



MASTER IN ANALYTIC PHILOSOPHY
TFM/Final Research Paper
September 2024

The Problem of Expert Appraisal: Metacognition and the Rationality of “Thinking for Yourself”

by Cathlene Cruz Centeno

Under the supervision of:
Dr. Thomas Sturm



UNIVERSITAT DE
BARCELONA



Universitat
Pompeu Fabra
Barcelona



Acknowledgements

I would first like to thank my parents, Arlene and Catalino, for their unwavering love, support, and patience. I hope you guys are proud that in pursuing epistemology, I have finally managed to put my trust issues to (hopefully, good) use. Next, I would like to both thank and apologize to my good friend, Aslak Larechibera, for his encouragement and insight, having been at the receiving end of my late night rants on my topic. I consider myself very lucky to have you in my corner. Long overdue is my thanks to my undergraduate professor, Dr. Nathan Ballantyne, whose lessons on epistemic humility have been invaluable to me all these years and well beyond the classroom. Finally, I would like to thank my supervisor, Dr. Thomas Sturm, for his encouragement, patience, and many enlightening conversations. Most of all, thank you for imparting to me that philosophy is most rewarding when done in the service of helping others.

Abstract

In our current age of hyperspecialization, we often turn to experts to help us settle many of our important questions. But faced with uncertainty of when to trust experts and which ones to trust, we are led to ask, when and to what extent is it rational for novices to think for themselves? In response, I will argue that 1) there are some specific cases where the novice would not be rational in deferring to expert judgment, 2) but that for most others, the rationality of thinking for oneself becomes doubtful ultimately due to problems of metacognition, and 3) that this does not support the pessimistic conclusion that that the novice is essentially “blind” in her trust of experts. Ultimately my aim is to defend the novice’s rationality in thinking for herself even in difficult cases of expert deference, and against the claim that such deference cannot be rationally grounded. In Section 1, I develop a more realistic understanding of expertise and highlight cases where the novice would be rational in thinking for herself when there is empirical evidence of expert unreliability. In Section 2, I show that even in deferring to experts, novices must still think for herself by engaging in some degree of inquiry about which source and expert to believe. Although second-order appraisal strategies have been proposed to guide this inquiry, they ultimately run into problems concerning metacognition. In Section 3, I argue against Neil Levy’s (2024) pessimistic argument that these problems render the strategies unemployable and that the novice cannot be rationally justified in thinking for herself via their use. I then propose further strategies the novice may use, how this may be incorporated into the Boost program, and conclude by suggesting further avenues for future research.

Keywords: expert, novice, psychology, rationality, metacognition, education

The Problem of Expert Appraisal: Metacognition and the Rationality of “Thinking for Yourself”

0. Introduction: A Few Figs from Thistles?

For many of our important questions, we often turn to the experts for answers that will help us to make more informed and accurate judgements. Typically, this is regarded as the most epistemically responsible course of action. After all, virtually all of us are novices who lack specialized knowledge on many subjects and cannot practically inquire into every question ourselves—and even those of us who do possess expertise in some fields are novices in a disproportionately greater number of others. This cognitive division of labor between experts and novices, however, is also a source of tension: On the one hand, it is granted as the kind of “default” that deference to those who know more than us is the rational response (DiPaolo, 2021; Hardwig, 1985; Fricker, 2006). On the other hand, we also extoll the values of autonomous, critical thinking. Such is what motivates the popular imperative to “think for yourself” or, as more recently formulated as an internet slogan bearing more directly on inquiry, “Do Your Own Research” (Reid, 2024; Ballantyne et al., 2022). These oppositions consequently raise the following questions: 1) When is it rational to defer to experts? And 2) when and to what extent is it rational to think for yourself?

These questions of deference and discernment are age-old. Plato, in one of his dialogues, *Charmides*, questions whether a man who is aware of his own ignorance can correctly judge whether another person truly knows what he claims to know, and discern a true doctor from a mere fraud (170d-e). In the Gospel of Matthew, Jesus addresses the problem of discerning true prophets from the false, advising his followers that “Ye shall know them by their fruits. Do men gather grapes of thorns, or figs of thistles?” (Matthew 7:16). Given that neither thorn nor thistle bear fruit, this topological metaphor signifies that credible emissaries of truth can be identified by our recognizing the very truth of their messages. Augustine interrogates these questions in even greater depth in “The Advantage of Believing,” where he asks “But if truth itself forces you to conclude that only a few have it, yet you do not know who those few are, what then?” (Section 16, p.128). Ultimately, he responds that it is reasonable for the ignorant to defer to the knowing few, but notes the difficulty of identifying who these knowing few are, given their own ignorance (Section 24, p.135). On the other side, we famously have René Descartes’ 17th-century “Discourse on the Method of Rightly Conducting One’s Reason and of Seeking Truth in the Sciences,” advising us not to blindly follow others and to pursue truth ourselves. In the following century, we are famously told by Emmanuel Kant in “What is Enlightenment?” to “have the courage to use

your own understanding!” (Sapere aude!). Already we seem to have a case of (philosophical!) expert disagreement.

Of course, these claims are best understood when historicised, with particular attention to controversies concerning the intellectual authorities of their time. Our age is no different, and we can also ask— for as long as these questions and iterations of expert deference and appraisal have been around— why the urgency to address them, now? For one, we live in a time where the proliferation of expertise has brought us to an age of hyperspecialization. Ironically, that there are now more putative experts than ever, along with the corresponding rise in expert dependence, has also provoked an alarming counterculture of expert skepticism and populist distrust (Grundmann, 2017, p. 30). Second, beyond encountering expert counsel when we are actively in its pursuit, we are now bombarded with alleged expert statements in the news, social media, and advertising. Third and even more sobering, is that this increased potential for misinformation and expert misguidance poses a threat to deliberative democracy and participation in the legislative process—particularly on the pressing matters of climate change and health-crisis management (Anderson, 2011; Goldman, 2016; Ballantyne, 2022; DiPaolo 2022; Reid, 2024; Viehoff, 2023).

Epistemologists have taken note of the urgency and practical salience of addressing the expert deference problem. Much of the traditional epistemological literature has erred towards the default view that deference to experts is the rational choice and has focused on whether the novice can reasonably choose between two disagreeing experts. This question, however, has largely been “obliquely addressed,” with the focus on whether novices’ capacities for appraisal are possible, at least in theory (Ballantyne, 2022, p. 229). On the large, the concluding call is for a happy-medium via the intuitively attractive position that we “need to defer to experts in some cases, but not do so uncritically;” but what such cases would be and what extent of critical scrutiny looks like more concretely, is left open (Reid, 2024, p. 5). Finally, while much of the recent literature has recognised the potential cognitive impediments to novices’ appraising experts and so has called for corrective educational measures, how these could practically be implemented or incorporated into an ameliorative program has again, been underdeveloped.

In this paper, I hope to address these lacunae by way of responding to the question: Faced with uncertainty of when to trust experts and which ones to trust, when and to what extent is it rational for novices to think for themselves? Here, thinking for oneself is to be understood in both a stronger and weaker sense: The stronger sense refers to cases where the novice makes judgements directly on a given question that she may otherwise defer to experts

for. The weaker sense refers to cases where the novice has decided to defer to the expert, but nevertheless must engage in some form of inquiry to obtain and sift through the expert responses, and to judge which experts are credible. Ultimately, I will argue that 1) there are some specific cases where the novice would not be rational in deferring to expert judgment, 2) but that for most others, the rationality of thinking for oneself becomes doubtful ultimately due to problems of metacognition, or judging the reliability of one's own judgment, and finally 3) that neither of these would amount to the "epistemic trespassing" it has recently been charged with, nor the pessimism that the novice is essentially "blind" in her trust of experts. Ultimately my aim is to defend the novice's rationality in thinking for herself even in difficult cases of expert deference.

In Section 1, I will begin by expanding the traditional understanding of expertise to be more reflective of how expert deference occurs in reality, and how we should distinguish between interactional and contributory expertise. This distinction will be helpful in then circumscribing the cases where the novice would not be justified in deferring to experts in interactional domains, in light of empirical evidence that expert judgment is no better or even worse than novices'. This insights of this section will also be necessary for the development of further appraisal strategies in the following sections. In Section II, I will discuss the more difficult cases in which we would be more rational in deferring to experts, but for which we are still left with some onus of thinking for ourselves due to the research and appraisal of expert credibility that deference nevertheless involves. Addressing how we might responsibly go about such inquiry, I will offer a brief overview of the second-order appraisal strategies typically proposed by epistemologists, followed by the objection to their employability from the recent literature on metacognitive bias. I will end this section by introducing an argument from Neil Levy (2024) which takes the above findings to its pessimistic end that the novice is ultimately "blind" in her deference to experts, thereby threatening the rationality of the novice thinking for herself. In Section III, I will argue against Levy's position by addressing each of his three main claims. In the process, I will address the problem of charging novices with "epistemic trespassing," offer corrective strategies for countering the problems of metacognitive bias described in the previous section, dispute the detection of bias as a defeater in the first place, and propose further strategies for expert appraisal that the novice can deploy without herself possessing expertise. Finally, I will conclude this section by suggesting how the above can be incorporated into the Boost Approach, an ameliorative program in rationality.

To support my arguments, I will be drawing from the literature in epistemology, sociology of expertise, cognitive psychology, and behavioral economics. While epistemologists have already begun to take note of problems of metacognition in individual judgment and inquiry, I hope to make some modest contributions to the literature by developing a more realistic and comprehensive account of expertise, addressing whether epistemic trespassing can apply to novices as it has thus far been exclusively applied to experts, identifying more concretely the cases where expert deferral is unwarranted, and linking proposed solutions for metacognitive problems to an existing education proposal which—despite having much overlap with the existing literature and its repeated calls for educational corrections—has surprisingly gone unmentioned in said literature. Hopefully, venturing towards hitherto underaddressed and disconnected areas will help us to find a few figs from thistles.¹

A few points of clarification before proceeding: First, my discussion will initially feature much mention of “bias” and “debiasing” strategies. Although much of this discussion will refer to bias in a context where it does indeed obstruct human judgment in inquiry and expert appraisal, it is not an endorsement of the tradition that exists in mainstream psychological literature which understands bias in exclusively negative terms. On the contrary, this tradition will be discussed in greater detail in Section III, and disputed in favor of a more charitable understanding of bias. Second, it is important to acknowledge that there already exists a classic argument against deferring to experts, which relies on the difference in goals between the expert and novice. This is notably advanced by Steve Fuller (1986) who argues that while experts can provide factual information, this is useful for instrumental rationality where the novice can consider how such information may or may not be conducive to her ends; however, the greater consideration of substantive rationality—what ends are worth pursuing—are ultimately up to the novice, such that she is the final authority on whether it is (ir)rational to defer to an expert. Under this understanding, a novice would not be rational in deferring to experts in instances where their goals differ (Pierson, 1994). I will not be appealing to this argument, as my discussion will focus more on empirically supported cases of unreliable expert judgment, even in cases where the expert and novice share the same goal. Moreover, I will be challenging the underlying assumption in the above argument that experts merely offer facts.

¹ The phrase “a few figs from thistles” is a reference to poet Edna St. Vincent Millay 1922 poetry collection of the same name. Millay inverts the phrase’s original formulation from the Gospel of Matthew, to signify that we can indeed find truth or wisdom in places where we least expect.

1. Updating our Understanding of Expertise: Interactional vs. Contributory Domains

1.1. The Traditional Account of Expertise

The definition of expert that has been most prominent in mainstream epistemology is that of Alvin Goldman's "veristic" or "truth-based" account (Hardoš, 2018, p.272; Hjørland, 2024, Section 2). According to this definition, experts are those who 1) possess a "substantial body of truths" in a given domain and 2) have the methodological competence or "know-how" to engage in inquiry that answers new questions that arise in said domain (Goldman, 2001, p.91). This account is motivated by several constraints, the first of which is most relevant to our discussion due to its more problematic downstream effects. This constraint is that Goldman explicitly excludes performance-based experts as "these are not the kind of experts with which epistemology is most naturally concerned;" rather, of epistemological concern are "cognitive or intellectual experts" who are able to offer true beliefs (p. 91). Over the next twenty years, Goldman revisits the subject of expertise, maintaining his veristic definition and reiterating the distinction between performative experts and cognitive/intellectual ones. For example, he writes in a 2016 article, "Let us put the performance domain to the side and focus exclusively on the informational or cognitive domains" (p. 4). His 2021 article does make an important amendment to include both intellectual and performance domains in his treatment of expertise, the latter more precisely defined as those who can help the novice with their "ability to produce actions or results, such as curing physical ailments or identifying profitable investment plans" (p.87). Still, however, the categorisation of expertise as either intellectually-based or performance-based, is still preserved.

On the other side of the coin, novices (also referred to as laypeople) are characterized as those who do not have the sufficient knowledge or competence to directly settle questions in a domain. Although the possibility of upward mobility is possible, such as when a novice improves their epistemic station by acquiring expertise herself, it is normally assumed that the novice remains a novice. This is fairly reflective of the reality that one's becoming an expert is usually less common if not unfeasible due to practical constraints such as time and resources. Erring on the side of caution, though at the risk of uncharitability, the bottom line is that the gap between the novice and expert is insurmountable. This understanding is widely "taken as orthodoxy in social epistemology" as evaluations about first-order claims, or claims directly about the evidence in an expert domain, "requires being an expert

oneself” (Miller, 2021, Section 2.2). For the novice to remain within her ken, she must take recourse to either consulting the experts about the question at hand, and in the event of receiving conflicting answers, using indirect second-order evidence bearing on the experts’ credibilities to choose which one to trust. The transactional picture this paints is thus of the following: The novice “selects a candidate from the available pool,” “poses the question to the expert” who is then expected to offer a “true answer,” so that the novice then “decide[s] whether to accept the answer.” (Goldman, 2016, p.8).

1.2. Problems with the Traditional Account

This traditional understanding of expertise, however, has trouble accounting for how we turn to experts in reality. Its first deficiency is that it is limited to cases where novices have epistemic ends and is searching for a true, factual response. In effect, the traditional understanding is most applicable to scientific domains, as evidenced by the way the literature typically uses scientific questions when offering examples. Questions in comparatively “messier” fields such as politics and economics, are not typically addressed. And yet we also accept that there are experts in non-scientific domains, and similarly practice turning to them for questions which matter to us and which bear on our lives going well. In such cases, our questions do not necessarily have epistemic ends, but seek answers which are prescriptive, action-guiding, and sensitive to the stakes at play. Note the difference, for example, between asking (to use an example to which Goldman often returns) “When will there be a solar eclipse in the future?” versus “Will it be advisable for my family to immigrate to that country?” A comprehensive account of expertise should also be able to account for the latter. Second, the transaction imagined under the traditional account no longer reflects how novices seek and receive expert counsel in the digital age. Instead of finding an individual expert to pose a question, we more typically rely on technology which has afforded us the ability to do a quick online-search that pulls up a plethora of websites containing expert statements— all within seconds. Much of the time, such statements are no longer credited to a particular individual expert, but to a body or established organization of experts. For example, we often attribute and take as credible, expert statements to bodies like the World Health Organisation or Center for Disease Control (eg. According to the CDC...” “In a recent statement from the WHO...”).

More recent authors have similarly recognized these shortcomings of the traditional account of expertise. A synthesis of their criticism will be helpful to making more positive

claims about the different kinds and functions of expertise. Asher Miller (2021), for example, locates the foundations of the traditional account in social epistemology, and the “core assumptions” on which it rests. Since the traditional account extends and reflects these core assumptions, it inherits problems that result in the limitations discussed above. Particularly consequential are the assumptions of “the objectivity of facts” and “veristic idea,” as these contribute to the “orthodox assumption in social epistemology that experts simply transmit ‘truth’” that does not need interpretation (Miller, 2021, Sections 2.1, 1.2).² Hence the silence of the traditional account on “messier” domains like politics and economics, which are not understood to deal in truth or fact. Contrary to the traditional “one-size-fits all account of expert testimony” that has since been adopted, Miller concludes by advising that the novice first distinguish between experts who squarely relay facts and those who do not.

1.3. Interactional vs. Contributory Expertise

Sociologists of expertise have issued a similar call for finer grained distinctions of expertise. Pavol Hardoš (2018) agrees with Goldman’s veristic definition, but builds on its clause that an expert is someone who is capable of engaging inquiry to produce new answers in the domain. He highlights that this skill should not be divorced from the social implications that the expert does so for the purpose of offering such answers to the public. As such, our understanding of expertise should be “more sensitive to the interplay of contextual, institutional, and communicative cues that what she [the expert] says has relevance to us” (p. 273, 277). Again rejecting the idea that experts merely dispense facts, our understanding and differentiation of experts should also consider how they relate to and interact with laypeople. Reiner Grundmann (2017), also noting how the literature on expertise has neglected to consider the way experts are socially embedded, offers a more precise categorisation of expertise that goes beyond the intellectual performance distinction. Developing the points of Miller and Hardoš, Grundmann highlights the advisory nature expertise, via what he calls “relational” experts:

However, expertise is essentially something delivered at the request of someone else who wants it. This makes expertise relational in a double sense: it relates to clients and it relates to their needs, which often is the need for guidance in decision-making...This relational concept can be summarized as follows: experts mediate

² “Truth” and “fact” are used interchangeably in Miller’s original paper.

between the production of knowledge and its application; they define and interpret situations; and they set priorities for action. (pp. 25-26; emphasis added).

Again, beyond merely relaying facts, experts are also typically called upon for prescriptive counsel, which sometimes requires them to act more as interpreters tasked with making “abstract knowledge actionable, or advising clients on how to act in the absence of certain knowledge” (p.31; emphasis added). From this Grundmann establishes a distinction between interactional and contributory expertise: Interactional expertise is that described above, whereas contributory expertise refers to experts who produce the knowledge in a domain, via research and experimentation (p.33). For example, in the medical field, scientists are the contributory experts who produce knowledge, but doctors are the interactional experts who must make use of this knowledge to treat and advise patients; in economics, contributory experts are those who are able to perform retrospective analyses or probabilistic calculations, while interactional experts are more like financial advisors, telling their clients where to invest.

The interactional vs. contributory distinction more accurately reflects the diversity of novice-expert transactions, and adds to the traditional account a better contextualization of experts’ social functions. Modifying the traditional understanding of expertise to include Grundmann’s distinction, but preserving the definition of novice as one unable to make first-order judgements, we can return to the question of whether the novice would be rational in thinking for herself by beginning to ask whether she would be doing so in an interactional or contributory domain. If it is the former, where there experts engage in greater degrees of mediation between facts and the novice, the novice has somewhat more leeway in thinking for herself and in questioning whether the interpretation (not the facts) the expert offers is appropriate. In what follows, I will enumerate cases where the novice would indeed be justified in thinking for herself in cases situated within interactional domains, either due to the unreliability of expert judgment, or because expert judgment fares no better than that of laypeople. The distinguishing feature of these cases is that they are supported by empirical evidence of expert unreliability. It is important to qualify that I still maintain that deference to experts is still the rational course of action most of the time. That it is easier to enumerate the select instances when it is not, acts as testament. Highlighting these cases is thus not meant to lend support to movements from the greater public calling for sweeping skepticism of expertise, nor to populist rejections of science. On the contrary, identifying cases where the

novice is rational to think for herself can help to curb expert skepticism to the limited and precise situations where it is justified by evidence.

1.4. The Case for Algorithms

Michael Bishop (2001) addresses expert deliberation regarding “prediction problems people face” everyday, offering examples of a university admissions board deciding on the best candidate, or of medical experts trying to arrive at the correct diagnosis. These cases thus fall outside the realm of contributory expertise. They do not concern the production of hard facts, but the judgment of what action is best given the interpretation of evidence. Bishop demonstrates that expert predictions are consistently bested by actuarial, statistical predictions produced using a “mechanical procedure” which applies an equation, formula, or algorithm (p.183). To be thorough, he even demonstrates that this was the case whether one compared expert performance to algorithms of ideal models, improper models, and even randomized models. He thus argues that in these cases, we would be more epistemically responsible in placing our trust in algorithms rather than human experts. A possible objection to this case, is that it does not exactly call for the novice to think for herself, but rather to shift her trust to a different source of epistemic authority. Nevertheless, it presents a situation where turning to expert judgment is not the most reliable choice, and is useful to our discussion as it offers a more important lesson: That we should rethink the epistemic practices that we conventionally take to be the most responsible, and measure responsibility based on its reliability and the accuracy of the judgment it produces.

In the case of algorithms, the intuitively epistemically responsible practice was that of laboriously reviewing every piece of evidence. In short, the assumption was that harder work meant greater accuracy. The evidence of the algorithms’ greater accuracy, however, refuted this and pointed to the more counterintuitive conclusion that here, less work (plugging numbers into an equation) produced more accurate results. The use of algorithms was thus the more epistemically reliable practice. The lesson that we should move past our discomfort to be more open to the counterintuitive idea that quick(er)-and-easy(ier) practices can sometimes be the most optimal, will be especially relevant towards the development of expert appraisal strategies in Section III. For now, we may turn to two more cases of empirically demonstrated expert unreliability. For although the case for algorithms’ accuracy is solid, the reality is that they must be developed for specific scenarios, such that the novice does not

have access to them nor can develop them on their own. The following are thus cases which present more promising, practical alternatives for the novice.

1.5. Expert Political Forecasts

Psychologist Philip Tetlock (2005) is concerned with the reliability of expert predictions or “forecasts” in the field of politics. Once again, this concerns the interactional domain as these expert statements are not in the business of producing facts, but interpreting current affairs to offer judgment and advice about the future. Compared to contributory, scientific domains, the field of politics is notorious for being “messy” with respect to its methodology, or rather, lack of. This was precisely Tetlock’s problem in trying to get his study off the ground. In attempting to develop a quantifiable way to measure the accuracy of political experts’ predictions, Tetlock highlights the difficulty of trying to establish “a norm for judging good judgment” in a domain that is notorious for applying “self-perpetuating worldviews with their own self-serving criteria for judging judgment and keeping score” (p.4). The problem is that political experts are infamous for offering vague statements that are virtually impossible to prove were correct– or for reputation’s sake, to prove were incorrect. Tetlock eventually succeeds by asking experts to formulate their predictions in such a way that is precise and falsifiable, sharpening statements like “The Gross Domestic Product will improve very soon,” into “The GDP will increase within a 5% margin, sometime within the next week.” What Tetlock then found was that experts’ predictions were no better than those of laypeople. He concludes, “collapsing across all judgments, experts on their home turf made neither better calibrated nor more discriminating forecasts than did dilettante trespassers” (note that the “trespassers” here refer to laypeople) (p. 54; emphasis added).

Tetlock, however, does not leave laypeople unequipped and offers insight into how they can improve the accuracy of their own forecasts. He explains that another fruit bore of his study was finding a correlation between cognitive styles and accuracy: A “fox”-like cognitive style, where one’s thinking is more flexible and capable of synthesizing pieces of information from many sources, regularly outperformed the predictive abilities of “hedgehogs” who think more rigidly and try to square every piece of information as part of the “one big thing”– be it a theory or ideology– that they know (p. 72). Accordingly, he recommends that laypeople think like a fox in order to improve their own forecasting skills. In a later publication, Tetlock and Gardner (2014) specifically address the capacities of laypeople to become accurate forecasters themselves. Though they qualify that

“Superforecasting does require minimum levels of intelligence, numeracy, and knowledge of the world,” they note that such competences are reasonably within the reach of most laypeople. More generally, he advises that everyone improve the accuracy of their judgment by being “open-minded, careful, curious, and above all– self-critical” (p.19). We thus have another case where a novice’s deference to experts would not privilege her to more accurate judgments than if she had used her own. Returning to the question of whether it is rational for the novice to think for herself, we can at the very least answer in the lukewarm affirmative. That is, the novice would be rational in thinking for herself because it is not as though” deferring to experts would help her fare any better. More strongly, however, we can also respond that the novice would be more justified in thinking for herself were she to follow Tetlock’s advice to use empirically supported ways to improve her own forecasts, as this would open the possibility of the novice outperforming the judgements of the experts.

1.6. Experts and Statistical Illiteracy

Gerd Gigerenzer is another notable psychologist dedicated to improving the public's critical thinking, and is especially motivated by his findings about experts’ poor performance in regards to risk and statistical literacy. In his book *Risk Savvy* (2014), he blankly states in the introduction that he will be arguing for the following claim:

2. Experts are part of the problem rather than the solution. Many experts themselves struggle with understanding risks, lack skills in communicating them, and pursue interests not aligned with yours. Giant banks go bust for exactly these reasons. Little is gained when risk-illiterate authorities are placed in charge of guiding the public.
(p.14; emphasis in original)

Focusing his concern on the interactional, advice-giving domains of medicine and finance, he laments that all too often, experts 1) confound situations of risk, wherein all variables and respective probabilities are known, for situations of uncertainty, wherein the above are not all known, and 2) regularly pass on this misinformation to laypeople. To emphasize the stakes, one example that Gigerenzer often uses is the occurrence of false positives in HIV tests. Because much of the public is still ignorant that false positives can occur, and because medical experts fail not only to communicate, but also calculate and understand, false positives themselves, people are mistakenly led to believe that a positive result for HIV on the first clinical test is 100% certain (Gigerenzer, 2020, pp.200-204). In light of this,

Gigerenzer advises that one think critically: Pausing to ask herself whether she is in an at-risk group in the first place, gathering information about what and how the particular test was conducted, calculating the rate of false positives herself, and going in for a second test. He summarizes: “Do not assume that health professionals understand health statistics. Be skeptical and have the courage to think about the numbers yourself.” (2020, p. 205; emphasis added). Even more imperatively, he writes “ You have no choice but to think for yourself” (2014, p.2). It is worth emphasizing the importance of contextualizing these statements. Behind them are decades of research that have amply and consistently established expert unreliability with respect to particular competences. The rationality of a novice’s thinking for oneself in light of these findings— and those of Bishop and Tetlock’s described above— are thus well substantiated by consistent empirical evidence. But these fall in the minority of cases; most other times, we do not have evidence of systemic expert unreliability to such a thoroughly measured and recorded degree. Having identified three cases where it is relatively less-ambiguous that the novice would be rational in thinking for herself, we now turn to more complex cases.

2. The Problem of Metacognition and Epistemic Trespassing

Lacking evidence to suggest otherwise, most of the time we find that the rational course of action is to turn to the experts. But even when choosing the path of deference, the novice is not completely relieved of having to think for herself, given that she must still engage in some form of research and inquiry. As mentioned above, the traditional picture of deference that novices selects a candidate, poses her question, and decides whether to accept an answer, has been replaced by taking recourse to digital resources and performing a websearch of one’s question. In doing so, we often find a staggering amount of resources that purport to offer expert statements. The kind of “thinking for ourselves” we are tasked with now is that of assessing these statements and choosing which ones to accept. This question of which statements we should choose to accept ultimately leads us to the further question of which expert is credible enough for us to trust— especially when the experts disagree on an answer. To this end, we must resort to gathering indirect evidence of the expert’s credibility, and resort to second-order assessments of expertise in order to decide what and whom to trust.

2.1. Strategies and Criteria for Second-Order Appraisal

Epistemologists have recognized the difficulty of deciding whether an expert is credible or trustworthy, and so have proposed strategies that the novice can deploy to ground

her appraisal. To date, Alvin Goldman (2001) and Elizabeth Anderson's (2011) proposals have been among the most paradigmatic and frequently cited in the literature. In his seminal paper, Goldman tackles the question of how a novice can justifiably choose between disagreeing experts and proposes five strategies: 1) indirectly evaluate the experts' competence via their dialectical performance, 2 & 3) consult meta-experts, checking for expert consensus where experts in the domain have independently come to agree with either expert 4) check whether the expert is subject to any biases or conflicts of interests, 5) verify the expert's past "track-record" of giving accurate responses, or verify the correctness the an expert's claim via direct perception herself (2001, pp. 92-108).³ It is important to note that Goldman's motivation for addressing this problem was to push back against the testimonial skepticism by John Hardwig that novices' deference to experts is essentially "blind" or "groundless," (Hardwig, 1985; Goldman, 2001, p.86). As such, it was enough for his task to show that novices are capable of appraising expertise at least in principle. The rational justification he hopes to provide via his strategies should be understood in a minimal sense; that is, that their purpose is merely to help the novice arrive at a judgment based on reasons that are more justified than those they had before. Goldman himself modestly qualifies that his paper is exploratory.

Anderson also addresses the same problem of choosing between disagreeing experts as part of a broader campaign to defend the democratic participation of ordinary citizens in the policy-making process. Rather than blindly defer to scientific testimony that they cannot directly assess, she argues that ordinary citizens do have "the second-order capacity to judge [the] trustworthiness" of the scientific claims that underwrite policy, such that they have a hand in its making. Sharing some overlap with Goldman's proposals, she suggests that one approach the second-order appraisal of expert testimony by 1) assessing their credentials according to her proposed hierarchy of expertise, 2) assessing their honesty and record of epistemic responsibility, and 3) assessing whether there is a consensus of credible experts, ie. those having met criteria 2 (pp. 147-149). What is notable about Anderson's proposal is that she also claims for the above strategies to be within a layperson's reach, and can be reasonably expected to be performed by someone with a highschool education and access to the internet (pp. 150-151).

³ Strategies 2 and 3 are listed together, as Goldman also combines them in his original paper on the grounds that they both involve looking to other experts in the field.

Epistemologists writing on the expert-deference problem have largely received Goldman and Anderson's papers as a step in the right direction, with many having been inspired to propose their own amendments to sharpen the above proposals. Naturally, the papers have also prompted criticisms about their feasibility. These critiques can be categorized into two kinds. The first concerns the practical availability of the evidence required to deploy these strategies. For example, it may be that it is difficult to observe or find a source confirming any expert consensus, or that experts' track-records are non-existent. This is especially the case when the novice's question bears on a still-volatile situation like the Covid-19 pandemic, "subject to a high degree of disagreement amongst experts, and marked by rapid changes" (Reid, 2024, p.1). Similarly, Goldman and Anderson's call to identify conflicts of interest or bias may be unemployable, given that not all conflicts of interest are observable; while this strategy may work for financial conflicts of interest for which there is a money-trail, conflicts of interest resulting from one's personal beliefs or preferences may not be outwardly expressed and hence detectable (Almassi, 2016). Finally, the changing dynamics of a field may render a strategy obsolete. For example, to require that consensus only be legitimate if the experts of that domain independently arrived at the same conclusion (so as to avoid the agreement of mere 'yes-men') would not be a practical requirement to impose on scientific fields where coauthorships and collaboration on large, costly, irrepeatable experiments are becoming the norm (Levy, 2024, Section 2). Anderson's proposals and especially her claim that the novice is capable of conducting the necessary research has also been characterized as "quite optimistic," and as applying to matters like climate change for which there does exist a clear consensus and trove of resources (Reid, 2024, p.7).

The above criticisms are founded on external factors, and still leave the proposed strategies relatively sound in the sense that they would still be employable should the right evidence be available. The second kind of criticism, however, cuts more deeply and involves the problem of metacognition. Metacognition addresses the ability of the agent to accurately judge whether she is a reliable judge, at all. To illustrate the way this is more fundamental to second-order judgment, consider Ballantyne's helpful analogy of a small aircraft pilot: Due to weather conditions obscuring visibility and his plane's navigational system malfunctioning, the pilot cannot make judgements on what is the correct path to take. In other words, he is unable to make direct, first-order judgments about the problem. He thus defers to asking two air traffic controller stations, but they give him conflicting information on what path to take.

Trying to decide between the two, he considers how factors like the closer proximity of one station, or the more advanced technology of the farther but larger station, could bear on one controller's judgement being more accurate than that of the other. Here, he is considering second-order evidence— the kind that Goldman and Anderson's strategies are concerned with. However, the pilot recalls the phenomenon of hypoxia. That when flying above a certain altitude for too long, the pilot begins to suffer from oxygen deprivation that impairs his cognitive ability and increases the likelihood of error. Unless he has some way of gauging his oxygen levels to rule out his being in a hypoxic state, the pilot now doubts whether even his second-order assessments were not subject to memory impairment or reasoning errors (2022, pp. 232-253). This tertiary level of judgment is the concern of metacognition.

2.2. The Problem of Metacognitive Biases

The argument from poor metacognition is that, like pilots without oxygen-level detectors, we are not competent judges of whether we make good judgements. Compared to the criticisms about the practical availability of evidence, the problem of poor metacognition poses an even more fundamental threat to the use of second-order appraisal strategies. For if we suffer from metacognitive impairments that skew our judgment and of which we are not even aware, then even when we do possess the necessary evidence, we would not be trusted to make sound judgements about them. Unfortunately, findings from cognitive psychology seem to corroborate the antecedent, with studies demonstrating that we are indeed subject to systematic metacognitive biases. Though without using the language of metacognitive bias, Anderson already broaches the problem in discussing how "cultural cognition" can pose a problem to her proposal. Studies of cultural cognition have demonstrated that the way one's beliefs and means of processing new information, especially on contentious political issues, are filtered through the lens of one's cultural values and membership in ideological groups (Kahan and Braman, 2006; Anderson, 2011, p.158; Brennan, 2020, p.11) The consequence is that when people engage in research, they overrate the quality or importance of evidence that merely corroborates their pre-established values and memberships. Conversely, they perceive counterevidence as threats, defensively processing them with greater skepticism, assigning them less importance, and being less likely to be able to recall them. This explanation also extends to the phenomenon of recalcitrant beliefs, and the way that the willingness to change one's mind about a belief is inversely proportional to that belief being "logically or psychologically connected to other beliefs and values that are important to them" (Jost, 2015, p. 618). This threat of cognitive dissonance is analogous to a game of

Jenga; the entire tower is in greater danger of toppling the lower and more foundational the piece that is removed.

But even when our personal identities are not at risk, a similar phenomenon occurs for more contained and lower-stake situations, just out of our preference for explanatory coherence. This has been consistently demonstrated in studies about the correction of misinformation. These studies employ tests where subjects are presented with an initial story, then a retraction of a piece of that story, and then given a comprehension test where they must recall basic facts about the event. It was found that even in “neutral scenarios, in which people have no reason or motivation to believe one version of events over another...retractions rarely, if ever, have the intended effect of eliminating reliance on misinformation, *even when people believe, understand, and later remember the retraction*” (Lewandowsky et al., 2012, pp. 113-114; emphasis added). Even when controlled for motivated reasoning, psychologists found that the mere fact that a story had already been formed makes it resistant to change. These findings offer a few immediate takeaways: One is that a greater onus falls on the experts and the media to make careful and accurate statements the first time around. On the part of the novice, we find that more evidence does not necessarily mean more accuracy and nuance given the potentially distorted way we process the evidence; moreover, this can happen even without our awareness, in neutral situations, and even when we do not mean for them to. The latter thought that our judgment may be flawed despite our best intentions is particularly disturbing, and all the more because there is again psychological evidence in its support.

As if the final nail in the coffin, it appears that we are also subject to a “bias blindspot” which, as the name implies, 1) prevents us from accurately identifying when our own judgment has been biased 2) makes us more likely to recognize when others are biased, and to overestimate how much (Pronin, 2007, p.37). This blindspot is in turn attributed to the “illusion of introspection” and “naïve realism.” The former is when one overestimates their introspective capacity to consciously reflect on and understand their own cognitive process, while the latter is when people take their own perception of reality to be the most veridical account, such that any one who disagrees must be the in the wrong due to their own biases. Returning briefly to the strategies for appraising expert credibility, this does not bode well particularly for the strategy of bias detection. More recent studies forge an even closer link between this blindspot and the appraisal of competence. Consistent with the above findings,

people who lack knowledge tend to be ignorant of the fact that they do, and this misappraisal of self-knowledge in turn results in the misappraisal of others. As with naive realism, people tend to use themselves as the benchmark of competence such that they judge others who act differently to be mistaken. This reasoning occurs in both directions: When asked to evaluate those less competent than them, such reasoning leads one to the correct evaluations of others' lesser competence. But when asked to evaluate those more competent than them (ie. experts), they again—but this time mistakenly—misjudge these people to be incompetent (Dunning & Cone, 2022). Keeping with the recurring theme of overestimation, not being aware of one's own ignorance also makes her more likely to overrate and be over confident in her competence (Elga 2005, Dunning 2014). Ironically, the result is that those who would stand to benefit the most from epistemic correction and education, are also those who do not see themselves as needing it.

2.3. Levy's Account and the Challenge of Pessimism

Given the accumulation of this evidence, the rationality of the novice thinking for herself, even under the provision of epistemic guidelines, is naturally cast into doubt. The degree of doubt has varied among epistemologists, who often offer their prognoses on the spectrum of optimism or pessimism. Brennan (2020), for example, calls for “reserved optimism,” endorsing Anderson's strategies but proposing that we add a fourth metacognitive reliability condition that asks the novice to “look for disqualifying evidence that you have been fooling yourself about the above criteria” (p.232; emphasis in original). He suggests that a possible way of detecting metacognitive bias is by externalizing and keeping record of one's patterns of judgment and experiences of certainty, so as to identify any tendencies indicative of bias. In contrast to this attempt to salvage the second-order appraisal strategies, Levy (2024) reaches what he admits is a strongly pessimistic conclusion and asks us to do away with them. In his article “No Trespassing! Abandoning the Expert/Novice Problem,” Levy argues that all of the expert-appraisal criteria that philosophers have hitherto proposed are ultimately unemployable; that even when we try to use the criteria, they “won't reliably move us from doubt or unjustified belief to a better justified judgement” (Introduction). His account thus threatens to comprehensively deny the rationality of the novice who would be trapped in a position of unjustified belief.

Levy's argument can be broken down into three main claims: The first is that the criteria proposed do not work because “we need deep knowledge” of the domain and its

structure (Section 2). The second is that using the criteria only amounts to us extending the ways in which we already tend to trust, and so does not improve our justificatory station. This claim can also be understood as his concluding call that we first need better ways of trusting before we can correctly use the criteria. Third, in light of the first two claims, any attempt by the novice to deploy these strategies would amount to her committing epistemic trespassing because she would be making judgements on matters which fall outside of her competence. Levy's objection is especially troubling as it not only threatens any attempt at second order appraisal, but also gets at the heart of why the expert-appraisal problem was important to address in the first place. Recall that in Goldman's seminal paper, he is motivated to address the problem to argue against Hardwig's testimonial skepticism that novice's trust in experts is "blind," as this would have the disturbing implication that "a layperson (or scientist in a different field) cannot be *rationally* justified in trusting an expert" (2001, p.86, emphasis in original). Levy directly responds to Goldman's objective, making it clear he fully endorses Hardwig's position:

Goldman hopes that his criteria can allow us to avoid the kind of testimonial scepticism that threatens if we say, with Hardwig, that novice trust in experts is groundless. He's wrong: the criteria don't underwrite the contrastive judgment he hopes for. (Section 4).

Levy's criticism is thus the latest incarnation of the testimonial skepticism which threatens the novice's rationality in the strongest sense.

Of particular interest is Levy's second claim that we are merely extending the ways in which we already trust, as this is the claim most evocative of the metacognitive biases discussed above. Levy's paper is mainly epistemological and theoretical, and so does not rely much on psychological literature. The most explicit mention he makes of metacognition is when he notes that he will especially be responding to the articles of Anderson, Goldman, Guerrero, and Brennan, and explicitly rejects the metacognitive reliability criteria of the latter. With respect to what he means by "the ways in which we trust," for most of the paper he speaks of this only in the general sense; though towards the end of his paper, he offers the example that we extend the trust we already have in certain institutions, media, or peers, and explains this using the language of bias. It is thus not difficult to see how there is a natural link between Levy's second claim and the problem of metacognition. After all, metacognition is the study of how we already tend to judge; and as we have seen above, it is concerned with

the way that our judgment has been shaped by the institutions, media, and peers we already endorse. Levy's account is thus of interest because it can only be strengthened by the evidence on metacognitive bias, and because it provides us with a vehicle for articulating the threat of metacognitive bias to the rationality of the novice. For if Levy is right that using any of the proposed criteria for second-order appraisal amounts only to extending the ways we already judge—which, as the evidence on metacognition suggests is flawed—then the novice would comprehensively have no rational grounds for thinking for herself.

3. Defending the Rationality of the Novice: A Response to Levy

3.1. Novice Epistemic Trespassing?

Most immediately, we can dispute Levy's third claim that the novice would be engaging in epistemic trespassing, by responding that the term cannot apply to novices at all. Originally coined by Ballantyne (2019), epistemic trespassing occurs when “thinkers who have competence or expertise to make good judgments in one field, but move to another field where they lack competence—and pass judgment nevertheless,” and occurs “in three varieties—we are experts in one field, but we lack evidence from another field, or its skills, or both” (pp. 367, 376). A paradigmatic go-to example has been that of Linus Pauling, the two-time nobel prize winning chemist who trespassed onto the field of medicine in claiming that ingesting large doses of Vitamin C could cure cancer. Epistemic trespassing has thus been restricted to applying to experts. Levy's charge, however, understands it in a more general sense of making judgements which fall outside one's competence, and so raises the question of whether it can be extended beyond experts to apply to novices. DiPaolo's (2021) development of epistemic trespassing, however, demonstrates that it cannot.

In his paper, “What's wrong with epistemic trespassing?” DiPaolo answers the titular question by arguing that epistemic trespassing is an abuse of authority where experts neglect the vulnerabilities of novices. What he finds missing in Ballantyne's original account, is that it raises the possibility that other kinds of trespassing exist, but does not comment on their potential asymmetries. For example, Pauling claiming that Vitamin C can cure cancer is intuitively a graver transgression than a car mechanic making a judgment about international relations despite not being an expert of foreign policy. Accounting for this asymmetry thus amounts to identifying “what's wrong when it is wrong” (p.226). DiPaolo begins by turning to norms of assertion to more practically ground epistemic trespassing a speech act, and to show that epistemic trespassing is wrong when someone is speaking as an expert, or makes it

unclear that they are not. This is wrong due to the relationship of authority that experts have over novice, and the understanding that “expertise is inherently social and comparative” (echoing the account of expertise developed in Section I). Because experts have novices who depend on them and are in a position of greater vulnerability, their word puts “substantial rational pressure” on the novice to accept their claims, “imposing an undue burden of responsibility on the listening novice” (pp. 235, 236, 239). Moreover, experts typically have a larger media platform through which their statements can be made public. Thus, the harm and misinformation that could potentially ensue because of their expert trespassing occurs on a greater scale. As the psychological literature on the resistance of misinformation mentions above, they have a greater responsibility to publish accurate statements. Epistemic trespassing is thus rightly restricted to experts, given that what is wrong with it is the abuse of epistemic authority to harm vulnerable dependents. Because this is an authority that only experts possess, epistemic trespassing is not applicable to novices.

Even if we refrain from using the term epistemic trespassing, however, Levy’s two remaining claims—wherein the thrust of his pessimism lies—are what truly undermine the rationality of the novice thinking for herself. For even without using the term of epistemic trespassing, one can still object that the novice is irrational in making judgements beyond her competence. To argue against the first claim that using the strategies for second-order appraisal themselves require expertise, it will be necessary to show how we can still use these strategies with relatively superficial knowledge of an expert domain, which can be reasonably expected of the novice to possess or obtain. Levy’s second claim is that we are in need of a better foundation of judgment (or ways of trusting) before we can use second-order appraisal strategies, and is the one most connected to the findings about our metacognitive biases. Although the claim is pessimistic in the sense that it is saying the proposed strategies are not actually helpful to us, at least as it stands, the silver lining is that this objection also opens the door to asking what can we do to practically improve the way we judge and trust? That is, if we need a better foundation to reliably do so, what are the corrective efforts we can undertake? To undermine the pessimism of this claim, we can thus show that there are indeed ways that we can improve our judgment and to curb the metacognitive biases discussed above. This approach follows a familiar strategy in the literature, where it is enough for our purposes to show that there is at least a feasible “way out” of the problem, such that pessimism is not warranted. More fundamentally, we can also respond to the second claim by disputing the treatment of biases as a defeater of good judgment, at least to the extent that it

has been taken to be in the literature. In what follows, I will address this second claim first; logically, it will be necessary to show that we should not abandon second-order appraisal strategies at all, before arguing that they can be used with non-expert knowledge.

3.2. Metacognitive Debiasing Strategies

The psychological literature has thankfully not only been devoted to pointing out our deficiencies, but have used the insights gleaned from studying their causes to point us towards possible remedies by the use of debiasing strategies. One approach has been to address how we can mobilize people's personal identities and social memberships to render them more receptive to evidence. Rather than divorcing one's identity from the evaluative process, studies indicate that *affirming* one's identity and reassuring its preservation is a more effective means of getting people to be more objective in their reception of evidence (Lewandowsky et al., 2012, p.122). Responding to the problem of cultural cognition, for example, Anderson cites psychologists' findings to suggest that "policies drawn up in response to scientific findings can be invested with multiple meanings that affirm the values of diverse ideological groups" (2011;p. 158). That is, when we present evidence (arguments, positions, etc). in such a way that they can readily lend themselves to various identities, they are less likely to be perceived as a threat and trigger defensive metacognitive biases. Along these lines, Jost (2015) suggests a more targeted "framing" approach where the presentation of evidence is tailored to appeal to a specific audience. He uses the example of a study done in the United States, where pro-environmental initiatives were presented to climate change deniers, most of whom tend to be Republican. It was observed that when the initiatives were presented using patriotic rhetoric such as the United States "leading the world in green technology," or when carbon pricing was referred to as an "offset" rather than a "tax," Republicans were much more receptive to the proposals (pp.625-626). A potential problem with this approach, however, is that it can be seen as a form of manipulation given the way it targets a specific audience; Jost does not raise its ethical implications, but himself refers to the tactic as "labeling manipulation." In this case, Anderson's proposal of more neutrally presenting a position as being compatible with many ideologies may be more ethical, though at the risk of being less effective.

Brennan already suggested that we can become more conscious of our metacognitive biases by externalizing our judgements, keeping score of them in order to identify any recurring tendencies that indicate bias. The difficulty with this strategy, however, is that it is

not effective in cases where we do not have a track-record of our judgements, or when we are faced with an unprecedented situation for which we have not made analogous judgements in the past. For these situations, we can turn to another strategy recommended in the psychological literature, of externalizing our judgments via argumentation and deliberation with others. Returning to the studies addressing the correction of misinformation, later studies found that longer group discussions about the false information that was initially presented, led to greater acceptance of the correct ones (Lewandowsky et al., 2012, p.121). Mercier and Sperber (2011), in their well-known proposal of the Argumentative Theory of Reason, corroborate that group deliberation leads to better reasoning than does individual reflection, and identify the optimal conditions for this to occur. They specify that such effects occur under optimal settings where members hold diverse beliefs and are forced to critically evaluate their arguments— something which would not be the case if all group members already agree and are thus less motivated to deliberate (p.63). This finding is reiterated by Fisher and Keil (2014) in their work on the “illusion of argument justification” or the tendency for people to overrate the quality of their arguments. They conclude from this study that “the most effective interventions were those that asked participants to consider alternative positions,” and that social interaction worked to optimize one’s reasoning (p.13). In sum, even if we do tend to be subject to blindspots in our judgments, their externalization via their expression in group deliberation allows others to point them out and in short, give us a reality check.

3.3. A More Nuanced Understanding of Bias

Thus far, the line of defense against Levy’s second claim has been to offer ways that we can indeed correct for metacognitive biases, in order to show that we can reliably improve our judgment of who to trust, before extending it to the second-order appraisal strategies. A second line of argument is to dispute whether the biases should be treated as a defeater to such an extent, at all. The mainstream epistemological literature has traditionally treated bias in an exclusively negative light. Its detection has been akin to hitting a wall. Understood as a strong defeater of rational judgment, efforts to defend rationality have mainly focused on its mitigation, if not, elimination. While we can agree in light of the psychological evidence above that this is a worthwhile endeavour towards improving our judgment to some extent, we should also rethink the extent to which we assign biases such weight as serious defeaters. The strongly negative understanding of bias in epistemology can be traced back to the history

of the term, and its being traditionally associated with the heuristics and biases tradition that has pervaded mainstream psychology since the mid-1970s. The tradition originates from Amos Tversky and Daniel Kahneman's (1974) tests of human rationality, which—measured against the norms of Bayesian probability, formal logic, and rational choice theory—yielded abysmal results which presented human reasoning as systematically flawed. Consequently, the tradition has been associated with the more pessimistic view of human rationality, where bias is understood more sweepingly as the equivalent of poor judgment.

There exists, however, the dissenting tradition of Simple Heuristics founded by Gigerenzer, which offers a more nuanced understanding of biases and heuristics. This tradition has contested the measurement of human reasoning according to the content-blind norms above, as these are not the norms that humans were evolutionarily poised to adapt to. Instead, biases and heuristics developed in humans in order to equip them with effective cognitive shortcuts they can use to quickly make decisions in suboptimal environments and in the face of uncertainty. Indeed, Gigerenzer and his colleagues have amply demonstrated the reliability, and even advantage, of using biases and heuristics in the environments and contexts they were developed to respond to (Gigerenzer, 1991; 2008). That biases can demonstrably reliable judgments when we are sensitive to the contexts in which they are used, points us towards a more charitable view of biases. Beyond investigating its origins, Gigerenzer also addresses the contemporary treatment of bias in psychology and behavioral economics. He identifies the presence of a “bias bias,” or “tendency to spot biases even where there are none;” this is especially unwarranted, however, given that “a systematic review of the literature shows little evidence that the alleged biases are potentially costly in terms of less health, wealth, or happiness” (2018, p. 1). The upshot is that we should be more prudent against treating biases as a “boogeyman” that gravely threatens our rationality and daily functioning.

This in turn has relevant consequences for epistemology. First, is that if biases are simply inherent to our cognitive systems given the way that we have evolutionarily developed, making the defense of our rationality contingent on the maximal elimination of bias only sets us up for a Sisyphean task. That is, the prospects of defending our rationality are doomed from the beginning, due to us having set the bar too high. Returning to Brennan's suggestion that we add a fourth metacognitive reliability condition to our criteria for expert appraisal, this condition should be developed so as not to require too much of the novice.

Second and practically speaking, a too-dogmatic and unconstrained effort to check ourselves for bias would also lead us to an infinite regress of doubt, questioning, for example, whether some judgment was merely the product of a bias, which in turn was only the product of a metacognitive bias, which we perhaps mistakently thought we avoided due to another metacognitive bias, and so on. We would be reduced to the proverbial centipede who, forced to think about the movement of each leg, was paralyzed into inaction and fell into a ditch. Third, even when the detection of bias is possible, there are reasons why using its presence as a defeater of others' judgements would be epistemically irresponsible. For if we take bias to be a reason to downgrade others' arguments, but it is the case that *everyone* is biased in some form, bias detection would amount to a scorched-earth strategy where no is left standing as being credible or rational (Solomon, 2014, p.254). Ballantyne (2015) arrives at a similar conclusion that we should temper our confidence in "debunking" strategies founded on bias attribution, at least before we have a more principled and systematized way of treating bias (p.159). Again, while we can agree that it is a worthy cause to try and curb our systematic metacognitive biases where we can, we can also endorse that should not be taken to an extreme, or as an absolute necessity when it comes to defending human rationality. Returning to Levy's pessimism, a more holistic and mitigated understanding of bias undercuts his claim that our biased ways of trusting would completely prevent us from making use of second-order appraisal strategies.

3.4. Domain Metaknowledge and Heuristics

Having argued that our current ways of trusting do not undermine the use of second-order strategies to the extent that he argues, we can go on to address Levy's first claim that using such strategies requires expertise. To demonstrate how the novice can still manage to use the strategies on relatively superficial knowledge, we can take inspiration from Bishop, Tetlock, and Gigerenzer's approaches in Section 1. For instance, the main lesson from Bishop was that sometimes, reliable epistemic practices can be those which are counterintuitively "lazier" or less demanding of information. One such practice could be the use of metaknowledge about the norms and practices of an expert domain. To illustrate how much of a difference such ostensibly superficial knowledge can make, take the real example of Ben Santer, a well-established scientist known for his work on anthropogenic climate change. Climate change denialists portrayed Santer as having committed "data fraud by revising his conclusions in light of new information and constructive criticism of peers" (Brennan, p.10).

The success of this propagandist tactic depended on the public's ignorance that the process of peer review is actually just a norm within the discipline.

In asking the novice to gather metaknowledge about the domain, the novice would not be required to possess first-order knowledge about that domain, but more superficial information regarding its methodology, extent of collaboration, whether and how the field is more interactional or contributory, and so on. Being aware of the degree of collaboration in within a field can help the novice better deploy Goldman's strategy of consensus; they would better understand the requirement for consensus to have been arrived at independently, would not be applicable to inherently dialogic fields in the humanities, or in certain sciences where collaboration is becoming more the norm. Distinguishing whether the domain is more contributory or interactional can also signal to the novice the degree of deference that is expected of her. For interactional domains where there is a greater degree of mediation between the novice and the facts of the domain, the novice would understand that she has more justificatory wiggle-room to factor in her goals and experiences in deciding whether to accept the expert's counsel. However, in contributory domains which deal more directly with the facts and first-order evidence, she would be much more strongly inclined to accept such experts' answers. For example, the novice would understand that she would comparatively be more justified in being skeptical of some experts' claims in politics than in a field of science, understanding the former to be "messier" and less concerned with the quantifiable, than the latter which follows a more rigorous and streamlined methodology of repetitive experimentation.

Moreover, metaknowledge of a domain can be supplemented by the empirical evidence on the reliability of certain cognitive styles. Here we return to Tetlock's findings on correlates of good judgment that fox-like thinkers, who are able to piece together bits and pieces of information from many sources, make better forecasters than hedgehogs, who try to square the evidence into one theory or ideology. Keil (2010) has since taken these findings to consider how they may apply to other fields of expertise beyond politics. He points out that, while a fox-like thinker would thrive in a changeable field like politics, a hedgehog-like thinker would thrive in a field like biology where evidence is squared to one paradigmatic overarching theory— in this case, evolution. Equipped with this information, the novice may be better poised to indirectly appraise experts' argumentative styles, say, when listening to them defend their positions. Goldman actually considers the strategy of evaluating an

experts' argumentative performance, again because novices should not be making first-order judgements by evaluating the *content* of the experts' claims. However, he dismisses this as dangerously superficial since the evaluation of argumentative style may lead us to favor a charismatic but incorrect orator, over an inarticulate but correct one (2001, p.96). But when backed by the above empirical evidence, the novice would be more justified in relying on an evaluation of argumentative performance. They would have better justificatory grounds for example, in choosing a fox-like orator in a political debate, or hedgehog-like scientist in a discussion over evolutionary theory.

Other authors have worked towards redeeming the strategy of appraising argumentative performance in a way more in line with Gigerenzer's work via the proposal of heuristics. Integral to Gigerenzer's crusade to equip people with sharper cognitive tools for critical thinking, is his work on identifying "fast and frugal heuristics," or cognitive shortcuts that we can quickly deploy and which have been empirically demonstrated to be reliable in certain environments (2008, p.208). That these are especially handy "for problems characterized by inevitable uncertainty, very little information, or very little available solution time," make them especially apt for application as second-order appraisal strategies (Sturm, 2012, p 77). One that has been proposed by Matheson (2005) is "The Dialectical Performance Heuristic." As the name suggests, this heuristics develops how we evaluate the experts' argumentative performance and comes in three versions: First, the "Rhetorical" version links dialectical superiority (ie. the possession and quickness of articulate rebuttals) to the possession of greater evidence and hence credibility. But as this falls subject to Goldman's initial worry mentioned above, Matheson adds a second "Moral" version. Here, the more credible expert is likely to be the one who responds more respectfully to their opponents' counterarguments, treating them with charitability and patience, affording them the fair amount of time to be expressed, and listening to them with attentiveness. The third "Explanatory" version is that the more credible expert is the one who displays superior "managerial" ability in synthesizing more pieces of evidence into a coherent explanation (pp. 150-151, 155). Similar to Matheson's moral criteria, Skipper (2020) proposes that we use a "humility heuristic." Formally, Skipper formulates the heuristic as a function of Bayesian probability, explaining that we have greater reason to trust the answer of someone to some question *q*, when in response to some previous question *p*, that person openly admitted their ignorance of the answer. In other words (and made more relevant to the expert appraisal problem), when an expert expresses epistemic humility by responding "I don't know" to

some questions, we have greater reason to take her as someone who is careful not to give false replies; thus, in the instances where she is able to offer us an answer, we have more reason to accept it as accurate. Matheson and Skipper's heuristics offer further promising second-order appraisal strategies— especially in light of the latest calls in epistemology for greater epistemic humility (Ballantyne, 2022; Brennan, 2020). But as attractive as they are in principle, they have yet to be empirically supported. It would be worthwhile to test the reliability of these heuristics, again following Gigerenzer's insight that we should investigate more precisely the environments or domains in which heuristics are most appropriate.

Although I have argued in defense of the novice's rationality and for the salvageability of second-order appraisal strategies, Levy's objections nevertheless offer helpful insights. First, he is correct that we do need to pay more attention to the foundations of the ways in which we trust, and particularly the influence of the institutions, media, and peers we already subscribe to. His articulation of his position also offers us a means of formulating the threat of metacognitive bias against the novice's rationality in thinking for herself, and to reevaluate the extent to which we should consider such biases as a defeater if we wish to avoid being doomed to a pessimistic conclusion. Finally, it offers a lesson on the kind of cognitive labor that is expected of the novice. That if she is to use second-order appraisal strategies, even without being required to have expert-level knowledge, some level of research will still be required of her. To consider how the proposed strategies above are not overly burdensome, recall that even Goldman's original strategies demanded research efforts. Compared to the difficulty of inquiring into which experts others agree with, or finding and verifying an expert's track record, educating oneself about the norms and practices of a domain or of argumentative styles would be equally, if not less, burdensome. Whether choosing to think for herself in the strong sense where she knows expert counsel is has been demonstrably unreliable, or choosing to defer to the experts and facing the problem of whom to trust, the novice must accept some onus of critically thinking for herself.

3.5. Boost: An Ameliorative Program in Rationality

As helpful as the above proposals may be, they would not be of any use to the novice if she is not made aware of them. Indeed, a common thread that repeatedly appears across both the epistemological and psychological literature is the call for educational initiatives that can practically equip the novice with these strategies. Actual, existing programs, however, are rarely mentioned by name. One cited by Ballantyne, is the Stanford History Education

Group (2022, p.15). Renamed the Digital Inquiry Group since 2023, their mission is to provide primary and secondary school teachers with material for a history curriculum which incorporates education in digital literacy and “Civic Online Reasoning” (DIG, 2024). Another educational program through which the above findings can be disseminated is that of the Boost program. Founded by Ralph Hertwig and Till Grüne-Yanoff (2016), the Boost program is an ameliorative project in behavioral public policy, whose goal is to use findings from psychology and behavioral economics to design policy interventions that educate citizens in critical thinking. Recognizing the problem that “people seem prone to making decisions that are detrimental to their own welfare or to public welfare,” Boost interventions are meant to improve people’s judgments such that they make decisions which are more rationally conducive to their goals and preferences (Hertwig & Grüne-Yannoff, 2015, p.156). That citizens improve their decision-making skills all while preserving their autonomy is a central tenet of the Boost program. It is thus especially apt for our purpose of addressing how people can responsibly think for themselves.

The interventions that Boosts are interested in are those which offer a “high return on a small amount of smart information in domains in which everybody makes choices” (Hertwig & Grüne-Yannoff, 2015, p. 159). For example, one of the main interventions the Boost program proposes is that of teaching risk and statistical literacy—a skill which, as described in Section 1, experts themselves lack, and which can be transferable across a multitude of domains. To determine what high-return interventions should be prioritized, Hertwig and Grüne-Yannoff call on policymakers to identify domains in which people “operate under highly consequential knowledge gaps and misconceptions” (2015, p.167). Given the salience and far reaching consequences of deferring to the correct experts, there is thus an opening for the strategies proposed above to be incorporated into the Boost program. Teaching media literacy, domain norms and methodology, and research skills all fall within the scope of Boosts’ concerns. There is even room for the inclusion of education about metacognition, in light of more recently proposed amendments to the program. For example, behavioral economist Malte Dold (2023), praises Boosts’ agency-centric approach but recommends that in addition to trying to improve people’s decision-making skills via “internal cognitive processes,” the program should also educate people on “the external social conditions (the norms, rules, institutions, and resources) that facilitate and constrain people’s efforts to reflect, act upon, and identify with their evolving preferences” (p.17). That is, beyond improving instrumental rationality by teaching the most effective means for pursuing

one's goals, we should also turn our sights towards substantive rationality and teach people to reflect on how the formation of their goals has been unconsciously shaped by the environment they are in. Since the latter is also a concern which drives the study of metacognition, there is an opportunity to address metacognition within the Boost program via Dold's proposal. Finally, Boost is well-poised to address the wider public because it is not limited to being taught in schools. Rather, "boosts aim to foster or develop new competences under conditions of limited time and resources...and typically in an adult citizenry that cannot be subjected to years of additional schooling" (Hertwig & Grüne-Yannoff, 2017, p.982). This point is especially attractive to epistemologists concerned with the political and democratic implications of the expert appraisal problem. Of course, we can anticipate the worry that educating the wider public is ambitious, or even optimistic. But what is important is that these proposals at least offers a possible start for practical, deployable interventions. For as has been argued, it is inevitable that some responsibility of thinking for ourselves will be necessary; we may as well undertake the steps—even if difficult—to do it well.

4. Conclusion

To summarize the development of this paper, I began in Section 1 by offering an understanding of expertise that was more reflective of its function in reality, and which included the distinction between interactional and contributory experts. This was followed by cases where in interactional domains, the novice was rational to think for herself where there was empirical evidence of experts' unreliability. The further purpose of this section was to gather insights to be later used towards the development of more expert appraisal strategies. In Section 2, I argued that even for the majority of cases where we choose to defer to experts, we are ultimately forced to engage in research that requires us to make judgements over which sources to trust, which in turn raises questions about which experts are most credible. This was followed by a review of the second-order appraisal strategies typically proposed in epistemological literature, and an explanation of how the most serious and recent threat to the use of such strategies was the problem of poor metacognitive awareness. I concluded this section by introducing the pessimistic argument from Levy that these problems render the strategies unemployable and that the novice cannot be rationally justified in thinking for herself via their use. In Section 3, I responded to each of Levy's three claims, arguing that the novice cannot commit epistemic trespassing, that there do exist debiasing strategies to

combat the problem of poor metacognition, and more fundamentally disputed the responsibility of using biases as strong defeaters in the first place. Upon arguing that we should not abandon second-order appraisal strategies, I then went on to propose further strategies as inspired by the insights gleaned in Section I, and proposed their incorporation into the educative program of Boost. Overall, my goal was to have defended the rationality of the novice thinking for herself by highlighting more clear-cut cases of expert unreliability, and by arguing against the pessimism that her deference to experts is essentially groundless.

Given its complexity and invocation of different disciplines, the tension between thinking for oneself and expert deference remains fecund with further avenues for research, two of which most immediately come to mind. First, is to address the role and responsibility of experts in responsibly communicating information to the public. Though the focus on arming the novice with tools for expert-appraisal is understandable given their vulnerability as dependents, we must be careful that this focus is not asymmetrical lest it imply a practice of *caveat emptor* where the novice carries a greater burden of evaluating the information she is given. Of particular interest would be how we can foster greater public trust in expertise by developing amongst experts, practices of transparency and epistemic humility. For example, we could work towards normalizing experts admitting to the public when they are ignorant of a subject, such as in volatile and unprecedented situations like the recent Covid-19 health crisis. Understanding admissions of ignorance as a sign of credibility rather than weakness would also help us to move away from an error culture that, in discouraging the open-discussion of mistakes, only prevents their correction. Second, we can turn to the burgeoning literature on epistemic distributive justice, which treats the participation in the production and reception of knowledge as “a limited good that needs to be justly distributed among putative members of an epistemic community” (Miller & Pinto, 2022, pp. 1-2). This is especially relevant to the discourse about how novices, despite the need to defer to experts, can still be active participants in the democratic legislative process. Moreover, as epistemic distributive justice is also concerned with the way knowledge is accessible and channeled to different communities and identity groups, its literature could also help to inform the way we enact educational initiatives.

To conclude, I would like to take a step back to talk more broadly about the reflections that interrogating this topic has produced. First is that the problem of rationally thinking for oneself and of expert deference is too complex to be addressed by epistemology

alone. It is important to turn to the insights of other disciplines—even beyond psychology, sociology, and behavioral economics— to develop a more comprehensive and realistic approach to the problem. As we have seen, we often need the help of others to check for our blindspots. Second is that, despite it being obvious, it is worthwhile to keep reminding ourselves that this is an inherently a difficult problem. And that it is difficult, precisely because it is real and it is here. Goldman was right in highlighting that unlike traditional, theoretical epistemology problems, for which we have the luxury of positing thought experiments and controlling for variables— the special difficulty of the expert deference problem is that it is founded on an unideal reality involving unideal citizens. Unable to start from scratch, we are more like passengers of Neurath’s boat, continuously repairing the vessel as we sail. As such, we should perhaps move away from the rhetoric of optimism and pessimism, and from offering too early prognoses of defeat which discourage attempts at finding realistic solutions. As thorny as the problem is, it will be by digging our hands into the thistle that we can hope to find some surprising fruit.

References

- Almassi, B. (2016). Toxic funding? conflicts of interest and their epistemological significance. *Journal of Applied Philosophy*, 34(2), 206–220.
<https://doi.org/10.1111/japp.12180>
- Anderson, E. (2011). Democracy, public policy, and lay assessments of scientific testimony. *Episteme*, 8(2), 144–164. <https://doi.org/10.3366/epi.2011.0013>
- Augustine, & Rotelle, J. E. (2012). *The works of Saint Augustine: A translation for the 21st Century*. New City Press.
- Ballantyne, N. (2015). Debunking biased thinkers (including ourselves). *Journal of the American Philosophical Association*, 1(1), 141–162.
<https://doi.org/10.1017/apa.2014.17>
- Ballantyne, N. (2018). Epistemic trespassing. *Mind*, 128(510), 367–395.
<https://doi.org/10.1093/mind/fzx042>
- Ballantyne, N. (2022). Novices and expert disagreement. *Reason, Bias, and Inquiry*, 227–253. <https://doi.org/10.1093/oso/9780197636916.003.0011>
- Ballantyne, N., Celniker, J. B., & Dunning, D. (2022). “Do your own research.” *Social Epistemology*, 38(3), 302–317. <https://doi.org/10.1080/02691728.2022.2146469>
- Bishop, M. A. (2000). In Praise of Epistemic Irresponsibility: How Lazy And Ignorant Can You Be? *Synthese*, 179–208.
<https://doi.org/10.1023/A:1005228312224>
- Brennan, J. (2020). Can Novices Trust themselves to choose trustworthy experts? reasons

for (reserved) optimism. *Social Epistemology*, 34(3), 227–240.

<https://doi.org/10.1080/02691728.2019.1703056>

Descartes, R. (n.d.). Discourse on the method of rightly conducting the reason, and seeking truth in the Sciences. The Project Gutenberg eBook of A Discourse on Method, by René Descartes. <https://www.gutenberg.org/files/59/59-h/59-h.htm>

Digital Inquiry Group. (n.d.). <https://inquirygroup.org/>

DiPaolo, J. (2021). What’s wrong with epistemic trespassing? *Philosophical Studies*, 179(1), 223–243. <https://doi.org/10.1007/s11098-021-01657-6>

Dold, M. (2023). Individual agency in behavioral public policy. <https://csgs.kcl.ac.uk/wp-content/uploads/2023/06/Malte-Dold.pdf>

Dunning, D. (2014, October 27). We are all confident idiots. *Pacific Standard*.
<https://psmag.com/social-justice/confident-idiots-92793/>

Dunning, D., & Cone, J. (2022). “The Cassandra Quandary: How Flawed Expertise Prevents People from Recognizing Superior Performance among Their Peers.” Department of Psychology, University of Michigan.

Elga, A. (2005). On overrating oneself. . . and knowing it. *Philosophical Studies*, 123(1–2), 115–124. <https://doi.org/10.1007/s11098-004-5222-1>

Fricker, E. (2006). Testimony and epistemic autonomy. *The Epistemology of Testimony*, 225–250. <https://doi.org/10.1093/acprof:oso/9780199276011.003.0011>

Fuller, S. (1986), *Social Epistemology*. Bloomington and Indianapolis: Indiana University Press.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond “heuristics and biases.” *European Review of Social Psychology*, 2(1), 83–115.
<https://doi.org/10.1080/14792779143000033>

Gigerenzer, G. (2008). Bounded and rational. *Rationality for Mortals*, 3–19.

<https://doi.org/10.1093/oso/9780195328981.003.0001>

Gigerenzer, G. (2014). *Risk savvy*. Penguin.

Gigerenzer, G. (2018). The bias bias in behavioral economics. *Review of Behavioral Economics*, 5(3–4), 303–336. <https://doi.org/10.1561/105.000000092>

Gigerenzer, G. (2020). When all is just a click away. *Critical Thinking in Psychology*, 197–223. <https://doi.org/10.1017/9781108684354.010>

Goldman, A. I. (2001). Experts: Which ones should you trust? *Philosophy and Phenomenological Research*, 63(1), 85. <https://doi.org/10.2307/3071090>

Goldman, A. I. (2016). Expertise. *Topoi*, 37(1), 3–10. <https://doi.org/10.1007/s11245-016-9410-3>

Goldman, A. I. (2021). How can you spot the experts? an essay in social epistemology. *Royal Institute of Philosophy Supplement*, 89, 85–98. <https://doi.org/10.1017/s1358246121000060>

Grundmann, R. (2016). The problem of expertise in knowledge societies. *Minerva*, 55(1), 25–48. <https://doi.org/10.1007/s11024-016-9308-7>

Grüne-Yanoff, T., & Hertwig, R. (2016). Nudge versus boost: How coherent are policy and theory? *Minds and Machines*, 26(1–2), 149–183. <https://doi.org/10.1007/s11023-015-9367-9>

Hardoš, P. (2018). Who Exactly is an Expert? On the Problem of Defining and Recognizing Expertise. *Sociológia*, 50(3), 268–288.

Hardwig, J. (1985). Epistemic dependence. *The Journal of Philosophy*, 82(7), 335–349.

<https://doi.org/10.2307/2026523>

Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science*, 12(6), 973–986.

<https://doi.org/10.1177/1745691617702496>

Hjørland, B. (2024, January 1). Social epistemology (isko encyclopedia of Ko). Social epistemology (IEKO). <https://www.isko.org/cyclo/se>

Jost, J. T. (2015). Resistance to change: A Social Psychological Perspective. *Social Research: An International Quarterly*, 82(3), 607–636.

<https://doi.org/10.1353/sor.2015.0035>

Keil, F. C. (2010). When and why do hedgehogs and foxes differ? *Critical Review*, 22(4), 415–426. <https://doi.org/10.1080/08913811.2010.541695>

Kahan, D. M., and D. Braman. 2006. “Cultural Cognition and Public Policy.” *Yale Law & Policy Review* 24: 149–172.

Kant, I., & Humphrey, T. (n.d.). What is enlightenment? [1] Immanuel Kant (1784).

https://www.nypl.org/sites/default/files/kant_whatisenlightenment.pdf

Levy, N. (2024). No trespassing! abandoning the novice/expert problem. *Erkenntnis*.

<https://doi.org/10.1007/s10670-024-00794-8>

Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012).

Misinformation and its correction. *Psychological Science in the Public Interest*, 13(3), 106–131. <https://doi.org/10.1177/1529100612451018>

Matheson, D. (2005). Conflicting experts and dialectical performance: Adjudication Heuristics for the layperson. *Argumentation*, 19(2), 145–158.

<https://doi.org/10.1007/s10503-004-7041-0>

Matthew 7:16-20. YouVersion | The Bible App | Bible.com. (n.d.).

<https://www.bible.com/bible/compare/MAT.7.16-20#>

Mercier, H., & Sperber, D. (2011). Why do humans reason? arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2), 57–74.

<https://doi.org/10.1017/s0140525x10000968>

St. Vincent Millay, E. (n.d.-c). A few figs from thistles. *A Few Figs From Thistles*.

<https://digital.library.upenn.edu/women/millay/figs/figs.html>

Miller, A. (2021). The Problem of Expert Testimony, and the Problems with Social Epistemology's Attempt to Resolve it. *Bilt.online*.

<https://bilt.online/wpcontent/uploads/2021/08/The-Problem-of-Expert-Testimony-and-the-Problems-with-Social-Epistemologys-Attempt-to-Resolve-it-Miller.pdf>

Miller, B., & Pinto, M. (2022). Epistemic equality: Distributive epistemic justice in the context of justification. *Kennedy Institute of Ethics Journal*, 32(2), 173–203.

<https://doi.org/10.1353/ken.2022.0011>

Pierson, R. (1994). The epistemic authority of expertise. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1994(1), 398–405.

<https://doi.org/10.1086/psaprocbienmeetp.1994.1.193044>

Plato, Charmides. *Plato, Charmides*, page 170. (n.d.).

<http://www.perseus.tufts.edu/hopper/text?doc=Plat.%2BCharm.%2B170&fromdoc=Perseus%3Atext%3A1999.01.0176>

Pronin, E. (2007). Perception and misperception of bias in human judgment. *Trends in Cognitive Sciences*, 11(1), 37–43.

<https://doi.org/10.1016/j.tics.2006.11.001>

- Reid, A. (2024). Does public justification face an ‘expert problem’? some thoughts in light of the COVID-19 pandemic. *Critical Review of International Social and Political Philosophy*, 1–21. <https://doi.org/10.1080/13698230.2024.2313953>
- Skipper, M. (2020b). The humility heuristic, or: People worth trusting admit to what they don’t know. *Social Epistemology*, 35(3), 323–336.
<https://doi.org/10.1080/02691728.2020.1809744>
- Solomon, M. (2014). Social Epistemology in Practice. In *Philosophy of Social Science: A New Introduction* (pp. 249–262). essay, Oxford University Press.
- Sturm, T. (2012). The “Rationality wars” in psychology: Where they are and where they could go. *Inquiry*, 55(1), 66–81. <https://doi.org/10.1080/0020174x.2012.643628>
- Tetlock, P. E. (2005). *Expert political judgment: How good is it? how can we know?* Princeton University Press.
- Tetlock, P. E., & Gardner, D. (n.d.). *Superforecasting : The art and science of prediction*.
https://esotericlibrary.weebly.com/uploads/5/0/7/7/5077636/philip_e._tetlock_-_superforecasting_the_art_and_science_of_prediction.pdf
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>
- Viehoff, D. (2023). *Challenging democratic commitments*. Oxford Studies in Political Philosophy Volume 9, 1–33. <https://doi.org/10.1093/oso/9780198877639.003.0001>