientific inquiry leads to *an ever-increasing accumulation of knowledge that approximates reality asymptotically* (Fig. 2). Because of our finite nature as human beings, we may never have definitive knowledge of reality, but we are undoubtedly getting closer and closer.

Complementary to this kind of formalisation, we have a universally accepted ethos of science, which provides a set of standards and norms. When properly applied, these standards and norms guarantee the *validity and objectivity of scientific knowledge*. Scientific method and practice become self-correcting filters that automatically detect and weed out erroneous or irrational beliefs or biases. In that sense, scientific inquiry is seen as independent of the identity or personality of the researcher. *It does not matter who applies the scientific method*. The outcome will be the same as long as its standards and norms are followed correctly. All we have to do to accelerate scientific progress is to crank up the pressure and increase the number of scientists (Fig. 2).

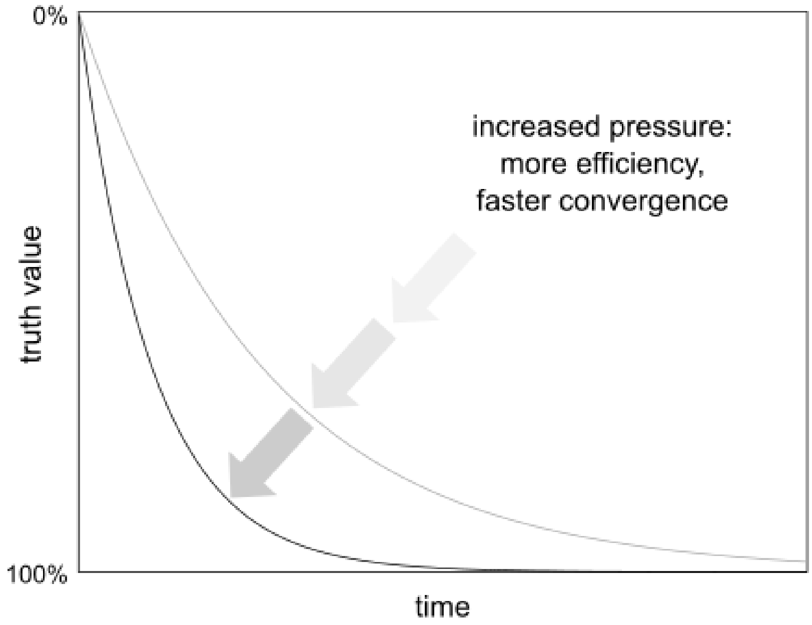

Figure 2: Naïve realism suggests that the universal scientific method leads to empirical knowledge that approximates a complete understanding of reality asymptotically (represented by exponential functions in this graph). Scientific progress does not depend in any way on the backgrounds, biases, or beliefs of researchers, which are filtered out by the proper application of the scientific method.According to this view, simply applying increased pressure to the research system should lead to more efficient application of the scientific method, and hence to faster convergence to the truth. See text for details.

This view has a number of profound implications:

- It sees researchers (once properly trained to adhere to the scientific method and ethos) as completely replaceable.
- It therefore fails to appreciate the diversity in researchers' experiences, motivations, interests, values, and philosophical outlooks.
- It leads to the idea that scientific inquiry can be optimized based solely on quantitative measurements of the productivity of individual researchers.

It is easy to see that all of these points are highly problematic, especially when considered in the context of democratic citizen science. A naïve realist is better off without it, since democratic citizen science values the individual's motivations and point of view, the diversity of citizen scientists, and takes into account a multiplicity of epistemic and non-epistemic goals beyond the efficient production of research output. All of these factors only slow traditional science down. Or do they?

In reality, the simplistic view of naïve realism outlined above leads to *a veritable cult of measurable productivity* (Muller, 2018), which is steering science straight into a game-theoretical trap. The short-term thinking and opportunism that is fostered in a system like this, where rare funding opportunities confer a massive advantage and heavily depend on a steady flow of publications with high visibility, severely limits creative freedom and prevents scientists from taking on high-risk projects. Ironically, this actually diminishes the productivity of a scientific community over the long term, since the process of scientific inquiry tends to get stuck in local optima within its search space. It lacks the flexibility to escape.

What we need to prevent this dilemma is a less mechanistic approach to science, an approach that reflects the messy reality of limited human beings doing research in an astonishingly complex world (Wimsatt, 2007). It needs to acknowledge that there is no universal scientific method. Scientific research is a creative process that cannot be properly formalised. Last but not least, scientific inquiry represents an evolutionary process combining exploitation with exploration that thrives on diversity (both of researchers and their goals). Not just citizen science, but science in general, deserves an updated epistemology that reflects all of these facts. This epistemology needs to be taught to scientists and the public alike, if we are to move beyond naïve realism and allow democratic citizen science to thrive.

Box 2: Beyond naïve realism

Our criticism of naïve realism and measurable productivity allows us to reflect on our practices within the Crowd4SDG project.

- Neither positivism nor falsificationism give us a clear understanding of how hypotheses are generated. For this reason, we will collaborate with the consortium partners to raise awareness as to how the methodology we propose to participants for coming up with the initial problem statement influences the solution design space. Specifically, we will stress that a project's design has a direct link to its potential for novelty, and will raise awareness about biases that may limit perspectives.
- The Crowd4SDG consortium has taken action from the start to prevent a unidimensional focus on productivity. Data generated by the tools developed by the partners, such as SDG in Progress and Decidim4CS, will be leveraged to assess the diversity in participants' experiences, motivations, and needs.
- The consortium plans to collaboratively revise metrics and descriptors to assess the projects at different phases of a GEAR cycle.

Our recommendation for future GEAR cycles of Citizen Science project innovation is to integrate the feedback of citizens participating in the program as part of the evaluation framework through a co-design strategy (see Section 5 "Science as Deliberation" below), and to assess the level and quality of outreach of these teams to impacted citizens, for example by tracking deliberations on Decidim4CS.

3. Science in perspective

The first major criticism that naïve realism must face is that there is no formally definable and universal scientific method. Science is quite obviously a cultural construct in the weak sense that it consists of practices that involve the finite cognitive and technological abilities of human beings which are firmly embedded in a specific social and historical context. Stronger versions of *social constructivism*, however, go much further than that. They claim that science is *nothing but* social discourse (see Zammito, 2004, for an historical overview). This is a position of *relativism*: it sees scientific truths as mere social convention, and science as equivalent to any other way of knowing, like poetry or religion, which are simply considered different forms of social discourse. We find this strong constructivist position unhelpful. In fact, it is just as oversimplified as the naïve realist stance. Clearly, science is neither purely objective nor purely culturally determined.

Perspectival realism (Giere, 2006; Massimi & McCoy, 2019; Wimsatt, 2007) and, similarly, *critical realism* (Bhaskar, 1975), provide a middle way between naïve objectivist realism and strong forms of social constructivism. These kinds of non-naïve realism hold that there is an accessible reality, *a causal structure of the universe*, whose existence is independent of the observer and their effort to understand it. Science provides a collection of methodologies and practices designed for us to gain trustworthy knowledge about the structure of reality. At the same time, perspectival realism also acknowledges that we cannot step out of our own heads: it is impossible to gain a purely objective "view from nowhere" (Anderson, 2017). Our access to the world, at all levels—from the individual researcher to the scientific community to the whole of society and humanity—is fundamentally biased and constrained by our cognitive and technological abilities, which we exercise under particular social and historical circumstances.

Each individual and each society has its unique perspective on the world, and these perspectives *do* matter for science. To use Ludwik Fleck's original terms, every scientific community is a Denkkollektiv (*thought collective*) with its own Denkstil (*thought style*), which circumscribes the type and range of questions it can ask, the methods and approaches it can employ, and the kinds of explanations it accepts as scientific (Fleck, 1935). All of these aspects of inquiry have changed radically, time and again, throughout the history of philosophy and science, the most famous example being the transition from Aristotelian to Cartesian and then Newtonian styles of inquiry during the Scientific Revolution (see the open-source book by Barseghyan et al., 2018 for an excellent overview). Our Denkstil is likely to evolve further in the future. In other words, there is no way to define science, or the scientific method, in a manner which is independent of social and historical context. Scientific inquiry is not formalisable in this way, and it never will be.

At this point, it is important to note that perspectives are not arbitrary opinions or points of view. Perspectivism is *not* relativism (see also Yucel, 2018). Instead, perspectives must be justified. This is the difference between what Richard Bernstein (1989) has called *flabby* versus *engaged pluralism*. In the words of philosopher William Wimsatt, perspectives are "intriguingly quasi-subjective (or at least observer, technique or technology-relative) cuts on the phenomena characteristic of a system" (2007, p. 222). They may be limited and context-dependent. But they are also grounded in reality. They are not a bug, but a central feature of the scientific approach. Our perspectives are what connects us to the world. It is only through them that we can gain any kind of access to reality at all (Polanyi, 1958). Popper was right in saying that it is impossible to obtain absolutely certain empirical facts. Our knowledge is always fallible. But we can still gain empirical knowledge that is sound, robust, and trustworthy (up to a certain degree) (Massimi, 2018; Wimsatt, 2007). In fact, science gives us knowledge of the world that is *more* robust than what we get from other ways of knowing. That is precisely its purpose and societal function. Let us elaborate a bit more.

If scientific inquiry is not a purely formal activity, then scientific methods do not work like algorithms which are guaranteed to yield an ever-closer approximation to reality, no matter who is using them. Real science, performed by real scientists, does not actually aim to come up with a perfect explanation of everything. Instead, researchers make use of *imperfect(ible) heuristics*—fallible short-cuts, improvisations that solve scientific problems (most of the time) in specific areas under specific circumstances (Giere, 1988; Wimsatt, 2007). Herbert Simon called this *satisficing* (Simon, 1955, 1969): heuristics are not perfect, but they help us achieve our epistemic goals within a reasonable amount of time, energy, and effort. This is a pragmatic view of science.

Yet, science is not just problem-solving either. As Aristotle already recognised, the ultimate goal of inquiry is to supply us with a structured account of reality (see Kitcher, 1992, for a contemporary discussion of this issue). This is possible, but not as easy as a naïve realist might think. Being good at solving a problem does not automatically imply that a heuristic also teaches us something about the structure of reality. It could work for all the wrong reasons. How can we find out whether we are deceiving ourselves or not? In order to do this, we need to assess the *robustness* (or *soundness*) of the knowledge that a heuristic produces in a given context. Remember that empirical insights are never absolutely certain, but they can be robust if they are "accessible (detectable, measurable, derivable, definable, producible, or the like) in a variety of independent ways" (Wimsatt, 2007, p. 196). It *is* possible to estimate the relative robustness of an insight—what we could call *perspectival truth*—by tracking its invariance across perspectives, while never forgetting that the conditions that make it true always depend on our own present circumstances (Massimi, 2018). Thus, multiple perspectives enhance robust insight, and a multiplicity of perspectives is what democratic citizen science provides. It is by comparing such perspectives that science provides trustworthy knowledge about the world—not absolutely true, but as true as it will ever get.

Having multiple perspectives becomes even more important when we are trying to tackle the astonishing complexity of the world. Science is always a compromise between our need for simple explanations that enable understanding, and the unfathomably complex causal structure of reality, especially in areas such as the life sciences (including ecology), or social sciences such as psychology, sociology, and economics (see, for example, Potochnik, 2017). Perspectival realism frees us of the naïve realist idea that science must provide a single unified account of reality—a theory of everything. As a matter of fact, unified accounts are only possible for simple systems. In contrast, complex systems (in a perspectival sense) are *defined* by the number of distinct valid perspectives that apply to them (Wimsatt, 2007). A complex system is not just a complicated mechanism, like a clockwork or a computer. The more valid perspectives, the more complex the system. Climate resilience is an excellent example of a scientific problem that is incredibly complex in this way, since a full understanding of its causes and consequences requires insights from a variety of actors (researchers, farmers, policy makers, technologists, and impacted populations), and from a variety of fields, ranging from biogeochemistry, ecology, agriculture and hydrology to economics and other social sciences. Without such diverse perspectives there can be no understanding. Democratic citizen science can be an essential tool to provide more diversity, and thus more robustness in climate research.

Finally, diversity of perspectives lies at the very heart of scientific progress itself. Such progress can occur in two qualitatively different ways: as the "normal" gradual accumulation and revision of knowledge, or in the form of *scientific revolutions* (Kuhn, 1962). In this context, it is important to notice that when a new discovery is made, the resulting insight is never robust at first (Wimsatt, 2007). Its soundness must be gradually established. This is where Merton's universal skepticism reaches its limitations: if applied too stringently to new insights, it can stifle innovation. As a new insight becomes accepted, other scientific theories may be built on top of it through a process called *generative entrenchment* (ibid.). The more entrenched an insight, the more difficult it becomes to revise without bringing down the growing theoretical edifice that is being built on its foundation. For this reason, entrenched insights should ideally also be robust, but this is not always the case. Scientific revolutions occur when an entrenched but fragile insight is toppled (Kuhn, 1962; Wimsatt, 2007). Classic examples are the assumptions that space and time are pre-given and fixed, or that energy levels can vary continuously. The refutation of these two entrenched yet fragile assumptions led to the twin revolutions of relativity and quantum mechanics in early 20th-century physics (see Barseghyan et al., 2018, for a recent review).

As we construct and expand our scientific knowledge of the world, more and more insights become robust and/or entrenched. At the same time, however, errors, gaps, and discrepancies accumulate. The detection of patterns and biases in those flaws can greatly facilitate scientific progress by guiding us towards new problems worthy of investigation. Wimsatt (2007) calls this the *metabolism of errors*. Basically, we learn by digesting our failures. For this to work properly, however, we need to be allowed to fail in the first place (see Firestein, 2015). And, yet again, we depend on a multiplicity of perspectives. To detect biases in our errors, we require a disruptive strategy that allows us to "step out" of our own peculiar perspective, to examine it from a different point of view. This is only possible if alternative perspectives are available. Scientific progress is catalysed by diversity in ways which a naïve realist cannot even begin to understand.

In summary, we have shown that the diversity of perspectives is essential for the progress of science and for the robustness of the knowledge it generates. This diversity of perspectives, in turn, depends on the diversity of individual backgrounds represented in the communities involved in designing, managing, and performing research. Of particular importance in this regard are individuals with a personal stake in the aims of a scientific project. Their perspectives are privileged in the sense of having been shaped by personal experience with the problem at hand, in ways which are inaccessible to a neutral observer. Such engaged perspectives are called *standpoints* (Haraway, 1988; Harding, 1995; Hartsock, 1983). Each individual standpoint can broaden the scope and power of the cognitive and technological tools being brought to bear on an issue. This is particularly important in the context of climate resilience, where local experiences and challenges must be considered as an essential part of any problem solution. Being engaged (contra Merton's principle of disinterestedness) is positive in this context, since it enhances scientific inquiry by promoting robustness and applicability, and democratic citizen science becomes a necessary condition for the production of adequate solutions. Therefore, it is of utmost importance that the relevant stakeholders are recognised and properly represented in the research process.

Box 3: Science in perspective

When considering the problem of climate resilience, the diversity of engaged standpoints within Citizen Science projects is of particular importance: local experiences and challenges need to be considered as an essential part of envisioned solutions. This translates into specific actions and evaluation schemes:

- The diversity of participants (countries, gender, disciplinary backgrounds) will be promoted through the use of outreach platforms such as Goodwall (going beyond academic institutions) and the constitution of novel teams during the so-called Gather phase of the GEAR cycles, where participants are recruited and teams are formed.
- To assess the outreach and integration of diverse standpoints at the level of the citizen science projects forged during future GEAR cycles, we will collect engagement data from tools developed by Crowd4SDG, such as Decidim4CS, from Google Analytics data (geolocation, gender) on the usage of the platforms used by the projects, and from surveys on citizen engagement within the projects.
- Beyond diversity metrics, in other words whether a diverse array of standpoints was represented, we will also assess the strength of the diverse contributions through involvement in the deliberation process, using tools such as Decidim4CS and Slack, as well as through surveys.

As such, the diversity of participants will be examined both at the global level of the Crowd4SDG project, but also within the projects generated within GEAR cycles, through profile information and contribution levels.

4. Science as process

The second major criticism that naïve realism must face is that it is excessively focused on research outcomes, thereby neglecting the intricacies and the importance of *the process of inquiry*. Basically, looking at scientific knowledge only as the product of science is like looking at art in a museum. However, the product of science is only as good as the process that generates it. Moreover, many perfectly planned and executed research projects fail to meet their targets, but that is often a good thing: scientific progress relies as much on failure as it does on success (see section 3). Some of the biggest scientific breakthroughs and conceptual revolutions have come from projects that have failed in interesting ways. Think about the unsuccessful attempt to formalise mathematics, which led to Gödel's Incompleteness Theorem (Nagel & Newman, 2001), or the scientific failures to confirm the existence of phlogiston, caloric, and the luminiferous ether, which opened the way for the development of modern chemistry, thermodynamics, and electromagnetism, respectively (Barseghyan et al., 2018). Adhering too tightly to a predetermined research plan can prevent us from following up on the kind of surprising new opportunities that are at the core of scientific innovation. Research assessment that focuses exclusively on deliverables, and outcomes and does not integrate considerations about the process of inquiry, can be detrimental to scientific progress.

Sometimes, and especially in democratic citizen science, the goal *is* the journey. Such citizen science projects put a strong emphasis on facilitating their participants' individual learning and their inclusion in the process of inquiry at the level of the research community. Furthermore, the problems of how to manage collaborations, data sharing, and quality control are no longer peripheral nuisances, but become themselves a central part of the research focus of the project. Democratic citizen science is as much an inquiry into the natural world, as it is an inquiry into how to best cultivate and utilize humanity's collective intelligence (see Nielsen, 2011). The most valuable outcome of a citizen science project may very well be an improved learning and knowledge-production process. We now turn our attention to this dynamic. In this section, we look at the cognitive activities and research strategies that individual researchers use to attain their epistemic goals. The role of interactions among scientists and their communities will be the topic of section 5.

The first thing we note is that scientific knowledge itself is not fixed. It is not a simple collection of immutable facts. The edifice of our scientific knowledge is constantly being extended (Wimsatt, 2007). At the same time, it is in constant need of maintenance and renovation. This process never ends. For all practical purposes, the universe is cognitively inexhaustible (Rescher, 1996, 2009). There is always more for us to learn. As finite beings, our knowledge of the universe will always remain incomplete. Besides, what we can know (and also what we want or need to know) changes significantly over time (Barseghyan et al., 2018). Our epistemic goalposts are constantly shifting. The growth of knowledge may be unstoppable, but it is also improvised and messy—anything but the straight line of naïve realism depicted in Fig. 2.

Once we realise there is no universal scientific method, and once we recognise the shifting nature of our epistemic goals, the process of knowledge production becomes an interesting and intricate object of study in itself. The aim of our theory of knowledge must adapt accordingly. Classic epistemology, going back to Plato and his dialogue "Theaetetus" (Chappell, 2021), considered knowledge in an abstract manner as "justified true belief," and tried (unsuccessfully) to find universal principles which allow us to establish it beyond any reasonable doubt (Gettier, 1963; Hetherington, 2016). *Naturalistic epistemology*, in contrast, aims to understand the epistemic quality of actual human cognitive performance (Kitcher, 1992). It asks which strategies we—as finite beings, in practice, given our particular circumstances—can and should use to improve our cognitive state: what are the processes that robustly yield reliable and relevant knowledge about the world? The overall goal of naturalistic epistemology is to collect a compendium of *cognitively optimal processes* that can be applied to the kinds of questions and problems humans are likely to encounter. This is a much more modest and realistic aim than any quixotic quest for absolute knowledge, but it is still extremely ambitious. Like the expansion of scientific knowledge, it is a never-ending process of iterative and recursive improvement—an ameliorative instead of a foundationalist project (ibid.). We must always build on the imperfect basis of what we have already constructed.

Just like scientific perspectivism, naturalistic epistemology leads to context-specific strategies that allow us to attain a set of given epistemic goals. What is important in the context of our discussion is that different cognitive processes and research strategies will be optimal under different circumstances. There is no universally optimal search strategy for inquiry (or anything else)—there is no free lunch (Wolpert & Macready, 1997). What approach to choose depends on the current state of knowledge and level of technological development, the available human, material, and financial resources, and the epistemic goals of a project. These goals may be defined in terms of solving a particular problem, in terms of providing new insights into the structure of reality, and/or in terms of optimising the research process itself. Choice of strategy is in itself an empirical question. Naturalistic epistemology must be based on history and empirical insights into error-prone heuristics that have worked for similar goals and under similar circumstances before (Kitcher, 1992). We cannot justify scientific knowledge in a general way, but we can get better at appraising its epistemic value by studying the process of inquiry itself, in all its complexity.

One central insight from this kind of epistemology, which is supported by empirical and theoretical evidence, is that evolutionary search processes such as scientific inquiry are subject to what Thomas Kuhn (1977) has called *the essential tension* between a productive research tradition and risky innovation, which has since been recast and popularized as the strategic balance between *exploration* (gathering new information) and *exploitation* (putting existing information to work) (for an accessible introduction, see chapter 2 of Christian & Griffiths, 2016). It is important to note, however, that we are not really talking about a balance in the sense of an equilibrium here. The optimal ratio between the two strategies cannot be precisely computed for an open-ended process with uncertain returns such as scientific inquiry (ibid.). Instead, we need to switch strategy based on local criteria and incomplete knowledge. The situation is far from hopeless though since some of these criteria are known. For instance, it pays for an individual researcher, or an entire research community, to explore at the onset of an inquiry. This happens at the beginning of an individual research career, or when a new research field opens up. Over time, as information accumulates, exploration yields diminishing returns. At some point, it is time to switch over to exploitation. Imagine moving to a new city. Initially, you will explore new shops, restaurants, and other venues, but eventually you will settle down and increasingly revisit your favorite places. This is an entirely rational meta-strategy, inexorably leading people (and research fields) to become more conservative over time (see Fortunato et al., 2018; Wu et al., 2019, for evidence on this in science).

Here, we have an example where the optimal research strategy depends on the process of inquiry itself. A healthy research environment provides scientists with enough flexibility to switch strategy dynamically, depending on circumstances. Unfortunately, industrial science does not work this way. The fixation on short term performance, measured through output-oriented metrics, have locked the process of inquiry firmly into exploitation mode. Put differently, exploring does not pay off in such a system. Exploration requires time, effort, and a willingness to fail. It may be bad for short-term productivity, but is essential for innovation in the long run. This is the game-theoretic trap we discussed in section 2. It is sustained by the view that the attainment of the epistemic goals of science can be accelerated by maximising research output.

In this section, we have argued that naturalistic epistemology, an empirical investigation of the process of inquiry itself, could lead us out of this trap. But it is not enough. We also need a better understanding of the social dimension of doing science, which is what we will be discussing next.

Box 4: Science as process

In the Crowd4SDG context, the generated early phase projects are exploratory by design. The tracking of the process of exploration is possible with the use of the Crowd4SDG tool SDG in progress. In this platform, the project is documented through bricks corresponding to outputs and milestones achieved during the project course. These bricks can be commented on and re-used by other projects to foster open innovation.

In the context of Crowd4SDG, we will analyse the use of the SDG In Progress platform for project documentation. We will monitor platform usage data (in particular evolution of number of bricks and comments) and interrogate its relation with the originality of the solution provided (evaluation grid) as well as with self-reported assessment of the ability to foster and structure the exploration process through surveys.

As such, the enquiry process will be examined within the projects generated during the GEAR cycle 2, through documentation data and surveys.

5. Science as deliberation

The third major criticism that naïve realism must face is that it is obsessed with consensus and uniformity. Many people believe that the authority of science stems from unanimity, and is undermined if scientists disagree. Ongoing controversies about climate science or evolutionary biology are good examples of this sentiment (see, for example, Melo-Martín & Intemann, 2018). To a naïve realist, the ultimate aim of science is to provide a single unified account—an elusive theory of everything—that most accurately represents *all* of reality. This kind of thinking about science thrives on competition: let the best argument (or theory) prevail. Truth is established by debate, which is won by persuading the majority of experts and stakeholders in a field that some perspective is better than all its competitors. As Robert Merton (1973) put it: competing claims get settled sooner or later based on the principle of universalism. There can only be one factual explanation. Everything else is mere opinion.

However, there are good reasons to doubt this simplistic view. In fact, uniformity can be pernicious (Beatty & Alfred, 2010). This is because all scientific theories are *underdetermined by empirical evidence*. In other words, there is always an indefinite number of scientific theories able to explain a given set of observed phenomena. For most scientific problems, it is impossible to unambiguously settle on a single best solution based on evidence alone. Even worse: in most situations, we have no way of knowing how many possible theories there actually are. Many alternatives remain unconsidered (Stanford, 2010). Because of all this, the coexistence of competing theories need not be a bad thing. In fact, settling a justified scientific controversy too early may encourage agreement where there is none (Beatty & Alfred, 2010). It certainly privileges the status quo, which is generally the majority opinion, and it suppresses (and therefore violates) the epistemic equality of those who hold a minority view that is not easy to dismiss (ibid.). In summary, too much pressure for unanimity leads to a dictatorship of the majority, and undermines the collective process of discovery within a scientific community.

Let us take a closer look at what this process is. Specifically, let us ask which form of information exchange between scientists is most conducive to cultivating and utilizing the collective intelligence of the community. In the face of uncertainty and underdetermination, it is *deliberation,* not debate which achieves this goal (Beatty & Alfred, 2010). Deliberation is a form of discussion that is based on dialogue, rather than debate. The main aim of a deliberator is not to win an argument by persuasion, but to gain a comprehensive understanding of all valid perspectives present in the room, and to make the most informed choice possible based on the understanding of those perspectives (see, for example, Bone et al., 2006). What matters most is not an optimal, unanimous outcome of the process, but the quality of the process of deliberation itself, which is greatly enhanced by the presence of non-dismissible minorities. As Popper already pointed out, the quality of a scientific theory increases with every challenge it receives. Such challenges can come in the form of empirical tests, or thoughtful and constructive criticism of a theory's contents. The deliberative process, with its minority positions that provide these challenges, is stifled by too much pressure for a uniform outcome. As long as matters are not settled by evidence and reason, it is better—as a community—to suspend judgment and to let alternative explanations coexist.

It is not difficult to see how deliberation—with its choice-making based on the understanding of multiple perspectives—is particularly important for interdisciplinary and transdisciplinary projects. Such projects boost scientific innovation when they manage to integrate different perspectives into a cohesive solution (reviewed in Fortunato et al., 2018). They help science break out of the inexorable tendency of research fields to become more conservative over time (see section 4). They are key to generate and enhance epistemic exploration. But, like other exploratory processes, they need time and effort to establish. Deliberative processes cannot be rushed. To integrate them into our research environment, we need to assess their quality directly. In the context of Crowd4SDG, this is done through the use and further development of decidim4CS as a digital deliberation platform (WP2 ; https://decidim4cs.iiia.csic.es).

Deliberative processes that facilitate collective intelligence work best with relatively small groups of deliberators, each with an engaged and non-dismissible standpoint on the matter at hand. However, many scientific projects—especially those of democratic citizen science—require human and material resources that go beyond the capabilities of small groups. This is particularly relevant in the field of climate resilience, where the number of impacted citizens reaches the planetary scale. In such cases, the deliberation process needs to be based on a suitable community structure in order to scale. This is why an increasing amount of science is done by teams (Fortunato et al., 2018). There is empirical evidence that small teams of investigators are more innovative than isolated individuals or large-scale consortia (Wu et al., 2019). This is because they strike a delicate balance between a diversity of standpoints and the ability of its members to productively engage in deliberation. The deliberative process can then be rescaled as an interaction between teams, resulting in a hierarchy of interactions that enable collective intelligence at multiple levels. This is an area of investigation that needs much more attention than it currently receives.

Box 5: Science as deliberation

In the context of Crowd4SDG, the deliberation process will be supported by the dedicated Decidim4CS tool, as well as using the Slack communication platform for community discussion.

The digital traces of these platforms allow to gather insights on team communication and deliberation dynamics. During the first GEAR cycle, we have shown how Slack communication data allows us to evaluate the dynamics of inter- and intra-team communication, as well as interactions with the organizing team. This is summarized in D4.3. In future GEAR cycles, the use of the Slack platform will be harmonized across the phases of the cycle in order to facilitate data analysis throughout the program.

In the context of the upcoming GEAR cycles, we will complement this analysis with engagement data from Decidim4CS to assess the deliberation process. The platform has been in development in the first year of the project. We will study its usage in future GEAR cycles in order to foster its ability to engage all relevant stakeholders, and monitor the deliberation process to complement the analysis of communication channels from Slack.

6. An ecological vision for Citizen Science

In sections 3–5, we have outlined the three main pillars of an emerging epistemology that is tailored to the needs of democratic citizen science, but is applicable to all academic research. We see the kind of citizen science it envisions as paradigmatic for a more participatory research environment, adequate for the complex planetary-scale problems humanity is facing today. Its highest aim is to foster and put to good use the collective intelligence of humanity. In order to achieve this, *we need research communities that are diverse, engaged, representative, and democratic*. What we propose here is an "ecological" vision for a science which supports diversity, inclusion, and deliberation. This vision stands in stark contrast to our current industrial model of doing science (see section 1). The two approaches are compared in Table 1. Note that both models are highly idealised. They represent different ideals of how research ought to be done—two alternative ethos for science.

"Industrial Science"	"Ecological Science"
Control	Participation
Maximized output	Maximised reproducibility
Competitive & closed	Open & collaborative
Intellectual monoculture	Diversified perspectives
Risk averse/exploitative	Open to exploration
Fixated on short-term optimisation	Focus on long-term progress

Table 1: Two idealised models for scientific research. This table compares different emphases exhibited by "industrial science" vs "ecological science." Note that both visions represent ideals, which are rarely attainable in practice. Most scientific projects will come to lie somewhere along the spectrum between these two extremes. See text for details.

We have argued that the naïve realist view of science is not, in fact, realistic at all. In its stead, we have presented an epistemology that adequately takes into account the needs and capabilities of limited human beings, solving problems in a world of planetary-scale complexity. The ecological research model proposed here is less focused on direct exploitation, and yet, it has the potential to be more productive in the long-term than the current industrial system. However, its practical implementation will not be easy, due to the game-theoretic trep we have maneuvered ourselves into (see section 2). Escaping this trap requires a deep understanding of the social and cognitive processes that enable and facilitate scientific progress for all. Finding such processes is an empirical problem, which is only beginning to be tackled and understood today. The Crowd4SDG project is just one example of such an empirical investigation, which must be grounded in a suitable epistemological framework, and a correspondingly revised ethos of science, able to provide philosophical and ethical guidance for our attempts to improve our methods for scientific project management, monitoring, and evaluation through experience and experimentation. These methods must acknowledge the contextual and processual nature of knowledge-production. They need to focus directly on the quality of this process, rather than being fixated exclusively on the outcome of scientific projects. They need to encompass multiple levels—from the individual investigator to their research community to the context of society in general. And they need to account for a diversity of epistemic goals.

Unfortunately, such explorative efforts are likely to fail unless we break out of the restrictive framework we have built around ourselves through an ever stronger focus on measuring research output, detached from any consideration of the cognitive and deliberative processes that generate it. Before we can achieve anything else, we must use our new appreciation of the process of inquiry to move beyond our metric fixation, beyond the cult of productivity (Muller, 2018). As a first step, this requires a broader awareness of the underlying philosophical issues. While the epistemological arguments we have presented here are well-known among philosophers of science, they are virtually unheard of among practicing scientists, science stakeholders, and the general public. This urgently needs to change before we can have the kind of conversations that lead to sustainable changes in mindset and policy. Democratic citizen science is one of the most important initiatives towards increasing diversity, representation, and participation in science today. In addition, it is one of the main sources for new insights into the process of inquiry, and its process-oriented assessment. For these reasons, citizen science must play a key role in the upcoming transition from an industrial to an ecological model of doing research. In the final section of our paper, we will discuss the kind of measures we could experiment with to improve the assessment of citizen science projects along the lines of the philosophical argument we have presented above.

7. Beyond metric fixation: implications for project evaluation

Our philosophical analysis points to a central conclusion: any proper evaluation of a scientific project must include *an epistemic appraisal of its process of inquiry*, including an assessment of the material, cognitive, deliberative, and organizational practices involved in knowledge production. It is not enough to judge a project by its outcome alone—the number of scientific publications it has produced, let's say, or the amount of factual knowledge its participants are able to regurgitate at a final debrief or exam. This central insight underlies a recently proposed multidimensional evaluation framework for citizen science projects, which makes a fundamental distinction between process-based and outcome-based aspects of assessment (Kieslinger et al., 2018; Schaefer et al., 2021). It identifies three core dimensions to citizen science: scientific, participant, and socio-ecological/economic. For each of these, it defines criteria of evaluation concerning *both* aspects of "process and feasibility" as well as "outcome and impact" (Fig. 3). Such a framework can not only be applied to strategic planning, the selection of specific projects to be funded, and impact assessment after a project is finished, but also to monitor and, at the same time, to mentor participants and facilitate the progress of a project while it is running. Evaluation itself becomes a learning process—learning about learning—that supports participatory self-reflection and adaptive management practices (Schaefer et al., 2021).

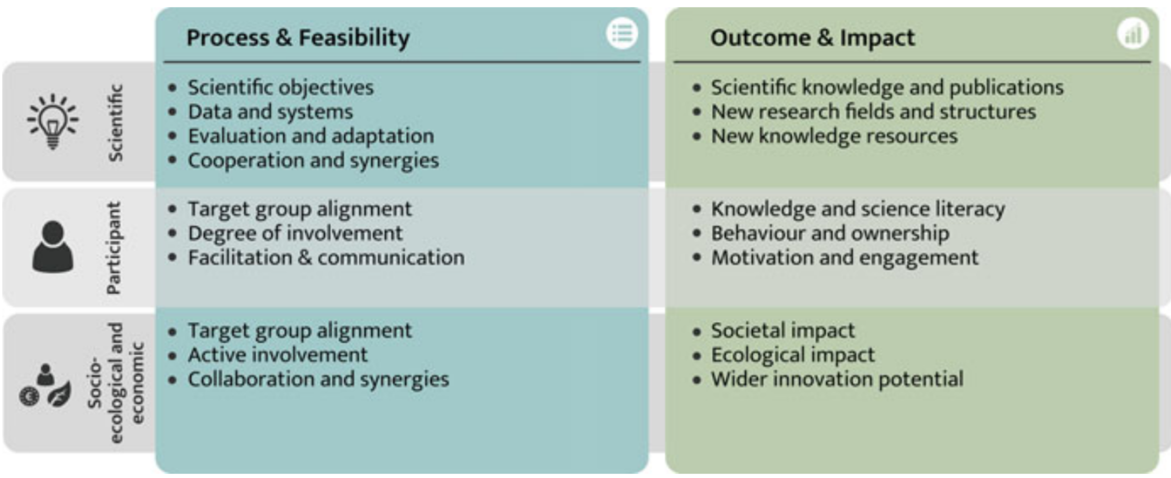

Figure 3: An assessment framework for democratic citizen science. From Schaefer *et al.*, 2021.

Due to the epistemological nature of our argument, we focus mainly on the scientific-knowledge dimension of this evaluation framework here, although epistemic processes underlying individual and collective learning and their wider societal and ecological impact are also subjects highly deserving of closer philosophical attention. In this context, it is worth mentioning again that not all citizen science projects have their main focus on the production of new scientific knowledge. Non-epistemic goals—changes in individual attitudes and behavior, cultural practices, or policies, for example—can be equally or even more important in some cases. For this reason, the evaluation framework in Fig. 3 is designed to be flexible and adaptive in terms of weighting different criteria. Moreover, while we limit our discussion to process-based aspects of scientific knowledge production, we do not want to leave the impression that evaluation of outcome is unimportant. *Both* aspects need to be considered together. What we do want to do here is to highlight the fact that process-based evaluation remains undervalued and underdeveloped in the current system of academic research. Our analysis provides epistemological reasons for addressing this problem. Developing adequate approaches to process-based assessment requires an improved understanding of suitable practices of individual and community-level knowledge production that can actually be carried out in today's research environment.

Beyond emphasising processual and participatory methods of evaluation, there is another fundamental point that arises from our analysis: many of the features that make democratic citizen science (and science in general) worthwhile and productive are impossible to capture by standardised metrics. For example, the originality, relevance, and value of a scientific insight cannot be quantified objectively, because notions of "originality," "relevance," and "value" contain fundamentally subjective and radically context-sensitive facets that are crucial to their meaning. Similarly, there is no standardised algorithm to assess the robustness or soundness of a piece of scientific knowledge. Instead, proper robustness analysis requires a careful comparison of scientific perspectives and an assessment of their independence from each other, which cannot be done without deep insight into the research topic and all the approaches that are being compared (Wimsatt, 2007). Standardised measures can support, but never fully replace judgment based on experience. Similarly, there is no metric for the generalisability or the adaptiveness of a scientific result. The range of circumstances under which some theory or insight may be usefully applied is impossible to predict, or even prestate (Kauffman, 2016; Rescher, 2009). Discovery cannot be planned in this sense. Much of scientific inquiry is driven by serendipitous coincidences, historical accidents, which cannot be captured by any predictive measure based on past evidence alone.

Thus, discovery cannot be forced, but it can be facilitated by providing an environment that is conducive to it. Our epistemological framework implies that this can be achieved by incentivizing collaborative processes and deliberation based on a diversity of standpoints. Obviously, this same argument also applies to the assessment of the wider socio-ecological implications of a project, its stakeholder engagement, its social embeddedness, and so on (Schaefer et al., 2021). Each research project should be assessed under consideration of its particular scientific and societal circumstances, as well as its particular epistemic and non-epistemic goals. Even so, much of its value will only become evident in hindsight. Trying to define one-size-fits-all metrics or numerical indicators for qualities such as originality, relevance, robustness, adaptedness, or generalizability is bound to be counterproductive, because each and every scientific project, and the knowledge it generates, is different. Generalised abstraction ignores situation-dependent nuances, which may be essential for the success of a project, which can only be assessed qualitatively and in retrospect.

Finally, there is another problem that arises in systems where rewards and punishments no longer depend on professional judgment—based on personal experience, honesty, dedication, and talent—but on quantitative indicators implemented as standard metrics of comparative performance. Such systems become vulnerable to *metric gaming* (Muller, 2018). When a metric becomes the target of the measured system, Goodhart's Law applies, which states that such metrics are no longer good indicators for the system's original purpose. Efforts become channeled into optimizing performance as measured by the metric, often in ways that are not conducive to the system's wider goals. This happened, for example, to the U.S. school system after the introduction of standardised testing, which led to widespread teaching to the test (ibid.). Similarly, surgeons who are rated on the number of their successful operations often refuse to take on difficult cases (ibid.).

Metric gaming is also taking over the academic research system, where an unhealthy fixation on publication metrics leads to risk avoidance and the short-term optimization of personal research output to the detriment of community-level, long-term progress. Somewhat, ironically, this trend *is* measurable: while the content of individual scientific publications is progressively diminishing, approaching what has been called the minimal publishable unit of information, the number of authors per paper is rapidly increasing (see, for example, Fire & Guestrin, 2019). These trends are empirical signs of an academic system that is being manipulated. Such a system no longer rewards those who do the best work, but those who are most efficiently gaming the metric.

All of this poses a formidable challenge for scientific project evaluation. On the one hand, we really do need methods to compare the quality of scientific projects: how else are we going to implement a fair and rigorous system for strategic planning, funding, monitoring, and assessment in research? On the other hand, we know that the value of a scientific project is radically context-dependent, and that standardised metrics make a system vulnerable to being gamed. As we have seen in section 3, this does not necessarily have to lead us into relativism, considering any project as good as any other. There *are* criteria by which we can assess the promise and importance of a project, or the robustness of the knowledge it produces. What we need then, if we want to adopt an ecological model of citizen science, is an approach to evaluation, grounded in a perspectival, naturalistic, deliberative epistemology that is flexible and adaptable to the specific needs and circumstances at hand, and yet rigorous in its approach to epistemic appraisal.

A first step towards such an approach is to overcome our current *metric fixation* (Muller, 2018). Instead of being based on a set of fixed standards and metrics, project assessment should be grounded on shared values and procedures, themselves constantly subject to evaluation. To attain that goal, scientific assessment should not only evaluate the quality of the deliberative process of inquiry, but must itself become a deliberative, participatory, and democratic process.

Second, we need to carefully choose appropriate procedures to evaluate both quantifiable and non-quantifiable aspects of a project, and how they compare with alternative approaches in terms of achieving its specific goals. These procedures should be adapted to context, transparent, flexible, and they should include an element of self-evaluation. One suitable model is *co-evaluation* (Mayer et al., forthcoming), an approach to assessment that includes all actors involved in or affected by a project in an iterative process and is based on methods from participatory action research. On top of this, there must be a meta-level process that evaluates the evaluators as they assess a project, guided by deliberative procedures. Finally, assessment must include an evaluation of the quality of this deliberative process itself. More than resembling the hierarchical mechanism of a clockwork, this method of project assessment imitates the self-regulatory and homeostatic dynamics of a living organism.

8. Implications for Crowd4SDG

In the previous sections, we have exhibited the benefits of moving towards perspectival realism and focusing on the research process. We have highlighted three fundamental pillars supporting an ecological approach to citizen science: the diversity of standpoints involved, their inclusion in the research process, and the deliberation by which they dialogue towards shared solutions. These matter both to the Crowd4SDG project as a whole, and at the level of Citizen Science projects being generated by the Crowd4SDG innovation cycles.

In this section we focus on implications for Citizen Science projects being generated by the Crowd4SDG, the evaluation of the Crowd4SDG project as a whole being dealt with in D3.4.

8.1. Operationalization of the epistemological analysis

We propose to operationalize the three epistemological pillars introduced in this report the following way:

Diversity of standpoints: in our context, a standpoint can be characterized by socio-economic and demographic variables. These consist for instance in information provided by participants at registration: the *disciplinary backgrounds* of participants (diversity of scientific perspectives), their *skills* (diversity of approaches), *gender*, *country of origin* (cultural diversity), *geographical location* (diversity of local experiences), and their *institutional affiliation* or lack thereof (diversity of work experiences). In addition, some of these observables can be obtained through the use of Google Analytics when considering the citizens engaged within a project through one of the proposed tools in the Citizen Science Solution Kit.

We note that these measures of diversity are promoted in the Crowd4SDG design, through the use of outreach platforms such as Goodwall (going beyond academic institutions) and the constitution of novel teams during the “gather” phase of GEAR cycle 2. In addition, the Crowd4SDG consortium has already worked intensely to consider and facilitate a diversity of perspectives for tackling climate change and urban resilience issues during GEAR cycle 1 by engaging with relevant stakeholders and authorities and with a wide and varied range of participants.

Inclusion in the research process: related to the former diversity measure, here we consider the *equality of representation* of the various standpoints involved. The level of inclusion of the various parties within a project can be assessed through the density and structure of the *interaction network* of project participants, using self-reports of collaborative tasks (through the CoSo “Collaborative Sonar” app) or communication data (examples are showcased in D4.3 using Slack communication data), and the *level of engagement* of participants in the research process, of relevant stakeholders, and the perceived inclusion of diverging standpoints within the project using surveys directed to participants during the GEAR cycle. Note that this information can also be extracted from the Decidim4CS platform of deliberation, which we elaborate below.

Deliberation process: we examine the quality and strength of engagement of the relevant stakeholder in the co-design process. To that end, Decidim4CS has been included as a deliberation tool in the citizen science toolkit for Crowd4SDG participants to leverage for their project. The data generated within Decidim4CS is of prime interest to assess the role of deliberation in the developed projects. Measures of the *number of citizens involved in the deliberation process*, *number of messages exchanged*, or *length of discussion threads* can be used to operationalise our assessment of the deliberation process. Though crude, such data is a first step in accounting for the complexity of the deliberation process, which could in the future incorporate deeper textual analyses out of the scope of the current measurement. Moreover, we note that such insights are fundamentally unique to the Crowd4SDG context, since deliberations happening in traditional research and innovation setups (through oral or written communication) are rarely visible, or if available are not accessible in a standard format and based on the original account of authors (*e.g.* written correspondences).

8.2. Assessment of project success

In this report, we showcase the limits of predetermined metrics when these don't align with the goal of the project nor take into account its context. Crowd4SDG initially proposed that a democratic citizen science approach fosters the originality, relevance, robustness, adaptedness, or generalizability of the research outcomes. The current assessment grid used by Crowd4SDG judges at the end of each phase to select the projects that will move forward contains the following criteria: *novelty*, *relevance*, *feasibility*, and *crowdsourcing*. These can be used to partly operationalize the above criteria, for example originality and relevance. While robustness, adaptedness and generalizability necessitate a longer-term observation that is not directly measurable for the early stage projects considered, criteria of feasibility and crowdsourcing can provide information supporting their eventual achievement. In particular, we will provide a clearer description of the crowdsourcing outcome to suggest that it relates to engagement, inclusion, diversity, and the deliberative process.

Finally, as we have shown in the report, some criteria to quantify the "success" of a given project are highly context-dependent, and cannot therefore be fully determined in advance. As such, we propose to include the citizens partaking in the Crowd4SDG project in a deliberation process to propose ad-hoc measures of success or processes to assess it within the context of each particular GEAR cycle. For this, we propose to use Decidim4CS at the project level to develop a deliberative meta-process on metrics, descriptors, and procedures for assessing the projects at the different phases of the GEAR cycle and generating potentially overlooked measures.

8.3. Relevance for the Crowd4SDG consortium

Finally, our insights have a direct impact on the CrowdD4SDG project implementation.

We propose that a summary of this report should be shared with participants to stress how the methods employed by the Crowd4SDG teams have a direct impact on the project design and their potential for novelty, and raise awareness about biases that may limit a team's perspectives. We also want to provide an epistemological context for encouraging teams to iterate and prototype, a process that is already promoted by the coordinating team, which in turn fosters risk taking and increases the chances for elaborating innovative solutions.

In addition, in order to foster the inclusion of all actors, in particular vulnerable or marginalized populations, the coordinating team will work closely with the relevant local actors and organizations working on the topic of each GEAR cycle. Moreover, we encourage a higher involvement of local mentors to increase the local integration of solutions and their sustainability.

9. General conclusion

In this report, we have introduced an epistemological framework that will serve as the foundation not only for the development and adaptation of new metrics and descriptors for project evaluation in the GEAR cycles of the Crowd4SDG project (WP4, see section 8), but also for democratic citizen science and academic research in general. The framework is based on the three pillars of perspectival realism (section 3), process thinking in the form of natural epistemology (section 4), and deliberative practice (section 5), leading to what we have called an "ecological model" of doing research (section 6). Perspectivism implies that the range of backgrounds and motivations of individual researchers in a community greatly influences the kind of questions that can be asked, the kind of approaches that can be used, and the kind of explanations that are accepted in a given research and innovation field. Naturalistic epistemology focuses our attention on the quality of the cognitive processes leading to a given research output, while deliberative practice emphasises the community-level social dynamics that are required to enable collective intelligence. Together, these pillars lead to a new research and innovation ethos that values diversity, inclusion, and good communication much more than the traditional Mertonian approach to science (see sections 2 and 6).

We have described the implicit amalgamation of positivist, Popperian, and Mertonian ideas in the minds of scientists and stakeholders as "naïve realism" (section 2). It could be argued, though, that our own vision of democratic citizen science is itself naïve. In fact, Mirowski (2018) has characterised open science (and citizen science with it) as something even worse: a pretext to extend neoliberal free-market thinking, with the aim of enabling platform capitalism (as exemplified by online giants such as Google and Facebook, or publishing corporations such as Elsevier) to build commercial monopolies on the systems of knowledge production. We are sympathetic to Mirowski's criticism, but emphasise that what he describes is a citizen science as it exists (and struggles) in the current status quo of the industrial system. Our attempt to sketch a more "ecological" epistemological framework for academic research could be seen as an attempt to provide the philosophical foundations for the new "political ontology" and the "economic structure" Mirowski is calling for (ibid.). We are in no way naïve enough to think this will be easy to implement under the current socio-political circumstances, or that it will be achieved in some sort of utopian way. Instead, we see the new ethos of science we are outlining here as something that can guide and inspire us while working pragmatically towards a more humane and sustainable research system based on more democratic values and procedures.

The main feature of our ecological model of research—what makes it resilient towards attempts at gaming the rules—is its adaptive flexibility: it adjusts itself to the circumstances of each project to be evaluated—its epistemic and non-epistemic aims, the backgrounds and motivations of its participants, and the nature of its particular research question and methodology. It employs a situated process-based quality assessment that relies on shared values and procedures, rather than standard metrics (which may still be used to support it, of course, but are no longer the only evaluative tool). Its adaptive nature renders it more resilient against attempts at gaming the system. The assessment process becomes a learning process itself, which can dynamically react to novel circumstances (see section 7).

Our framework requires that we pay much more attention to the process of inquiry than in a traditional system, where evaluation is largely based on immediate and measurable research outcome. In particular, *we recommend quality assessment to focus on the aspects of diversity, inclusion, and deliberation*. The evaluation of the potential of a project should be combined with constant monitoring and facilitation of the research process. Are all relevant standpoints of impacted stakeholders represented in the community? Do project participants feel they are heard and can make a relevant contribution to the project? Is the deliberative format properly facilitated? Does it enable high-quality cognitive engagement of participants with the research problem at hand? Do participants understand the ethos of doing scientific research and innovation? Do they understand the criteria by which they will be evaluated? Are they given enough autonomy? Are they allowed to fail, while still having their efforts appreciated? Can they disagree with the majority view during deliberation? Can they comment on and contribute to the evaluation of their efforts themselves? This kind of process-focused assessment and facilitation allows a project to be deemed a success, if its process was properly implemented, even if the desired output may not have materialized at the end of the project. It allows participants and evaluators to jointly learn from their successes and (often more importantly) failures. And it generates a more collaborative and positive atmosphere in which to undertake creative work. Such a system cannot compete with industrial science on short-term efficiency. It takes time and effort to implement, and the deliberative process is optimised for participation and learning, rather than production. In the long run, however, this system has the potential to be more productive and innovative than the present one. It provides a way for exploration to reenter the world of academic research, allowing us to escape the local search maxima that the game-theoretic trap of the cult of productivity has gotten ourselves stuck on.

Bibliography

Anderson, R. L. (2017). Friedrich Nietzsche. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2017/entries/nietzsche/

Bai, H. (2020). Epistemic Injustice and Scientific Knowledge Distribution. *Filosofija. Sociologija*, *31*(3), Article 3. https://doi.org/10.6001/fil-soc.v31i3.4269

Barseghyan, H., Overgaard, N., & Rupik, G. (2018). *Introduction to History and Philosophy of Science*. Pressbooks. http://hakobsandbox.openetext.utoronto.ca

Beatty, J., & Alfred, M. (2010). Should We Aim for Consensus? *Episteme*, *7*(3), 198–214. https://doi.org/10.3366/E1742360010000948

Bernstein, R. J. (1989). Pragmatism, Pluralism and the Healing of Wounds. *Proceedings and Addresses of the American Philosophical Associations*, *63*(3), 5–18.

Bhaskar, R. (1975). *A Realist Theory of Science*. Routledge.

Bone, Z., Crockett, J., & Hodge, S. (2006). *Deliberation forums: A pathway for public participation*. http://www.regional.org.au/au/apen/2006/refereed/1/2918_bonetz.htm

Chappell, S.-G. (2021). Plato on Knowledge in the Theaetetus. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2021). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2021/entries/plato-theaetetus/

Christian, B., & Griffiths, T. (2016). *Algorithms to Live By: The Computer Science of Human Decisions*. Henry Holt and Co.

Couch, J., Theisz, K., & Gillanders, E. (2019). Engaging the Public: Citizen Science. In K. L. Hall, A. L. Vogel, & R. T. Croyle (Eds.), *Strategies for Team Science*. Springer Nature.

Creath, R. (2021). Logical Empiricism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2021). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2021/entries/logical-empiricism/

Dewey, J. (1927). *The Public and Its Problems*. Holt Publishers.

Eitzel, M. V., Cappadonna, J. L., Santos-Lang, C., Duerr, R. E., Virapongse, A., West, S. E., Kyba, C. C. M., Bowser, A., Cooper, C. B., Sforzi, A., Metcalfe, A. N., Harris, E. S., Thiel, M., Haklay, M., Ponciano, L., Roche, J., Ceccaroni, L., Shilling, F. M., Dörler, D., … Jiang, Q. (2017). Citizen Science Terminology Matters: Exploring Key Terms. *Citizen Science: Theory and Practice*, *2*(1), 1. https://doi.org/10.5334/cstp.96

Fire, M., & Guestrin, C. (2019). Over-optimization of academic publishing metrics: Observing Goodhart's Law in action. *GigaScience*, *8*(6), giz053. https://doi.org/10.1093/gigascience/giz053

Firestein, S. (2015). *Failure: Why Science Is So Successful*. Oxford University Press.

Fleck, L. (1935). *Entstehung und Entwicklung einer wissenschaftlichen Tatsache: Einführung in die Lehre vom Denkstil und Denkkollektiv*. Benno Schwabe & Co.

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D., & Barabási, A.-L. (2018). Science of science. *Science*, *359*(6379), eaao0185. https://doi.org/10.1126/science.aao0185

Franzoni, C., & Sauermann, H. (2014). Crowd science: The organization of scientific research in open collaborative projects. *Research Policy*, *43*(1), 1–20. https://doi.org/10.1016/j.respol.2013.07.005

Frickel, S., Gibbon, S., Howard, J., Kempner, J., Ottinger, G., & Hess, D. J. (2010). Undone Science: Charting Social Movement and Civil Society Challenges to Research Agenda Setting. *Science, Technology, & Human Values*, *35*(4), 444–473. https://doi.org/10.1177/0162243909345836

Gettier, E. L. (1963). Is justified true belief knowledge? *Analysis*, *23*(6), 121–123.

Giere, R. N. (1988). *Explaining Science*. University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/E/bo3622319.html

Giere, R. N. (2006). *Scientific Perspectivism*. University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/S/bo4094708.html

Grayling, A. C. (2019). *The History of Philosophy*. Penguin Books.

Haklay, M. (2018). Participatory Citizen Science. In S. Hecker, M. Haklay, A. Bowser, Z. Makuch, J. Vogel, & A. Bonn (Eds.), *Citizen Science: Innovation in Open Science, Society and Policy* (pp. 52–62). UCL Press.

Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, *14*(3), 575–599. https://doi.org/10.2307/3178066

Harding, S. (1995). "Strong objectivity": A response to the new objectivity question. *Synthese*, *104*(3), 331–349. https://doi.org/10.1007/BF01064504

Hartsock, N. C. M. (1983). The feminist standpoint: Developing the ground for a specifically feminist historical materialism. In S. Harding & M. B. P. Hintikka (Eds.), *Discovering Reality: Feminist Perspectives on Epistemology, Metaphysics, Methodology, and Philosophy of Science* (pp. 283–310). Springer Netherlands.

Hetherington, S. (2016). *Knowledge and the Gettier Problem*. Cambridge University Press. https://doi.org/10.1017/CBO9781316569870

Kasperowski, D., & Hillman, T. (2018). The epistemic culture in an online citizen science project: Programs, antiprograms and epistemic subjects. *Social Studies of Science*, *48*(4), 564–588. https://doi.org/10.1177/0306312718778806

Kauffman, S. A. (2016). *Humanity in a Creative Universe*. Oxford University Press.

Kieslinger, B., Schäfer, T., Heigl, F., Dörler, D., Richter, A., & Bonn, A. (2018). Evaluating citizen science – towards an open framework. In S. Hecker, M. Haklay, A. Bowser, Z. Makuch, J. Vogel, & A. Bonn (Eds.), *Citizen Science: Innovation in Open Science, Society and Policy* (pp. 81–94). UCL Press. https://doi.org/10.2307/j.ctv550cf2

Kitcher, P. (1992). The Naturalists Return. *The Philosophical Review*, *101*(1), 53–114.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. University of Chicago Press.

Kuhn, T. S. (1977). *The Essential Tension: Selected Studies in Scientific Tradition and Change*. University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/E/bo5970650.html

Leydesdorff, L. (2005). Evaluation of research and evolution of science indicators. *Current Science*, *89*(9), 1510–1517.

Mac Domhnaill, C., Lyons, S., & Nolan, A. (2020). The Citizens in Citizen Science: Demographic, Socioeconomic, and Health Characteristics of Biodiversity Recorders in Ireland. *Citizen Science: Theory and Practice*, *5*(1), 16. https://doi.org/10.5334/cstp.283

Mahr, D., & Dickel, S. (2019). Citizen science beyond invited participation: Nineteenth century amateur naturalists, epistemic autonomy, and big data approaches avant la lettre. *History and Philosophy of the Life Sciences*, *41*(4), 41. https://doi.org/10.1007/s40656-019-0280-z

Massimi, M. (2018). Four Kinds of Perspectival Truth. *Philosophy and Phenomenological Research*, *96*(2), 342–359. https://doi.org/10.1111/phpr.12300

Massimi, M., & McCoy, C. D. (2019). *Understanding Perspectivism: Scientific Challenges and Methodological Prospects*. Routledge. https://www.routledge.com/Understanding-Perspectivism-Scientific-Challenges-and-Methodological-Prospects/Massimi-McCoy/p/book/9781138503069

Melo-Martín, I. de, & Intemann, K. (2018). *The Fight Against Doubt: How to Bridge the Gap Between Scientists and the Public*. Oxford University Press.

Merton, R. K. (1973). *The Sociology of Science*. University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/S/bo28451565.html

Mirowski, P. (2018). The future(s) of open science. *Social Studies of Science*, *48*(2), 171–203. https://doi.org/10.1177/0306312718772086

Muller, J. Z. (2018). *The Tyranny of Metrics*. Princeton University Press.

Nagel, E., & Newman, J. R. (2001). *Godel's Proof* (D. R. Hofstadter, Ed.; Revised Edition). NYU

Press.

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

Ottinger, G. (2017). Reconstructing or reproducing? Scientific authority and models of change in two traditions of citizen science. In D. Tyfield, R. Lave, S. Randalls, & C. Thorpe (Eds.), *The Routledge Handbook of the Political Economy of Science* (pp. 351–363). Routledge.

Polanyi, M. (1958). *Personal Knowledge: Towards a Post-Critical Philosophy* (Corr. Ed edition). University Of Chicago Press.

Potochnik, A. (2017). *Idealization and the Aims of Science*. University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/I/bo27128726.html

Rescher, N. (1996). *Process Metaphysics: An Introduction to Process Philosophy*. SUNY Press.

Rescher, N. (2009). *Unknowability: An Inquiry Into the Limits of Knowledge*. Lexington Books.

Schaefer, T., Kieslinger, B., Brandt, M., & van den Bogaert, V. (2021). Evaluation in Citizen Science: The Art of Tracing a Moving Target. In K. Vohland, A. Land-Zandstra, L. Ceccaroni, R. Lemmens, J. Perelló, M. Ponti, R. Samson, & K. Wagenknecht (Eds.), *The Science of Citizen Science* (pp. 495–514). Springer International Publishing. https://doi.org/10.1007/978-3-030-58278-4_25

Seibt, J. (2020). Process Philosophy. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2020). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2020/entries/process-philosophy/

Sigmund, K. (2017). *Exact Thinking in Demented Times—The Vienna Circle and the Epic Quest for the Foundations of Science*. Basic Books.

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, *69*(1), 99–118. https://doi.org/10.2307/1884852

Simon, H. A. (1969). *The Sciences of the Artificial*. MIT Press. https://mitpress.mit.edu/books/sciences-artificial

Smart, P., Holmes, S., Lettice, F., Pitts, F. H., Zwiegelaar, J. B., Schwartz, G., & Evans, S. (2019). Open Science and Open Innovation in a socio-political context: Knowledge production for societal impact in an age of post-truth populism. *R&D Management*, *49*(3), 279–297. https://doi.org/10.1111/radm.12377

Stanford, P. K. (2010). *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. Oxford University Press.

Stephan, P. (2015). *How Economics Shapes Science* (1st edition). Harvard University Press.

Strasser, B. J., Baudry, J., Mahr, D., Sanchez, G., & Tancoigne, E. (2019). “Citizen Science”?

Rethinking Science and Public Participation. *Science & Technology Studies*, *32*(2), 52–76. https://doi.org/10.23987/sts.60425

Wimsatt, W. C. (2007). *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Harvard University Press.

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, *1*(1), 67–82. https://doi.org/10.1109/4235.585893

Wu, L., Wang, D., & Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. *Nature*, *566*(7744), 378–382. https://doi.org/10.1038/s41586-019-0941-9

Yucel, R. (2018). Scientists' Ontological and Epistemological Views about Science from the Perspective of Critical Realism. *Science & Education*, *27*, 407–433. https://doi.org/10.1007/s11191-018-9983-x

Zammito, J. H. (2004). *A Nice Derangement of Epistemes: Post-Positivism in the Study of Science from Quine to Latour*. University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/N/bo3634623.html

Annex : List of abbreviations

Abbreviation	Description
AI	Artificial Intelligence
CS	Citizen Science
GEAR	Gather, Evaluate, Accelerate, Refine
NSO	National Statistical Office
SDG	Sustainable Development Goal
WP	Work Package