1

QUESTIONING TRUST

Onora O'Neill

It is a cliché of our times that trust has declined, and widely asserted that this is a matter for regret and concern, and that we should seek to "restore trust." Such claims need not and usually do not draw on research evidence. Insofar as they are evidence-based – quite often they are not – they are likely to reflect the findings of some of the innumerable polls and surveys of levels of trust that are commissioned by public authorities, political parties or corporations, or other organizations, are carried out by polling companies,[1] and whose findings are then published either by those who commissioned the polls, by the media or by interested parties.

Even when polls and surveys of public attitudes of trust and mistrust are technically adequate, the evidence they provide cannot show that there has been a *decline* in trust. That would also require robust comparisons with earlier evidence of public attitudes of trust and mistrust to the same issues – if available. And even when it is possible to make such comparisons, and they indicate that trust has declined, this may still not be a reason for seeking to "restore trust." A low (or reduced) level of trust can provide a reason for seeking to "restore" trust only if there is *also* evidence that those who are mistrusted, or less trusted, are in fact *trustworthy*: and this is not generally easy to establish (see Medina, this volume).

In short, polls and surveys of attitudes or opinions do not generally provide evidence of the trustworthiness or the lack of trustworthiness of those about whom attitudes or opinions are expressed. Trust may be misplaced in liars and fraudsters, in those who are incompetent or misleading, and in those who are untrustworthy in countless other ways. Equally mistrust and suspicions may be misplaced in those who are trustworthy in the matters under consideration. Judgments of trustworthiness and of lack of trustworthiness matter greatly, but attitudinal evidence is not enough to establish them. Trustworthiness needs to be evidenced by establishing that agents and institutions are likely to address tasks and situations with reliable honesty and competence.

Evidence of attitudes is therefore not usually an adequate basis for claiming that others are or are not trustworthy in some matter. Yet such evidence is much sought after, and can be useful for various other purposes, including persuasion and reputation management. Here I shall first outline some of those other uses, and then suggest what further considerations are relevant for placing and refusing trust intelligently. Broadly speaking, my conclusion will be that doing so requires a combination of epistemic and practical judgment.

1.1 The Limits of Attitudinal Evidence

Investigations of trust that are based on opinion polls and surveys can provide evidence of respondents' generic attitudes of trust or mistrust in the activity of types of institution (e.g. banks, companies, governments), or types of office-holder (e.g. teachers, scientists, journalists, politicians). Responses are taken as evidence of a *trust level*, which can be compared with trust levels accorded to other office-holders or institutions, and these comparisons can be tabulated to provide *trust rankings* for a range of types of institutions or office-holder at a given time. Repeated polling may also provide evidence of changes in trust levels or in trust rankings for types of institution or office-holder across time.

However, attitudinal evidence about trust levels and rankings has limitations. Most obviously, a decline or rise in reported levels of trust in specific types of institution or office-holder across time can be detected only by *repeated* polling, using the same or comparable methods. However, for the most, part polling was less assiduous and frequent in the past than it is today, so reliable comparisons between past and present trust levels and rankings are often not available. And where repeated and comparable polls have been conducted, and reliable comparisons are feasible, the evidence is not always that trust levels have declined, or indeed that trust rankings for specific types of institution and office-holder have changed. For example, in the UK journalists and politicians have (for the most part) received low trust rankings in attitudinal polls in the past and generally still do, while judges and nurses have received high trust rankings in polls in the past and generally still do.

Even where polling suggests that levels of trust have declined, the evidence they offer must be treated with caution. Polls and surveys collate evidence about informants' *generic attitudes* to *types* of institution or to *types* of office-holder, but cannot show whether these attitudes are well-directed. But while the evidence that polls and surveys of trust levels provide cannot show who is or who is not trustworthy, trust rankings can be useful for some other quite specific purposes. I offer two examples.

One case in which trust rankings are useful is in summarizing consumer rankings of the quality and performance of *standardized products or services*, when these rankings are based on combining or tabulating the views of *a wide range of consumers*. For example, trust rankings of hotels or restaurants, of consumer durables or retail outlets, can be usefully informative because they are *not* mere expressions of generic attitudes. Such rankings reflect the *experience* of those who have bought and used (or tried to use!) a specific standardized product or service, so can provide relevant evidence for others who are considering buying or using the same product or service.

However, it is one thing to aggregate ratings of standardized products and services provided by those who have had some experience of them, and quite another to crowd-source views of matters that are not standardized, or to rely on ratings of standardized products or services that are provided by skewed (let alone gerrymandered) samples of respondents. Reputational metrics will not be a reliable guide to trustworthiness if the respondents whose attitudes are summarized are selected to favor or exclude certain responses (see Origgi, this volume).

A second context in which the attitudinal evidence established by polling can be useful is for persuading or influencing those who hold certain attitudes. Here attitudinal evidence is used not as a guide for those who are planning like purchases or choices, but for quite different (and sometimes highly manipulative) purposes.

18

Some uses of polling evidence are entirely reputable. For example, marketing departments may use attitudinal evidence to avoid wasting money or time cultivating those unlikely to purchase their products. Some are more dubious. For example, political parties may use evidence about political attitudes, including evidence about attitudes of trust in mistrust in specific policies and politicians, to identify which "demographics" they can most usefully cultivate, how they might do so, and where their efforts would be wasted. In some cases this is used to manipulate certain groups by targeting them with specific messages, including misinformation and disinformation.[2] Information about generic attitudes, and in particular about generic attitudes of trust and mistrust, can be useful to marketing departments, political parties and other campaigning organizations *whether or not those attitudes are evidence-based* because the evidence is not used to provide information for others with like interests or plans, but as a basis for influence or leverage. I shall not here discuss whether or how the use of polling results to target or orchestrate campaigning of various sorts may damage trust – including trust in democracy – although I see this as a matter of urgent importance in a digital world.[3]

1.2 Judging Trustworthiness

Once we distinguish the different purposes to which attitudinal evidence about trust levels can be put, we have reason to reject any unqualified claim that where trust is low (or has declined) it should be increased (or "restored"). In some situations increasing or restoring trust may be an improvement, and in others it will not. Nothing is gained by raising levels of trust or by seeking to restore (supposed) former levels of trust unless the relevant institutions or office-holders are *actually* trustworthy. Placing trust well matters because trustworthiness is a more fundamental concern.[4]

Seeking to gain (more) trust for institutions or individuals that are not trustworthy is more likely to compound harm. The point is well illustrated by the aptly named Mr. Madoff, who made off with many people's money by running a highly successful Ponzi scheme that collapsed during the 2008 banking crisis. It would have been worse, not better, if Madoff had been more trusted, or trusted for longer – and better if he had been less trusted, by fewer people, for a shorter time, since fewer people would then have lost their money to his scam. Aiming to "restore" or increase trust will be pointless or damaging without evidence that doing so will match trust to levels of trustworthiness. By the same token, mistrust is not always damaging or irrational: it is entirely reasonable to mistrust the untrustworthy (see D'Cruz, this volume).

Aligning trust with trustworthiness, and mistrust with untrustworthiness is not simple. It requires intelligent judgment of others' trustworthiness or untrustworthiness, and the available evidence may not reveal clearly which agents and which institutions are trustworthy in which matters. Typically we look for evidence of reliable competence and honesty in the relevant activities, and typically the evidence we find underdetermines judgments of trustworthiness or lack of trustworthiness.

There are, I think, two quite distinct reasons why this is the case. The first is that available evidence about trustworthiness may be epistemically complex and inconclusive. Judging any particular case will usually require two types of epistemic judgment. Judgment will be needed both to *determine* whether a given agent or institution is trustworthy or untrustworthy in some matter, and to *interpret* evidence that could be taken in a range of ways. Both *determining* (alternatively *determinant*, or *subsumptive*) and *reflective* judgment are indispensable in judging whether some institution or office-holder is trustworthy in some matter.[5]

19

But determinant and reflective judgment are only part of what is needed in judging trustworthiness and lack of trustworthiness. Judgments of trustworthiness also usually require *practical* judgment. Practical judgment is needed not in order to classify particular agents and institutions as trustworthy or untrustworthy, but to guide the placing or refusal of trust where evidence is incomplete, thereby shaping the world in some small part. Practical judgment is not judgment of a particular case, but judgment that guides action and helps to shape a new or emerging feature of some case.

Decisions to place or refuse trust often have to go beyond determining or interpreting existing evidence. For example, where some agent or agency has meager powers it may make sense to place *more* trust in them than is supported by available evidence or any interpretation of the available evidence, for example because the downside of getting it wrong would be trivial or because the experience of being trusted may influence the other party for the better. In other cases there may be reason to place *less* trust than the available evidence or any interpretation of that evidence suggests is warranted, for example because the costs of misplacing trust would be acutely damaging.[6]

Making practical judgments can be risky. If trust and mistrust are badly placed, the trustworthy may be mistrusted, and the untrustworthy trusted. Both mismatches matter. When we refuse to trust others who are in fact trustworthy we may worry and lose opportunities, not to mention friends and colleagues, by expressing suspicions and by intrusive monitoring of trustworthy people. Those who find their trustworthiness wrongly doubted or challenged may feel undermined or insulted, and become less sure whether the effort of being trustworthy is worthwhile. And when we mistakenly trust those who are in fact untrustworthy we may find our trust betrayed, and be harmed in various, sometimes serious, ways. So in placing and refusing trust intelligently we have to consider not only where the evidence points and where it is lacking and where interpretation is needed but also the costs and risks of placing and misplacing trust and mistrust. Judging trustworthiness is *both* epistemically *and* practically demanding.

1.3 Aligning Trust with Trustworthiness in Daily Life

The epistemic challenges of placing and refusing trust well despite incompleteness of evidence are unavoidable in daily life. A comparison with the equally daily task of forming reliable beliefs is suggestive. Both in scientific inquiry and in daily life we constantly have to reach *beliefs* on the basis of incomplete rather than conclusive evidence, and do so by seeking *relevant* evidence for *specific* claims and by accepting that further check and challenge to those beliefs may require us to change them. Both in institutional and in daily life we constantly have to place or refuse *trust* on the basis of incomplete rather than conclusive evidence of others' trustworthiness.

However trust and mistrust can be placed intelligently by relying on *relevant* evidence about *specific* matters, by allowing for the possibility that further evidence may emerge and require reassessment, and also by addressing *practical questions* about the implications of trusting or refusing to trust in particular situations. In judging others' trustworthiness we often need to consider a fairly limited range of *specific* evidence. If I want to work out whether a school can be trusted to teach mathematics well, whether a garage can be trusted to service my car, or whether a colleague can be trusted to respect confidentiality, I need to judge the trustworthiness of that particular school, garage or colleague *in the relevant matter* – and any generic attitudes that others hold about average or typical schools, garages or colleagues will be (at most) marginally

relevant. We are not lemmings, and can base our judgments of trustworthiness in particular cases on relevant evidence, rather than on others' generic attitudes.

Typically we need to consider a limited range of quite specific questions. Is A's claim about an accident that damaged his car honest? Is B's surgical competence adequate for her to undertake a certain complex procedure? Assuming that C is competent to walk home from school alone, and honestly means to cross the road carefully, can we be sure that he will *reliably* do so when walking with less diligent school friends? Is he impetuous or steady, forgetful or organized?

In these everyday contexts, judging trustworthiness and lack of trustworthiness may be demanding, but may also be feasible, quick and intuitive. Indeed, judging trustworthiness and untrustworthiness is so familiar a task that it is easy to overlook how complex and subtle the epistemic and practical capacities used in making these judgments often are. Placing and refusing trust in everyday matters often relies on complex and subtle cultural and communicative capacities, such as abilities to note and respond to discrepancies between others' tone, words and action. These capacities are not, of course, infallible but they are often sufficient for the purpose, and can sometimes be reinforced by additional and more intrusive checks and investigation if warranted.

1.4 Aligning Trust with Trustworthiness in Institutional Life

Placing trust and mistrust well can be harder in institutional settings, and particularly so in the large and complex institutions that now dominate public and corporate life (see Alfano and Huijts, this volume). Here too our central *practical* aim in placing and refusing trust is to do so *intelligently*, by aligning trust with trustworthiness, and mistrust with untrustworthiness. But since much public, professional and commercial life takes place in large and complex institutions, and involves transactions that link many office-holders and many parts of many institutions to unknown others, it can be much harder to judge trustworthiness. Indeed, the assumption that there is a general decline in trust may not reflect greater untrustworthiness, but rather the current domination of institutional over personal connections, of the *system world* over the *life world*.[7] Some standard ways of addressing the additional demands of placing and refusing trust intelligently have been well entrenched in institutional life, but they also raise problems.

Two types of approach have been widely used to improve the alignment of trust with trustworthiness in complex institutional contexts. Some approaches aim to raise levels of trustworthiness *across the board*, typically by strengthening law, regulation and accountability. If this can be done the likelihood that trust will be placed in untrustworthy institutions or office-holders can be reduced. Other approaches seek to support capacities to place and refuse trust intelligently by making evidence of others' trustworthiness – or lack of trustworthiness – more available and more public.

Approaches that aim to improve trustworthiness often combine these forward-looking and retrospective elements. Forward-looking measures include establishing clearer and stricter requirements for trustworthy action, and for demonstrating trustworthiness, as well as stronger measures to deter and penalize untrustworthy action. The rule of law, a non-corrupt court system, and enforcement of contracts and agreements and penalties for their breach are standard ways of supporting and incentivizing trustworthy action. More laws are enacted; legislation in different jurisdictions is better coordinated; primary legislation is supplemented with additional regulation and with

copious guidance and (where relevant) by setting more precise technical standards (see Gkouvas and Mindus, this volume).

In parallel with these forward-looking measures for improving trustworthiness, retrospective measures for holding institutions and office-holders to account have also often been strengthened. These retrospective measures include a wide range of measures to secure accountability by monitoring and recording compliance with required standards and procedures. However, approaches to strengthening account-ability have a downside. They are often time-consuming, may be counterproductive, and at worst may undermine or damage the very performance for which office-holders and institutions are supposedly being held to account.[8] Over-complex ways of holding institutions and office-holders to account are widely derided by those to whom they are applied, and sometimes undermine rather than support capacities to carry the central tasks of institutions and of professionals. At the limit they generate perverse incentives or incentives to "game" the system. Even when incentives are not actually perverse, they may offer limited evidence that is useful for placing or refusing trust intelligently.

A second range of measures that supposedly improve trustworthiness focuses neither on regulating institutions and their office-holders, nor on holding them to account for compliance, but on making information about their action and their shortcomings more *transparent*, thereby enabling others to judge their trustworthiness – or untrustworthiness – for themselves. Transparency is generally understood as a matter of placing relevant information in the public domain. This can provide incentives for (more) trustworthy action, since untrustworthy performance *may* come to light, and *may* be penalized. The approach is not new: company accounts and auditors' reports have long been placed in the public domain, and similar approaches have been taken in many other matters.

However, transparency is often not particularly helpful for those who need to place or refuse trust in specific institutions or office-holders for particular actions or transactions. Material that is placed in the public domain may in practice be *inaccessible* to many for whom it might be useful, *unintelligible* to some of those who find it, and *unassessable* for some who can understand it.[9] Transparency does not require, and often does not achieve, communication – let alone dialogue – with others. Placing information about performance in the public domain may provide (additional) incentives for compliant performance by institutions and office-holders, but its contribution to "restoring" trust is often meager. Many people will have too little time, too little knowledge and too many other commitments to find and make use of the information.

So while additional law, additional regulation, and more exacting demands for accountability and transparency can each provide *incentives* for (more) trustworthy performance, they are often less effective than their advocates hope. Where account-ability requires too much of institutions and office-holders, one effect may be that excessive time is spent on compliance and on documenting compliance, sometimes to the detriment of the very matters for which they are being held to account. In some cases measures intended to improve accountability create perverse incentives, which encourage an appearance of compliance by supplying high scores on dubious metrics and by ticking all the right boxes. The idea that multiplying requirements and assembling and releasing ever more ranking and other information about what has been done or achieved will always improve performance is not always borne out, and is sometimes no more than fantasy.[10]

1.5 Trust and Mediated Discourse

Given the difficulty of measuring, monitoring and publicizing, let alone communicating, evidence of trustworthiness in institutional contexts, it is reasonable to ask whether further or different means could be used to support judgments of trustworthiness and untrustworthiness in institutionally complex contexts. What is needed is neither complete evidence nor conclusive proof, but adequate evidence of reliable honesty and competence that allows others to judge with reasonable assurance which office-holders and institutions are likely to be trustworthy in which specific matters, combined with a sufficient understanding of the practical implications of placing and refusing trust.

Often we do not need to know a great deal about the relevant institutions and office-holders, any more than we need detailed knowledge about others in everyday situations. Every day we manage with some success to place trust intelligently in drivers (whom most of us do not know) not to run us over, in retailers (whom most of us do not know) to provide the goods we purchase, in doctors to prescribe appropriate treatment (which most of us cannot identify for ourselves). However, in these everyday cases the task is feasible in large part because judgments of trustworthiness can focus on a limited range of *relevant* and *specific* matters and do not require comprehensive judgments of others' honesty and competence, or of their reliability.

However where action is mediated by complex systems it may be harder to find analogues of the cultural and communicative capacities that support the intelligent placing and refusal of trust in everyday life. Has the complexity of institutional structures perhaps now undermined capacities to judge trustworthiness? Or are there better measures which could support the intelligent placing and refusal of trust in institutional life? Although we seldom need to make across-the-board judgments of others' trustworthiness, difficulties can mount when we need to judge the trustworthiness of claims and commitments that depend on complex institutions or arcane expertise, and more so when communication and evidence pass through numerous intermediaries whose trustworthiness is not assessable.

Difficulties are seemingly compounded if there is no way of telling who originates or controls the claims that are made, or even whether content has been produced by persons or by automated microtargeting, and whether claims and commitments are evidenced or invented. The difficulty of placing and refusing trust in institutions and office-holders has recently been compounded by widespread assertions that "experts" are not to be trusted, by the easy proliferation of "fake news," by the fact that originators of claims can remain anonymous and that some content may have been produced, multiplied and targeted by artificial rather than human intelligence.

In a remarkably short time we have moved from hoping that digital technologies would support a brave new world of universal participation in discussion that would be helpful to trustworthy relations with others, and even to democracy, to something entirely different. Rather than finding ourselves in a quasi-Habermasian world, in which citizens can check and challenge one another's claims and reach reasonable views of their truth and their trustworthiness – or otherwise – we now find ourselves in a world in which these technologies are often used to undermine or limit abilities to assess the trustworthiness of others' claims. However, these problems may arise not from the technologies that now provide communications *media*, but from the fact that they allow content to travel via large numbers of unidentified and unknown *intermediaries* whose ability to modify, suppress and direct content is often unknown and undiscoverable.

It is easy to imagine that the problem lies in the *media* we use, rather than in the role of *intermediaries*. That is exactly the worry that Plato articulated in his account of Socrates' concerns about written communication. In *Phaedrus* Plato sets out the issues in these words:

> You know, Phaedrus, writing shares a strange feature with painting. The off-spring of painting stand there as if they are alive, but if anyone asks them anything, they remain most solemnly silent. The same is true of written words. You'd think they were speaking as if they had some understanding, but if you question anything that has been said because you want to learn more, it continues to signify just that very same thing forever. When it has once been written down, every discourse roams about everywhere, reaching indiscriminately those with understanding no less than those who have no business with it, and it doesn't know to whom it should speak and to whom it should not. And when it is faulted and attacked unfairly, it always needs its father's [i.e. its author's] support; alone, it can neither defend itself nor come to its own support.[11]

The worry about writing that Plato ascribes to Socrates is that texts can become separated from their authors, with the result that nobody stands ready to interpret or explicate the written word, or to vouch for its meaning, its truth or its trustworthiness. The passage contrasts writing with face-to-face, spoken communication in which hearers can ask speakers what they mean, and why they hold certain views, or act in certain ways. In doing this speakers provide "fatherly" support and help hearers to understand what they mean and allow them to check and challenge claims and commitments, and to reach more intelligent judgments about speakers' honesty, competence and reliability, and so about their trustworthiness. In face-to-face communication evidence is provided by the context of speaking, by speakers' expressions and gestures, and by the testimony of eyewitnesses. Hearers can use this immediate evidence to assess speakers' honesty, competence and reliability – and to detect failings. The relation between speaker and hearer, as Plato describes it, allows us to place and refuse trust in the spoken word, but is missing when we have only the decontextualized written word and its author cannot be identified, let alone questioned.

Yet while face-to-face speech has these – and other – merits, writing has advantages for judging trustworthiness. The fact that texts can be detached from writers and the context of writing can provide lasting and transmissible records that can be used by many, over long periods, and often provides indispensable evidence for judging trustworthiness and untrustworthiness. Although readers can seldom see writers' expressions and gestures, or judge the contexts in which they wrote, they have advantages that listeners lack. Writing supports intelligent judgment of the truth and trustworthiness of past and distant claims and commitments because it can provide a lasting trace that permits back references, reconsideration, reassessment and the creation of authorized versions. This is why writing is essential for many institutional processes, including standard ways of supporting and incentivizing trustworthiness, such as the rule law, reliable administration and commercial practice. By contrast, appeals to what is now known about ancient sayings may be no more than hearsay or gossip, and may offer little support for judging others' trustworthiness or untrustworthiness. The fact that the written word can bridge

space and time rather than fading with the present moment contributes hugely to possibilities for placing and refusing trust well.

Moreover, the contrast that Plato drew between spoken and written communication is obsolete. The communication technologies of the last century enable the spoken word and images to be recorded and to bridge space and time, and contemporary discussion about placing and refusing trust no longer debates the rival merits of speech and writing, or of other communications media. We live in a multimedia world in which speech, like writing, can be preserved across time and can be recorded, revisited or transmitted, rather than fading as it is spoken, so can often be checked or challenged, corroborated or undermined, including by cross-referring to written sources, images or material evidence. Although face-to face communication has distinctive and important features, these reflect the fact that speaker and listener are present to one another, rather than their reliance on oral communication. For us ancient debates about the rival merits of the spoken and the written word, of orality and literacy, are interesting but remote.[12] And yet Plato highlighted a genuine problem.

1.6 Media, Intermediaries and Cultures

While the *medium* of communication may not be the key to judging trustworthiness and untrustworthiness, the role of *intermediaries* in communication is fundamental. The new communication technologies of the last 50 years not only support a wide variety of *media* for communication (see Ess, this volume). They also make it possible – and sometimes unavoidable – to route communication through complex *intermediaries*. Some of these intermediaries are institutions and office-holders; others are components or aspects of communication systems and internal institutional processes including algorithmic processes. Not only can these intermediaries shape and reshape, redirect or suppress, communication, but they can often do so while remaining invisible, and without this being apparent or their contribution being known to or understood by the ultimate recipients of the communication that they (partly) shape. Those intermediaries that are institutions or office-holders may be honest in some respects and dishonest in others, competent for some matters but not for others; reliable in some contexts and unreliable in others: and each of these is relevant to the trustworthiness of mediated communication. Often intermediaries are shaped by the structure of communication systems, which most will find hard to assess, and which may not be open to any scrutiny.

It is this proliferation of *intermediaries*, rather than the differences between various communications *media* that shapes and modifies communicated content, and that can support or disrupt processes for checking or challenging, corroborating or undermining, mediated claims and commitments, and so also affect capacities to make judgments that bear on the placing or refusal to trust. Where intermediaries not merely transmit communicated content, but can edit or alter, suppress or embellish, insert or omit, interpolate or distort content, both the intelligibility and the assessability of claims and commitments, and capacities to judge trustworthiness and untrustworthiness, truth and falsity, may be affected and may falter or fail.

However, failure is not inevitable. Where mediated communication is entirely linear and sequential, as in the children's game of "Chinese whispers," and messages pass sequentially between numerous intermediaries, judging the truth or the trustworthiness of mediated communication may face insurmountable obstacles. Many approaches to judging others' claims and commitments would founder if we had to compare mediated

messages with originals in order to judge their truth or trustworthiness.[13] However, this image of serial communication, in which trustworthiness requires those at later stages of transmission to judge claims and commitments by reference to earlier stages, indeed to originals, seems to me mistaken.

Both communication and judging trustworthiness and lack of trustworthiness can be helped where *multiple* intermediaries are linked by *multiple* pathways with *varied* capacities to originate or modify content. Where messages can travel by many paths, abilities to note and respond to evidence of discrepancies in tone, words and action, and so to judge trustworthiness and untrustworthiness, can be supported by cultural as well as by formal legal and institutional measures. Institutional culture can then supplement the evidently incomplete approach to judging trustworthiness that formal institutional and digital processes offer.

However, institutional cultures also vary. Some are trustworthy and others are not. Some provide ways of judging trustworthiness and lack of trustworthiness, others damage capacities to judge trustworthiness and lack of trustworthiness. Cultures, like institutions and their office-holders, may be corrupt or untrustworthy. Neither a corrupt culture, nor a culture of mere compliance with institutional rules and requirements (law, regulation, accountability, transparency and internal rules) is likely to provide sufficient cultural support for judging or for fostering trustworthiness.

So if we conclude that culturally mediated ways of judging trustworthiness are necessary to augment those provided by institutional systems (law, regulation, accountability, transparency), it is worth working out which sorts of cultures might best be incorporated into institutional life. Rather than inflating and expanding formal systems for securing compliance and accountability yet further, it may be more effective to build and foster cultures that support trustworthiness and capacities to judge trustworthiness. Doing so would evidently not be easy, but there may be gains to be had by rejecting cultures of fear, intimidation or corruption as well as fragmented cultures that compartmentalize institutions into silos or enclaves,[14] and by considering how good cultures can contribute robustly to institutional trustworthiness.

Notes

1 Many organizations conduct attitudinal polls and surveys. Some – such as Gallup, IpsosMori or Eurobarometer – have become household names; some are less known or more specialized; some are covertly controlled by varying interest groups.
2 Moore (2018).
3 Taplin (2017).
4 O'Neill (2018a); Hawley (2019).
5 Kant (2000), 5:180.
6 O'Neill (2018b).
7 Habermas (1981); Baxter (1987).
8 I once heard the effects of excessive demands for documenting compliance nicely illustrated by a midwife, who told an inquiry into the safety of maternity care in England and Wales that it now took longer to do the paperwork than to deliver the baby.
9 Royal Society (2012).
10 Consider for example the unending debates about the metrics used to rank schools and universities in many developed countries, such as *PISA* rankings of schools or the Times Higher Education World University Rankings.
11 Plato, *Phaedrus*, 275d-e.
12 Ong (1980).
13 Coady (1992).
14 Tett (2016).

References

Baxter, H. (1987) "System and Life-World in Habermas's 'Theory of Communicative Action,'" *Theory and Society* 16(1): 39–86.

Coady, C.A.J. (1992) *Testimony: A Philosophical Study*, Oxford: Oxford University Press.

Habermas, J. (1985) *A Theory of Communicative Action*, T.J. McCarthy (trans.), Boston, MA: Beacon Press.

Hawley, K. (2019) *How To Be Trustworthy*, Oxford: Oxford University Press.

Kant, I. ([1790] 2000) *Critique of the Power of Judgement*, P. Guyer and E. Matthews (trans.), Cambridge: Cambridge University Press.

Moore, M. (2018) *Democracy Hacked: Political Turmoil and Information Warfare in the Digital Age*, London: Oneworld.

O'Neill, O. (2018a) "Linking Trust to Trustworthiness," *International Journal of Philosophical Studies* 26(2): 1–8.

O'Neill, O. (2018b) *From Principles to Practice: Normativity and Judgement in Ethics and Politics*, Cambridge: Cambridge University Press.

Ong, W.J. ([1980] 2002) *Orality and Literacy: The Technologizing of the World*, 2nd edition, New York: Routledge.

Plato (1973) *Phaedrus*, W. Hamilton (trans.), Harmondsworth: Penguin Books.

Royal Society (2012) *Science as an Open Enterprise*. https://royalsociety.org/-/media/policy/projects/sape/2012-06-20-saoe.pdf

Taplin, J. (2017) *Move Fast and Break Things: How Facebook, Google, and Amazon Cornered Culture and Undermined Democracy*, New York: Macmillan.

Tett, G. (2016) *The Silo Effect: Why Every Organisation Needs to Disrupt Itself to Survive*, New York: Simon & Shuster.