

ARTICLE

The Value of Openness in Open Science

Carlos Santana 

Department of Philosophy, University of Pennsylvania
Email: csantana@upenn.edu

Abstract

Open science is good for both epistemic and social reasons, but in nonobvious ways, it can have detrimental epistemic side effects. Drawing on case studies and the social epistemology of science, I show how practices intended to increase transparency, communication, and information sharing in science can backfire. We should not reject Open Science, just implement it carefully. I argue that we can do so by treating openness as a governing value in science, and thus, that our pursuit of openness needs to be balanced against our pursuit of the whole scheme of values that govern science.

Keywords: open science; values in science; applied epistemology; data infrastructures; science policy

1. Introduction

Science, pretty much everyone agrees, should be open. Promoting the principles and practices of open science (OS) has become a central aim of contemporary science policy, and it is not unusual to hear people say that OS is “what science should be,” to quote one scientist (Skríp 2023).

My method in this article is to see what happens when we consider OS not merely a set of practices and policies but a commitment to the value of *openness*. Two lessons fall out of this approach. First, although openness is an epistemic good, the optimal amount of openness is not maximal openness. That raises the question, of course, about how to balance the goods of openness with the goods it trades off with. An answer to that question begins by understanding openness as one of the plural values in the governing value schemes of science alongside Kuhnian values such as accuracy and fruitfulness and feminist values such as diffusion of power and social utility. Once we understand OS not as a set of universal principles and practices that we can unreservedly apply to any and all scientific activity, but instead as a value that must be weighed against the other core values of science, we can start to make rational decisions about how to pursue openness in different scientific contexts. This leads to the second lesson: we should embrace a plurality of OS practices reflecting the value schemes of different scientific communities, rather than promoting acontextual “best practices.”

Here is the plan. First (Sec. II), I will make sure we are all on the same page about what OS is and how central it is becoming to contemporary scientific institutions and practices. Then (Sec. III), I will present a series of arguments that, although OS is a good thing, it can have detrimental epistemic side effects. Optimal openness, therefore, is not always maximum openness. We can make sense of this, I argue, by (Sec. IV) seeing openness as one of the core values that govern science. Thinking in these terms allows us to take openness as a fundamental scientific good, while also understanding that it needs to be constrained by the other fundamental scientific goods. I will conclude (Sec. V) by making a case for implementing OS in a measured, rational way, which means adopting OS practices on the basis of both good empirical evidence and careful application of our

best theoretical understanding of the social epistemology of science. I worry that we are on track to make a dogma out of openness in science, and science should not be guided by dogmas, even those that probably point in the right direction. Let us not get closed-minded as we weigh the value of OS.

2. Openness and OS

OS encompasses too many different principles and practices to be straightforwardly definable. Roughly speaking, it is about increasing transparency, communication, and data sharing in science. Open-access publication is OS. Inviting a journalist to visit your laboratory is OS. Implementing metadata standards to facilitate cross-platform interoperability is OS. Preregistration requirements at journals are OS. Intergovernmental science policy interfaces such as the IPCC (Intergovernmental Panel on Climate Change) and IPBES (Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services) are OS, citizen science is OS, and so on.

For present purposes, then, let us adopt the maximally inclusive definition of OS used in UNESCO's *Recommendation on Open Science* (UNESCO 2021), which encompasses everything on that list and more. The full definition is rather long, so I will just recap the highlights. According to UNESCO, OS includes:

- 1 making knowledge available and usable;
- 2 increasing collaboration;
- 3 information sharing;
- 4 transparency at each stage of the scientific process.¹

The UNESCO definition stresses that none of these components are purely intramural. OS is about increasing each of 1–4 within science, yes, but it is also about achieving all four principles between science and society as well as between science and other knowledge systems. Furthermore, it stipulates that the relevant sense of “science” is a broad one that includes social sciences and humanities, as well as the natural sciences.

While early understandings of OS focused on data availability and open-access publishing (Guédon 2001), UNESCO's definition is typical of more recent definitions in emphasizing that usability is just as important as accessibility. It is not enough, for example, to allow publics and policymakers free access to scientific publications; they need representations of that scientific knowledge in ways that they can digest and use to inform their decisions. Similarly, it is not enough to require that researchers publish their data alongside their articles. Those data need to be in commonly used formats and accompanied by the kind of metadata that facilitates reuse and interoperability. Following Leonelli (2023), we could characterize this shift as a move from an “object-oriented” view of OS, which is mostly about making scientific products accessible, to a “process-oriented” approach, which understands that science is an activity undertaken from diverse epistemic positions. Effective process-oriented OS facilitates joint activity across those diverse positions.

Therefore, OS is an omnivorous concept that is swallowed up a wide array of aims, practices, and infrastructures. Holding this diversity together is a commitment to the value of openness. I am not going to set myself up for failure by trying to give an analytic definition of *openness*. Beyond the traditional shortcomings of analytic definitions, OS is a quickly evolving phenomenon as well as an idea that belongs to a varied set of communities. Those are not features amenable to neat definitions. Therefore, take the following characterization of *openness*² for the rough and flexible account it is meant to be.

¹Transparency turns out to be a complicated concept in its own right: see Elliott (2022).

²For sake of clarity in argument, I'll be using *openness* to refer to the value and *Open Science* (OS) to refer to the practices aiming at realizing that value. Both terms, of course, could be (and sometimes are) used to refer to either facet of the phenomenon.

OPENNESS: A concept representing the way epistemic communities assign value, both epistemic and social, to increasing the flow of usable information within and between communities, through practices including transparency, translation, data and communication infrastructure, and increased inter-community engagement, among other things.

Openness in this sense has quickly become a significant aim of both science and science policy. If you are reading this, you probably do not need much convincing, but just in case, here are a few telling facts:

In 2013, the White House Office of Science and Technology Policy directed all U.S. Federal agencies with R&D budgets over \$100 million to develop a strategy to increase public access to federally funded research (Holdren 2013). By 2022, they wouldd decided to stop using the “we’ll make a plan” stall tactic, and explicitly directed that after December 31, 2025 (almost),³ all federally funded research publications and data must be publicly accessible without embargo (Nelson 2022).

This is in line with general trends. In the 10-year period from 2002 to 2012, the number of journals indexed in the Directory of Open Access Journals increased from 35 to 7889 (Poltronieri 2016). At the time of writing, it is now 20,773⁴. That is quite the growth curve. Similar trends apply to both the number of journals encouraging preregistration of experimental design and the total number of registrations (Nosek and Lindsay 2018). The Corpus of Contemporary American English tracks a more than tenfold increase in the frequency of the term “Open Science” when you compare 2005–2009 to 2015–2019, and mentions of OS in the corpus are about as likely to be in magazines or blogs as they are in academic publications, suggesting that the concept is not trapped in the academy (Davies 2008–).

By 2023, the majority of researchers in every subject area “strongly agreed” that open-access publishing and open data practices should be academic standards (Hahnel et al. 2023). The same study found less agreement on open peer review practices and the use of preprint repositories. For those practices you have to count the “agrees” in addition to the “strongly agrees” to get a majority in favor of making them standards—but that is still a pretty significant endorsement of openness by the academic community.

Clearly, OS is surging, and there’s little indication that the swell will abate. On the contrary, the 2021 UNESCO recommendation encourages all 194 member states of UNESCO to facilitate OS “by taking appropriate steps, including whatever legislative or other measures may be required” to implement the objectives of the recommendation. It takes a lot of consensus to get “whatever measures may be required” into a global policy document like the UNESCO recommendation. That the global community can agree so readily to such an expansive vision of OS attests to how central openness has become to science policy.

3. Crack the Window Open, but Leave the Screen in Place

The OS train is chugging ahead full steam, and people are justifiably excited. But excitement risks becoming hype, so it is worth slowing down for a second and considering what side effects we might be overlooking. In this section, I will describe a few ways that openness yields epistemic bads alongside its goods.

My focus on the *epistemic* side effects of OS is a departure from most existing criticisms of OS, which generally focus on negative social consequences. Most prominently, friendly critics of OS have pointed to how it can reproduce inequities in science (Bahlai et al. 2019; Ross-Hellauer 2022). For example, open-access publishing often costs the authors several thousand dollars, which creates a barrier to equal participation in science for researchers in less wealthy countries, at less wealthy

³Exceptions include, for instance, when open data would violate privacy law.

⁴<https://doaj.org/> accessed 12 August 2024.

institutions, or in fields where grants tend to be smaller (Chan et al. 2020; Berger 2021). Similar pay-to-play concerns have been raised about the costs of data archiving (Bahlai et al. 2019).

Elliott and Resnik (2019) reviewed several other social concerns about transparency practices in science. Among others, they noted that researchers worry about being “scooped,” that is, having credit for scientific work stolen by someone who, say, accessed their preprint or preregistration data (see also Hahnel et al. 2023). They noted that the time and energy required to comply with transparency requirements might be a burden on research funding that could be better spent elsewhere. Worse, transparency requirements have been used by bad actors to harass scientists, slow down evidence-based regulation, and cast doubt on the validity of legitimate scientific results (McGarity and Wagner 2008; Oreskes and Conway 2011).

These concerns about inequity and injustice are important factors in how we should evaluate OS, but they are also relatively well-publicized. Furthermore, OS advocacy generally acknowledges that there are political and ethical constraints on openness. For instance, the UNESCO recommendation lists “human rights, national security, confidentiality” and several other political considerations as potentially justifying restrictions on openness. Otherwise, it advocates that “[a]ccess to scientific knowledge should be *as open as possible*” (UNESCO 2021, emphasis added). Since ethical and political considerations are broadly recognized as potential limiters on OS, we do not need to spend more time addressing them here.

Therefore, instead of piling on about potential ethical risks of openness, let us instead explore *epistemic* problems OS can raise. My target is the idea that science should be “as open as possible” within the bounds of ethical and political restrictions; call this idea *maximal openness*.⁵ Maximal openness, I am going to argue, often will not produce the best science from an epistemic standpoint. Unlike ethical and political pitfalls for OS, its epistemic pitfalls are not widely discussed or acknowledged. In what follows, I will supplement the meager literature about epistemic pitfalls of OS by applying research about the epistemology of science in general to the case of OS. The results of that application are that in several ways, maximal openness does not maximize the epistemic returns of science.

3.a Issue 1—standards standardize

The principles of OS tend toward being implemented as standards: rules, best practices, and formal conventions that cover everything from requirements for publication to proper experimental design to how to comment on your code. Valuing openness pushes us toward these sorts of standards because standards facilitate information transfer. The more we all use the same scientific instruments and model organisms, represent our data in the same ways, present our findings in the same places, and so on, the more easily we will be able to comprehend and make use of each other’s research activities. That is a real social epistemic benefit and so standardization has become an integral part of OS practices.

But standards standardize. The intended effect of implementing a standard is to create homogeneity. Homogeneity is not always an epistemic good because diversity can be an epistemic resource (Kitcher 1993; Solomon 2007). This means, as Leonelli (2022) identified, that OS is potentially a “foe” of epistemic diversity. Leonelli focused on the ramifications of the inequity issues mentioned above. For example, the potential for researchers from less wealthy countries to either be exploited (scooping) or excluded (financial barriers) by OS standards could mean less participation by researchers from those countries. If so, more openness sometimes decreases the intellectual diversity of the research community.⁶

⁵Technically, the ethical and political restrictions mean that this is *qualified* maximal openness. For brevity’s sake, however, I’ll just use ‘maximal openness’ as shorthand. Thanks to a referee for pushing me to be clear on this.

⁶I’m focusing on the epistemic costs of openness, but it’s worth mentioning in passing that there’s a similar ethical/political cost to standardization. Elliott (2016) shows how standardization can invisibly crystallize value-laden judgments, leading to

But the problem runs deeper. As Guzzo et al. (2022) put it, a risk of OS is that its standards will “narrow in favor of a specific style of research.” For example, preregistration requirements result in the “privileging of confirmatory research” and exclusion of exploratory studies. If our commitment to openness results in the substantial reduction of exploration, we have taken our eye off the epistemic ball. Similarly, data standards might privilege quantitative methods at the cost of qualitative methods, which would be a major epistemic loss for many disciplines.

To drive the point home, let us examine the influential FAIR data principles⁷ (Wilkinson et al. 2016). FAIR data maximize openness by being findable, accessible, interoperable, and reusable. These principles apply to data, metadata, and data infrastructure. Meeting them involves creating all sorts of (domain-specific) standards: standards for data formatting so data are machine-readable, standards for what belongs in the metadata, standardized vocabularies and ontologies to ensure comparability of datasets, and centralized registries or indexes so all the data can be searched in one place. With FAIR data standards in place, we can do more reliable data synthesis because we will find more of the relevant data and it will be in commensurable forms. Homogenized metadata, vocabularies, and ontologies facilitate machine learning (ML) and AI methods in science. There is a lot to like about the FAIR data principles, which is why they have been widely endorsed in science policy contexts.

However, there are real epistemic costs as well. A standardized vocabulary or ontology, for instance, becomes a conservative force in science since the costs of replacing the standard might be impractically high. Mineralogical science fell into this trap. When the International Mineralogical Association decided to implement new standards for mineral nomenclature and classification in 1959, they realized that they would just have to “grandfather” in much of the preexisting nomenclature, even though it violated the new standards, leading to an inconsistent classificatory scheme (de Fourestier 2002; Santana 2019). Sometimes grandfathering is just what a science has to do, but there is a cautionary tale in there about not rushing to fossilize a standard out of a rush to achieve maximal homogeneity. The epistemic benefits of having a metadata standard *right now* need to be weighed against the epistemic benefits of taking the time to develop a standard that is more likely to be evergreen.

A different set of issues is raised by data standards meant to facilitate machine-readability. Machine-readability is a procrustean goal. We can make all the data easily visible to the robots, but often only by mangling the data so they all fit in the same bed. That could mean distorting it, but more often would mean leaving things out—important things like context. Consider what a researcher performing a meta-analysis should be doing. They will need to determine whether the sets of data they are synthesizing are meaningfully comparable, which involves making judgment calls about whether they are drawn from equivalent populations, result from experiments testing the same hypothesis, and so forth. These are tricky questions since whether two populations or experimental designs are equivalent is a matter of context and scale. Hypotheses, for instance, exist in nested hierarchies (Heger et al. 2021), and whether the two are equivalent depends on which level of the hierarchy matters in context. The expert meta-analyst can make the judgments necessary to answer those questions if they have access to the data *in context*. However, if the data have been decontextualized to facilitate machine-readability, the result might be unreliable data syntheses.

Therefore, we want to weigh the benefits of increasing machine-readability and data interoperability against the epistemic risks. As scholars of scientific practice have long emphasized, scientific communities are capable of information transfer even if the objects and concepts at the boundaries that bridge those communities are not standardized (Star and Griesemer 1989).

problematically value-laden research practices that can be exploited by bad actors. This, as Elliott notes, could then yield further epistemic problems similar to the ones I discuss in what follows.

⁷Reminder: the goal is not a takedown of the FAIR principles, which are likely a net positive. It’s just to show that they incur epistemic costs.

Standardization is not always necessary. The right approach to OS standards is to ask questions about whether the efficiency introduced by standardization is worth paying the cost in loss of epistemic diversity and contextual information. The answer will not always be “yes,” so openness is the value that needs to be tempered.

3.b Issue 2—social epistemology is not additive to individual epistemology

Individual and social epistemology diverge, and more openness between individual researchers does not lead to better social epistemic outcomes under all conditions. A well-known example is Zollman’s (2007) result that maximal communicativeness can lead to premature consensus, but we can see that result as one instance of a more general phenomenon in which social transmission amplifies the error. Think of it this way: the more the merrier at a party, until there is a cold going around, at which point you want less interaction. Openness is the opposite of epistemic social distancing. And if you are in an environment with transmissible errors, you want some social distancing to give you time to address them before they overwhelm your community.

Epistemic risk of that infectious sort is something we have to worry about only under specific conditions (Rosenstock et al. 2017) but still should be enough to give us pause about dogmatic commitment to openness. Furthermore, I am making a broader point about the relationship between individual and social epistemology. The emergent epistemic outcomes on the social level are not necessarily just whatever’s going on at the individual-level scaled up. This is what Mayo-Wilson et al. (2011) call the “Independence Thesis.” Social and individual epistemology diverge, in that rational individual epistemic behavior might not produce the best social epistemic outcomes and vice-versa.⁸ The problem is probably even worse than what they identify because it is likely a general issue of scale and emergence (Reymondon 2024). Bad epistemic hygiene on the part of a laboratory tech might end up being good for the laboratory’s epistemic hygiene⁹ but that might be bad for the scientific field it belongs to¹⁰ which might be good for science taken as a whole which might be bad for the state of knowledge in broader society. All of which is to say, openness is an epistemic good, but epistemic goods at one scale can simultaneously be epistemic bads at other scales. That is a reason to do some careful, holistic thinking about when and where to increase openness in science.

3.c Issue 3—scientists are experts

Good science often involves the application of expert judgment. Transparency, which is a key aspect of openness, restricts the application of expert judgment because it requires experts to make their judgments sensible to nonexperts (Nguyen 2022). Therefore, transparency often restricts or distorts the application of expertise.

Nguyen (2022) called this effect the *epistemic intrusiveness* of transparency and catalogued three ways in which “expertise can be distorted by the pressure of public transparency,” all of which we can apply to the scientific context. First is deception, in which the experts’ reported justifications are not their actual justifications. When Nobel Prize winner Peter Medawar (1963) answered his own question, “Is the scientific paper a fraud?” with a ready “yes,” this is what he was referring to. The scientific paper is a fraud not because it generally contains fabricated data, but rather because it contains fabricated justifications: “it misrepresents the processes of thought that accompanied or

⁸I’m particularly fond of the example used by Fleisher (2018), in which a researcher intentionally endorses a hypothesis they do not believe in order to improve the cognitive division of labor in their field. Fleisher tries to keep this case from looking like an instance of the Independence Thesis, but has to radically redefine ‘epistemic rationality’ to do so.

⁹Maybe it breaks them out of a rut, or forces them to re-evaluate their protocols.

¹⁰Perhaps in breaking out of their rut, they have unknowingly adopted the same approach as a rival lab, reducing epistemic diversity and introducing an inefficient redundancy.

give rise to the work that is described in the paper” (Medawar 1963). Driving this deception is epistemic intrusiveness. Scientists are hyperspecialists (Millgram 2015), and even most readers in their own field will not be able to fully grasp the reasons for a researcher’s conclusion. So, in the article, we get justifications that are intelligible to a broader audience, and these may or may not match the actual process of discovery. More openness makes the audience even broader and so deepens the mismatch.

Alongside deception, Nguyen (2022) identified “limitation” and “guidance” as forms of epistemic intrusiveness. Limitation is when experts, scientists for our purposes, only act within the bounds of what is intelligible to an expert audience. For example, a field ecologist might only select study sites that they can describe in articles as being analogous to controlled laboratory settings (Kohler 2002),¹¹ knowing that their laboratory scientist readers will not be able to understand the epistemology of fieldwork (Halm and Santana 2024).

Guidance is when experts internalize publicly available reasons as their own preferred reasons. For instance, nonexperts generally cannot judge the true epistemic quality of a scientific paper, so instead they judge on the basis of publicly intelligible reasons, like citation metrics. The phenomenon of guidance means that some scientists start to do the same, treating a number of citations as actually indicative of how good a paper in their field is, rather than applying their domain-specific expertise to make that judgment. Both limitation and guidance tend to be epistemic bads because they leave an epistemic resource—expertise—underutilized.

I have heard skepticism from both in-person audiences and an anonymous reviewer that Nguyen’s worries are as applicable to scientists as they are to, say, politicians and policymakers. My first response is that a whole lot of scientists are also policymakers. In the United States, for instance, Federal agencies employ close to 300,000 scientists (NCSES 2020). That is nearly as many as the ~350,000 PhD scientists employed at academic institutions (National Science Board, NSF 2023). If the ratio is anything similar worldwide, a significant chunk of science occurs in the administrative and policy contexts where Nguyen’s arguments have their best toehold.

Moreover, limitations and guidance do occur in academic science, most obviously in practices like peer review and the presentation of results. Consider the following referee report, quoted in Vellend (2018):

“while the authors are careful to state that they are discussing biodiversity changes at local scales, and to explain why this is relevant to the scientific community, clearly media reporting on these results are going to skim right over that and report that biological diversity is not declining if this paper were to be published in *Nature*. I do not think this conclusion would be justified, and *I think it is important not to pave the way for that conclusion to be reached by the public*” (emphasis added).

Here, the referee is explicitly engaging in limitation, arguing that *Nature* should reject the submission because the scientific reasoning involved would be misapprehended by the public. *Nature* rejected the paper. It was eventually published in another venue, but as Vellend (2018) suggests, making it harder to publish legitimate research when it might be misunderstood by the public still distorts our scientific knowledge. I do not know how common limitation of this sort is in peer review, but suspect it is not vanishingly rare when it comes to research on politicized topics (see Hobbs 2017 for a similar case).

For an example of guidance at play, take Fahrig’s (2017) analysis of habitat fragmentation research. Drawing on the data behind 40 years of research, Fahrig shows that research on fragmentation overwhelmingly supports the conclusion that, when fragmentation is distinguished from habitat loss, fragmentation tends to increase biodiversity. Researchers conducting these

¹¹As Kohler’s excellent historical work shows, this is not a mere hypothetical example. It’s a significant part of the actual history of how ecology developed as a discipline.

studies, however, have tended to “emphasize” negative effects, “down-play” positive ones, warn against extrapolating from positive results, and label positive results as effects of “heterogeneity” rather than “fragmentation” (Fahrig 2017). After exploring potential explanations for this mismatch between results and the reasoning presented in abstracts and discussion sections, Fahrig concludes that it is driven by a “fear of misinterpretation” that researchers end up internalizing as confirmation bias (Fahrig 2017). If Fahrig’s explanation is right, this is a case of guidance.

I recognize that a couple of case studies does not support a broad claim about limitation and guidance being widespread in academic science. But they do, I hope, help you see how it is plausible that they could occur, and you can probably adduce some examples of your own of scientists considering how publics might (mis)interpret them when making decisions about publication or how they interpret their results. Epistemically responsible science requires a good deal of autonomy since interference from extrascientific actors¹² can distort epistemic processes. Because transparency results in limitation and guidance, it can reduce that autonomy, incurring epistemic costs. Maximizing transparency also maximizes the risk of reduction in autonomy, so maximal openness is not the optimal way to trade-off the epistemic pros and cons of transparency.

Again, the point here is not to vilify openness. Healthy science requires both a great degree of transparency and significant accountability to nonexpert audiences. But Nguyen (2022) was right that, when expertise matters, transparency alone does not lead to the ideal social epistemic outcomes. Healthy epistemic communities involve a good mix of transparency and trust, and despite what you might infer from some of the rhetoric surrounding OS, more transparency does not always lead to more trust (de Fine Licht and Naurin 2015). Transparency is a way to monitor those we do not fully trust, and so we sometimes take it as evidence that actors are untrustworthy (de Fine Licht 2011). Even with the increase in replication work and other OS practices, science will always require a great deal of just trusting that scientists are generally honest actors. That openness can reduce as well as increase trust is thus an important reason not to treat openness as an unmitigated epistemic good.

3.d Issue 4—context matters

Make your code downloadable by anyone—that is OS!

But... be aware that doing so might also facilitate sloppy science by people in other fields who do not really understand all the assumptions and modeling decisions that went into that code. They will download it, use it to model something in a totally different domain, and no one in its new domain of application will be qualified to critique it effectively. That might not be a decisive reason to avoid publishing your code, but it is an epistemic tradeoff of openness. And obviously this applies not only to code, but also to datasets, experimental protocols, and everything else.

Here is a quick example. Phylogenetic comparative methods were revolutionary when they were introduced in evolutionary biology, and they remain an important and powerful tool. But they are a tool for some jobs and not others, and phylogenetic algorithms and the trees they output are often misused by scientists who are not phylogeneticists because they do not understand the biases and limitations of the statistical methods involved in phylogenetics (Cooper et al. 2016). For instance, Pozzi et al. (2014) showed how primatologists have sometimes misused phylogenetics, not because the trees they were relying on were bad science, but because they did not understand that certain types of gene-oriented phylogenetics are not reliable for species trees. The primatologists saw their phylogeneticist colleagues using a fancy screwdriver, and excitedly borrowed it, only to use it to hammer in some nails poorly. Sometimes it is better to have the speedbump of not leaving your tools just laying around to be borrowed by anyone—only making your code available by request, for instance. Not only does that prevent epistemic misuse, but it also secures the additional good of

¹²Or even scientists who aren’t experts in the right domain (Clark and Walsh 2009).

creating more engagement between practitioners. For example, emailing a colleague to request their code might lead to follow-up questions, instructions on how to properly use the code, or even collaboration—all stuff that does not happen if you can just download it from an open repository. These are real goods that can be lost if we pursue maximal openness.

In a similar vein, in some domains of science, knowledge is knowledge only when contextually situated, but openness decontextualizes and de-situates it, leading to epistemic losses. We do not necessarily just want to slap a metadata label on some research on introduced predators in New Zealand that makes it look like it is comparable to, say, research on management of invasive plants in Texas, or even those same introduced predators on some other island. That leads to hasty generalizations, to meta-analyses and other forms of data synthesis that look statistically powerful and inferentially valid, but which have ignored the fact that the type of research is largely time and place specific (Elliott-Graves 2016).

This goes double for the aspects of OS where the knowledge systems interacting are not just different breeds of academic science, but science and other forms of knowledge, such as Indigenous and local knowledge. Openness between academic science and these other knowledge systems is increasingly an important part of science policy (UNESCO 2021; Prabhakar and Mallory 2022). But Indigenous and local knowledges are deeply tied to place and context (Burkhart 2019; Liboiron 2021). Integrating them into OS frameworks thus bleaches them of much of their epistemic value, in addition to often being ethically problematic (Lopez-Huertas and Santana 2024). Effective knowledge co-production across science and other knowledge systems thus requires balancing openness against other values.

3.e Issue 5—misinformation

As Elliott and Resnik (2019) pointed out, groups who can benefit from epistemic sabotage are motivated to misuse open data. They might, for instance, “spread misleading reanalyses of the data” (Elliott and Resnik 2019). I think the problem is likely worse than Elliott and Resnik suggest, because not only will these special interest groups be incentivized to misuse open data, they will often be better positioned to use the data than epistemically responsible actors, leading to a net-negative epistemic effect of open data access. Suppose we conduct a successful citizen science project to collect data on industrial pollution and health outcomes in a county hosting several large manufacturing plants, then use the data to request regulating those plants more heavily. Assume that, because we value openness, all the data have gone into a public repository. Who is best positioned to use that data? The poor communities affected by the pollution? Under-resourced and politically embattled government environmental agencies? Academic scientists who need to chase the next grant? Most likely, it will be the polluters themselves, who can hire a statistician to massage the data until enough doubt comes out to delay regulation. My point here is that data should often be freely available, but not always—which is something some databases with sensitive environmental information have begun taking steps to address (Quinn 2021). And this is not just a moral point about evil polluters (or whoever) getting away with hurting people. It is an epistemic issue since making the data open increases the risk of the spread of misinformation. We should be cautious, then, about letting our enthusiasm for openness in science lead us to create mandatory open data regimes, at least in some domains.

Misinformation can be fueled not merely by misuse of open data, but also by making the inner workings of science too visible when that science is controversial. Drawing on a case where leaked emails between climate scientists were used to feed climate skepticism, John (2018) showed how “just as publicising the inner workings of sausage factories does not necessarily promote sausage sales, so, too, transparency about knowledge production does not necessarily promote the flow of true belief throughout the population.” As John demonstrated, propagandists can exploit openness to portray even epistemically responsible scientific practices as misconduct. This is especially true in cases where public audiences have a poor understanding of how the science works (Kovaka 2021).

This is not to say that science should generally be cultish and secretive, but might be a reason to be judicious rather than dogmatic about what features of scientific practice are made open.

3.f Recap—the epistemology of OS

By applying existing work in social epistemology and the philosophy of science to the context of OS, I have catalogued five ways in which openness can incur serious epistemic risks for science. Where do these arguments lead us? Not, I will continue to insist, toward a rejection of OS. Openness is a saguaro cactus: a beautiful ecological keystone that we should love, celebrate, and occasionally water, but should not embrace in a full-on bearhug.

If I am right, we are faced with the question of how to appropriately care for our prickly OS cactus. To that question I now turn.

4. The Role of Openness as a Value in Science

The question is how to implement OS carefully, sensitive to the epistemic risks and trade-offs it incurs, while still recognizing openness as a key scientific good. Part of the answer, of course, must be to experiment with different OS practices, standards, and infrastructure, and assess their effects empirically. That will only get us so far, however. Practical limitations and the complexity of the sociology of science will often make measuring the effects of OS interventions intractable. But more to the point, normative assessments of OS are underdetermined by evidence because they are questions of value.

Looking back on the past several decades of research about values in science, Kitcher (2024) reminded us that the issue is less one about which values belong in science, and more about the sets of values scientists and scientific communities adopt, and how those values are ordered and weighed against each other. Following Kitcher, let us call that combination of a set of values and their ordering and weighting a *value scheme*.

Value schemes are a familiar idea. Kuhn (1977) famously provided a nonexhaustive list of five values (that he calls “virtues”), including accuracy, consistency, broad scope, simplicity, and fruitfulness. Note that this list is not yet a value scheme in our present sense, because it does not specify any ordering or weighting. As Kuhn observed, values on the list can conflict, and so communities that weight them in different ways might come to different scientific conclusions. Longino (1996) revised Kuhn’s list by drawing on virtues identified in feminist science studies. Her list includes empirical adequacy, novelty, ontological heterogeneity, mutuality of interaction, applicability to human needs, and diffusion of power. Longino’s revisions are motivated in part by feminist aims but also because Kuhn is focused mostly on theory choice, while Longino is considering scientific practice more broadly. She is concerned not only with which theories scientific communities adopt but also how those communities structure themselves and how they pursue research.

This tradition of thinking about value schemes, I am suggesting, will help us make sense of openness as a value in science. Openness has become a core consideration in how many scientific communities structure themselves and pursue research. In some cases, it may even be a consideration in theory choice. Scientists may be favoring theories and models that are more accessible and usable to external parties, whether those parties are publics, other scientists, or algorithms.¹³ Whatever we may think of the specific values on Kuhn and Longino’s lists, the idea that science has some core governing values is a useful one, and the OS movement is succeeding at making openness one of those values.

¹³I’m convinced this is the case, in part because I’ve witnessed scientists use appeals to openness to argue for adopting particular theories or models. But these are in conversation, and not the sort of reasons that tend to appear in formal publications, so I’m not going to try to provide evidence for it here.

I have been engaging in some loose talk about “governing” values in science that I need to tighten up. Scientific value schemes include a panoply of values, including things like economic efficiency, collegiality, sustainability, and fair distribution of credit. Only a few values, however, are ones that we would expect to see consistently being ranked at the top (or most heavily weighted) in scientific value schemes. Accuracy and applicability to human needs, for instance, are going to be among the most important values across a good proportion of value schemes. It would be very surprising, however, to see a community holding a scheme that ordered collegiality above accuracy in model selection, or gave it equal weight to applicability to human needs in determining pursuit worthiness. Those values that are consistently placed among the highest in value schemes are governing values, and it is the governing values that make it onto lists like Kuhn’s and Longino’s. Openness, even if it was not a governing value in science in 1977 or 1996, has become one in 2024.

The fact that science works according to schemes of values that often conflict means that we have to weigh the values in our scheme against each other. Consequently we rarely want to maximize any single value. We value simplicity and social utility, for instance, but few of us think that science should be maximally simple or purely utilitarian. Most of us have yet to realize the same fact about openness, and OS policies often aim for maximal openness. Taking a values-in-science perspective toward openness shows that this is a big mistake.

It also shows us how to manage the tradeoffs involved in OS. One way science handles conflicts between its governing values is by selecting the appropriate value scheme for the circumstances. In times of urgency, such as a pandemic, a community might shift some weight from accuracy to applicability to human needs, for example. Similarly, policy-adjacent scientific disciplines might select value schemes in which diffusion of power is ordered higher than in the schemes used by more policy-distant disciplines. The lesson here is that there is not a one-size-fits-all solution to the question of how to rationally implement OS. It will depend on the value scheme which fits the present needs of the relevant communities.

Another way science handles value conflicts is by allowing subcommunities to adopt different value schemes, which results in scientific pluralism (Kellert et al. 2006). This is not an “anything goes” sort of pluralism. A scientific community must have good, nonarbitrary reasons to adopt its scheme, which serves as a strong constraint keeping pluralism from collapsing into some sort of epistemic relativism (Veigl 2021). Nevertheless, there may be good, nonarbitrary reasons sufficient to justify more than one value scheme, making the ongoing coexistence of plural scientific communities reasonable. Since openness is a governing value that sometimes conflicts with other governing values, we may need to accept that different communities can and should vary in how important it is in their value schemes.

With that in mind, we can start to apply our understanding of openness as a governing value in science to the details of OS policy. For example, taking pluralism seriously means that we should prefer a diverse scientific infrastructure. Consider OS practices in publishing. Maximal openness means that every journal should practice open access, open data, preregistration (where applicable), and open peer review. We can now see how that might be an inappropriate prioritization of openness over all other scientific values. Since communities might legitimately select value schemes that prioritize openness differently, we should expect a diverse publishing ecosystem. Some communities might adopt schemes that make universal preprint archiving optimal, while others do not. Some reasonable value schemes might suggest that open peer review is unnecessary, other schemes will entail that it is mandatory, and yet other schemes will yield the result that we want options for both open and closed peer review. That means that rather than homogenizing the publishing landscape, taking openness seriously as a value will suggest that a landscape filled with a variety of publishing practices, some more open than others, is a sign of well-functioning science.

Similarly, we should treat top-down OS mandates with caution. Perhaps, there are ethical or political reasons that justify mandates like those in which governments require open access and open data for all publicly funded research, but in general, this is likely to involve politics overriding scientific values, not implementing them. And while it makes sense for governments, professional

societies, and foundations to encourage openness—just like they should encourage accuracy, fruitfulness, novelty, and the rest—to mandate specific practices top-down is a dangerous way to do so since it does not allow scientific communities to implement openness according to its place in their community’s well-motivated value schemes. As an alternative to strict mandates, governments could encourage OS practices in more flexible ways, such as by using incentives rather than mandates, or by issuing general directives that allow scientific subcommunities to implement them according to their community’s judgments. Either approach would, unlike strict mandates, allow scientific communities to value openness as merely one constituent in a value scheme.

As a final example of how the theoretical machinery I have introduced has applied payoff, consider the connection between OS and ML and AI. ML and AI have incredible epistemic potential. Already they can perform many epistemic tasks more efficiently and effectively than unaided human researchers, and their capabilities will probably continue to increase. But consider an analogy to the automobile. Cars can perform many transportation tasks more efficiently and effectively than unaided human researchers. But they also require car-specific infrastructure like roads and parking lots. In response, many communities went all-in on creating car infrastructure, razing neighborhoods for freeways, turning urban centers into massive parking lots, and so on. The result is that many communities have become car-friendly, and utterly hostile to human beings. It is dangerous to be a human on the road in a car, and even more dangerous to be a human on, or crossing, the road if you are not in a car. The most fundamental aspect of transportation, moving one’s body around one’s own community, has become more expensive, difficult, and dangerous due to the wide implementation of car-friendly infrastructure.

Much of OS is about razing scientific neighborhoods to facilitate machine access to research, or about building massive parking lots for data in quantities that only AI will be able to make use of. Dogmatic embrace of openness risks taking science, one of the great humanistic achievements, and making its infrastructure friendly to machines at the expense of good-old-fashioned squishy, sloshy brains. Human access to science could become more epistemically expensive, difficult, and dangerous if we recapitulate the mistake cities made when they let traffic engineers dominate their planning departments. Recognizing that openness sometimes conflicts with other values in scientific value schemes can serve as a check on this process of dehumanizing epistemic space. Particularly effective will be schemes that include values like Longino’s (1996) feminist values of ontological heterogeneity, mutuality of interaction, applicability to human needs, and diffusion of power. Openness is still a governing value in science, and the idea is not to put a halt to all the OS infrastructure that will facilitate ML- and AI-driven research. But we do need to think about how to build infrastructure for machines *and* people that ensures the preservation of whatever the epistemic equivalents are of protected bike paths, walkable urban centers, and public transit.

5. Conclusion—Values and the Applied Epistemology of Science

I hope that what I have accomplished above is to make complicated something that seemed simple. OS seems to many people to be an unquestionable good, and they see its shortcomings as mere technical issues to be solved with better infrastructure. It *is* an unquestionable good. But even unquestionable goods in science must be tempered against the other goods in the value schemes of science. This makes what a healthy OS looks like more complicated—and more pluralistic.

Asking whether openness should be a governing value in science is the wrong question. As Kitcher (2024) observed, to ask of a value “Is it legitimate for this value to play this role in science?” puts too much ontological weight on *values* as universal, abstract things, when the more fundamental phenomenon is *valuing*, which is something particular, concrete people and communities do. I have borrowed the “value scheme” framing from Kitcher as a way to avoid making that mistake. Science policy is not a gladiatorial match, where at the end we are required to give a thumbs up or thumbs down verdict on whether a particular value gets to fight on in science or meet a gory end. If we took that approach to thinking about openness in science, we would be forced to give an

enthusiastic thumbs up, and so we would lose all the nuances I catalog above. Instead, we are forced to ask more difficult—and more useful—questions about how to balance openness against other governing values in science. Or, more precisely, we should require science and science policy communities to be explicit about why they balance the values in their value schemes the way they do. If they do so, they may find that their relationship to OS is more complicated than they might have thought.

To close, I would like to summarize what I take to be some of the action items that follow from the arguments I have marshaled. Most obviously, we need to scale back the maximalist attitudes, directives, and language we see about OS in science policy. Relatedly, we should develop approaches to implementing OS that allow for decentralization, so that diverse scientific communities can adopt the versions of OS practices that work for their needs and values. Determining those needs and values should be a collaborative process driven by the researchers involved in the scientific community, as well as key stakeholder groups in policy-adjacent sciences. It is a mistake to abandon the reins of science policy to whichever engineers have developed the most exciting new information technology, or to administrative officials who value OS as a surveillance tool more than a scientific boon. Similarly, we need some serious research and development to understand how to preserve a healthy epistemic ecosystem for human beings in a scientific landscape that is increasingly structured to facilitate the movement of data and machines.

Finally, we need more critical thinking about the roles OS can play in the epistemic relationships between science and the publics. We have been quick to assume that openness secures trust and spreads knowledge and understanding, but some of the issues I have discussed highlight how this is not always the case. I am confident that OS has the potential to provide these epistemic goods, but only if it is the right kind of openness, appropriately shaped and tempered by the other governing values in science.

References

- Bahlai, C., Bartlett, L. J., Burgio, K. R., Fournier, A. M., Keiser, C. N., Poisot, T., & Whitney, K. S. (2019). Open science isn't always open to all scientists. *American Scientist*, 107(2), 78–82.
- Berger, M. (2021). Biodiversity at the centre: Decolonizing open access. *Development and Change*, 52(2), 383–404.
- Burkhardt, B. (2019). *Indigenizing philosophy through the land: A trickster methodology for decolonizing environmental ethics and Indigenous futures*. MSU Press.
- Chan, L., Hall, B., Piron, F., Tandon, R., & Williams, L. (2020). *Open Science beyond open access: For and with communities. A step towards the decolonization of knowledge*. The Canadian Commission for UNESCO.
- Clarke, S., & Walsh, A. (2009). Scientific imperialism and the proper relations between the sciences. *International Studies in the Philosophy of Science*, 23(2), 195–207.
- Cooper, N., Thomas, G. H., & FitzJohn, R. G. (2016). Shedding light on the 'dark side' of phylogenetic comparative methods. *Methods in Ecology and Evolution*, 7(6), 693–699.
- de Fine Licht, J. (2011). Do we really want to know? The potentially negative effect of transparency in decision making on perceived legitimacy. *Scandinavian Political Studies*, 34(3), 183–201.
- de Fine Licht, J., & Naurin, D. (2015). Open decision-making procedures and public legitimacy. *Secrecy and publicity in votes and debates* 131. Cambridge University Press
- de Fourestier, J. (2002). The naming of mineral species approved by the Commission on New Minerals and Mineral Names of the International Mineralogical Association: A brief history. *The Canadian Mineralogist*, 40(6), 1721–1735
- Davies, M. (2008-). *The corpus of contemporary American English (COCA)*. Available from <https://www.english-corpora.org/coca/>.
- Elliott, K. C. (2016). Standardized study designs, value judgments, and financial conflicts of interest in research. *Perspectives on Science*, 24(5), 529–551.
- Elliott, K. C., & Resnik, D. B. (2019). Making open science work for science and society. *Environmental Health Perspectives*, 127(7), 075002.
- Elliott, K. C. (2022). A taxonomy of transparency in science. *Canadian Journal of Philosophy*, 52(3), 342–355.
- Elliott-Graves, A. (2016). The problem of prediction in invasion biology. *Biology & Philosophy*, 31, 373–393.
- Fahrig, L. (2017). Forty years of bias in habitat fragmentation research. In *Effective conservation science: Data not dogma* (pp. 32–38). Oxford University Press, USA.
- Fleisher, W. (2018). Rational endorsement. *Philosophical Studies*, 175(10), 2649–2675.

- Guédon, J. C. (2001). In *Oldenburg's long shadow: Librarians, research scientists, publishers, and the control of scientific publishing*. Association of Research Libr.
- Guzzo, R. A., Schneider, B., & Nalbantian, H. R. (2022). Open science, closed doors: The perils and potential of open science for research in practice. *Industrial and Organizational Psychology*, 15(4), 495–515.
- Hahnel, M., Smith, G., Scaplehorn, N., Schonenberger, H., & Day, L. (2023). *The state of open data 2023*. Springer Nature
- Halm, D., & Santana, C. (2024) [forthcoming]. Debiasing collection in field biology. *Philosophy, Theory, and Practice in Biology*
- Heger, T., Aguilar-Trigueros, C. A., Bartram, I., Braga, R. R., Dietl, G. P., Enders, M., ... & Jeschke, J. M. (2021). The hierarchy-of-hypotheses approach: a synthesis method for enhancing theory development in ecology and evolution. *BioScience*, 71(4), 337–349.
- Hobbs, R. (2017). Novel ecosystems: can't we just pretend they're not there?. In *Effective conservation science: Data not dogma* (pp. 45–50). Oxford University Press.
- Holdren, J. (2013). *Memo, February 22, 2013: Increasing Access to the Results of Federally Funded Scientific Research*. White House Office of Science and Technology Policy.
- John, S. (2018). Epistemic trust and the ethics of science communication: Against transparency, openness, sincerity and honesty. *Social Epistemology*, 32(2), 75–87.
- Kellert, S., Longino, H., & Waters, C. K. (2006). The pluralist stance. In S. Kellert, H. Longino, & C. K. Waters (Eds.), *Scientific pluralism* (pp. vii–xxix). Minneapolis: University of Minnesota Press.
- Kitcher, P. (1993). *The advancement of science: Science without legend, objectivity without illusions*. Oxford University Press, USA.
- Kitcher, P. (2024). The scientist, qua scientist, is an ethical agent. *Filozofia*, 79(3), 231–243.
- Kohler, R. E. (2002). *Landscapes and labs: Exploring the lab-field border in biology*. University of Chicago Press.
- Kovaka, K. (2021). Climate change denial and beliefs about science. *Synthese*, 198(3), 2355–2374.
- Kuhn, T. S. (1977). objectivity, value judgment and theory choice. In: T. S. Kuhn(Ed.), *The essential tension* (pp. 320–339) Chicago: University of Chicago Press,.
- Leonelli, S. (2022). Open science and epistemic diversity: Friends or foes?. *Philosophy of Science*, 89(5), 991–1001.
- Leonelli, S. (2023). *Philosophy of open science*. Cambridge University Press.
- Liboiron, M. (2021). *Pollution is colonialism*. Duke University Press.
- Longino, H. E. (1996). Cognitive and non-cognitive values in science: Rethinking the dichotomy. In *Feminism, science, and the philosophy of science* (pp. 39–58). Dordrecht: Springer Netherlands.
- Lopez-Huertas, J., & Santana, C. (2024). *Contextual Empiricism as coexistence: Braiding Indigenous and feminist epistemologies for collaborative planning*. Unpublished manuscript.
- Mayo-Wilson, C., Zollman, K. J., & Danks, D. (2011). The independence thesis: When individual and social epistemology diverge. *Philosophy of Science*, 78(4), 653–677.
- McGarity T., Wagner W. 2008. *Bending science: How special interests corrupt public health research*. Cambridge, MA: Harvard University Press.
- Medawar, P. B. (1963). Is the scientific paper a fraud? *The Listener* 70 (12 September): 377–378. Reprinted as Chapter 3 in Medawar, P. B. 1996. *The Strange Case of the Spotted Mice and Other Classic Essays on Science*. New York, NY: Oxford University Press.
- Millgram, E. (2015). *The great endarkenment: Philosophy for an age of hyperspecialization*. Oxford University Press.
- National Science Board, NSF. (2023). Academic Research and Development. *Science and Engineering Indicators 2024*. NSB-2023-26. Alexandria, VA. Available from <https://ncses.nsf.gov/pubs/nsb202326/>.
- NCSES. (2020). *Special tabulations of Office of personnel management data*. National Center for Science and Engineering Statistics. Available from <https://ncses.nsf.gov/>.
- Nguyen, C. T. (2022). Transparency is surveillance. *Philosophy and Phenomenological Research*, 105(2), 331–361.
- Nelson, A. (2022). *Memo, August 25, 2022: Ensuring Free, Immediate, and Equitable Access to Federally Funded Research*. White House Office of Science and Technology Policy.
- Nosek, B., & Lindsay, S. (2018). Preregistration Becoming the Norm in Psychological Science *Association for Psychological Science Observer*, February 28, 2018. <https://www.psychologicalscience.org/observer/preregistration-becoming-the-norm-in-psychological-science> Accessed 28.3.2024
- Oreskes, N., & Conway, E. M. (2011). *Merchants of doubt: How a handful of scientists obscured the truth on issues from tobacco smoke to global warming*. Bloomsbury Publishing USA.
- Pozzi, L., Bergey, C. M., & Burrell, A. S. (2014). The use (and misuse) of phylogenetic trees in comparative behavioral analyses. *International Journal of Primatology*, 35, 32–54.
- Poltronieri, E., Bravo, E., Curti, M., Ferri, M., & Mancini, C. (2016). Open access publishing trend analysis: statistics beyond the perception. *Information Research: An International Electronic Journal*, 21(2), n2.
- Prabhakar, A., & Mallory, B. (2022). Memo: Guidance for Federal Departments and Agencies on Indigenous Knowledge. Office of Science and Technology Policy, Executive Office of the President
- Quinn, A. (2021). Transparency and secrecy in citizen science: Lessons from herping. *Studies in History and Philosophy of Science Part A*, 85, 208–217.

- Reymondon, M. (2024). Articulating homogeneity and diversity among scientific communities to favor the emergence of minority standpoints. Unpublished manuscript.
- Rosenstock, S., Bruner, J., & O'Connor, C. (2017). In epistemic networks, is less really more?. *Philosophy of Science*, 84(2), 234–252.
- Ross-Hellauer, T. (2022). Open science, done wrong, will compound inequities. *Nature*, 603(7901), 363–363.
- Santana, C. (2019). Mineral misbehavior: Why mineralogists don't deal in natural kinds. *Foundations of Chemistry*, 21(3), 333–343.
- Skrip, M. (2023). What Is Open Science? An Intro, Common misconceptions and advice. *Center for Geospatial Analytics News*, Nov 16 2023, NC State University. <https://cnr.ncsu.edu/geospatial/news/2023/11/16/what-is-open-science/>. Accessed 23.3.2024.
- Solomon, M. (2007). *Social empiricism*. MIT press.
- Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, 19(3), 387–420.
- UNESCO. (2021). *UNESCO Recommendation on Open Science*, UNESCO General Conference (Paris: UNESCO, 2021), <https://doi.org/10.54677/MNMMH8546>
- Veigl, S. J. (2021). Notes on a complicated relationship: Scientific pluralism, epistemic relativism, and stances. *Synthese*, 199(1), 3485–3503.
- Vellend, M. (2018). Are local losses of biodiversity causing degraded ecosystem function. In *Effective conservation science: Data not dogma* (pp. 27–31). Oxford University Press, USA.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *ScientificData*, 3, 160018.
- Zollman, K. J. (2007). The communication structure of epistemic communities. *Philosophy of Science*, 74(5), 574–587.