Transparency is Surveillance

C. Thi Nguyen

Abstract

In her BBC Reith Lectures on Trust, Onora O'Neill offers a short, but biting, criticism of transparency. People think that trust and transparency go together but in reality, says O'Neill, they are deeply opposed. Transparency forces people to conceal their actual reasons for action and invent different ones for public consumption. Transparency forces deception. I work out the details of her argument and worsen her conclusion. I focus on public transparency – that is, transparency to the public over expert domains. I offer two versions of the criticism. First, the epistemic intrusion argument: The drive to transparency forces experts to explain their reasoning to non-experts. But expert reasons are, by their nature, often inaccessible to non-experts. So the demand for transparency can pressure experts to act only in those ways for which they can offer public justification. Second, the intimate reasons argument: In many cases of practical deliberation, the relevant reasons are intimate to a community and not easily explicable to those who lack a particular shared background. The demand for transparency, then, pressures community members to abandon the special understanding and sensitivity that arises from their particular experiences. Transparency, it turns out, is a form of surveillance. By forcing reasoning into the explicit and public sphere, transparency roots out corruption — but it also inhibits the full application of expert skill, sensitivity, and subtle shared understandings. The difficulty here arises from the basic fact that human knowledge vastly outstrips any individual's capacities. We all depend on experts, which makes us vulnerable to their biases and corruption. But if we try to wholly secure our trust — if we leash groups of experts to pursuing only the goals and taking only the actions that can be justified to the non-expert public — then we will undermine their expertise. We need both trust and transparency, but they are in essential tension. This is a deep practical dilemma; it admits of no neat resolution, but only painful compromise.

In her 2002 BBC Reith Lectures on Trust, Onora O'Neill offers a biting criticism of

one of the ethical presumptions of modern life. People think that trust and transparency

go together, says O'Neill. And they think that transparency is wholly good. But actually,

she says, trust and transparency are deeply opposed.

> Transparency can encourage people to be less honest so increasing deception and
> reducing reasons for trust: those who know that *everything* they say or write is to be made
> public may massage the truth. Public reports may underplay sensitive information; head
> teachers and employers may write blandly uninformative reports and references; evasive
> and uninformative statements may substitute for truth-telling. Demands for universal
> transparency are likely to encourage the evasions, hypocrisies and half-truths that we
> usually refer to as 'political correctness', but which might more forthrightly be called either
> 'self-censorship' or 'deception.' (O'Neill, 2002a, 73).

These observations strike me as deadly accurate. [1] But O'Neill's discussion has seen

very little uptake. In much public life and academic discussion, transparency is still often

presumed to be an unalloyed good.

My aim in this paper is to develop O'Neill's criticism in some detail, filling in some of

the argumentative gaps from her brief presentation. At the same time, I want to

emphasize the desperate complexity of the situation. I think O'Neill is entirely right that

transparency is in tension with trust — and that too much transparency undermines trust.

---

[1] The Reith lectures occurred in 2002, the same year as the publication of her groundbreaking book, *Autonomy and Trust in Bioethics* (2002b). However, this transparency argument does not occur in such a bold form in the book — though there are quiet hints in that direction. My own guess is that, given the schedules of publication, the book was written some years earlier, and that the critique of transparency crystallized after the book was finalized. O'Neill returned to themes of accountability and transparency in later work (2005, 2006) but I find this later work less useful, for the topic at hand. In the later work, her eye had turned to worries about the managerial state and the structure of bureaucratic accountability – that is, formal schemes by which we impose sanctions to hold people responsible. She did not, so far as I can tell, return to the specific argument quoted here, which concerns the implications of transparency – that is, simply of making information available.

But I also think we also need transparency, precisely because we often have good reason to distrust. The result is a genuine practical dilemma, for which there is no painless solution. Transparency and trust each have their own costs and benefits, and we must decide which sacrifices to make in each particular circumstance.

I also want to argue that O'Neill actually underestimates the harms of transparency. She worries that, since an expert's actual reasons are inaccessible to non-experts, experts under regimes of transparency will be forced invent fake reasons for public presentation. According to O'Neill, transparency won't change the experts' actions, only their recorded justifications. Call this the *deception argument*.

The deception argument is certainly plausible. It is not, however, the end of the story. Transparency can have an even more intrusive effect. It can change what experts do, pressuring experts to only act in ways which readily admit of justification in non-expert terms. The demand for transparency can undermine the practical application of expertise itself. Call this the *epistemic intrusion argument* against transparency.

When we trust experts, we permit them to operate significantly out of our sight. There, they can use the full extent of their trained sensitivities and intuitions – but they are also freed up to be selfish, nepotistic, biased, and careless. On the other hand, we can distrust experts and demand that they to provide a public accounting of their reasoning. In doing so, we guard against their selfishness and bias – but we also curtail the full powers of their expertise. Transparency *pressures experts to act within the range of public understanding.*

The demand for transparency can end up *leashing* expertise.

There is another way to think about it that doesn't center expertise. Some considerations can only be understood by members of a group, based on their shared experience. Call these *intimate reasons*. Intimate reasons can include everything from the experience of a life lived under a particular form of oppression to the specific aesthetic experiences that arise around a particular shared sensibility. But transparency can demand that groups justify their actions in more publicly accessible terms — which pressures those groups to conceal or avoid using those intimate reasons. Transparency can threaten certain kinds of community life – the kinds that depend on mutual, experience-based, shared understanding. Call this the *intimate reasons argument*.

Both the intimate reasons argument and the epistemic intrusion argument concern the ways in which transparency can undermine the special kinds of sensitivity and awareness that can come with expert training or particular life experiences. This is not to say that transparency has only bad effects or comes from only bad motivations. Transparency can be good; it is often reasonable and sometimes necessary. It brings to light biases, errors, and hidden malfeasance.[2] But the point is that transparency emerges from distrust. Jeremy Bentham makes this quite explicit in his arguments for governmental transparency. Public officials are supposed to act in the public interest – but, Bentham worries, they are granted power, and power is essentially corrupting. Declares Bentham:

---

2 For an excellent overview of past arguments for political transparency along these terms, see (Kogelmann, 2021, Ch 1).

"…In the framing of laws, suspicion cannot possibly be carried to too high a pitch… every man ought to be presumed disposed to be guilty and endeavouring to be guilty to the purpose of legislative enactment" (Bentham, 1989, 15). Only by forcing the actions of public officials into the perpetual light of public awareness can we protect ourselves against public officials' abuses of power.[3]

Bentham's account does give us good reason to want transparency in political life. What he misses — and what O'Neill exposes — are the downsides of transparency. Transparency is a kind of surveillance — an intrusive form of monitoring. When sufficiently pervasive, it creates a kind of justificatory paranoia. To be under the eye of such bureaucratic surveillance is to be constantly being overseen by a distant, unsympathetic, uncharitable outsider. Such surveillance can be appropriate, when the consequences are sufficiently weighty and the threat of corruption sufficiently high. And in the matters of highest importance – and in cases where we have good reason to suspect corruption – we may genuinely need pervasive and aggressive surveillance. But surveillance has a significant cost, and we should decide if, and how much, of that cost we are willing to bear.

Surveillance is most justified when the overseen are likely to be corrupt or biased, and the overseers careful, sensitive, and well-intentioned. Surveillance is at its worst when the overseen are skilled and good-hearted, and the overseers unthinking, insensitive or

---

3 I owe my understanding of Bentham, and this particular quotation, to Brian Kogelmann's excellent history of the political theory of transparency (Kogelmann, 2021, Ch 1).

inexpert. If we accept Bentham's presumptions — that power accumulates in political leaders, and that it inevitably corrupts — then transparency over political leaders will usually be appropriate. But not all forms of transparency share those particular power dynamics. For every case of transparency of the politically powerful to the dedicated public, there is another case of, say, the transparency of a publicly funded LBTQ+ support program to a hostile public; or the transparency of a community arts non-profit to an indifferent state legislature; or the transparency of a small humanities department to a profit-obsessed board of trustees.

Transparency works to eliminate the non-explicit and the private. That is where corruption and bias live — but also sensitivity, expertise, and intimacy. There is no getting around this tension. Transparency can sharply curtail its subjects' freedom of action, which limits both their ability to do bad and to do good. We should not drop all forms of transparency, naively giving our absolute trust to every corporation, politician, and institution. But neither should we not try to eliminate trust entirely and attempt to render all processes utterly transparent. We need to find some workable compromise between trust and transparency. Transparency is not an unreserved good. It is rough medicine which we must take with caution.

Kinds of transparency

The term "transparency" is used in all sorts of contexts in very different ways, so let me start with a brief taxonomy. "Transparency" in its broadest sense, indicates any process in which some entity makes information about its own activities available, to be used in further decisions and actions. [4] Let's call this *broad transparency*.

Why engage in broad transparency? Schemes of transparency may be set up to serve at least two very different functions. First, we may wish to make the details of some process be made available so that the process may be assessable to outsiders. Let's call a scheme, designed for such a purpose, *assessment transparency*. Assessment transparency includes metrics and measures set up to check for corruption or measure performance. Second, we may want to make information available for borrowing, repurposing, and further usage. Let's call this *re-use transparency.* We see re-use transparency in the open-source movement over computer code, where a major goal is to allow other programmers to freely take and adapt code. We can also set up systems of transparency to serve both purposes. Scientific transparency often serves both aims: it makes research data available to enable scientists to check up on one another's methodology, and also to permit the re-use of that data.[5]

Let's distinguish transparency from another closely related concept: accountability. Transparency concerns those processes in which information is made available for some

---

[4] I adapt this definition from Turilli and Floridi (2009).
[5] I am drawing here on Kevin Elliot's (2020) extremely useful, and detailed, taxonomy of scientific notions of transparency. My discussion here simplifies some of his taxonomy for brevity's sake.

further decision; accountability concerns those formal processes for evaluating and imposing sanctions on organizations and individuals for their failures. Transparency and accountability are often deeply interrelated, but we can have transparency without a formal system of accountability. We can simply publicize information and then let the chips fall where they may. My concern here centers on the harms of transparency itself. In some of the cases we will consider, the harms of transparency will depend on the workings of formal systems of accountability; but in other cases, transparency can harm without the need for formal systems of sanction. This is why it is useful to see transparency as a form of surveillance. Surveillance may be more burdensome when it is tied to some formal system of sanctions – but the harms of surveillance do not require them. Constant scrutiny is, in and of itself, an intrusion.

Much depends, however, on who that information is made available to. In *expert-aimed transparency*, information about a process is made available to other experts in that process – like when one epidemiologist makes their data available to other epidemiologists. In *public-aimed transparency*, information is made available to the public in general or their non-expert representatives.[6] Public-aimed transparency will turn out to be the most intrusive form.

What does it mean to make information "available" to the public? Obviously availability requires at the very least, the simple ability to access the information. But recall

---

6 This is a spectrum distinction, rather than a binary decision, since the difference between experts and non-experts is itself a spectrum distinction.

that transparency is information availability *for the sake of making decisions.* In order for the public to act on that information, the information must be comprehensible to them. Suppose an economist dumps their spreadsheets onto a public server. If those spreadsheets are only decipherable to another expert, they will be extremely difficult for the public to use in for any kind of decision-making. This information is "available" in some minimal sense, but it isn't readily usable by the public. This is why, as we shall soon see, real-world processes of public-aimed transparency often translation and simplification.

I will focus, for the core of this paper, on one type of transparency: where *expert domains are made transparent for the purpose of assessment by the public.* The full name for this, in my taxonomy, is *"*public-aimed assessment transparency over expert domains", but for brevity's sake, I will just refer to it as "public transparency". In particular, I will focus on those cases in which public transparency involves significant informational transformation for accessibility – often in the form of metrics and indicators.

I focus on such cases for two reasons. First, that sort of public transparency is actually extremely common. It visible everywhere in public life, from non-profit charities to the management of higher educational institutions. Second, it is where we see O'Neill's effect most vividly. Metrified public transparency turns out to be most sharply intrusive form of transparency.[7] And studying transparency's most intrusive form will help us see how we

---

[7] In focusing on public oversight over expert domains, my discussion will differ sharply from the standard

might manage its harms. Departures from the stringent demands of public transparency, towards other forms of transparency, will mitigate – but not wholly dissipate – those burdens.

The argument I'm making is, I hope, not an elitist one. I am not suggesting that there is some separate class of essentially better-educated experts who "know better" than the masses. Rather, we must admit that human knowledge is so vast that no person can master more than a sliver of it. An expert nuclear scientist is also likely inexpert in medicine, carpentry, aeronautics, chemistry, epidemiology, agriculture, and art history. Perhaps one human being can straddle a handful of expert domains, but nobody can come close to comprehending the whole expanse of human knowledge. The vast and sprawling nature of the modern epistemic landscape requires us to become hyperspecialists in narrow domains (Millgram, 2015). The problem with public transparency is not that the "public" are somehow ill-educated or unintelligent. We would

---

philosophical discussion about political transparency. This discussion has largely concerned oversight by members of the public over their duly elected officials and representatives. According to that discussion, though public officials may gain some expertise, the matters with which they are concerned are not usually esoteric. As some have put it, public officials are not supposed to be distant experts, but simply proxies for the public. In that conversation, the primary problem is this: how do we get officials to behave as public proxies — to actually serve the public interest — rather than pursuing their private interests? (Again, see Kogelmann [2021, Ch. 1 and 2] for a thorough overview. I owe the distinction between expert cases and proxy cases to discussion with Kogelmann.) Though such questions are central to the discussion in political philosophy, they are not the central cases for our present discussion. I am most concerned with cases where experts — and not public proxies — are put under regimes of public transparency. Key cases for me, then, will be those of expert scientists, doctors, art historians, educators, and economists, put under the demands of public transparency. I take my usage of "esoteric" from Alvin Goldman, who defines "esoteric statements" as those statements which "belong to the relevant sphere of expertise, and their truth-values are inaccessible to [a novice]" (Goldman, 2001, 94).

still face the same problem in a world of experts. The nature of public transparency puts oversight in the hands, not of the relevant expert within the public, but of the aggregated public as a whole. And the aggregated public is inexpert about most things, even if every member of the public is expert in something.[8]

Some case studies

To better understand the effects of public transparency, we can draw on a wealth of empirical work on the effects of transparency and accountability on institutions, bureaucracies, and the people who work inside of them. This works comes to us from fields such as sociology, anthropology, and history. Much of this work focuses on one particular form of public transparency – the imposition of quantified metrics and indicators as part of oversight – exposing its painful consequences. Let me offer a few case studies drawn from this literature. Again, my goal here is not to discuss every form

---

8 One might think that the aggregated public is, in aggregate, an expert at everything. Helene Landemore has suggested as much: that cognitively diverse groups trump non-diverse groups since they aggregate knowledge (Landemore, 2012). But, in order for that group to effectively leverage its diversity, it must be able to properly recognize a good solution. What happens when a good solution involves expertise to identify? The general public, when aggregated by a simple vote, is an aggregate non-expert — because most people are non-experts on most things. In order for group deliberation to yield the right judgment of expert matters, the group will have to be convinced by experts in the right direction. This works in those cases where experts can successfully explain their reasons to non-experts, given the time permitted for group deliberation. But in any terrain that is genuinely esoteric, non-experts will not apprehend those reasons. Any plausible, real-world group deliberation won't provide for enough time for, say, climatologists to teach the public enough statistics, climatology, and computer modeling theory to properly appreciate, for themselves, the forcefulness of the models for climate change. So the deliberative/aggregative mechanism results in aggregating a lowest common denominator of understanding, rather than highlighting the appropriate expert. See also (Gaus and Hankins, 2017).

of transparency, but to give the starkest illustration of transparency in its most intrusive form, so that we may better manage it.

The first case concerns transparency for charitable non-profits. Donors are often concerned that the charities to which they donate are legitimate and well-managed. This has led to the rise of charity watchdog organizations, the most famous of which is Charity Navigator. Charity Navigator performs extensive audits of various non-profit charities, so that the public can make choose the most efficient and the least corrupt. Charity Navigator operated, for much of its existence, with one key metric: low overhead. A non-profit's overhead is defined as the amount of funds, donated by the public, which are spent internally. The more a non-profit spends on internal overhead — administration, employing workers, setting up its own base of operations — the worse its score. As many experts in the non-profit world have pointed out, however, overhead turns out to be a lousy metric. It forces non-profits to compete for better Charity Navigator rankings — and through them, more donations — by reducing internal costs. Past a certain point, reducing internal costs actually damages the ability of non-profits to function effectively. Non-profits under this mode of transparency have been pushed to operate on shoe-string budgets, employing an inadequate number of workers. But, importantly, the metric *seems good* on quick inspection — good enough that many external donors use that metric as a guide to their donations. (Brandt, 2013; Coupet and Barrett, 2018; Wellar, 2018).

The second case comes from sociologist Jennifer Lena's history of the American arts –

and who controlled what counted as "art". Early in its life, the National Endowment for the Arts came under significant political criticism for possible corruption, pork-barreling, and nepotism. Congressional oversight made the following accusation: that the NEA was mishandling funds because the theatrical productions which it funded had brought in poor box-office returns (Lena, 2019). Notice how the demand for publicly comprehrensible justification structures the exchange. The NEA could have responded by offering various specialist justifications: that they chose to fund a particular theatrical production because it was groundbreaking, or brilliant, or daring. But those justifications all would have required some degree of expertise to assess. Judgments of artistic quality require significant expertise. And, insofar as the public doesn't have that expertise, they can't distinguish between sincere expert evaluations of aesthetic qualities, and, say, nepotism and old-boy-network elitism masquerading behind a facade of expertise. Thus, Congressional oversight demanded justification in terms of a metric accessible to the general public and its representatives: profits.

Lena's case is a particularly glaring example of a larger trend. In *Trust in Numbers,* Theodore Porter offers a historical study of why political institutions tend to seize on oversimplified, objective-looking measures. The centerpiece is Porter's painstaking analysis of the history of the cost-benefit analysis. Cost-benefit analyses are attempts to offer a quantified justification for a decision, by weighing the estimated financial benefits weighed against the estimated costs. The cost-benefit analysis, says Porter, tends to

replace richer forms of analysis of weighing complex and conflicting values. The cost-benefit analysis tends to neglect any values that cannot be expressed in terms of financial benefit — reducing a difficult multidimensional value decision to a single, mechanical calculation.

The cost-benefit analysis, as a mode of justification, originated with the American Army Corps of Engineers. One might be tempted here, says Porter, to offer a simple explanation: that engineers like numbers, so they tend to offer quantitative justifications. But, as a matter of fact, the American Army Corps of Engineers (and its French equivalent) operated for over a century without cost-benefit analyses, opting instead to make decisions by considering complex, plural, qualitatively-expressed values. The cost-benefit analysis became important only when the Army Corps was put under direct Congressional oversight, and the Congressional demand for transparency. Congresspeople would constantly accuse the Army Corps of pork-barreling — of choosing to pursue one project, in one state, over another county in another state, out of nepotism. The qualitative weightings of the Army Corps were insufficient to dispel those accusations. Thus, the Army Corps began to produce cost-benefit analyses in order to publicly defend against charges of nepotism. Congress trusted the objective, mechanical and auditable reportage of the cost-benefit analysis over less auditable analyses which drew on more complex renditions of value — often arrived at through experienced, expert assessments which were inaccessible to Congressmembers (Porter, 1995, 114-190).

The third case study is drawn from Sally Engle Merry. Merry is an anthropologist specializing in human rights issues. She took part in the United Nations efforts to generate various indicators and metrics, like the Human Development Index, which scores every country on their support for human rights and quality of life. In *The Seductions of Quantification*, Merry turns her anthropologist's eye on the committees and institutes which produce such human rights indicators. Merry offers a particularly devastating study of the US State Department's *Trafficking in Persons (TIP) Reports*, which rates various countries on their performance in reducing sex trafficking. Says Merry: since the *TIP*'s formulation reflects the State Department's policing mindset. The *TIP* measures the reduction of sex trafficking by counting the number of arrests and convictions of sex traffickers. But, says Merry, this particular metric forced one particular approach to the reduction of sex trafficking: policing. But there are other approaches. Sex trafficking typically arises in areas of great poverty. The systematic reduction of poverty, and the creation of better job opportunities, will reduce sex trafficking. But such economic interventions drive out sex trafficking without police action. Reducing poverty actually reduces the conviction rate of sex traffickers. So successful economic interventions are reported by the *TIP* failures to control sex trafficking.

The *TIP* is a particularly heinous instance of bad metrics, but it reflects a larger issue. Sex trafficking is incredibly difficult to measure, in part because what counts as sex

trafficking is complex and ambiguous,[9] and because much of the trafficking occurs far away from any institutional eye. Attempts to get clear on the success of interventions into sex trafficking require the complex application of expertise in economics, sociology, and political science. Conviction rates are obvious, public, and easily entered into institutional record-keeping — so they get used in judging the outcomes of large-scale interventions (Merry, 2016, 112-160).

The last case study is drawn from my own experience with university bureaucracy. In the governance techniques of modern universities, departments are expected to perform "assessment". Assessment involves providing evidence of how well the department achieves various specified "learning outcomes". Assessment reports are made available to upper administration and various oversight groups. In publicly funded institutions, these assessment reports are usually intended for use by state legislators and the public at large, for the purposes of public oversight. Some assessment techniques involve oversight of one department's outcomes by another department. One philosophy department might, say, send writing samples from its students to a member of a different philosophy department — who would rate those samples on their clarity and acumen. This is a form of expert transparency. Such expert transparency might satisfy one kind of worry: if we trusted the profession of philosophers in general, we could use one philosophy

---

9 Merry's example: suppose one woman voluntarily crosses a border to work in a brothel, to avoid starvation. The next year she brings her sister, who also faces starvation. Is the first woman now an international sex trafficker?

department to check if another one had gone off the rails. But it will not satisfy a broader worry — that of, say, a state legislator who suspects that all philosophy is bullshit. Some universities have, in response to such worries, began moving towards assessment metrics that focus on more accessible measures of student success like graduation rate, graduation speed, and post-graduate employment rates and salaries. Notice, once again, that such measures certainly present themselves as objective. Their application is entirely auditable to outsiders. They involve mechanically collectable data, whose importance can be understood without any particular expertise. But they also don't track what most philosophy professors probably care about: teaching critical thinking skills; cultivating intellectual virtues such as reflectiveness, intellectual humility, and curiosity. This is by design. Recognizing the successful cultivation of those virtues is a matter of some expertise. Depending on that mode of assessment would put the process of evaluation of a department outside the full understanding of non-experts. It would force the non-experts to trust. In order to eliminate the need for trust, we need to find an evaluative mechanism whose workings are entirely available to non-experts.

Here are some opening observations about these case studies. First, the desire for public oversight is perfectly understandable. Often, public funds are being invested into some group of experts. And the worries about bias and corruption among experts — or even just disconnection from the public interest — are also quite reasonable. But in each case, when we open up some expert activity to public inspection, we can distort the goals

of those experts. In order for the assessment procedure to be made available to the public — in order for it to be auditable by outsiders to the discipline — it must be put in terms that are accessible to the public. But insofar as the most appropriate terms of assessment require expertise to comprehend, then the demand for public accessibility can interfere with the application of expertise.

There are two variations of the problem. The primary problem lies in the overseers' lack of expertise. Let's call this the *inexpert oversight* problem. A secondary problem is that, in most circumstances, overseers can commit relatively few cognitive resources to each oversight task. For example, administrative oversight is usually performed by upper-level administrators, CEOs, or the public at large — all of whose attention is divided among a vast number of tasks, including a large number of oversight tasks for different departments. Let's call this the *information pipeline* problem. Thus, even if the overseers might in principle have the capacity to properly evaluate the domain, time pressure forces them to resort to simpler modes of evaluation. The scientist who becomes a CEO may have the scientific background to properly evaluate his company's research departments, but not the time. The result can be functionally similar to the inexpert oversight problem. Notice, too, that public transparency suffers from both problems: the public is both inexpert and rushed. And, in societies deeply invested in public transparency, the public will be especially rushed – since it must split its attention between managing a vast number of transparent expert processes.

The problem here is deep. The public has a reason to want experts to do things in the public interest, especially when the public is funding the venture. The public has a very good reason to suspect that often, experts will wander away from the public interest. But a sufficiently intrusive form of direct surveillance by the public will prevent well-intentioned experts from succeeding in adequately applying their expertise in the public interest.

This problem does not admit of a neat solution. Proper trust is a delicate dance. Annette Baier famously argued that trust involved making oneself vulnerable to others (Baier, 1986). I am suggesting, here, that vulnerability approaches us from both ends. When we trust, we leave ourselves vulnerable to exploration. But the failure to trust adequately leaves oneself vulnerable in another way — not to betrayal by the trusted person, but to all sorts of errors and misapprehension. A person might refuse to trust doctors, for fear of making themselves vulnerable to manipulation — but this sort of machismo leaves them more vulnerable to illness.

To properly balance trust and transparency, I think we need to understand the dangers involved with both. The dangers of corrupt or biased experts have been exceedingly well-charted.[10] So I will focus on the dangers of transparency, particularly its most intrusive form: public transparency.

---

10 For excellent contemporary surveys, see (Landemore, 2012, 1-88; Kogelmann, 2021, Ch. 1; Elliott 2020).

Epistemic intrusiveness

Here is the epistemic intrusiveness argument, in full. A reminder: "public transparency" is short for "public-aimed assessment transparency over expert domains."

The Epistemic Intrusiveness of Public Transparency

1. Public transparency demands that the experts justify their actions to non-experts.

2. Many of the actual reasons for expert action are not available to non-experts.

3. The demand for transparency creates pressure for experts to offer justifications for their action in terms available to the non-expert.

4. This pressure distorts, to varying degrees, expert action and communication.

The linchpin of the argument is 2 — that expert reasons are often unavailable to non-experts. Any adequate understanding of expertise, however, supports the claim. First, expert reasoning often requires expertise to assess. This is true even when the reasoning is fully explicit. The climate change scientist's evaluation of computer climate change models; the statistician's assessment of whether a statistical argument was improperly p-hacked: these evaluations may proceed from arguments and evidence that have been made fully explicit, and the evaluation process itself may be fully explicable. However, proper evaluation depends on a significant degree of expertise.

Even more telling, however, are cases where the successful deployment of expertise requires the application of tacit or implicit knowledge. Expertise is acquired through long training and enormous experience, often emerging in the form of skilled perception or trained intuition. A skilled pediatrician can tell at a glance when a baby is ailing and needs immediate attention, through an intuitive synthesis of perceptions of skin tone, movement, eye contact, and more. The skilled surgeon has an intuitive feel for which cutting angles would be more dangerous. And a skilled teacher can tell the difference between genuine reflective inquiry and a mere parroting of ideas.[11]

Instances of tacit knowledge make the cost of epistemic intrusiveness especially clear. As Hubert Dreyfus puts it, experts don't operate by mechanically applying clear, simple, easily explainable rules. That's what a beginner does. Such rules help one make the first stumbling moves in a domain. But as one acquires experience and collects feedback, one's sensitivities, perceptions, and intuitions begin to break away from those simple, explicit rules for action. However, such intuitions are hard to justify in explicit terms and tidy arguments. They depend on a vast storehouse of experience and training.[12] To the extent

---

11 The literature on skills is vast. Though many of the details in the philosophical literature are still being debated, I am relying here here on minimal claims about the nature of implicit knowledge and skill, which are generally accepted. The locus classicus for modern discussions of skill is in (Dreyfus and Dreyfus,1986). (Selinger amd Crease, 2002) offers a useful criticism of the Dreyfusian approach, pointing out that they fail to account for the possibility of expert bias. I take my own account to be an attempt to fuse Dreyfus' confidence in experts with Selinger & Crease's worries. For a modernized version of the key insights, which integrates a philosophical account of skill with contemporary research in cognitive science, see (Fridland and Stichter, forthcoming).

12 Dreyfus' approach has been criticized for presuming that all expert deployments of skill are non-explicable in this way. Obviously, critics point out, expertise involves some explication. That observation, though it may damage the strength of Dreyfus' original argument, won't affect the epistemic intrusiveness argument. For my argument only relies on there being *some* expertise that is tacit in this way. Transparency demands block trained expert perceptions and intuitions when they occur. Such

that an expert is confined to using reasons that can be made explicit, they cannot fully deploy those trained abilities. The more intrusive a regime of public transparency, the more it will force experts to reason like beginners.

There are a variety of ways and degrees to which expertise can be distorted by the pressure of public transparency. Let's consider some basic options:

Deception: Experts will come to *invent*, for public consumption, justifications for actions that are quite different from the actual justifications used in their deliberations.

Limitation: Experts will come to *limit their actions* to those for which they can find some genuine public justification.

Guidance: Experts will come to *seek or prefer* publicly available reasons in their own deliberations about how to act.

Deception is the least intrusive effect. An expert might be forced only to Deception, when external oversight is sufficiently lax. O'Neill only considered the possibility of Deception. She had been focused on bioethics cases, and I suspect that, at the time she wrote, the regimes of transparency were for less strict for doctors, so Deception was still an available response. A doctor might make a medical decision, acting on their trained instincts. The doctor might then, for the purposes of the paperwork, invent some reason

---

perceptions and intuitions need not compose the whole of expertise for the transparency demand to be damaging, so long as some important part of their expertise involved such implicit skill.

that fits the official diagnostic criteria. If the system relies on the doctors themselves to record the relevant evidence, then there's plenty of room for them to fudge it on the paperwork. Such a looser structure of oversight affords significant *leeway*.

But Deception becomes harder as leeway diminishes. When external overseers gain the ability to perform independent checks on the experts' claimed justifications, experts are forced to Limitation. In Lena's discussion of the Congressional oversight over the NEA, Congress didn't have to take the NEA's word for it. They were evaluating NEA funding in terms of box office receipts, and those receipts were a matter of public record. The NEA had no room to fudge. The same hold of university assessment of departments by student graduation rates and speeds. Those metrics are recorded and maintained by the university in general — so once again, an individual department has no wiggle room with those metrics. Insofar as their actions are required, through formal regimes of accountability, or pressured, through the informal public gaze, to meet those external oversight criteria, they must limit their actions to those that will match those criteria.

But Limitation still leaves some room for expert reasoning. Limitation doesn't require that the experts' actual reasons for selecting something align with their publicly declared justification. It simply demands that the expert produce some public justification. So experts could narrow the option space of possible actions to what they can provide public justification for, but then choose from within that options pace using expert reasoning. We could imagine, for example, that the NEA knows that they can only fund those

theatrical performance which will have decent box office returns. They could limit their candidate pool to those shows that promised minimally decent returns, in order to satisfy the demands of the overseers. But then they could choose from within that candidate pool using their own expert sense of artistic merit. Limitation places an exterior constraint on the variety of possible actions, but it doesn't force experts to tak, as their primary target, the maximization of box office returns.

Things are different with Guidance. In Guidance, the experts come to use those public justifications centrally in their own reasoning, as positive guides for choice. In Guidance, the NEA would begin to pick projects directly for their box office potential. Philosophy departments would make hiring and curriculum decisions in direct pursuit of higher graduation rates. Anti-sex-trafficking activists would start chasing conviction rates. In Limitation, the demands of public justification create an external constraint, preventing certain actions. In Guidance, public justifications are internalized as sources of positive reasons. To the extent that those inexpert-available reasons dominate the deliberation processes of experts — to the extent that experts divest themselves of their expertise in deliberation — then their expertise will be undermined.

Empirical research indicates that this sort of motivational capture does, in fact, occur. Clear, simple metrics, imposed from the outside, can become incredibly motivating. Law students and law school administrators are captured by the *US News & World Report* law school rankings; Army Corps Engineers come to use cost-benefit analyses as their primary

guide; politicians become entirely focused on moving their countries up the UN's various rankings (Espeland and Sauder, 2016; Porter, 1995; Merry, 2016).

The degree of intrusiveness of Guidance, however, depends on the precise relationship that experts have to those public criteria. In *incentivized guidance*, experts are offered some incentives which correlate to their performance on those publicly available metrics. Law school administrators might be offered bonuses for moving their school up the rankings; school districts get extra funding for performance on a standardized test. Academics are more likely to be promoted if their citation rates are high. Notice that in incentivized guidance, the external metric provides some direct reason for action. It is no longer a mere limitation to be met, but a target to be directly pursued. But since it is only an incentive, experts still have room to trade off between the incentive and their other values. A professor can still choose to trade off the career value of more citations against their love of interesting, but deeply unpopular, topics. But in *internalized guidance*, experts take on those public criteria as their primary goal in the space. If an academic begins to take on citation rates as their primary target for publication, then there is nothing left to trade off against.

According to the empirical research, both incentivized guidance and internalized guidance do occur — most clearly in cases with metrified targets. Espeland and Sauder say that when the *USN&WR* started publishing its rankings, administrators were often incentivized to pursue the rankings, but treated it as a complex trade-off decision,

weighing the financial importance of the rankings against their loyalty to other legal values. But within a decade, most administrators seem to have fully internalized the rankings as their primary goal. It sets the terms of their success (Espeland and Sauder, 2016, 60-163). Elsewhere, in my discussion of gamification, I've offered an account of why we might internalize such metrics, in a process which I call"value capture". When we internalize such metrics, we shield ourselves from the painful complexities of dealing with values in their true, subtle, and inchoate form. Life becomes something like a game, with a clarity of value and purpose. And we gain, not only internal clarity, but external clarity – it becomes effortless to justify ourselves to others, when we have all gotten on the same value standard. This is why gamification is so effective. When something like Twitter offers us a simple score for our communicative acts, we can get the bliss of games by taking that score as our main goal in the activity (Nguyen, 2020a, 189-215; 2021a; 2021b). These explanations work for any sort of metric or gamification, not just ones created for transparency. But we can see at least one pathway by which public transparency could lead to internalized guidance: transparency generates metrics for public use, and then experts internalize those metrics for the pleasures of value clarity.

Crucially, in value capture cases, transparency can undermine expertise without the need for formal regimes of accountability. Experts simply internalize metrics because of, say, their beguiling clarity. This then changes their behavior. What might this look like? Consider the existence of various aggregations of academic citation rates, like the H-

index, which provides a quick way to quantitatively assess any academic's career.[13] Every active scholar's H-index is publicly available, and lets any onlooker quickly compare the success of one academic to the success of another. In some university environments, success on those metrics is formally tied to various incentives and sanctions. In other institutional invornments, those metrics are not part of the formal evaluation procedure – but the mere existence of these public metrics seems to be deeply motivating to many a scholar, anyways.

To sum up: in Guidance, experts limit themselves to using publicly available reasonings in their actual deliberation. Since publicly available reasoning is inexpert reasoning, then, to the extent that experts confine themselves to such reasoning, then they deprive themselves of some of the key powers of their own expertise. To the degree that Guidance takes hold, then expertise is undermined.

This may sound absurd to some. I think it not only plausible, but so familiar and pervasive that it easily passes beneath notice. The clearest cases are those where the demand for transparency brings experts to abandon the use of their trained intuition, and to instead take action by applying some clear policy based on public criteria. In many bureaucratic circumstances, what this looks like is, say, experts ignoring their own skilled

---

[13] The H-index is calculated in the following way: a scholar's H-index is n, when they have authored n articles that each have at least n citations. For example, a scholar that has published 6 articles, with 200, 5, 3, 1, and citations has an H-index of 3 – they have published 3 articles with at least 3 citations. At the time of writing, every active scholar's H-index is automatically computed by Google Scholar and available to all,with a quick Google search.

perception and trained intuition about how to behave and acting from "best practices" —

which are often clear and mechanically applicable guidelines for action, which offer

deterministic recommendations for action based on publicly accessible criteria.

The degree of intrusiveness here depends on the specifics of a particular regime of

transparency. For one thing, there are two different locations in a process where we could

demand transparency: during the process, or at its outputs. In *output transparency,* we

demand that experts offer justifications for their actions in terms of comprehensible,

monitorable outputs for action. In *deliberative transparency*, we go further and demand

that experts offer comprehensible, assessable reasons for particular actions. Output

transparency includes assessing doctors in terms of lives saved, or assessing police in

terms of cases closed, or assessing a university department in terms of its graduate rate

and alumni salary. Deliberative transparency includes checking on doctors' diagnostic

criteria for providing a particular drug to a patient; or asking that department's budgetary

requests be justified, line by line, in publicly understandable ways. Output transparency

offers more leeway. Experts can take whatever pathway their expertise deems best, so

long as they hit a publicly visible target. Deliberative transparency is more intrusive. It

puts, not just the expert's end-products, but also their practical reasoning, under the

microscope.

We can now get a better sense of some of the dimensions of intrusiveness possible in

different arrangements of public transparency. Public transparency undermines expertise

to the extent that it pressures experts to employ non-expert reasoning. We've seen two dimensions along which transparency can intrude. The first dimension: how far does transparency intrude into the expert's thinking? We have seen four levels of increasing intrusion: Deception, Limitation, Incentivized Guidance, and Internalized Guidance. The second dimension: how much of the process does transparency monitor? We have seen that Output Transparency is moderately intrusive, but Deliberative Transparency is deeply intrusive. The most intrusive form of public transparency, then, is when we get Internalized Guidance Deliberative Transparency. And I think that form is actually quite common

I am not claiming that public transparency is all bad. My claim is that public transparency simultaneously exposes incompetence, corruption, and bias *while at the same time* limiting experts' freedom to use their trained expert abilities. Public transparency involves a trade-off. We will often want to take the trade, to some degree, but we should understand what we are losing when we slide towards more transparency. The deep problem here is that some very different things can be happening when experts act beyond our understanding – when they pursue goals that we cannot fully comprehend. Perhaps the experts have gone off the rails. Perhaps these politicians have become corrupt; perhaps these arts non-profits administrators are just handing out the grants to their art school buddies; perhaps these academic philosophers have drifted away any matters of real public interest in their pursuit of dusty obscurities. Or: perhaps the experts are responding to reasons only they can see. Perhaps the politician's experience

lets them see why this apparently simple and easy solution won't work, because of the

particular history of hostility between two neighborhoods. Perhaps the arts administrator

sees the promise of this indie comic artist and recognizes they need some funding and

time to develop, before the rest of the world can catch on. Perhaps these philosophers

understand that these complex and arcane technical issue in metaphysics aren't really

obscure at all, but are deeply relevant to thorny problems of great public interest, like the

socially constructed nature of race and gender. The problem is that expert action is often

deeply incomprehensible to the rest of us. In expert realms, corruption and the genuine

pursuit of subtle goodness are very hard for the non-expert to tell apart.

Can't We Have Clear Outputs?

The taxonomy of public transparency I've offered might now give one some hope for

an easy solution. Perhaps we just need to restrict ourselves to the right form of

transparency. Deliberative transparency seems too intrusive, but perhaps we can settle on

output transparency as a decent compromise. In particular, the methodology of output

transparency pairs nicely with other considerations about public control. Consider the

case of publicly funded institutions like city governments and universities, and publicly

funded research and art. Since the public is providing the funding, shouldn't the public

get to set the goals of that funding?[14] And shouldn't the public get to judge the use of their funds, by measuring how well those goals have been achieved?

It is plausible to think that the public should have a significant hand in determining the goals of publicly funded work. But there is a great difference between the public setting the goals of an action in some broad sense, and the public setting the precise criteria for evaluating the achievement of those goals. It is one thing for the public to set as a goal for a university "a good education"; it is another thing to specify that the university's progress towards that goal will be measured in terms picked out and fixed by the public, such as student graduation rate, graduation speed and average alumni salary.

So, here are two possible relationships a public could have with the experts it funds. First, the public could engage in *direct specification*. That is, the public could specify in detail what their funded experts are to pursue, leaving no room for expert interpretation of the goal. Second, the public could engage in *value handoff*. They could specify some value or goal to be pursued in a more general or schematic manner. Experts could be told to support, say, a good education, or the arts, for the sake of the public good. And then experts could fill in the details of that goal — using their expert understanding to help specify what counted as, say, a good education, or valuable art.

The difference matters insofar as we think expertise involves, not just on picking the

---

14 Philip Kitcher argues convincingly against the idea that science is a pure pursuit of knowledge which should be conducted free of public oversight. He argues, rather, that the public can and should set the goals for publicly funded scientific work (Kitcher, 2001).

right means to an end, but in properly describing and understanding that end. A goal, one might think, is best understood by the relevant experts. Experts better grasp how to flesh out those goals. They better see the values that are on offer and the targets that should be pursued. (In many cases, with artistic experts, one might think that part of their aesthetic expertise involves a refusal to set exact criteria of success – that aesthetic expertise involves, instead, a sensitive openness to being surprised by an artwork or a creative process.)[15]   But, once again: permitting experts to flesh out our goals for us involves a substantial act of trust. It makes the full and detailed specification of a goal inaccessible to non-experts.

There's a significant tension, then, between the subtlety of certain goals and our desire for public transparency. In order to make the outputs of a process fully transparent, we need a publicly available method for evaluating the outputs. This will be fine for those processes whose successful outputs can be readily specified in non-expert terms. But it constitutes a real limitation in those cases where the proper targets of expert action are best understood by the experts themselves. And such subtle values are quite common. As Tal Brewer puts it, in many cases, it takes long engagement with an activity to fully grasp its value. Beginning art appreciators don't see the full value of art appreciation. They start

---

15 In analytic aesthetics, this thought is often expressed in the view that aesthetic properties — and thus thick aesthetic value — are not inferentially derivable from the non-aesthetic properties on which they supervene (Sibley, 1959; Dorsch, 2013). Or, as Jerry Levinson puts it, we turn to real art to be surprised by it and learned from it. The stuff that we turn to expecting a reliable effect — that's sentimental art, or porn (Levinson, 2005).

because they trust others that there is something of value there, and the value of each art form is only unveiled through years spent grappling with it.[16]

This is easiest to see when you think of how outsiders tend to hastily appraise your beloved activities. I'm a rock climber; outsiders tend to think that the purpose of rock climbing is raw displays of strength or courting deadly danger. It takes long experience with rock climbing to fully see its various values — the way it opens up graceful motion and new possibilities of fine and beautiful balance. Think, too, of the experience of teaching any introductory class, be it literature, physics, or philosophy. New students are so unfamiliar with the area that they have a poor feel for its value. Simple explanations will do nothing; only time spent enmeshed in the activity will fully reveal its value. And properly understanding the value of an activity is required to properly articulate its appropriate aims.[17] Many people come to exercise thinking its sole value lies in, say, losing weight or looking muscular. Time spent with physical fitness can slowly reveal the fullness of its value — its effects on mental health, on bodily well-being, the joys of movement itself. Long experience is often required to fully recognize the value of an activity and to recognize its appropriate aims. To the extent that this is true for an activity, then, then the

---

16 See (Brewer, 2011). (Callard, 2018) also addresses the issue of how experience with an activity is required to understand the value of the activity. For a fuller discussion of the relationship between trust and aesthetic growth, see (Nguyen, forthcoming b).

17 Here I differ from Brewer. Brewer distinguishes between production-oriented accounts of activity, which find the value in the output of the activity, and accounts of activities which are intrinsically valuable. His argument is that expertise is required to see the intrinsic value of activities. But I take my argument and examples to show that in many (though not all) cases, expertise is also required to see the values of an activity's product.

demands of public output transparency will distort that activity's aims, and its practitioners' ability to achieve that value.

The attraction to direct specification, then, turns out to be premised on the view that there is *nothing to learn* about the nature of appropriate goals from the relevant experts. We, the public, set the targets, and the experts just apply their skill to hitting the targets that we set. With value handoff, on the other hand, we are trusting experts to be scouts into the nature of value — to be sent with a vague gesture into some complex terrain, and then to come back and enlighten us, not just about better instrumental methods for achieving what we already are already valuing, but with a more refined understanding of what there is to value. And in many domains, this seems quite plausible. Ecologists better understand the value of dynamic ecosystems; marine biologists better understand the value of the oceans; jazz experts better understand the value of complex improvisation. Experts, we might think, are not just tools to get a job done; they are educators that teach us something more about the nature of the job we should wish them to do. We might call this the expert-ladenness of value.

This expert-ladenness, however, does not hold for all expert domains. It applies to those areas where the values in question are complex and relatively independent of personal peculiarities — like the value of natural ecosystems. The argument is least applicable to those cases where the experts are trying to help members of the public achieve their own, autonomous values. For example, consider the issue of patient

autonomy in matters of their own health. We might think that the value — the health of the patient — is being pursued here is not exclusively the domain of the medical expert, but something much more complicated. The core value to be pursued should be set autonomously by the patient — though understanding how various interventions will impact patient health is a matter of much expertise. In that case, the entity with the best grasp of the value is some complex union of the patients and the medical experts. In those cases, the leashing of expert knowledge to a non-expert's judgment is not a distortion of the true value, because the non-expert has the best access to key parts of that value. Or, we might say: in such cases, each member of the public is actually the relevant expert on their own personal values.

This helps us understand why some instances of transparency seem far more palatable than others. Consider a crucial recent case: the development of COMPAS, an algorithmic system that predicts the likelihood of recidivism in a given subject and uses that to generate sentencing recommendations. As many philosophers of science have pointed out, transparency is important in cases like this. For one thing, the cost of failure is so high, that we can justify the burdens of increased vigilance. But perhaps more importantly, such projects have complex relationships to the public good.[18] Developing machine

---

[18] For some useful recent discussions, see Douglas (2017) and Schroeder (2018). There is, however, a tension between those discussions and mine. For example, Douglas' suggestion is that scientists should disclose their values, so that citizens can decide which values align with their own. But notice that the considerations I've adduced -- about how experts have a better understanding of value – pose some potential problems for Douglas' suggestion. Insofar as experts are explaining their values for the sake of influencing funding, and public policy, then we get the same pressures to deceive or simplify.

learning algorithms involves a series of complex, value-laden choices. For one, the developers must choose between letting the algorithm factor in socioeconomic status. Using socioeconomic data increases accuracy in predicting recidivism, but also seems to violate the principle of equality under the law. More accurate systems are likely to recommend higher sentences for people with lower socioeconomic status – thus perpetuating socioeconomic inequity.

In this choice, and many others, algorithm developers are choosing between substantively different conceptions of fairness (Biddle 2020, 7-13). But there is little reason to think that machine learning experts are also experts on fairness. In particular, machine-learning experts are unlikely to be among the population most impacted by the systems they are developing, and unfamiliar with the ways in which different concepts of fairness play out in the lives of socioeconomically marginalized groups. This turns out to be intriguingly similar to the doctor-patient case. Doctors aren't experts on the patient's personal values and machine learning scientists aren't experts on fairness and socioeconomic inequity. Transparency is particularly justified here over a particular group of experts, we might say, because their expertise is crucially incomplete for their task. The truly relevant expertise extends far beyond a particular institution or group. A thorny version of the original problem remains, however. Machine learning experts may not be experts on fairness, in all its complex forms, but they are experts in how to encode different conceptions of fairness into algorithms. How we are to negotiate cases in which

the relevant expertise is complexly distributed is an enormously difficult matter.

We should expect to find the clearest cases of the expert-ladenness of value in those domains where the values are more confined to a particular expert group. Plausibly, the NEA has a more complete expertise over the values of the arts they are considering than the machine learning scientist has over the values of different conceptions of fairness. In many cases of education, too, the student isn't aware of the forms of value that are available — the educator's role is to start to bring the student into connection with that value.[19] The role of the arts educator, for example, is to show students the way into forms of value that they do not grasp on their own. The publicly funded arts are best conceived of as explorers of a realm of value, who we fund to scout, discover, and then show us the way to new and surprising forms of value — and not as, say, a simple means pump out more of some kind of pleasure of which we are already fully conversant (Levinson, 2005).

Thus, in cases where a localized group of experts have the best access to a value, stringent demands for public output transparency will distort the expert's aims and functioning. We've already seen examples of such distortion. Public oversight over the NEA leads to judging arts funding in terms of box office returns; public oversight over universities leads to thinking of education in terms of graduation rates and employment

---

19 I am denying here the "consumer" model of education, where the student is purchasing a clearly specified product with an outcome that they understand, and already desire. One might note that the consumer model of education has a strong relationship with various stringent transparency regimes in education, that do tie educational successful with such knowable outcomes — like a salary, or measurable competence in a salable skill.

and salaries. Expert educators in a particular discipline have a far deeper and distinctive sense of the goals of education in their discipline than outsiders would. The problem is this: if non-experts are put in oversight roles over expert activities, how will they pick the right metric to judge that activity?

What we've found here is modernized update of a very old problem: the expert identification problem. If you are a novice and you want to learn a domain, how do you pick out the right expert? You need to sort the posers and fakers from the real thing — but that itself requires expertise.[20] The problem is, I think, quite hard to solve, especially for public deliberation. Empirical research has shown, for example, that juries are notoriously bad at picking real experts, typically putting more trust in the outward show of confidence than any real expertise. Alex Guerrero has argued convincingly that the expert identification problem is unsolvable on a public scale (Guerrero, 2014, 137-153).

Public output transparency simply pushes the expert identification problem up a level. Public output transparency does not ask that non-experts directly identify which experts to trust. But it does ask non-experts to successfully identify the metric or evaluative method by which we will judge expertise. But if we didn't think that non-experts could reliably identify experts, why should we think that non-experts could reliably identify the

---

20 (LaBarge, 1997) provides a wonderful explication of the problem's occurrence in Socratic dialogues. (Cholbi, 2007) provides a useful application to the problem of moral expertise. For an optimistic take on the expert identification problem, see (Goldman, 2001). For a convincing response to Goldman and generally pessimistic take on the expert identification problem, see (Guerrero, 2014). I take Guerrero's pessimistic argument to be largely successful.

right evaluative criterion for identifying experts? Shouldn't that be as out of reach as identifying the experts themselves?

The answer, I think, will vary depending on the expert domain in question. There is a way out of this dilemma for certain forms of expertise. In earlier work on the expert identification problem, I put it this way: Some expert domains have a clear *litmus test*. A litmus test is some successful expert outcome that is identifiable as such by non-experts. Here are some plausible examples: Expertise in basketball is identifiable to non-experts because it yields a higher score. Expert rock climbers are identifiable because they can get to the top of the cliff. Expert bridge-builders are identifiable because they can build bridges that don't fall down. But other domains don't have litmus tests. In those domains, expertise is required to properly judge a good outcome. I call these domains "cognitive islands", and I suggested that the moral and aesthetic domains are particularly clear examples of such cognitive islands. You need some moral sensitivity to recognize good moral action; the morally insensitive will fail to recognize moral failings in others (Nguyen, 2020b).

Notice that the notion of a litmus test includes two requirements. First, it needs to reliably detect real experts. Second, non-experts need to be able to properly identify the litmus test as such. The problem here is that, in many terrains, non-experts will mistakenly identify something as a good litmus test. The Charity Navigator and NEA cases are useful examples here. To the non-expert, it certainly seems like the best charity is the one that

sends the most money through to its targets. (The best charity is actually the one with the most positive impact, but that is a matter far more difficult to assess.) To the non-expert, it might seem like popularity and profit was a good indicator of artistic merit. (It isn't; artistic merit is a complex and many-layered thing). The non-expert mistakes these simplistic metrics for good litmus tests — and they make the mistake because of their non-expertise.

We can see this effect in quick attempts to discredit whole realms of expertise. Take, for instance, the following much-touted result: that science has shown that there's no such thing as wine experts, because wine experts can't tell red wine from white wine. This result consists of a proposal of a litmus test for an entire domain of expertise, and a demonstration of general expert failure.[21] But, as it turns out, this is not a good litmus test, because red wine and white wine are not always gustatorily distinctive. Non-experts typically subscribe to a simple view of wines: white wines are light and crisp, red wines are thick, powerful, and heavy. But in fact, there are plenty of rich, powerful white wines, like viogniers. And there are many crisp, delicate red wines, like cold-climate pinot noirs (Goldstein, 2006). The belief that a good litmus test for wine expertise is the ability to distinguish, by taste alone, between red and white wines is itself a mistake which arises out of inexpertise. To adapt an old saying: the problem isn't just that the non-expert

---

21 I have described the experiment as it was reported in the popular media. However, as it turns out, popular reporting significantly misrepresented the facts. For one thing, the experiment was actually run, not on experienced expert wine tasters like professional sommeliers, but on undergraduate students in wine science. http://sciencesnopes.blogspot.com/2013/05/about-that-wine-experiment.html

doesn't know, it's that they don't know what they don't know.

And, as it turns out, some of the apparently good litmus tests we've seen turn out not be quite so clear-cut on closer inspection. Let's return to the case of rock-climbing. We might think that there is a clear litmus test for expertise in rock climbing: a good rock-climber can get to the top of the cliff. But different climbs vary in difficulty, and an observer's assessment of a climber's ability will depend on their assessment of the difficulty what that climber has climbed. And here's the thing: non-rock-climbers are notoriously bad at assessing the difficulty of rock climbs. Typically, those without expertise in rock climbing tend to think that extremely overhanging climbs, with big and dynamic leaping motion, are the most difficult. But, as a matter of fact, many such climbs are relatively easy for any decent climber. Many of the most difficult climbs are actually less steep, requiring delicate and fine motion — but they are far less impressive to outsiders. This difference has actually shown up in a very public venue. The Olympics has included, as one of the main disciplines of rock climbing, speed climbing — climbing a fixed pattern of very large holds very quickly. It looks spectacular to the uninitiated. But speed climbing is largely considered a kid's game among serious rock climbers, for whom the real interest involves delicate and subtle movement, improvised in response to novel climbing problems.

Of course, there is another way to judge rock-climbing expertise. In rock climbing, there is an established scale of difficulty. Those who can climb the highest on the scale

are the true experts. But, importantly, the difficulty scale is set *by consensus among expert rock climbers*, based on their experience of the difficulty of climbs. A non-expert cannot establish where a climb sits on that scale. Since the rating scale was established internally by the community of experts, using it as a metric of expertise requires placing trust in the opaque judgments of those experts. Such internally established scales are not litmus tests and cannot provide public transparency.

This helps us understand why public transparency often takes the form of metrics. As Theodore Porter says, the basic process of quantification involves narrowing down information to make it more portable. Qualitative expressions of information are capable of great nuance and context sensitivity. But often, such expressions require enormous amounts of shared background knowledge and context to appreciate. Quantification strips out the variable parts — which typically includes much of the context-sensitive information — leaving those components which can travel stably between contexts, and be interpreted without shared background. Quantification creates portable, aggregable information — which is why it's the preferred tool of large-scale institutions and bureaucracies (Porter, 1995). This fits precisely with our story about public transparency. The problem with many of the metrics we've seen is that they are formulated around publicly available inputs and processes, so that they are comprehensible — and auditable — to a wide audience of people from widely varying backgrounds.

Notice that a similar problem arises if we try to use, not output metrics, but human

intermediaries like regulators. Suppose we try to create a cadre of expert regulators over, say, the pharmaceutical industry. If those expert regulators substantially exercise their expertise in making their decisions, then their decisions occur out of public view. We might then worry that those regulators aren't doing their job properly. There is even a term for it: "regulatory capture", which is when regulators' interests become more aligned with those they are regulating, rather than the public interest. This often happens because employees of regulatory institutions know that they can eventually retire to a well-paying job with the regulated institutions. Using expert regulators is often a good strategy – but notice that, insofar as we are substantively trusting in their expertise, then it is not public transparency. It is expert transparency, where the public must then trust one batch of experts to regulate another batch of experts. Now the public has to decide whether to trust these expert regulators or force transparency on them. But if we put the actions of those expert regulators under a scheme of public transparency, then their actions are now restricted by that scheme in the same way. Using expert regulators simply moves the "transparency or trust" question up a level – from the regulated to the regulators.

So, in order to successfully achieve public output transparency, we need an evaluative standard that both a.) is publicly available and b.) properly tracks the value of the activity. This evaluative standard is available for some activities, whose output is publicly evaluable — like bridge-building. But for a great number of human activities, the true value of the activity is subtle, and so the judgment of successful output requires expertise. We are then

faced with a choice: insofar as we wish to pursue that value through the division of labor, we must trust experts not only to perform the activity, but also to judge that activity's success. This puts them beyond the reach of direct public assessment. Insofar as we want public assessment, we will want to impose some scheme of public output transparency and judge those experts in terms of their successful pursuit of a goal we can understand and measure. But insofar as the true value of the process is subtle, then the stringent imposition of public output transparency will undermine the pursuit of that value. Stringent public output transparency will block the pursuit of any value which is not directly apprehensible to the inexpert public.[22]

A second worry involves the magnetism of clear litmus tests. In many cases, an activity will have a genuine litmus test, but only because the activity has already been reformulated around such an available litmus test. For example: the activity of "getting healthy" is complex and multidimensional, and success at getting healthy is hard to judge. And it is hard to judge who the right experts, gurus, and coaches are who help us get healthy. But if we recast the activity slightly — if we reconceive of the activity of getting

---

[22] These concerns will likely impact some other, transparency-adjacent schemes for managing public/expert relations. For example, Heather Douglas (2017) suggests that scientists should disclose their values, so that citizens can choose to trust those scientists who espouse values similar to their own. This can work so long as the value can be specified in a way so as to be recognizable to the public. But insofar as expert's actual values are substantially beyond the public's understanding, then some version of the problem re-occurs. Notice, however, that Douglas' solution does not fall prey to some of the problems I've described. In Douglas' solution, citizens use an approximate similarity of value to decide who to trust. Douglas does not demand that the public actually examine and evaluate specific actions, justifications, and successes from scientists – that is, Douglas doesn't demand that the public operate in a way that eliminates the need for substantive trust. She only suggests that the public use value similarity to figure out who to trust.

healthy as, say, centered around losing weight — then success becomes extremely easy to judge. Fitness gurus can now make publicly assessable claim to be experts in health because they can demonstrate a track record of helping their clients lose weight. (It is much harder to document one's claim to improve people's sense of long-term physical connection with their body.) Or, say, if we reconceive of the activity of teaching as one aimed at creating satisfied students, as measured in higher aggregate scores in student evaluations, or of getting students good jobs, as measured in hiring rates and salaries, then it will be easy to mechanically demonstrate which universities, disciplines, and departments are good at education.

These reformulations are, I think, perhaps the worst-case scenario for public transparency. In this case, the public can now successfully identify the experts in some activity, but only because we have already recast the entire activity so that we may easily identify good outputs, and, accordingly, the real experts. The demands of public transparency have been baked into the very essence of the reformulated activity and into the essence of what counts as expertise. The activity's nature has been reconciled with the pressure of public transparency, so as to exclude any form of expertise that might have recondite outputs. This reformulating tendency will tend to exclude, from our institutional lives, any activities that aim towards subtle values, which take time and experience to appreciate. And so the reformulating tendency helps to exclude subtle values from our awareness – and exclude, from our trust, experts in such subtle values.

What might the right answer be? Let me offer one solution, which offers some hope, at least in the scientific fields. Philip Kitcher has, in tangling with the difficulty of identifying highly specialized scientific experts, suggested a solution: we can adopt the methodology of "indirect calibration", instead of direct transparency to the public. I may not be able to figure out which theoretical physicists to trust, but I can trace a line of trust through overlapping disciplines, to those disciplines which I can effectively evaluate (Kitcher, 1993, 320-3). I trust the engineers that build bridges and airplanes, because there is a litmus test — those things don't fall down. Those engineers trust various material scientists, who in turn trust a certain bunch of applied physicists and statisticians, who in turn trust the theoretical physicists. Since each of adjoining fields overlaps substantially, experts in one field have some traction on successfully identifying experts in another. So I can trace a line of trust through overlapping fields, from fields with a litmus test into extremely distant and recondite fields.

The problem, of course, is that this form of trust forces me to give up a significant part of my intellectual autonomy. I end up trusting medical researchers and biologists with my life, even though the workings of their expertise are incomprehensible to me. This is a position of profound vulnerability. But the point is that this profoundly vulnerable, trusting position is the best that we cognitively finite beings can manage, when faced with the massive growth of specialist fields. Once the body of practical knowledge has grown beyond a certain length, the best we can hope for is this fractally iterated trust — a trust

that expands far beyond our grasp.[23] And the best we can do is to manage our trust's starting point with care — but there is no way to do it with guaranteed success, precisely because the nature of expertise puts the endpoints of our trust beyond our grasp. Our limited cognitive capacity creates the need to trust specialists and makes that trust impossible to wholly manage. And direct transparency turns out to simply quash expertise, in the attempt to generate the illusion of control and intellectual safety.

Intimate reasons

Let me now offer a second argument, which will approach many of the same themes as the epistemic intrusiveness argument from a slightly different angle. This can only be a brief sketch of a discussion which I hope to develop in greater detail.

This is the argument from intimate reasons:

Intimate reasons

1. Some reasons are only available to people in some community or group.

2. So many of the actual reasons for action for members of that community are not available to outsiders.

3. The demand for transparency placed on that community — and its associated

---

23 I have discussed the problem, and some possible solutions, to fractally iterated trust, in (Nguyen, 2018).

institution — asks that community members justify their actions in terms available to outsiders.

4. That demand distorts the actions and justifications of those community members.

Let's call *intimate reasons* those reasons which are only available to members of some community or subgroup. By "available", I mean that the reason is either: a.) *comprehensible* only to members of some community, or b.) has adequate *justificatory force* only directly recognizable to members of that community. By the former, I mean to indicate those reasons that can only be minimally comprehended, or seen as presenting any justificatory force, to members of a community. Examples might include reasons for action which are expressible only in community-exclusive jargon — such as scientific or academic jargon, religious esoterica, or the special language of a community. By the latter condition, I mean to specify those cases where a reason may be minimally understood by outsiders, but is inadequately moving or forceful. For example, the importance of protecting victims of trauma from further triggers might only be directly grasped by those who have also experienced similar trauma.

Importantly, intimate reasons are those that are not *directly* recognizable by outsiders to the community. Such intimate reasons may come to have some practical weight on outsiders when mediated by testimony. An intimate reason is still intimate if outsiders can come to give that reason weight, via trust in the testimony of a community insider. It is an

intimate reason so long as an outsider couldn't adequately grasp that reason, in its full force, on their own.

Why might there be intimate reasons? First, there might be such reasons if members of a community share some specific experience, where having that experience is required for a reason to be comprehensible or to have adequate force. For example, members of an oppressed community may, through their lifelong experience of oppression, become sensitive to a wide range of considerations. They may, say, be more painfully aware of the precise hurt of being intellectually ignored because of one's race or gender. They may be more acutely sensitive to what is lost in, say, the cultural appropriation of the community's traditions and objects.[24] They may be far more sensitive to the harms caused by various microaggressions, that those without that experience will usually see as less important – or not see at all.[25] And, for an entirely non-political example: members of an aesthetic community may develop sensitivities and sensibilities which give them special access to certain experiences. A listener who has spent the time to really understand, say hard bop jazz, can hear things in the music that others cannot. These experiences grant a more

---

24 For a further discussion of the intimate reasons involved in cultural appropriation claims, see (Nguyen and Strohl, 2019),

25 Examples of this sort are common in feminist and critical race theory. Excellent overviews and discussions can be found in (Wylie, 2012; Medina, 2013; Toole, 2019). See especially Alexis Shotwell's discussion, drawing on resources from feminist epistemology and critical race theory, of implicit understanding as it arises under circumstances of oppression (Shotwell, 2011). There is a strong parallel between my account of intimate reasons and the claims of standpoint epistemology. But note one key difference: standpoint theorists usually put their claims in terms of members of oppressed group having privileged access to *knowledge,* where my discussion here involves positing that members of oppressed groups have better appreciation of the force of certain *reasons*. I plan to further explore the relationship between intimate reasons and standpoint epistemology in future work.

powerful apprehension of that value, which will increase the felt force of those reasons which draw on that value. A critic's comment that Thelonious Monk gives us extraordinary layers of jagged, abrupt textures, will get one to listen to Monk if one has had the kind of aesthetic experiences that would lead one to value jagged, abrupt textures (Nguyen, forthcoming). Similarly, only through direct access to such experiences and their related values, or trust in somebody with such experiences, might an exclusive venue book Monk, or an arts funding institute offer him a grant.

Intimate reasons aren't just raw expressions of brute shared experiences. They can also take the form of expressions that make sense within the complex schemes of justification that have been collectively built on top of a foundation of such shared experiences. Consider the complex systems of communication that arise in oppressed communities, to talk about the particularities of their experience of oppression and resistance. Consider, also, the complex critical language that specific aesthetic communities generate. The inaccessibility of intimate reasons to the wider public may not be a permanent feature. Rather, the intimacy can be a contingent feature of the current dominant language and conceptual scheme. When an oppressed group has not been permitted to exercise adequate control over the creation of linguistic and conceptual resources, then their resonant experience — and the reasons and justifications that depend on those experiences — may be hard, or impossible, to make available to others.[26]

---

26 I am drawing here on Miranda Fricker's discussion of hermeneutic injustice (Fricker, 2007, 147-175)).

As Audre Lorde puts it, poetry is a form of "illumination, for it is through poetry that we give name to the ideas which are — until the poem — nameless and formless, about to be birthed, but already felt" (Lorde, 1984, 36).[27] But some groups have not had the time to make that poetry, because of the very circumstances of that oppression. And other groups may have had the time to make that poetry, but not yet had it reach a broader audience. The poetry often finds an expression in an intimate language, and provides illumination and language for people in that community, but not yet had the chance to reach and inform the language of outsiders. In all cases, group insiders can have expressions of justification whose force is not adequately felt by outsiders.

I recognize that the idea of intimate reasons goes against the grain of many theories, which take reasons to be essentially public.[28] I think, in fact, that proper appreciation of the lessons of standpoint epistemology – and aesthetic epistemology – will show the problems with the view that reasons are essentially public. This point, however, will require much further argument. For the moment, let me offer a quick sketch. Kyla Ebels-Duggan has persuasively argued that some reasons can be rational, but inarticulable. Some reasons only have force if one has had the appropriate experience, and some experiences lie beyond articulation (Ebels-Duggan 2019). Ebels-Duggan's emphasis here is on private experience and private reasons. But the same argument shows that there can be intimate

---

27 Alexis Shotwell offers a particularly useful analysis of how Lorde's insights fit into the scholarly discussion of feminist epistemology (Shotwell, 2011, 24-28).

28 An interesting echo of this argument may be found in Kogelmann and Stitch (forthcoming) criticism of the demand for public reason.

reasons. Some reasons depend, for their force, on an agent's having had some experience. Some experiences are inarticulable *but shared*. Aesthetic experiences offer an excellent example. It is impossible to convey in words the particular aesthetic glory of Monk's jagged textures to somebody who hasn't heard it and felt it themselves — but the experience is shareable. Those who have shared in it can talk about it and use that reference as a collective point of reference to talk about other aesthetic wonders. In fact, the standard view in aesthetics is that aesthetic properties are only graspable through direct experience — but that others can share in that experience.[29] This is why so much aesthetic talk involves comparison or reference to specific artworks.

If intimate reasons exist, then transparency can cause the same kind of intrusive damage to intimate reasoning as it does with expert reasoning. Community members who are pressured to be transparent to a broader public will either need to deceive, limit, or transform their justifications for action, in light of the demands of transparency. This becomes particularly clear when we think of publicly funded institutions made to support communities with intimate reasons. I am most concerned by cases of too-stringent requirements for transparency, imposed upon publicly funded institutions or groups intended to protect oppressed groups. Here is an example adapted from my own experience. University leadership had come to recognize that there the needs of some

---

29 I am relying here on the standard view that direct contact is required for aesthetic experiences, or to ground aesthetic judgments of value. A classic exposition of the view can be found in Frank Sibley, especially (Sibley, 1959). For more recent discussion, see (Hopkins 2001; Gorodeisky, 2010; Hopkins 2011; Gorodeisky, 2019; Nguyen, 2019).

oppressed group are not being adequately served by university resources. They recognized that their LGBTQ+ students experienced constant harassment and were under enormous strain. The university funded a LGBTQ+ student support program, which is appropriately staffed with members and allies of the LGBTQ+ community. The sort of staffing decisions reflects, I think, the widespread recognition of intimate reasons. People who have had a particular range of experiences, and lived in a community which shares those experiences, have a special sensitivity to a whole range of considerations that people who have not, do not. But that simple observation is deeply in tension with many arrangements of oversight and transparency. In one clash which I was a witness to, the LGBTQ+ support program wanted to set up safe spaces only accessible to LGBTQ+ students, but that decision had to pass through university oversight. University administrators understood the reasons in some minimal sense — but they obviously didn't appreciate their full force. Administrators took the reasons which weighed in favor of setting up safe spaces to be far less important than the reasons to avoid them as a discriminatory practice. The justification for safe spaces involved references to all kinds of reasons and considerations whose importance was keenly felt to people who had lived under the lifelong pressure of a specific form of harassment — but which were unavailable (or appeared at a muted distance) to the university administration and to the general public, who had mostly not. The point of staffing an LGBTQ support group with members of the community – with people intimately familiar with that particular form of suffering

and harassment – comes from recognizing the importance of those experiences. But, rather than trusting those experiences, too-stringent schemes of transparency can undermine that intimate understanding.

This example makes clear a particular odd feature of transparency. Transparency is usually treated as an anti-oppressive process, whose goals are to rein in powerful elites and bring them to serve the interests of the public at large. And some instances of transparency do serve that purpose, such as the transparency of powerful political figures to their constituents. But transparency, even transparency for publicly funded institutions, doesn't always have that power structure. In many cases, it is the general public — and the bluntness of understanding imposed by the general public's very generality — that can oppress public institutions based in, and designed to protect, vulnerable communities.

Transparency is surveillance

Transparency, as it turns out, is a kind of surveillance. Often, it is bureaucratic surveillance, which demands that the appropriate justifications be made to the overseers, or else, say, funding will be cut. And as with other forms of surveillance, it can be both deeply intrusive and deeply necessary.

Surveillance, under one standard account, is the act of paying close and sustained attention to some entity, for a particular reason. The most familiar, and paradigmatic form

of surveillance is performed on the suspicion of wrongdoing. But other motives are possible. Kevin Macanish points out, for example, that grocery loyalty cards are a kind of surveillance, whose purpose is to increase profits.[30] And notice that, under many familiar uses of "surveillance", what counts as "attention" is quite broad. Suppose that London is blanketed by closed-circuit television cameras which automatically record everything, but which are largely unmonitored by humans or any analysis system – until police come to investigate some particular crime. This is surely surveillance. What's important is not the actual constant presence of active, directed, conscious attention, but the potential for such active attention. Pressure comes from the fact that somebody *could* access those records – that that the records are sitting around quietly in storage, but available for access into the unpredictable future.

I've focused on forms of public transparency that are explicitly set up for the purpose of assessment. Such assessment transparency is surveillance by definition. We are making data available, not for the sake of the data itself, but to make some evaluation of the entity which produced that data. But what about re-use transparency? At first glance, it doesn't look quite like surveillance. Surveillance is attention paid to an *entity*. But in re-use transparency, we're not paying attention to an *entity* behind the data, but to the data itself. When I use your open-source code, I may not by judging you in any way – I just

---

30 I take this definition from Kevin Macanish's survey of surveillance ethics (Macanish, 2020). I am not distinguishing, as some have, between surveillance as attention from powerful entities above, versus sousveillance as attention from the people below (Thomsen, 2019).

want the code. But data made available for the purpose of re-use can easily be repurposed for assessment. Even if the original intent of the transparency is re-use, insofar as the data is entered into the public domain, then it is also available for use in assessment. And insofar as it is sitting around, available for use in assessment, and is so used, then it is also a kind of surveillance.

Imagine I installed closed-circuit television cameras in my university, and recorded everything – every meeting, every conversation, every class – and then dumped it all on YouTube. Suppose I did it because I genuinely thought our university was just so awesome, and wanted the world to learn from our methods. My intentions are to set up a re-use transparency. But, obviously, all that information can be easily repurposed for assessment. And that very process of publicity, no matter the intent, still creates that chilling effect, and opens the door to vicious misinterpretation. And such repurposing happens in the real world. Scientific transparency is often set up for the sake of data re-use, but once that data is made public, it can easily be used for assessment – and is, by hostile actors. Famously, scientific transparency about climate change research, and research about the health impacts of tobacco, has been used by political opponents to harass scientists and challenge their findings (Elliott 2016; 2020, 2; Oreskes and Conway, 2010). And I participated in a number of terrified conversations among progressive professors, as we all crashed into the era of Zoom teaching during the COVID pandemic, about recording lectures. Our students all wanted readily available recordings of our

lectures – but we were worried about snippets being taken out of context and distributed on social media, especially the alt-right social media, in social attacks. Public re-use transparency may not be set up for assessment, but precisely because that information has been made available, it can function as surveillance.

Thus, insofar as the information, which has been made public for re-use, is likely to be re-purposed for assessment, then re-use transparency is functionally equivalent to assessment transparency. Re-use transparency may not have been originally set up as for surveillance, but it functions as surveillance.[31] What is harmful here is that some distant, uncharitable observer is watching you and judging you, and that you have to think about how your activities will look and be explained to them. It matters little whether their avenue of access was originally set up for judgment, or for some other purpose. They're still watching.

We usually associate surveillance with large institutions monitoring individuals. What I hope to have made clear, with my examples and analysis, is that public transparency is a still type of surveillance, even when it inverts the usual power structure. Public transparency is the surveillance of institutions by the public at large – but this inverted surveillance still brings with it many of the same ills. Public transparency trades trust for monitoring; it has a chilling effect on free action. It destroys intimacy and the capacity for

---

[31] For a very useful discussion of how the function of an artifact can differ from its creators' intent, see Eaton (2020).

action from a shared, tacit understanding. Open surveillance produces, in its targets, a particular kind of paranoia — the constant sense that one must justify one's actions to somebody who has no sympathy. It is the paranoia that arises in constantly needing to justify oneself to a distant, unsympathetic overseer, and to imagine one's actions and results as evaluated by them. The more distant the overseer is from the overseen's particular context, the more likely the possibility of misunderstanding — and the more intense the paranoia. And who is a more distant, context-less overseer, than the rushed public?

I find that I can get a better grip on the problem by thinking about how differently things appear to me in my different roles. For most of us are sometimes cast in the role of the overseer, and sometimes in the role of the overseen. In my everyday life, transparency seems utterly sensible when I occupy the role of overseer, and it seems most blinkered and intrusive when I am in the role of the overseen, struggling to justify myself to a rushed and inattentive public. Thinking about this gap helps me to get a better grip on the possibility of my own false confidence, when I occupy the role of overseer. I find it useful to recall how it feels to be overseen – to think about how much of the rich realm of expert reasoning and shared communal understanding that I can use in my own domain – but which must be stripped out of my reports to that distant, harried overseer. When we are overseer, the faults of transparency are often hidden by the shallowness of our understanding of the overseen.

And the conceptual connection to surveillance suggests a larger point. Theorists of surveillance, following Foucault's analysis, have been worried about how we might internalize surveillance — and bring the external evaluative gaze inside our own reasoning and values.[32] And this clearly occurs with transparency. Under regimes of transparency, some individuals come to permit themselves to deliberate only using terms and reasons that would be publicly understood. Internalizing, as one's basis for practical reasoning, the demand that one be able to produce an explanation for any action to any conceivable questioner, turns out to be a kind of self-surveillance. It is one in which one internalizes the cold, distant, insensitive, contextlessness of the distant observer, and takes that on as in internal standard for reasoning. One is, then, constantly imagining how one would justify oneself if called before an unsympathetic court of inquiry.[33] One takes oneself to be beholden to explain all of one's own actions to such a distant overseer — on pain of feeling unjustified to oneself. And that, I think, is an appropriate diagnosis of a certain bureaucratic mindset, in which an actor which seeks an articulable policy in which to ground each and every one of their actions. And again, notice that this self-surveillance can occur without any authoritarian regime. It can arise naturally, in even the most well-

---

32 The modern discussion of surveillance in the various critical theories seems to follow largely in the footsteps of Foucault's discussion of the panopticon (Foucault, [1975] 1995; Lyon, 2006).

33 Sophia Dandelet (2021) offers an extremely compelling account of how we might be led to change our internal epistemic standards on the basis of unfair external justificatory pressure. Her core case is one in which a woman believes she has been sexually harassed, but when she thinks about the kinds of skeptical questions she might face if she told anybody, she raises her internal epistemic standards, and loses grip on her belief. Dandelet argues that, insofar as the degree of external skepticism is unfair, then this counts as a form of epistemic coercion. I suspect that, insofar as the pressure for transparency was unfair, then transparency can result in epistemic coercion.

intentioned of democracies, simply through the demand for transparency, when applied with sufficient stringency.

Ta-Nehisi Coates offers a brief, but compelling, account of what it is to be asshole. An asshole, say Coates, is "a person who demands that all social interaction happen on their terms" (Coate, 2013). Coates describes the experience of being in a bar while black, when somebody else — who is "invariably white" — will "stumble over drunkenly and decide that we should be engaged in conversation with them." The asshole, says Coates, is somebody who is insensitive to the delicacies of another's understanding, who expects the world to be perpetually open to conversation with them, in terms that they can readily understand. And so we could, with only the slightest stretch, come to see that the demand for public transparency is a kind of institutional assholery, which fails to respect the incredibly rich ways in which different communities and groups of experts can see the world in particular and sensitive ways. It fails to respect the fact that other people might speak and think in justificatory language deeply distant from the general public's.

Contemporary human knowledge is essentially scattered. Human knowledge is made of a vast number of hyper-specializations, each of which is mostly inaccessible to the rest of us. As Baier says, trust makes us vulnerable. And the painful truth is that, in the modern, socially extended form of knowing, we cannot secure our trust entirely. It is impossible for each of us to pick out the right experts for ourselves. This leaves us profoundly vulnerable. We might resist this vulnerability, by trying to impose systems to secure all that expert

knowledge, to bind it to our own understanding. But that attempt is based in an underestimation of the radically scattered and specialized nature of expert knowledge. The history of Western European philosophy has been marked by an epistemic individualism, by a dream of total intellectual autonomy, in which each of us was capable of ascertaining, for ourselves, all the justifications for all the truths that we need to use. But that is mere fantasy. And the movement towards total transparency is simply that Cartesian fantasy writ into public policy. It fails to respect the fact that the breadth of human knowledge far outstrips the capacities of any individual.

Seeing transparency as surveillance helps us think about how we should structure that surveillance – and to think about when that surveillance might be justified. Much of the pain of surveillance comes from the potential for misunderstanding. So, surveillance is least harmful when the surveillors are mostly likely to understand, like fellow experts, and it is the most harmful when the surveillors are most likely to misunderstand, like the public at large. The demand for public transparency arises from an entirely comprehensible democratic impulse; but reflection on the nature and opaqueness of expertise weigh in favor of a more constrained form of surveillance, by relevant experts.
But this constrained form of surveillance would constitute a departure from direct public transparency. It asks the public, as a whole, to trust, something beyond their ken.

But the shift from public to expert transparency still isn't a perfect solution. The harms of surveillance will be significantly ameliorated, but not eliminated. Why? One might think

that, if experts only have to explain themselves to other experts, then the problem of epistemic intrusiveness will disappear. But this depends on a crucial simplification, to which I have been a party. I have been speaking as if there are simple, unvariagated domains, where experts are attached to the whole domains. But in the real world, expertise is not so monolithic; every expert is a particular expert of a particular patch. And even expert transparency will typically involve movement between some distinctive patches. One group of chemists is being transparent to another group of chemists – but each working group has its own particular specialties, its own history of projects, and its own particular set of instruments, and its own version of expertise. So the potential for misunderstanding is lessened, but not eliminated, with expert transparency.

Suppose we could take care of that. Suppose, for the moment, that we could find two groups of experts with precisely the same expertise. Perhaps they are two anthropology departments at different universities, which happen to have the exact same set of sub-specialties. We can (and do) create systems of transparency between such departments, so that one such department can help oversee the other. Even then, there's still some amount of understanding that can be lost in translation between those two groups, because of the rich vein of subtle understanding that any community develops. In other words: here, the epistemic intrusiveness argument may lose its grip, but the intimate reasons argument still bites. Each set of people has its own history and culture, their own developed modes of communication, special to that particular community. These are

easily misunderstood by other groups, even if they are part of the same expert domain. There are in-jokes, complex friendships and enmities, and special understandings. Every group has its own intimate reasons, which arise from its own history of shared experience – and these reasons travel poorly. Expert transparency, then, still opens the door for misunderstanding. Whatever special knowledge or understanding is held by one group of experts, whatever is peculiar to a long-standing group's communal life, won't be available in its attempts to justify itself to another group of experts. Expert transparency is the gentlest form of surveillance, but it is still a form of surveillance.[34]

Transparency is surveillance. The harms of surveillance, as I've sketched them, are in the pressure exerted from oversight. That pressure can intrude, in various ways, on reasoning and goal-setting. The more distant the overseers in context, the more pressure to simplify and remove nuance from one's reported justifications – and from one's actions, which imply justifications. What we've learned is that the harmful effect scales with contextual distance, and one very crucial form of contextual distance lies in the degree of expertise. Public transparency is the most intrusive form of surveillance. But so long as there is any distance between the overseen and the overseer, we will get some form of that harm. We might try to close the gap by finding overseers as close as possible to the overseen – like intermediary expert regulators – but the closer those specially designated

---

[34] This paragraph draws heavily on James Scott's (1998). discussions of what's lost when we shift from local experts and expert communities, to standardized and centralized forms of expert administration. My discussion here is too brief, and I plan to expand on it in future work.

overseers are to the overseen, the less reason the public will have to trust those overseers.

There is a tension, O'Neill suggested, between trust and transparency. We can now flesh this out in far greater detail. As Millgram puts it, the contemporary world is a patchwork of hyperspecializations, each a domain of expertise mostly opaque to most of the others. To that, we can now add: the world is also a patchwork of intimate communities, each with their own rich and particular experience, and the intimate reasons that rise from them. If we try to hook up this patchwork with pure transparency, then we will push out expertise and intimacy. If we try to hook up this patchwork with pure trust, then we will open the doors for corruption, bias, and value drift. There is no easy solution here. This turns out to be the basic dilemma of navigating a world in which the amount of knowledge we need vastly outstrips any individual's capacities. There is no resolving the tension — only settling where we want to compromise, and which sacrifices we want to make, between the goods of trust and the need for transparency.

Baier makes an observation in a similar key, in a remarkable – and mostly neglected – passage at the end of her famed "Trust and Anti-trust". Some theorists, she says, take the contracts to be the paradigm of trust. But contracts, she says, are actually a peculiar and specialized instantiation of trust – an institutionally controlled trust. She says:

Contract enables us to make explicit just what we count on another person to do, in return for what, and should they not do just that, what damages can be extracted from them. The beauty of promise and contract is its explicitness. But we can only make explicit provisions for such contingencies as we imagine arising… Another functional excellence of contracts, which is closely connected with the expressness that makes breach easily established and damages or penalty decidable with a show of reasonable justice, is the security they offer the trusting party. They make it possible not merely for us to trust at will but to trust with minimal vulnerability. They are a device for trusting others enough for mutually profitable future-involving exchanges, without taking the risks trusters usually do take. They are designed for cooperation between mutually

suspicious risk-averse strangers, and the vulnerability they involve is at the other extreme from that incurred by trusting infants. Contracts distribute and redistribute risk so as to minimize it for both parties, but trusting those more powerful persons who purport to love one increases one's risks while increasing the good one can hope to secure. Trust in fellow contracters is a limit case of trust, in which fewer risks are taken, for the sake of lesser goods. (250-1)

In full-blooded trust, says Baier, we grant the trusted significant discretionary powers. We trust them to act in ways we could not have forseen, to problems that we could not expect. To extend Baier's thought: this full-blooded trust expands our ability to act and do, because it lets us rely and employ forms understanding and skill which transcend our own. But when we insist on delegating tasks, but refuse to trust in this rich manner – when we substitute trust in another's discretion, with oversight in terms that we can understand, we profoundly limit our ability to expand our agential powers through trust. In the interest of security, we prevent others from exercising their own special discretion, and limit what we delegate to the confines of our own limited understanding.[35]

And we should not forget that the problem that transparency is trying to solve is quite real. The problem with trust in the inarticulate is that it often leads to a kind of pernicious conservatism. Trust can enshrine bias inside a reification of the intuitive. And we do need clear, simple measures to expose that bias. University hiring practices should make this clear. Once, philosophy faculty members made their hiring based on largely intuitive senses of who was "good at philosophy" — which, as has been overwhelmingly demonstrated, leads to the perpetuation of various biases, and the perpetuation of the dominance of white males in philosophy, and the exclusion of other interests and modes.

---

[35] I've discussed the relationship between trust and agential integration in (Nguyen, forthcoming b).

As Merry — a staunch critic of the excesses of metrics — says, simple metrics are sometimes exactly the brute hammer required to break open entrenched bias. In every case she has studied, demonstrating the existence of systematic institutional bias did require clear, quantified demonstrations: undeniable numbers about the radical lack of diversity in the profession.

On the other hand, being guided by nothing but the simple numbers — making objective-seeming hiring decisions based on evaluations of publication numbers and citation rates — excludes something else very important: the developed sensitivity to philosophical quality present in trained experts. Similarly, it is a real and legitimate worry that insulated groups can develop specialized interests, putting them out of touch with the general public interest. But if we leash groups of experts to pursuing only the goals and taking only the actions that can be justified to non-experts, then we will undermine their expertise. Perhaps the answer is: we do need a leash, but a long and loose one.

Martha Nussbaum (1992) has offered a complex picture of the relationship between sensitivity and explicit argumentation. Moral sensitivity, she says, is a delicate and cultivated virtue, and cannot be captured in any explicit rule-set. Contextually attentive moral action is often the result of trained moral perception and cultivated sensitivity. At the same time, because of that inarticulateness, the deployment of moral sensitivity in the form of intuition is rife with bias. This is what, she says, philosophical ethics is for. Ethics can force us to articulate our principles and find areas of brute bias and insensitivity. But

Nussbaum's larger point is that we cannot live moral life, in the day to day, with such philosophical ethics. Daily moral life depends on sensitivity, trained emotional perception, and rich moral attunement. Philosophical ethics has a very specific role — it is an occasional intervention and occasional check, which can point out significant problems, and set one on the road to re-development. The cold abstractions of philosophical ethics can show us that our sensitivities are radically out of tune — that we are profoundly sexist or racist — and that we need to work to get to a place of better trained sensitivity. But we cannot live our daily moral lives guided only by those abstractions. We need to re-develop our sensitivities in the wake of radical re-tuning.

I think we something similar with transparency, especially in its most brutally public form. Transparency — the surveillance of justifications, forced into the most bare and publicly accessible of terms — is best as occasional intervention and not a constant regime. Sometimes we need to throw the books open and poke around for evidence of significant corruption and bias. But we should not conduct our daily business under a state of constant investigation. And we certainly should not internalize the terms of intrusive outsider surveillance, and use them as our daily guide to action.

I am not saying that some special group of people knows it all. Rather, I think that expertise, sensitivity, and various kinds of special awareness are distributed across the various people in the community. Some of us understand climate change models, others of us understand the value of experimental cinema or the particular musical value in the

Atlanta trap scene, others of us are sensitive to the pains of growing up biracial in America.

The goal is to *integrate and synthesize* these specialist sensitivities. We need a profound

division of cognitive labor – and a division of moral labor, and a division of valuing labor.

And to promote full the full flowering of that division, we need to trust each other. We

certainly need to check on each other — but the check is there as a safeguard in the

process of developing better relationships of trust, and not a replacement for trust.[36]

Bibliography

Baier, Annette. 1986. "Trust and Antitrust." *Ethics* 96 (2): 231–260. https://doi.org/10.1086/292745.

Bentham, Jeremy. 1989. *The Collected Works of Jeremy Bentham: First Principles of Preparatory to Constitutional Code,*. Edited by Philip Schofield. Oxford: Oxford University Press.

Biddle, Justin. 2020. "On Predicting Recidivism: Epistemic Risk, Tradeoffs and Values in Machine Learning." *Canadian Journal of Philosophy*. https://doi.org/10.1017/can.2020.27

Brandt, Julie. 2013. "Overhead Costs: The Obsession Must Stop." *Stanford Social Innovation Review*. October 3, 2013. https://ssir.org/articles/entry/overhead_costs_the_obsession_must_stop.

Brewer, Talbot. 2011. *The Retrieval of Ethics*. 1 edition. Oxford: Oxford University Press.

Callard, Agnes. 2018. *Aspiration: The Agency of Becoming*. New York: Oxford University Press.

Cholbi, Michael. 2007. "Moral Expertise and the Credentials Problem." *Ethical Theory and Moral Practice* 10 (4): 323–334. https://doi.org/10.1007/s10677-007-9071-9.

Coates, Ta-Nehisi. 2013. "How the Quiet Car Explains the World." The Atlantic. March 11, 2013. https://www.theatlantic.com/national/archive/2013/03/how-the-quiet-car-

---

explains-the-world/273885/.

Coupet, Jason, and Jessica L. Berrett. 2019. "Toward a Valid Approach to Nonprofit Efficiency Measurement." *Nonprofit Management and Leadership* 29 (3): 299–320. https://doi.org/10.1002/nml.21336.

Dandelet, Sophia. 2021. "Epistemic Coercion." *Ethics*. 131 (3): 489-510.

Dorsch, Fabian. 2013. "Non-Inferentialism About Justification – The Case of Aesthetic Judgements." *Philosophical Quarterly* 63 (253): 660–682. https://doi.org/10.1111/1467-9213.12063.

Douglas, Heather. 2017. "Science, Values, and Citizens." In *Oppure Si Mouve: Doing History and Philosophy of Science with Peter Machamer*. Ed. M.P. Addams et. al. New York City: Springer. 83-96.

Dreyfus, Hubert, Stuart E. Dreyfus, and Tom Athanasiou. 2000. *Mind Over Machine*. Simon and Schuster.

Eaton, A.W. 2020. "Artifacts and Their Functions." In *The Oxford Handbook of History and Material Culture*, eds. Sarah Anne Carter & Ivan Gaskell. Oxford: Oxford University Press.

Ebels-Duggan, Kyla. 2019. "Beyond Words: Inarticulable Reasons and Reasonable Commitments." *Philosophy and Phenomenological Research* 98 (3): 623–641. https://doi.org/10.1111/phpr.12463.

Elliot, Kevin C. 2020. "A Taxonomy of Transparency in Science." *Canadian Journal of Philosophy*. 1-14. https://doi.org/10.1017/can.2020.21

---. 2016. "Environment." In *Miseducation: A History of Ignorance Making in American and Abroad*, edited by A. J. Angulo. Baltimore, MD: Johns Hopkins University Press. 96–119.

Espeland, Wendy Nelson, and Michael Sauder. 2016. *Engines of Anxiety: Academic Rankings, Reputation, and Accountability*. 1 edition. New York, New York: Russell Sage Foundation.

Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Clarendon Press.

Gaus, Gerald, and Keith Hankins. 2017. "Searching for the Ideal: The Fundamental Diversity Dilemma." In *Political Utopias: Promise or Peril*, edited by Michael Weber and Kevin Vallier, 398–409. Oxford: Oxford University Press. https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780190280598.001.0001/acprof-9780190280598-chapter-9.

Goldman, Alvin I. 2001. "Experts: Which Ones Should You Trust?" *Philosophy and Phenomenological Research* 63 (1): 85–110. https://doi.org/10.1111/j.1933-1592.2001.tb00093.x.

Goldstein, Evan, and Joyce Goldstein. 2006. *Perfect Pairings: A Master Sommelier's Practical Advice for Partnering Wine with Food*. 1st edition. Berkeley: University of California Press.

Gorodeisky, Keren. 2010. "A New Look at Kant's View of Aesthetic Testimony." *British Journal of Aesthetics* 50 (1): 53–70. https://doi.org/10.1093/aesthj/ayp052.

———. 2019. "The Authority of Pleasure." *Noûs* 53 (4): 1–22. https://doi.org/10.1111/nous.12310.

Guerrero, Alexander A. 2014. "Against Elections: The Lottocratic Alternative." *Philosophy and Public Affairs* 42 (2): 135–178. https://doi.org/10.1111/papa.12029.

Hopkins, Robert. 2001. "Kant, Quasi-Realism, and the Autonomy of Aesthetic Judgement." *European Journal of Philosophy* 9 (2): 166–189. https://doi.org/10.1111/ejop.2001.9.issue-2.

———. 2011. "How to Be a Pessimist About Aesthetic Testimony." *Journal of Philosophy* 108 (3): 138–157. https://doi.org/10.5840/jphil201110838.

Kitcher, Philip. 1993. *The Advancement of Science: Science Without Legend, Objectivity Without Illusions*. Oxford University Press.

———. 2003. *Science, Truth, and Democracy*. Revised edition. New York: Oxford University Press.

Kogelmann, Brian and Stephen G.W. Stitch. Forthcoming. *Oxford Studies in Political Philosophy*. New York: Oxford.

Kogelmann, Brian. Forthcoming. *Secret Government: The Pathologies of Publicity*. Cambridge: Cambridge University Press

LaBarge, Scott. 1997. "Socrates and the Recognition of Experts." *Apeiron* 30 (4): 51–62. https://doi.org/10.1515/APEIRON.1997.30.4.51.

Landemore, Hélène. 2012. *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Princeton University Press.

Lena, Jennifer. 2019. *Entitled: Discriminating Tastes and the Expansion of the Arts*. Princeton: Princeton University Press.

Levinson, Jerrold. 2005. "Erotic Art and Pornographic Pictures." *Philosophy and Literature* 29 (1): 228–240. https://doi.org/10.1353/phl.2005.0009.

Lorde, Audre. 2020. *The Selected Works of Audre Lorde*. Edited by Roxane Gay. 1st edition. New York, NY: W. W. Norton & Company.

Lyon, David, ed. 2006. *Theorizing Surveillance*. Cullompton, Devon: Willan.

Macanish, Kevin. 2020. "Surveillance Ethics." In *Internet Encyclopedia of Philosophy*. https://iep.utm.edu/surv-eth/.

Medina, Jose. 2013. *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations*. Studies in Feminist Philosophy. Oxford, New York: Oxford University Press.

Merry, Sally Engle. 2016. *The Seductions of Quantification: Measuring Human Rights, Gender Violence, and Sex Trafficking*. University of Chicago Press.

Millgram, Elijah. 2015. *The Great Endarkenment: Philosophy for an Age of Hyperspecialization*. 1 edition. Oxford ; New York: Oxford University Press.

Nguyen, C. Thi. 2018. "Expertise and the Fragmentation of Intellectual Autonomy."

Philosophical Inquiries. 6 (2): 107-124.

————. 2019. "Monuments as Commitments: How Art Speaks to Groups and How Groups Think in Art." *Pacific Philosophical Quarterly* 100 (4): 971–994.

————. 2020a. *Games: Agency as Art*. New York: Oxford University Press.

————. 2020b. "Cognitive Islands and Runaway Echo Chambers." *Synthese*. 197 (7): 2803-2821.

————. 2021a. "How Twitter Gamifies Communication." *Applied Epistemology*. Ed. Jennifer Lackey. Oxford: Oxford University Press. 410-436.

————. 2021b. "The Seductions of Clarity". *Royal Institute of Philosophy Supplements*.

————. Forthcoming. "Trust and Sincerity in Art". *Ergo*.

————. Forthcoming b. "Trust as an Unquestioning Attitude." *Oxford Studies in Epistemology*.

Nguyen, C. Thi, and Matthew Strohl. 2019. "Cultural Appropriation and the Intimacy of Groups." *Philosophical Studies* 176 (4): 981–1002. https://doi.org/10.1007/s11098-018-1223-3.

Nussbaum, Martha. 1992. *Love's Knowledge: Essays on Philosophy and Literature*. Oxford: Oxford University Press.

O'Neill, Onora. 2002a. *A Question of Trust: The Bbc Reith Lectures 2002*. Cambridge: Cambridge University Press.

————. 2002b. *Autonomy and Trust in Bioethics*. Cambridge: Cambridge University Press.

————. 2005. "Gerechtigkeit, Vertrauen und Zurechenbarkeit," ("Justice, Trust, and Accountability") in *Gerechtigkeit: Auf der Suche nach einem Gleichgewicht*, ed. Otto Neumaier, Clemens Sedmak, Michael Zichy. Frankfurt: Ontos Verlag. 33-55.

————. 2006. "Transparency and the Ethics of Communication." In *Transparency: The Key to Better Governance?* Ed. Christopher Hood and David Heald. Oxford: Oxford University Press. 75-90.

Oreskes, Naomi and Erik Conway. 2010. *Merchants of Doubt*. New York: Bloomsbury.

Porter, Theodore. 1995. *Trust in Numbers*. Princeton: Princeton University Press.

Schroeder, S. Andrew. 2018. "Democratic Values: A Better Foundation for Public Trust in Science." *British Journal for the Philosophy of Science*. 10.1093/bjps/axz023

Scott, James. 1998. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven: Yale University Press.

Selinger, Evan M., and Robert P. Crease. 2002. "Dreyfus on Expertise: The Limits of Phenomenological Analysis." *Continental Philosophy Review* 35 (3): 245–279. https://doi.org/10.1023/A:1022643708111.

Shotwell, Alexis. 2011. *Knowing Otherwise: Race, Gender, and Implicit Understanding*. Penn State.

Sibley, Frank. 1959. "Aesthetic Concepts." *Philosophical Review* 68 (4): 421–450.

https://doi.org/10.2307/2182490.

Thomsen, Frej Klem. 2019. "The Concepts of Surveillance and Sousveillance: A Critical Analysis." *Social Science Information* 58 (4): 701–713. https://doi.org/10.1177/0539018419884410.

Toole, Briana. 2019. "From Standpoint Epistemology to Epistemic Oppression." *Hypatia* 34 (4): 598–618. https://doi.org/10.1111/hypa.12496.

Turilli, Matteo, and Luciano Floridi. 2009. "The Ethics of Information Transparency." *Ethics and Information Technology* 11 (2): 105–112. https://doi.org/10.1007/s10676-009-9187-9.

Wellar, Keenan. 2018. "About Nonprofit Waste, Overhead, and Financial Subservience." Nonprofit Quarterly. September 21, 2018. https://nonprofitquarterly.org/about-nonprofit-waste-overhead-and-financial-subservience/.

Wylie, Alison. 2012. "Feminist Philosophy of Science: Standpoint Matters." *Proceedings and Addresses of the American Philosophy Association* 86 (2): 47–76.