

This is a working copy (draft or preprint): It may differ slightly from final published version. Copyrights belong to the author and publishing permissions may have been transferred to the final publication venue.

Please consult <http://www.fredericbouchard.org> for links to the final version as well as other information about the author

Ceci est une version de travail (brouillon ou tiré à part) : cette version peut différer légèrement de la version finale publiée. Les droits d'auteurs appartiennent à l'auteur, mais certains droits ont pu être transférées à l'éditeur.

Veillez consulter <http://www.fredericbouchard.org> pour les liens à la publication finale officielle ainsi que pour avoir d'autres informations par rapport à l'auteur.

26/02/2016

THE ROLES OF INSTITUTIONAL TRUST AND DISTRUST IN GROUNDING RATIONAL DEFERENCE TO SCIENTIFIC EXPERTISE^{*}

FRÉDÉRIC BOUCHARD[§]

Debates about climate change, genetically modified crops and immunization have both strengthened for many the necessity to appeal to some form of scientific expertise while at the same time relativizing for others the epistemic authority of its claims. Scholars have offered diagnoses of this phenomenon. Their contribution has focused mostly on describing how expertise is marshalled, not why it is epistemically legitimate to invoke it. I will focus instead on two type of contributions that focus on the latter: some (e.g. Hardwig, Scheman, Grasswick) offer moral underpinnings for expertise based on trust in individual experts, while others (e.g. Goldman) offer an epistemic analysis of the reasons why it is sometimes reasonable to defer to scientific expertise regardless of trust in individuals. I will argue that both these “trust-centered” and “assessment-centered” accounts of expertise are partially right but that in order to develop a general theory of rational deference to experts, one needs to include the epistemic benefits of *both* trust *and* distrust. Using Blais’ and Merton’s work, I will argue that rational deference to experts based on institutionally organized skepticism about experts claims offers a better account of deference to expertise, than rational trust in individual experts: deference to expertise should be grounded in trust in scientific institutions, not trust in individual experts.

1- Introduction

Given the complexity of most phenomena, we have to delegate much epistemic work to other knowers and we must find reasons for relying on these specific knowers and not others. In our societies, these other knowers are often called experts and we rely on their epistemic authority more and more. For many complex phenomena such as climate change, genetically modified crops, and immunization, the experts that are called upon are *scientific* experts. For that reason, finding good reasons and justification for deferring to their knowledge is of paramount societal importance.

As we shall see, many have analysed this epistemic dependence on scientific experts and have tried to show in what contexts are there good reasons for epistemically depending on them. The salient feature of most of these takes on expertise is to highlight the role of trust in individual

*Accepted *Perspectives on Science* (Expected in Fall 2016)

http://dx.doi.org/10.1162/POSC_a_00224

§Frédéric Bouchard, Département de Philosophie, Université de Montréal, P.O. Box 6128, Station Centre-ville, Montréal, Québec, Canada H3C 3J7; email: f.bouchard@umontreal.ca

experts in grounding rational epistemic dependence to scientific expertise. In this paper I will show that, while trust may improve this epistemic relationship, it needs to be qualified in order to provide the means for rational deference. Rational deference should in fact be based on (temporary) *distrust* of individuals and how this distrust is implemented in scientific institutions and processes: this leads to the conclusion that, in the context of expert deference, one should trust institutions, not individuals. We will show why such an account may be better equipped to handle recognized limitations of trust-centered accounts of expertise. Such an account also has the advantage of recognizing the epistemic value of some of the skepticism held by laypersons toward science, but attempts to show that skepticism is more epistemically useful when utilized by the scientific institutions policing individual scientists' expert claims.

In the introduction and second section of the paper, I will describe briefly some aspects of the problem of expertise and epistemic dependence. In the third part of the paper we will shift to some notable treatments of expertise. We will compare some accounts that focus on the issue of trust in individual experts, "trust-centered" accounts (we will focus on Hardwig, Scheman and Grasswick), and accounts that do not focus on trust in individuals, "assessment-centered" accounts of expertise (Goldman and to some extent, Longino). As we will explain later, no account of expertise is completely 'trust-free'. A better label than 'trust-free' may be 'account where the trust relationship, if and when it obtains, does not ground the rational deference', or 'don't trust individuals, trust expert communities that distrust individuals'. For brevity's sake and because trust-centered accounts put so much weight on the moral aspect of the relationship between and individual expert and an individual layperson (while the latter accounts do not), we will provisionally adopt the shorter 'trust-free' label and later on in the paper I will show that a shift to an 'assessment-centered' may be the most fruitful. We will see that while "trust-centered" accounts are found lacking especially in the plausibility of its large scale applicability, "trust-free" accounts are found to be descriptively incomplete. Moreover neither account offer means of handling the problem of confirmation bias. In the fourth part of the paper, I will argue that "trust-free" accounts can be improved: for "trust-free" accounts to be more palatable, they have to offer a more refined understanding of the social epistemology it purports to develop, thereby offering an "assessment-centered" account of expertise. Reprising and developing some of Merton's work and developing some ideas of Blais on strategic trust (Blais 1987), I will argue that it is in fact the organized skepticism of scientific institutions (or to put it more provocatively, the inherent distrust found in scientific institutions) that can ground rational deference in scientific expertise in part because it explains how Goldman's source of evidence may work but also because it recognizes that some lay-communities' skepticism is not only reasonable but part of the scientific process itself. This will provide us with an "assessment-centered" theory of trust.

In our contemporary societies, *scientific* experts seem to play a special role in public discourse. The nature of expertise is fundamentally a collective or social problem since experts are *recognized* by others. You can call yourself an expert (university professors are often comfortable with this self-proclamation) but in an intuitive sense, you are not an expert until someone else recognizes you as an expert. *Scientific* expertise (the focus of this paper) appears to be recognized in our societies differently than other types of expertise. Although it is only a special type of expertise (one could be an expert dart thrower, an expert spin-doctor, an expert baker), I will argue that looking at this special case of expertise may help us understand certain features of expertise in general. I will argue that the justification of the reasonable deference to

experts lies with what Merton calls the “organized skepticism” of scientific institutions and how it maintains the conditions for “strategic trust” (Blais 1987). This may sound obviously true to many readers, but given various skeptical arguments found in the circles of climate-change deniers or anti-vaccination proponents, an analysis of the reasonability of expert-deference remains of some urgency.

Unless one adopts a constructivist or relativist analysis of expertise (i.e. experts are the actors recognized as such by their community [regardless of other epistemic features]), the fact that there seems to be an epistemic asymmetry between the layperson and the expert suggests that some agents are better than others (or better equipped) to describe or understand some phenomena in certain contexts. These phenomena may be a construction but in many cases there is a widely-shared belief that the expert has better access to some real facts about the world (e.g. is my water *really* contaminated?). This is the type of “modest realism” that Philip Kitcher invokes (2003) but one can find similar types of assumptions in other types of philosophy of science projects focused more on reformed notions of objectivity than on realism per se. Lloyd (1995) and Longino (1990) (among others) work on developing a clearer and more nuanced understanding of objectivity and the constraints that it operates under. This leads them to adopt a comparative and perfectionist understanding of our epistemic access to natural phenomena: there are better and worse ways of knowing, there are better and worse ways of explaining, and while the judgement of what counts as better or worse does not rest a naïve understanding of objective and unmediated access to the truth, it does not rest solely on an individual or group’s values or preferences either. Modest realism (or modest objectivity) assumes that there *really* are better ways of knowing. This assumption allows us to then ask whether there are reasons to believe that scientific experts have privileged access to some of these ways of knowing, and how is this privileged access generated and maintained.

2 Relativism and Expertise

Harry Collins and Robert Evans, who have been some of most influential writers on the topic of expertise, diagnose the issue as such in a text that aimed (and largely succeeded) at bringing the question of expertise to the forefront. These quotes give a good idea of the unease many have with regard with the topic of expertise.

The dominant and fruitful trend of science studies research in the last decades has been to replace epistemological questions with social questions, but we return to a rather old-fashioned approach, asking about the grounds of knowledge. What is different here, as compared with the debates about the grounds of knowledge that took place before the 'sociological turn' in science studies, is that we try to shift the focus of the epistemology-like discussion from truth to expertise and experience. (...) One of the most important contributions of the sociology of scientific knowledge (SSK) has been to make it much harder to make the claim: 'Trust scientists because they have special access to the truth'. Our question is: 'If it is no longer clear that scientists and technologists have special access to the truth, why should their advice be specially valued?' (Collins and Evans 2002, p.236)

They go on to highlight how the skepticism encouraged by the SSK has (unwittingly) weakened any and all epistemic authority.

By emphasizing the ways in which scientific knowledge is like other forms of knowledge, sociologists have become uncertain about how to speak about what makes it different; in much the same way, they have become unable to distinguish between experts and non-experts. (Collins and Evans 2002, p.239)

The aim of this paper, one might say, is to hammer a piton into the ice wall of relativism with enough delicacy not to shatter the whole edifice (the destruction that so many critics believe is the only solution). (Collins and Evans 2002, p.240)

So how does one reconcile relativism with genuine expertise or reasonable epistemic dependence? First let's note that "relativism" here is a methodological stance rather than a deeper claim about epistemological justification. "Relativism" relates to the principle of symmetry where one treats all theories and actors independently of whether it was later accepted that they were right or wrong. In other words, relativism in this context is akin to a methodological agnosticism about who is right or wrong rather than a claim about whether someone could ever be right or wrong about a given claim about a phenomenon. Even this deflationary understanding of relativism poses serious problems for any attempt to develop a general theory of expertise. Collins and Evans offer a complex proposition (that blossomed into a book, Collins and Evans 2007) that will not be fully examined here but a few points need to be highlighted for our purposes. While they provisionally reject the modest realist assumption, they also defend the view that some people are more credible than others over certain things. According to Collins and Evans, the traditional epistemic authority of experts derives not from epistemic virtues but from social features. They ask us to take seriously that other knowers *also* have epistemic authority. They do so by pointing to the importance of experience as core to expertise. A farmer has personal experience over his or her field in ways an agriculture expert from a Biology Department does not. It is difficult to see how methodologically relativist they truly wish to remain (see Jasanoff 2003, Wynne 2003 for comments); they argue that it's not that scientists are bad/wrong experts, it's that they aren't the *only* experts and they aren't epistemically privileged: Collins and Evans attempt to rehabilitate the scientific expert by integrating them in a broader and not exclusive epistemic community. Like most definitions for which the range of applications is broadened, there is a real risk of trivializing the concept of expertise beyond usefulness. How to avoid this trivialization of expertise ascriptions? If there are many other expert types beyond the scientific expert, who is right to do the attribution of expert roles? They argue that experts *about* expertise (i.e. scholars in HPS and STS) defined in terms of meta-expertise are the key to identifying the relevant experts (scientific or not). They argue that these meta-experts allow us to remain methodologically relativists while giving us a way of identifying the "right" (i.e. useful or credible) experts in any given context. Their view of expertise cannot ground fully expertise or rather cannot fully ground rational deference in scientific experts since they do not offer compelling criteria for how the meta-expert would establish when a scientific expert has legitimate epistemic authority. *How* are meta-experts to identify the *right* experts while retaining the methodological relativism implicit and explicit to Collins and Evans' approach?

This partial treatment of Collins and Evans should not dissuade the reader from looking at their approach (or that of their critics). But as they acknowledge, it remains a challenge to juggle both methodological relativism and genuine localized epistemic authority of certain knowers over certain knowledge objects. For that reason, other notable treatments of expertise have jettisoned the methodological relativism assumption while favouring more explicitly a modest realist assumption. These accounts will be our focus for the rest of the paper.

3 Trust and Expertise

For the rest of the paper, we will entertain theories of expertise that explicitly endorse some sort of modest realist or reformed objectivity assumption. Arguably, we could group modestly realist treatments of expertise into two broad families: trust-centered and trust-free accounts of expertise. Note that here we mean trust (or absence of trust) *in individual* experts (whereas later we will highlight the role of trust *in institutions*). As such, trust plays a role even in what I label ‘trust-free’ accounts, but this trust plays a secondary role in the rational deference to scientific experts. Arguments representing those two approaches that have had the most visibility are those of Hardwig, Scheman and others (trust-centered) and that of Goldman (trust-free). They both focus on epistemic dependence -the fact that an individual agent cannot know everything and therefore depends on other knowers for various knowledge propositions- but explain the dependence relationship in very different ways.

3.1 Trust-centered expertise

Hardwig (1985, 1991) develops a theory of expertise aimed mainly at understanding the epistemic asymmetry in biomedical research among researchers but also between medical experts and lay persons. This covers a broad range of settings, from the family physician giving medical advice to her patient, to pharmaceutical researchers and public health professionals. In all these cases, we are faced with the issue of trust between individuals. While the theory of expertise he develops could apply to other contexts of epistemic asymmetry, health related issues (especially for the example of the physician-patient relationship) impose the significance of trust. Hardwig’s is not the only trust-centered account of expertise (see Scheman 2001, Grasswick 2010, Whyte and Crease 2010 for notable accounts), but examining Hardwig’s account offers distinct advantages. Aside from the fact that it inspired many of the trust-centered accounts that followed, Hardwig examines the trust dependency of the individual, whereas other trust-centered accounts mentioned have focused on the relationship between science and society, i.e. how communities and groups interact with experts, not how individuals interact with experts. By focusing on biomedical expertise Hardwig’s account makes salient the personal and individual choices involved in deferring to an individual expert in a way that may be scalable to collective decisions about deferral. For that reason, Hardwig seems *prima facie* to have broader applicability to accounts focused on science-society interactions.

Hardwig (1985, 1991) offers a strong critique of epistemic individualism: he argues that it is often impossible to know by oneself. Given this fact, Hardwig argues that it would be *irrational* for the layperson not to defer to an individual expert’s pronouncements. The expert has a degree of epistemic autonomy (or epistemic independence) that the layperson does not have relative to

certain objects of knowledge and therefore, if the layperson wishes to have justified beliefs about things he or she is not epistemically autonomous about, he or she will have to defer to the expert. But since the layperson cannot assess the expertise of the expert she is relying on, there is a large amount of trust involved. As Hardwig starkly puts it, the layperson is trusting “blindly” (for the rest of the paper the expression will be replaced by “trusting on faith” or “trusting unquestionably” to avoid ableist language). The layperson must have some reasons for believing in the epistemic independence of the expert so it’s not completely on faith, but those reasons may be pretty thin. Hardwig’s argument has the advantage of taking the paradox of expertise head on: if one needed rich indications for evaluating the epistemic independence of the expert, the layperson would have to become an expert. According to Hardwig, we must sometimes trust experts relatively unquestionably because we cannot fully validate their claims independently. How could this trust as a sort of leap of faith ever be rational? How are we to argue for such faith-based initiatives if there are no initial reasons for trusting the expert in the first place?

Alamassi (2007) argues that we do in fact have some direct and indirect inductive evidence of expert reliability and credibility by considering that the expert’s testimony of belief in p is a form of evidence for p , and therefore that a weaker trust-centered account of expertise is possible and compatible with Hardwig’s account without arguing that a full suspension of judgement is necessary. Other accounts (e.g. Allchin 1999) highlight other indirect means of establishing and marshalling credibility that could be used to bolster expertise ascriptions. Instead of arguing for inferential reasons for depending on experts, Hardwig (1994) focuses on the character of individual experts and argues for an ethics of expertise. He proposes something akin to a virtue ethics for scientists and the moral education that should accompany such virtue ethics: scientists must be or must become trustworthy because of the importance of their work and the epistemic deference they must command. For Hardwig this is ultimately a question of character and moral education of individuals and, in this, Hardwig is somewhat optimistic about the feasibility of attaining the conditions of trustworthiness. Hardwig is not the only one to highlight the moral responsibility of individual experts (see Elliott 2006 for a different account of said moral responsibility at the individual expert level) but one can worry about the effectiveness of moral education to build character or about the content of the moral education itself. For those reasons we may wonder whether it would ever be reasonable to suspend one’s judgement, hoping that the expert is actually trustworthy in the way we care about. Also, even though Hardwig intends his account to be scalable from the individual expert to an expert community, it’s not clear how trustworthiness based on the character of individual experts could aggregate in a way to make expert communities trustworthy.

To answer this problem, other trust-centered accounts have focused on the establishment and maintenance of conditions of trust at the collective level. Some, such as Douglas (2005), have focused on the nature of the interaction between expert communities and the general public. She shows how experts can increase their trustworthiness by integrating the public into the construction of its objects. This makes the inclusion of expert advice more easily relevant in public discourse since society has already been mobilized in the investigation process itself. This approach integrates trust at a coarse-grain level and does not explicitly refer to the specific initial conditions of trust building (when there may be no trust at all to build upon). The biggest payoff of Douglas’ approach is that it describes some of the positive outcomes of a working trust relationship that have emerged (in this regard, she shares Hardwig’s optimism) and she gives

means of assessing whether successful relationships between experts and society have been achieved and to what degree.

The significance of trust in rational dependence on expert knowledge has been reprised cogently by other philosophers (such as Scheman 2001, Grasswick 2010, and Whyte and Crease 2010) who in various ways have explained how to rationally build and bolster trust in experts in the right contexts. Instead of focusing on partial success of expert-lay community relationships as do Douglas and Hardwig, Scheman, Grasswick and Whyte and Crease focus on failures and trust breakdown (e.g. Whyte and Crease adopt the metaphor of the “poisoned well”). How can we make experts trustworthy for collective decision making when there are good historical reasons for distrusting said experts because of past injuries that the expert community has caused or enabled? Focusing on examples where marginalized or disenfranchised groups’ knowledge of certain phenomena doesn’t cohere with experts’ claims (e.g. Grasswick give the example of Inuits’ experience and knowledge of fluctuation in Polar bear population), they argue that communication breakdown makes expert knowledge mobilization difficult or impossible. They show how there are various ways of building or mending those bridges such as insuring better representation of marginalized groups within the relevant epistemic communities and political bodies. In some sense, they all investigate just how far and in what conditions can trust be established and maintained (i.e. there can never be complete unquestionable trust but how can we sustain worthwhile trust relationships). In this regard, their appeals to trust is much more targeted and skeptical than Hardwig’s initial proposal.

While those trust-centered accounts reveal something essential about expert/lay-community relationships, there are genuine worries concerning their scalability (can they offer a general theory of expertise or do they instead focus on limit-cases of expertise?) and the robustness of their prescriptions (can we really build, rebuild and maintain trust in ways robust enough to ground rational deference in experts?). A few points: whereas one can wonder whether Hardwig’s and Douglas’ optimism is warranted, one can also wonder whether focusing on communication breakdown between marginalized groups and scientific experts is the best way to provide a general theory of expertise (note that the goal of the authors above wasn’t necessarily to provide such a general theory). It’s not clear that highlighting cases of the absence of trust (or complete distrust) is the best means of showing the necessity of some degree of trust in expert deference (possibly one of the justifications for the inclusion of the “Trusted Mediator” expertise case in Whyte and Crease, where trust is actively cultivated). Also, such accounts depend on a specific notion of the goals of science that may not be applicable to all cases of scientific expertise deference. Grasswick, reprising some aspects of Scheman’s account, highlights that one of science’s main function is to be trustworthy as a mode of knowledge acquisition that can inform collective action. In these frameworks, science should be socially relevant. The social relevance of *some* scientific results is undeniable, but it is unclear that focusing exclusively on the trustworthiness of the experts presenting these useful results, especially the trustworthiness *viz-à-vis* marginalized communities, can offer a general theory of expertise. Much of the history of science reveals that the identified goal of social relevance does not reflect many aspects of scientific practice and the expertise it generates. How does trust work for Paleobotanists, or experts of Babylonian mathematics, or String Theorists? They may still be socially mobilized as experts, but the *prima facie* absence of social relevance of their work makes moot the question of

trustworthiness in the way that it is framed by Scheman, Grasswick and Whyte and Crease. Grasswick identifies this problem and diagnoses it as such:

More “pure” sciences, with less obvious applications outside of simply the development of further scientific understandings, may not be as strongly impeded by a lack of trust of those situated outside the scientific community. But we needn’t decide whether or not there remain some “pure” sciences with no practical goals of inquiry at all in order to make the case that the trust of lay communities is likely an important factor in achieving the practical goals of many if not most scientific inquiries. If scientific results are going to be widely used, those outside of scientific communities need to trust scientific inquiry. (Grasswick 2010, p.393)

Trust-centered accounts often focus on cases where the social significance of the expertise is already recognized by the community. As Grasswick makes clear, this is obviously of great importance to the question of expertise but it excludes by definition cases where the practical goals are not yet clear. We do ask String Theorists to explain the universe to us, and many would defer to their expertise even though speaking of trust may not be the most appropriate description for the rational deference to their expertise. Trust is a strong value term that applies to situations of recognized great import to the layperson (which explains the moral underpinning of Hardwig’s account). Whyte and Crease define trust as such (but similar understandings could be found in Scheman and Grasswick): “For our purposes, trust means deferring with comfort and confidence to others, about something beyond our knowledge or power, in ways that can potentially hurt us” (Whyte and Crease 2010, p.412). Trust-centered accounts emphasize that the community *recognizes* that the knowledge or expert could potentially hurt it or some of its members. Yet many mobilizations of expertise do not meet this requirement. There are many cases of expertise where the significance or impact of the expertise is yet unclear or unrecognized. A general theory of expertise should encompass such cases as well. In other words, while I think Hardwig, Scheman, Grasswick, Whyte and Crease (and other trust-centered accounts) are right that trust is important and that the ways in which we must rebuild trust with marginalized and disenfranchised groups is of great moral and political importance, I remain unconvinced that trustworthiness is the best way to ground the initial rational deference in experts.

Aside from the problems that have been described so far, there are deeper epistemic reasons for being wary of a trust-centered account of rational deference to experts. According to Goldman (2001), trusting an expert (a crucial part of Hardwig’s argument) is problematic for many reasons. The most obvious and frequent is in cases of disagreements between equally trustworthy experts (Goldman 2001). Let’s assume that we have enacted all the prescriptions Hardwig, Scheman, Grasswick, Whyte and Crease have offered. How do we arbitrate between two equally trustworthy experts? Does this type of (frequent) conflict weaken trust in general? Ironically, by increasing the epistemic diversity of our expert communities (an important prescription of Scheman, Grasswick and Whyte and Crease’s account as well as Miller 2012) we may be increasing the probability that there will be broad disagreements among equally trustworthy experts. And trust-centered accounts of expertise do not offer the means to solve this issue.

Douglas (2012) offers hope that there are epistemic means independent of but compatible with trust-centered accounts to solve the issue of disagreements: a qualitative assessment of the completeness, transparency, communicability, scope and practicality of expert statements can help us assess the epistemic weight to be given to various kinds of expert claims. Even though Douglas does provide hope that there are ways of assessing the value of expert statements, the qualitative and imperfect nature of such a method also highlights that the limitations of trust-centered accounts cannot be completely remedied by other epistemic means.

The ultimate problem of trust-centered accounts is that while their remedies may improve the relationship between experts and laypersons in ways that are politically and morally necessary and in ways that may increase the epistemic diversity of viewpoints, it does so in ways which may reduce the epistemic grounding of the expertise that lay communities will accept. We need experts because our current knowledge does not seem to suffice. Experts need to tell us what we don't know, even sometimes what we do not want to hear. Trust-centered accounts argue that we will only listen to experts we trust *before* they offer their claims. Scheman and Grasswick suggest that part of this trust may come from having more people like us or with similar beliefs or values co-constructing the knowledge that grounds the expertise. But in doing so, trust-centered accounts run the risk of only feeding our confirmation bias with expertise that comforts our initial preferences and biases.

Confirmation bias may be the ultimate pitfall of trust-centered accounts of expertise. Beyond our own personal experience to this effect, there is much evidence that we accept much more easily what conforms to our held beliefs than evidence to the contrary. This “confirmation bias” (coined by Peter Wason 1960) is a robust finding that has launched many research projects in psychology and other social sciences, and proves deleterious in cases of expert dependence especially if trust (i.e. increased sympathy and number of shared beliefs and experience) is the bedrock to rational deference. One could reasonably trust an expert that may actually be wrong. If we maintain the modest realist assumption, we must have other means of rationally grounding deference that will give us reasonable means of choosing the right experts regardless of the confirmation of the statements to our previously held beliefs and values.

A lot of fruitful work on expertise has focused on the role of trust in epistemic deference and it is clear that trust is a crucial part of the equation for certain objects of expertise and its mobilization by society. But trust does not provide means of distinguishing competing experts, and it may increase the likelihood that we will overvalue the *wrong* expert. We need a theory of expertise that can rationally ground deference to expertise in ways that can handle disagreements among experts, that does not increase the likelihood of a confirmation bias, and ideally that is scalable across individuals and communities.

3.2 The sources of evidence of expertise

Goldman (2001) sees Hardwig's account of expertise (and indirectly other trust-centered accounts of expertise) as being unable to resolve the problem of competing trustworthy experts. For this reason, Goldman downplays the role of trust in favour of “sources of evidence” that

would help the layperson assess and even rank the experts that are presented. The sources of evidence are grounds of inference that the expert opinion is worthy of deference. In other words, this account of expertise focuses on identifying means of recognizing who is it reasonable to defer to in a given context. Trust-centered accounts argue that trust is necessary in expert-layperson relationships, but do not show that trust is sufficient. The problem is that trust alone cannot help us determine who are the *right* experts. One can therefore question whether trust is truly necessary or even desirable if the goal is not only to convince the public to take heed of expert advice but also to do so for sound epistemic reasons. Goldman offers a theory of expertise that does not depend as strongly on trust in individual experts but on grounds for rational deference (for brevity I will refer to such an account as “trust-free”, even though some form of implicit trust in the value of the interaction remains but comes to play only after the expert is already recognized as worth our consideration; this is part of the motivation for, later in this paper, shifting to the label “assessment-centered”). Goldman lists his sources of evidence to assess expertise as follows:

- (A) Arguments presented by the contending experts to support their own views and critique their rivals’ views.
- (B) Agreement from additional putative experts on one side or other of the subject in question.
- (C) Appraisals by “meta-experts” of the experts’ expertise (including appraisals reflected in formal credentials earned by the experts).
- (D) Evidence of the experts’ interests and biases vis-à-vis the question at issue.
- (E) Evidence of the experts’ past “track-records” (Goldman 2001, p.93).

All these indications are supposed to help the layperson establish whether it is reasonable to listen to the expert or not, and the hope is that when faced with conflicting experts, the differences in their experiences, experiments and conclusion will make it evident who we should defer to. It is assumed that it would be possible to do so without direct access to the knowledge of the facts under examination since various proxy can be used to assess the epistemic worth of the purported expert (diplomas, prizes, university affiliation and so on). These don’t prove the expert is right, but they act as indications that it’s reasonable to defer to their statements in for inferential purposes. Goldman describes the features of each source of evidence, but most readers will intuitively recognize the epistemic features that are highlighted by each and every one. A and B are about corroboration, C and E are about building a base of induction based on past successes, D is about identifying conflicts of interest.

While Goldman’s approach may cohere with traditional epistemological virtues and partly answer the problem of disagreement among experts (Goldman explains that two equally trustworthy experts will probably differ in one or more of these sources of evidence, but solutions such as the proposed by Douglas (2012) will probably also be needed), it does nothing to answer the skepticism that many marginalized groups may have with regards to these experts.

In other words, while it does not ground the initial deference on trust in individuals (thereby avoiding the problems highlighted in the previous section), it does not recognize the suspicion (in some communities) that will hinder any appreciation of these sources of evidence in the first place.

More importantly it does not offer a full answer to the problem of confirmation bias. Goldman is careful to show (for example for source B) that consensus or corroboration must be built in the correct way: individual experts must acquire their knowledge relatively independently for the aggregation of their advice to offer additional support to an expert's claims (see Miller 2013 and Beatty and Moore 2010). But his analysis leaves unresolved the problem that experts are worth our attention because they are recognized by other experts in various direct and indirect ways. This increases the likelihood that the herd will conform and that the chosen expert is 'merely' a well-recognized member of the herd. The risk of confirmation bias remains and may actually be higher than the risk of confirmation bias found in trust-centered accounts since Goldman's sources of evidence may actually increase the epistemic homogeneity of the expert class: the sources of evidence C and E may run the risk of increasing the intellectual and social capital of some experts that will be called upon. Said experts' epistemic authority may only reflect or confirm the opinions of the majority (or of the powerful) in the epistemic community.

Relatedly, what is lacking from Goldman's project is any indication why these criteria or their proxy are better than other types of justification of the epistemic authority of scientific experts. Aren't Goldman's sources of evidence merely reinforcing existing power structures, dishing out the social and intellectual capital in ways that respect the current order independently of added epistemic virtue? Let's assume for a moment that these sources of evidence or ground of inference are good means of identifying the better experts among a group of putative experts. Goldman provides no indications as to why we should expect to see these properties obtain in a reliable fashion in our world. *Why* do the credentials matter? *What* do they mean? *How* have they emerged and *how* are they maintained?

How could we improve Goldman's inferential take on expertise? I argue we need a better understanding of the external constraints under which the experts work, to assess when and where the conditions established by Goldman should prevail in a reliable fashion and why those constraints matter. We need a better understanding of the reward (and punishment) system that motivate scientist to conform to an epistemically desirable way of practicing science. We also need to take into account that there is for many laypersons and lay communities reasonable distrust and suspicion based on equally well established track-record of injury. While trust-centered accounts suffer from not being able to show who are the right experts to listen to, they are correct in their analysis of why the *right* experts are often ignored. Goldman's account is relatively trust-free, but it is also distrust-free thereby not giving its due to a significant hindrance to rational expert deference in our societies.

I will argue that the description of some institutional mechanisms of science can improve upon Goldman's account. Following Blais (1987) game-theoretical account of rational expert deference, I will show the reason that it's rational to listen to scientific experts is in part because the institutions they work in often *distrusts* them, or to put it less provocatively, the institutions

are concerned about the individuals' motivations, and polices their action to a high degree. This approach provides us on an *assessment-centered theory of expertise*. I will focus on how the structure of peer-review in particular grounds these institutional mechanisms and how they may maintain the payoff matrix described by Blais. Starkly put, while individual scientists trust each other on a regular basis (as Frost-Arnold 2013 reminds us), epistemic authority in science often works because of distrust which sets out collective cooperation games where defection of individuals comes at a high cost to everyone involved. So, whereas someone like Hardwig wishes to rationally ground trust in individuals, I think we should try to rationally ground deference based on distrust in individuals and trust in institutions. As we will see, contrary to trust-centered accounts, this gives us means of arbitrating between competing experts (using some of Goldman's sources of evidence), reducing the effect of confirmation bias (by actively trying to disprove claims in a Popperian and falsificationist approach by increasing epistemic diversity) while being relevant and scalable both for the individual layperson and lay-community interactions. Also, it recognizes that the skepticism of some laypersons is not an intrinsically misguided attitude.

4 Distrust, assessment, strategic trust, and expertise

We have seen some of the advantages and disadvantages of trust-centered and assessment-centered accounts of expertise. The preliminary conclusion is that trust-centered accounts are correct in their diagnosis of why experts are not listened to, but do not offer complete reasons as to why we should listen to them, and more importantly, they do not give us means of identifying the right experts. Trust-free accounts seem to give us means of remedying the situation but do not take into account how distrust hinders the possibility of ranking experts in the first place or do not provide the right tools for assessment of expertise in fraught contexts. My goal is to strengthen Goldman's sources of evidence approach with a sociological understanding of scientific institutions that takes into account some of the trust-centered accounts' concerns.

Goldman points to professional credentials and to corroboration by other experts as credibility granting mechanisms. But why do such mechanisms offer any epistemic virtues in an epistemically reliable way? Blais (1987) details the payoff matrix of the prisoner's dilemma under which experts and laypersons operate under in expert deference games. To put it simply, it's better (or less bad) for everyone (both the scientific community and the layperson community) for experts to be worth deferring to and for laypersons to defer to them. Trust in experts becomes a strategic matter and as such trust in experts and trustworthiness of experts become the less bad strategy. While this account solves many of the problems highlighted earlier, it does not detail how the payoff matrix is established and maintained in actual scientific practice. Blais does hint at rewards for cooperation and punishment for defection, but a deeper sociological understanding of the institutions involved is needed if we are to convince experts and laypersons that this strategy is optimal. I will briefly discuss famed sociologist of science, Robert K. Merton's views about science. As Richardson (2004) reminds us, Merton treats science as a special type of institution demanding its own independent type of analysis (see Cole 2004 for an assessment of the reception of Merton among some contemporary sociologists of science). For Merton, science is 'different' in how it knows, and it is 'better' in knowing (as in more successful in knowing the real world).

After recognizing how cultural contexts shape what scientists will investigate, Merton adds “The criteria of validity of claims to scientific knowledge are not matters of national taste and culture. Sooner or later, competing claims to validity are settled by the universalistic facts of nature which are consonant with one and not with another theory (Merton 1957 [1949]: 554).” So why should we expect scientific communities to be better equipped to identify these facts about the world? Merton offered what he viewed as a descriptive account of the norms guiding science. These norms or ethos *increase* the likelihood that the sort of epistemic virtues identified by Goldman (and in most traditional epistemology textbooks) will be actualized in various ways in scientific endeavours. For Merton, the “ethos of science” is the “affectively toned complex of values and norms which is held to be binding on the man [sic] of science. The norms are expressed in the form of prescriptions, proscriptions, preferences, and permissions” (Merton 1942, pp. 268-269).

In various works, Merton identifies four, and later five, norms of science that guide scientists in their work.

1-Communalism (earlier called communism): the idea that knowledge belongs and should be made available to all members of the scientific community. This reflects an ideal of Openness and Transparency.

2-Universalism: A scientific claim should be assessed independently of the properties of the scientist making it. For example, the nationality or race of a scientist should not affect our reception of her work

3-Disinterestedness: Scientific claims should be made independently of one’s interests. This is the ideal ‘pure’ science. It reprises intuitions most have about the epistemic toxicity of conflicts of interest.

4-Organized Skepticism: Scientists as individuals and as groups have obligation to scrutinize claims. For many, this norm was later understood in a Popperian fashion of falsifiability.

Later on he added a fifth norm:

5-Originality: Science has the norm that ‘better’ new claims are prized and valued.

Although Merton’s realism and progressivism muddle this question, he describes this approach as eschewing normative claims: he understands his account as a description of the norms of scientific institutions not whether these norms are a good or a bad thing. Merton describes the norms governing science, but intuitively we can see how these norms (if implemented) could provide some of the means for better knowing. Making claims available to others (for eventual tests or corroboration), not letting irrelevant externalities about agents (such as race or nationality) affect our evaluation of claims, controlling for conflicts of interest, and valuing corroboration or the process of falsification (let’s put aside originality for now since arguably it does not play an obvious role in expertise): all of these seem to describe values that could

produce better knowledge outcomes. This is not to say that they are necessary or sufficient conditions of good knowledge, but as Merton puts it, institutions or communities that have these values will likely produce better knowledge outcomes more reliably than communities that undermine those values (Merton's 1938 description about the decline of German science is striking in this regard). Insofar as good knowledge is helpful to our goals, we may wish to protect and encourage the norms that encourage knowing.

Many scholars critiquing Merton have marshalled much evidence to show counter examples or failures of Mertonian ethos in individual scientists or communities (e.g. Mulkey 1969, Barnes and Dolby 1970, Rothman 1972). Cases of frauds and monetary interests are easy to identify. The criticism of the Mertonian view of science is that although it aims for descriptive accuracy, it fails to describe many actual scientists' activities. Scientists are as corruptible as anybody else. But as many advocates of Merton correctly pointed out (e.g. Gaston 1978; Durbin 1984) these reproaches miss the point: Merton is describing institutional and community ethos, not actual individual motivations or behaviors. The norms are intended to reflect the regulatory ideals that collectively guide reward structures of scientific endeavour. Merton could be offering an accurate description of the norms in a community, even though the members of the community had deviant motivations or behaviors. But it is plausible that a community with the right norms may be more likely to produce the right corresponding outcomes. One could adopt a deflationist view and argue that communities that reject those norms may be missing out on good ways of improving knowledge outcomes. In this sense, Merton gives us the context that explains why the type of payoff matrix underlying the strategic trust identified by Blais (1987) may prevail: we have reasons for believing that those norms are often at work and we have reasons to believe that these norms do the policing that explains why individual experts wish to remain worth deferring to and not defect. Various studies have detailed Mertonian norms at work and the type of collective policing that they provide. Ranalli (2012) gives a Mertonian reading of how consensus was achieved by climate-scientists by in some sense following Mertonian Norms. "Good" climate science was identified by implicit and explicit evaluation of the character of the scientists according to Mertonian norms.

For our purposes, I wish to focus on some Mertonian Norms and how they could strengthen trust-free accounts of expertise. As Ranalli does, I would argue that mechanisms such as peer review show that scientific institutions have adopted the types of values described by Merton, especially "organized skepticism". The expression "peer-reviewed research" is trotted out in the scientific and public sphere as a marker of added epistemic value. We easily see how peer-review is an (imperfect) implementation of the organized skepticism norm that Merton identified, but we should not forget that it is also present to control for personal interest as well.

The initial idea of peer review (for the Royal Society) was to recognize authorship and precedence of claims, and to validate or reject new ideas (Wagner and Steinzor 2006). In this sense it is a form of skepticism. This model has then been copied to this day: peer review is structured in various ways (two or more reviewers, anonymous submission or not, anonymous reviewers or not, etc). But the actual structure of peer review hints that we are concerned not only about claims but about motivations and interests as well. It is more often than not done by multiple referees to control for individual biases and interests both on the submitter's side and on

the referees' side: if we believed we could really trust individual scientists, we wouldn't need peer review, or we could do it with a single evaluator. But we wish to scrutinize claims in part to control for personal biases on the author's part (i.e. we feel that we need a review process), and we are also worried about the interests of individual referees to accept or reject claims (i.e. we need more than one referee). The policing of all actors involved maintains Blais' payoff matrix and ground strategic trust. As Blais argues, the scientist making false claims or the one rejecting a paper for unwarranted reasons is in some sense defecting and losing out on the payoff of 'cooperating' in a 'healthy' and trust-worthy scientific community. Loss of respectability may entail among other things the inability to get research funds, promotions, the loss of prestige, and other benefits found in academic communities. Because of this payoff matrix, it is then strategic for the layperson to strategically trust experts coming out of those communities. What was lacking from Blais' account was a story of why and how the policing mechanisms emerged. Adding Mertonian norms to the picture helps us understand the emergence and the reliability of the policing mechanisms.

Peer-review is *skepticism* of one form or another, but we have good reasons to think that it is also *organized* as Merton claimed. We can see this in recent discussions in various scientific societies about the structure of peer review and the various epistemic merits of different structures of peer review (e.g. Jaschik 2011 for recent discussion in economics). The current professional societies' debates about the structure of peer review in their fields reflect many of Merton's norms at the institutional level and how they are real effective norms that constrain knowledge outcomes. Discussions about the structure of peer review for publications and grant attributions are about institutional regulatory norms and the main concern is about the effect of personal interest in claim making and claim evaluation. One could look at grant review panels (see Mallard et al 2009; Lamont 2010) and find similar Mertonian norms at work. The fact that scientific institutions have adopted these values and that various mechanisms try to impose these values on outcomes of scientific productions doesn't guarantee that science will be disinterested or communal or meet some other Mertonian ideals, but arguably it increases the probability that scientific production will be (even slightly) less biased than other claims of knowledge communities. Merton's initial theory was that the mechanisms that constrain scientific production are of a different sort than those found elsewhere by other knowers. In adopting a Mertonian view of scientific values but seeing them as an ideal enshrined in various scientific institutions (such as peer review), we can see why scientific experts rightly deserve our reasonable deference more than other knowers, not because the individual scientists are trustworthy but because science is generated by institutions that police scientists' epistemic authority more rigorously (or vigorously) than other types of institutions: Merton helps us see how the policing underlying the type of game-theoretical strategic trust in experts that Blais described could obtain. In turn, Blais' strategic trust and Merton's view of science explains why Goldman's sources of evidence obtain in a reliable fashion in scientific contexts and less so in others.

What is necessary for a more complete account of expertise is highlighting the positive epistemic role of distrust and skepticism about individual claims. Now ask yourself, if a broadly accepted feature of legitimate epistemic dependence is that the expert's claims be disinterested and that you can corroborate the results, what kind of institution can reliably generate some of these epistemic agents and claims? This is why it is easier to justify dependence on scientific experts

than other kinds of agents. It's not that the non-scientists are wrong or untrustworthy. It's not that scientific experts are necessarily right or trustworthy. It's that scientific experts' actions and claims are checked by institutions that have the useful critical epistemic values and ideals. Scientific experts are not trustworthy because they are objective and right about the truth of the universe, but because they've been given credentials by a system that is skeptical and critical about expertise. In addition to the sources of evidence identified by Goldman such as credentials or past track record, I argue that rational deference to scientific experts is grounded in part on the fact their claims are inspected by distrustful institutions that hand rewards and punishments according to the Mertonian norms they have internalized. Goldman and Blais offer reasonable grounds for deference and strategic trust, because scientists live in a skeptical (distrustful) Mertonian world.

This cynical account of expertise has the added advantage that it legitimizes the rational distrust that some communities have vis-à-vis expert communities. Building trust is hard, is failure prone, and may reinforce confirmation bias. Distrust is often a rationally and historically informed response to a negative interaction with experts that may be impossible to put aside (such as the infamous Tuskegee case and other cases discussed by Scheman 2001). Recognizing that parts of scientific institutions have a similar skepticism but with different motivations -the skepticism of peer review comes from Mertonian norms, whereas the distrust of certain communities stems from grievances- may be a way to offer rational deference to scientific expertise for trusting, "agnostic", and distrusting lay-persons and communities alike.

Organized skepticism as an ideal is also a means to attempt to control for confirmation bias. The organized skepticism found in peer-review and elsewhere in scientific institutions aims for skepticism often in falsifiability-like fashion. For this reason, it is better equipped at least in theory to handle and reduce confirmation bias. IF (and that is a big if) reviewers are chosen carefully, the possibility of confirmation bias is reduced relatively to other evaluation procedures. For this to be true, care should be taken in choosing credible reviewers that stand on different sides of a given debate, or that privilege different theoretical or methodological approaches. Reviewers and submitters that all agree may merely be forming an echo chamber of confirmation, thereby reducing the epistemic payoff provided by the review process. Diversity in scientific communities is key to their effectiveness.

Organized skepticism of this sort gives us some reason to believe that the expert claims that make it through peer review are worth one's consideration, more so than claims that have not been scrutinized in this way. In turn we can then consider that the individuals making those claims are worth our epistemic consideration. This is not to say that peer-review is perfect or even close to it. Many negative effects such as the "file drawer effect" encourage conservative consensus-like submission processes that downplay or silence negative or inconclusive results even though they may reflect the best empirical results on a given question. Self-censorship of scientists in is part fed by the conservative nature of peer-review. We can hope that the Mertonian norm of communalism encourages the disclosure of *all* research results thereby minimizing the conservative nature of peer-review and allowing for community scrutiny, but it is difficult to see how this degree of transparency can be made reality in practice. Given all these constraints, the policing that peer-review offers remains of some epistemic value in how it aims

at keeping everyone ‘honest’ in the knowledge game. With all its limitations, we should recognize that in many epistemic contexts, peer-review is better than nothing and more stable and effective than other internalized norms.

One of the things it does provide is a way of grounding rational deference to experts: one does not need to trust individuals if one thinks the institutions they are embedded in produce reliable results. The types of functional norms that make science work produce those reliable results, or rather we can have a reasonable expectation that it can produce them in a more consistent fashion than other types of institutional arrangements. Does that mean that scientific experts are reliable on every topic, or rather that it is rational to defer to them in all contexts? Of course not, but we now have the means to make sense of rational deference to them in certain contexts of public concern.

Some of the advantages of organized skepticism, such as its potential mitigation of confirmation bias, have been evoked but to get a fuller picture we now need to say a few words about disinterestedness. The ideal of disinterestedness reflects the hope that some sort of impartiality can emerge, that conflicts of interest can be identified and eliminated or quarantined, and that the truth (or as close as it gets) can emerge. One finds this concern among Goldman’s sources of evidence and in Merton’s norms of science. As Douglas (2008) correctly argues, experts’ work and claims reflect in various ways a large array of values. This value-ladenness weakens the plausibility that individual experts can ever be truly completely objective. Helen Longino’s work (especially Longino 1990) on the necessity for a pluralism of perspectives to counter the idiosyncrasies of individuals’ values and viewpoint, has inspired many research programs in this regard. Winsberg (2012) gives a powerful demonstration of this in his description of how climate models are generated and value-laden. Reprising one of Longino’s key points, he effectively argues that it’s not just the ‘big’ obvious values (ethical commitments, political ideologies, love of money and fame, etc), but also the values in the, as Winsberg put it, “nooks and crannies” of model choice that affect how science is conducted. This value-ladenness motivates Scheman and Grasswick to argue for better representation of marginalized groups in expert groups in part to find an echo of said-group’s values in expert claims. Miller (2012) offers a similar analysis arguing that meaningful consensus can only be achieved with the right social diversity (i.e. a broader array of values). Fehr and Plaisance (2010) go further in arguing that a specific array of values should be encouraged. In various ways, they all call for increased diversity within the expert community both to increase the potential for trust in experts and to possibly get to (as in Longino’s approach) a more objective knowledge by offering a broader set of perspectives.

Diversity plays a moral and a political role in how it aims at improving trust between various epistemic communities and it plays an epistemic role in how it functions to avoid individuals’ restricted viewpoints. But one should not underestimate the difficulty of achieving diversity in a fair and epistemically effective manner: what groups should be represented and how many representatives? At what point in the investigation? What do we do about segments of the population whose demographic representation decreases? Representation as a means to achieve the conditions of trust is difficult to establish and maintain in a fair way. This difficulty is not a reason to abandon the call for increased diversity but it may show that diversity depends on other criteria and justification as well. Increased diversity that fulfills the conditions for organized

skepticism are much easier to implement and maintain because they feed into the policing mechanism. As Longino (1990) argues, we do need increased representation of marginalized groups and social diversity, not primarily as a trust-building mechanism, but because this multiplicity of different values is the best way to control for the takeover by any particular values: diversity is part of the policing mechanism. In Mertonian terms, the diversity underlying the norm of Universalism helps sustain the system. The collapse of the whole system is possible if diversity is weakened, or if personal interest trumps Mertonian universalism. In Blais' terms, nobody wins by letting this collapse scenario play out. The better strategy is to cooperate in a healthy and credible scientific community instead of destabilizing it. With regards to the question of expertise, the epistemic virtue of diversity may be more robust and easier to implement than the political one. Science is value-laden. Expert advice is value-laden. But with increased epistemic diversity feeding Merton's organized skepticism, this value-ladenness can be assessed and expert claims evaluated in a way that attempts to control for individual interest. The ideal of pure science is obviously not reflected in all individual scientist's behavior and motivations, but it may be active at the institutional level in a way that controls for overt hijacking of evidence by certain values.

So *why* do marginalized groups distrust expert communities? Because they don't believe the experts share the same goals and values as they do. On some level, as Whyte and Crease highlight in their Poisoned Well cases (Whyte and Crease 2010) they distrust what they believe to be experts' values. What is often underappreciated is that certain norms of scientific institutions reflect a similar distrust of individual scientists' motives. The regulatory ideals of Merton should be encouraged so personal interests and values are mitigated (or quarantined) by a better functioning organized skepticism. This in turn may maintain the conditions making Blais' strategic trust possible. Such better functioning will depend on inclusion of a broader array of stakeholders but in this light stakeholder choice will depend less on political or demographic reasons than on epistemic diversity that can feed into Mertonian norms in order to provide more positive knowledge outcomes by providing the right assessment processes

Conclusions and future directions

I have argued that some sort of distrust and skepticism is necessary to provide an account of rational deference to experts that can handle the challenges of confirmation bias. To do so I have highlighted the epistemic importance of diversity to have functioning organized skepticism. *Scientific* experts, because of their institutions that aim to mitigate some of the preferences or values of individuals, are experts for whom deference is more rational than other types of expertise. The other types of expertise identified by Collins and Evans may be credible and trustworthy but the grounds for rational deference will have to be demonstrated by some other means.

An important point is that I have not said anything about how far this deference should go. Epistemic authority (even genuine) does not trump democracy. My point was rather that if democratic societies look for epistemic help in understanding some phenomena in order to make the informed decisions, it could do much worse than deferring to scientific expertise, for that kind of expertise offers the rational justification for its epistemic authority.

I have argued that since we are stuck with cases of epistemic asymmetry, depending on scientific experts is more reasonable than depending on other types of knowers. To understand why, we need to focus, as Blais (1987) suggests, on the ‘policing’ mechanisms of scientific inquiry, an “assessment-centered” view of expertise. What I provided here is a picture of how these mechanisms are grounded on some of the regulatory ideals internalized by scientific institutions. Mertonian values are not necessarily the motivation of all individual scientists, but they are embodied in many institutional mechanisms of science such as peer review. Scientific experts merit reasonable epistemic deference not because they are more objective and interest-free (they aren’t on both counts), but because, as Merton argued, their activity is regulated and assessed by institutions and processes with *ideals* of objectivity and disinterestedness. Their claims have a higher probability of being less tainted by interests and values incompatible with our quest for understanding.

The reasonable critique is that this wishful thinking about science impedes a real understanding of the interests molding scientific activity. Although I think this critique is saying something important, I think its cost is too high. We do depend on other knowers and we need some heuristics to determine who to listen to and when. If we are wrong about our heuristics, we will get hurt more often than not. I have argued that trust-centered accounts, while giving us means of convincing some groups to listen to some expert communities, do not provide the means of identifying the ‘right’ experts. Goldman provided general rules of inference (“sources of evidence”) to help us identify the right experts, but didn’t provide a description of the types of community they would reliably emerge from and does not address the lack of trust operating for many lay-communities. Merton’s ethos of science gives us means of addressing both lacunae. In unison with others before me (e.g. Gaston 1978; Durbin 1984) I have argued that one can find those Mertonian values in action (e.g. how various types of peer review processes exemplify most if not all of the Merton norms). Scientific experts have biases and interests, but the scientific enterprise, thanks to its regulatory ideals, tries to minimize the possible negative epistemic effect of these biases in a way that makes deference to scientific expertise more rational than deference to other types of knowledge where such means of surveillance over personal interests are not active. For that reason, rational deference in scientific experts is attainable *sans* trust or rather without trust in individuals, but with trust in distrustful institutions.

In short, scientific experts may have legitimate epistemic authority based on Goldman’s analysis and it may be strategic to trust them for reasons identified by Blais, but only because they are policed by Mertonian norms internalized by scientific institutions. One of the biggest gains of these arrangements is the fostering, via organized skepticism, of positive knowledge outcomes which in many contexts will increase the inferential grounds for deferring to scientific experts. It also addresses the lack of trust of certain communities since they can find an echo of their worries in the functioning of science itself. Trust-accounts try to reduce this distrust. I argue we should leverage it.

A significant implication of this argument is that because the legitimacy of scientific expertise is dependent on the values and ideals internalized in scientific institutions, it is wholly and fully

dependent on the health of these institutions and of its mechanisms. A lot of work has been done by political theorists on the conditions for healthy democratic institutions, but as Douglas and Fehr and Plaisance argue, philosophers of science and epistemologists now need to do similar work on scientific institutions. Understanding, preserving and improving the features of scientific institutions that buttress scientific authority (when it is legitimate) is of paramount societal importance. We need to do a better job in educating the broader public on the functioning of science to show that its regulatory ideals answer to worries similar to their own.

REFERENCES

- Allchin, Douglas. 1999. "Do We See Through a Social Microscope?: Credibility as a Vicarious Selector." *Philosophy of Science* 66, 3: 298.
- Almassi, Ben. 2012. "Climate Change, Epistemic Trust, and Expert Trustworthiness." *Ethics & the Environment* 17, 2: 29–49.
- . 2007. "Experts, Evidence, and Epistemic Independence." *Spontaneous Generations: A Journal for the History and Philosophy of Science* 1: 1.,
- Barnes, S. B., and R. G. A. Dolby. 1970. "The Scientific Ethos: A Deviant Viewpoint." *European Journal of Sociology / Archives Européennes de Sociologie* 11, 01: 3–25.
- Beatty, John, and Alfred Moore. 2010. "Should We Aim for Consensus?" *Episteme* 7, 03: 198–214.
- Blais, J. Michel, 1987. "Epistemic Tit for Tat" *The Journal of Philosophy*, 84, 7: 363-375.
- Coady, David. 2006. "When Experts Disagree." *Episteme: A Journal of Social Epistemology* 3, 1: 68–79.
- Cole, Stephen. 2004. "Merton's Contribution to the Sociology of Science." *Social Studies of Science* 34, 6: 829–844.
- Collins, H. M., and Robert Evans. 2002. "The Third Wave of Science Studies: Studies of Expertise and Experience." *Social Studies of Science* 32, 2 : 235–296.
- Crease, Robert, and Evan Selinger. 2006. *The Philosophy of Expertise*. Columbia University Press.
- Douglas, Heather. 2005. "Inserting the Public Into Science." In *Democratization of Expertise?*, edited by Sabine Maasen and Peter Weingart, 153–169. Netherlands: Springer, .
- . 2008. "The Role of Values in Expert Reasoning." *Public Affairs Quarterly* 22, 1: 1–18.
- . 2012. "Weighing Complex Evidence in a Democratic Society." *Kennedy Institute of Ethics Journal* 22, 2: 139–162.
- Durbin, Paul T. 1984. *A Guide to the Culture of Science, Technology, and Medicine*. Free Press.
- Elliott, Kevin C. 2006. "An Ethics of Expertise Based on Informed Consent." *Science and Engineering Ethics* 12, 4: 637–661.
- Fehr, Carla, and Kathryn S. Plaisance. 2010. "Socially Relevant Philosophy of Science: An Introduction." *Synthese* 177, 3: 301–316.
- Frost-Arnold, Karen. 2013. "Moral Trust & Scientific Collaboration." *Studies in History and Philosophy of Science Part A* 44, 3: 301–310.

- Gaston, Jerry. 1978. *The Reward System in British and American Science*. Wiley.
- Gingras, Yves. 2013. *Sociologie des sciences*. Paris: Presses universitaires de France.
- Goldman, Alvin I. 2001. "Experts: Which Ones Should You Trust?" *Philosophy and Phenomenological Research* 63, 1: 85–110.
- Grasswick, Heidi E. 2010. "Scientific and Lay Communities: Earning Epistemic Trust through Knowledge Sharing." *Synthese* 177, 3: 387–409.
- Haack, Susan. 2005. "Trial and Error: The Supreme Court's Philosophy of Science." *American Journal of Public Health* 95: S66–73.
- Hardwig, John. 1985. "Epistemic Dependence." *The Journal of Philosophy* 82, 7: 335–349.
- . 1991. "The Role of Trust in Knowledge." *The Journal of Philosophy* 88, 12: 693–708.
- . 1994. "Toward an Ethics of Expertise." In *Professional Ethics and Social Responsibility*, edited by Daniel E. Wueste. Rowman & Littlefield.
- Jasanoff, Sheila. 2003. "Breaking the Waves in Science Studies: Comment on H.M. Collins and Robert Evans, 'The Third Wave of Science Studies.'" *Social Studies of Science* 33, 3: 389–400.
- Jaschik, Scott. 2011. "Rejecting Double Blind | Inside Higher Ed." *Inside Higher Ed*, May 21, 2011.
- http://www.insidehighered.com/news/2011/05/31/american_economic_association_abandons_double_blind_journal_reviewing.
- Kitcher, Philip. 2003. *Science, Truth, and Democracy*. USA: Oxford University Press.
- Lamont, Prof. Michèle. 2010. *How Professors Think: Inside the Curious World of Academic Judgment*. Harvard University Press.
- Lloyd, Elisabeth A. 1995. "Objectivity and the Double Standard for Feminist Epistemologies." *Synthese* 104, 3: 351–381.
- Longino, Helen E. 1990. *Science as Social Knowledge*. Princeton: Princeton University Press, .
- . 2006. "Philosophy of Science after the Social Turn." In *Cambridge and Vienna*, edited by Maria Carla Galavotti, 12:167–177. Berlin/Heidelberg: Springer-Verlag.
- Lynch, Michael. 2007. "Expertise, Skepticism and Cynicism: Lessons from Science & Technology Studies." *Spontaneous Generations: A Journal for the History and Philosophy of Science* 1, 1.
- Mallard, Grégoire, Michèle Lamont, and Joshua Guetzkow. 2009. "Fairness as Appropriateness." *Science, Technology & Human Values* 34, 5: 573–606.
- Merton, Robert K. 1938. "Science and the Social Order." *Philosophy of Science* 5, 3: 321–337.
- . 1979. *The Sociology of Science: Theoretical and Empirical Investigations*. USA: University Of Chicago Press.
- Miller, Boaz. 2013. "When Is Consensus Knowledge Based? Distinguishing Shared Knowledge from Mere Agreement." *Synthese* 190, 7: 1293–1316.
- Mulkay, Michael. 1969. "Some aspects of Cultural Growth in the Natural Sciences." *Social Research* 36, 1: 22–52.
- Ranalli, Brent. 2012. "Climate Science, Character, and the 'Hard-Won' Consensus." *Kennedy Institute of Ethics Journal* 22, 2: 183–210.
- Richardson, Alan. 2004. "Robert K. Merton and Philosophy of Science." *Social Studies of Science* 34, 6: 855–858.
- Rothman, Robert A. 1972. "A Dissenting View on the Scientific Ethos." *The British Journal of Sociology* 23, 1: 102–108.
- Scheman, Naomi. 2001. "Epistemology Resuscitated: Objectivity as Trustworthiness" pp.23-52. In *Engendering Rationalities*, edited by Sandra Morgen and Tuana, Nancy. SUNY Press.

- Wagner, Wendy Elizabeth, and Rena Steinzor. 2006. *Rescuing Science from Politics: Regulation and the Distortion of Scientific Research*. Cambridge University Press.
- Wason, P. C. 1960 “On the Failure to Eliminate Hypotheses in a Conceptual Task.” *Quarterly Journal of Experimental Psychology* 12, 3: 129–140.
- Whyte, Kyle Powys, and Robert P. Crease. 2010. “Trust, Expertise, and the Philosophy of Science.” *Synthese* 177, 3: 411–425.
- Winsberg, Eric. 2012. “Values and Uncertainties in the Predictions of Global Climate Models.” *Kennedy Institute of Ethics Journal* 22, 2: 111–137.
- Wynne, Brian. 2003. “Seasick on the Third Wave? Subverting the Hegemony of Propositionalism: Response to Collins & Evans (2002).” *Social Studies of Science* 33, 3: 401–417.