

An Idle and Most False Imposition: Truth-Seeking vs. Status-Seeking and the Failure of Epistemic Vigilance*

Joseph Shieber
Lafayette College

1. Introduction

Herman Melville's last published novel, *The Confidence-Man: His Masquerade*, was not greeted positively when it appeared in 1857. Both the reading public and the critics found it too bleak and cynical (Bergmann 1969). The novel, which plays out over the course of a single April Fools' Day aboard a Mississippi steamboat, consists of a series of interlocking vignettes in which a variety of confidence men—who may or may not be one and the same person employing different disguises—encounter, and dupe, a number of figures representing different elements of 19th century American society.

In the intervening 150-plus years since its publication, *The Confidence-Man* has come to be seen as one of Melville's major works. More recent critics have recognized the novel as—among other things—a trenchant commentary on American social structures (Chase 1949), an investigation of the upheavals of gilded age capitalism (Knight 2016), an exploration of the rise of the law of contracts in 19th century America (Blumenthal 2019), or, in the aftermath of the US presidential election of 2016, a foreshadowing of the ascendancy of 21st century populist demagogues (Thurman 2017).

In the world of *The Confidence-Man*, there are only three types of people: trusting dupes, distrustful cynics, and confidence men. Melville introduces all three of these archetypes at the very outset of the novel. The book opens as passengers board a Mississippi steamer, walking past “a placard near the captain's office, offering a reward for the capture of a mysterious impostor, supposed to have recently arrived from the East—quite an original genius in his

* I am indebted to the audience at SUNY Brockport for stimulating and very helpful comments, and to the philosophy faculty at Brockport—particularly Joe Long and Gordon Barnes—for wide-ranging discussions on these and related topics.

vocation.” (Melville 1990, 49) Whether any—or all—of the hucksters who appear in the novel are this “mysterious imposter,” the placard is the first intimation of the centrality of the figure of the confidence-man in the action of the novel.

At this point in the opening of the novel a stranger appears, pushes himself through the crowd of boarding passengers and takes up a position just beside the placard. The stranger, who never speaks, carries a small slate, which he holds up, level with the placard, so that the passengers who read the placard will also read the messages that the stranger writes on the slate. The stranger’s messages are pleas for trust: “The words were these: ‘Charity thinketh no evil.’ ... ‘Charity suffereth long, and is kind’” (Melville 1990, 50). The passengers in the crowd, “perceiving no badge of authority about [the stranger], but rather something quite the contrary,” jostle him aside, and, “as if at last despairing of so difficult an adventure, wherein one, apparently a nonresistant, sought to impose his presence upon fighting characters, the stranger now moved slowly away, yet not before altering his writing to this: ‘Charity endureth all things.’ Shield-like bearing his slate before him, amid stares and jeers he moved slowly up and down, at his turning points again changing his inscription to—‘Charity believeth all things.’ and then—‘Charity never faileth’” (Melville 1990, 50).

Melville immediately places the passengers’ reaction to the appeals of the stranger in stark relief in comparison to their reaction to another character aboard the boat:

To some observers, the singularity, if not lunacy, of the stranger was heightened ... by the contrast to his proceedings afforded in the actions—quite in the wonted and sensible order of things—of the barber of the boat, whose quarters, under a smoking-saloon, and over against a bar-room, were next door but two to the captain’s office. ... With business-like dispatch, having rattled down his shutters, ... he concluded his operations by bidding people stand still more aside, when, jumping on a stool, he hung over his door, on the customary nail, a gaudy sort of illuminated pasteboard sign, skilfully executed by himself, gilt with the likeness of a razor elbowed in readiness to shave, and also, for the public benefit, with two words not unfrequently seen ashore gracing other shops besides barbers’:

‘NO TRUST’—an inscription which, though in a sense not less intrusive than the contrasted ones of the stranger, did not, as it seemed, provoke any corresponding derision or surprise, much less indignation; and still less, to all appearances, did it gain for the inscriber the repute of being a simpleton. (Melville 1990, 51)

Here, in quick succession, we see all three archetypes that Melville sketches in his novel. The placard warns the passengers against the threat of the confidence man. The stranger’s pleas represent the trusting dupe, who “thinketh no evil” and “believeth all things.” The barber—whose actions, unlike those of the stranger, meet with the approval of the passengers streaming on board the boat—represents the cynic, with his skillfully executed sign trumpeting “NO TRUST.”

Confidence men unquestionably exist. Given this, it might be easiest to follow the example of the passengers Melville describes in his novel, to adopt the motto of the cynic and to view as dupes those who, like Melville’s stranger, advocate trust. Even in Melville’s day, however, there were many who pushed back against such rampant cynicism.

To consider one prominent example, Melville’s near-contemporary, the notorious showman and self-promoter P.T. Barnum, repudiates the cynical option thus:

The greatest humbug of all is the man who believes—or pretends to believe—that everything and everybody are humbugs. We sometimes meet a person who professes that there is no virtue; that every man has his price, and every woman hers; that any statement from anybody is just as likely to be false as true and that the only way to decide which, is to consider whether truth or a lie was likely to have paid best in that particular case. ... Honor he thinks is a sham. Honesty he considers a plausible word to flourish in the eyes of the greener portion of our race, as you would hold out a cabbage-leaf to coax a donkey. ... Poor fellow! he has exposed his own nakedness. Instead of showing that others are rotten inside, he has proved that he is. He

claims that it is not safe to believe others—it is perfectly safe to disbelieve him. (Barnum 1866, 5)

Whatever reception Barnum’s repudiation of cynicism enjoyed in his own day, a recent theory in contemporary social epistemology suggests that, far from being the greatest humbug of all, the *typical* case is rather that of Barnum’s “man who believes . . . that any statement from anybody is just as likely to be false as true and that the only way to decide which, is to consider whether truth or a lie was likely to have paid best in that particular case.” Like Melville’s steamboat passengers, the proponents of this contemporary theory suggest that it is in fact correct to view the overly trusting as both dupes and outliers. Most people, according to these contemporary theorists, eschew gullibility for vigilance.

The classic statement of this theory is the 2010 paper “Epistemic Vigilance” (Sperber et al. 2010). In that paper, Sperber and his colleagues suggest that “humans have a suite of cognitive mechanisms for epistemic vigilance, targeted at the risk of being misinformed by others.” (Sperber et al. 2010, 359) Since its publication in 2010, the theory of epistemic vigilance has become extremely influential (representative works from proponents of the theory include Mercier & Sperber 2011; Sperber & Mercier 2017; Boyer 2018; Mercier 2020; other discussions include, e.g., Kornblith 2021; Levy 2021; Sterelny 2021).

In challenging the theory of epistemic vigilance, I will suggest that the theory is only partially supported by the evidence that its proponents cite in its favor. In fact, the supporters of the theory have fallen prey to a false dichotomy: that the only options are gullibility or epistemic vigilance. More particularly, I will argue that proponents of the theory of epistemic vigilance have conflated evidence for the idea that people are *vigilant* with evidence for the idea that their vigilance is *epistemic*—that is, that their vigilance is reliably responsive to the truth or evidential support of the information they are evaluating. Instead, I will argue that although people *are* indeed vigilant, their vigilance is sensitive to qualities that are *not* properly epistemic.

The argument proceeds as follows. In section two I present the evidence adduced by the supporters of the theory of epistemic vigilance and sketch an argument on behalf of the theory. Following that, in section three, I present a first major challenge to the theory: evidence from social psychology suggesting that people are in fact unreliable at evaluating their conversation partners for honesty and

competence. Then, in section four, I present a second major challenge to the theory: evidence from evolutionary psychology suggesting that it would be highly implausible to think that people would have evolved a capacity to be accurate at detecting truthfulness or competence. In section five I contrast one of the core pre-suppositions of the theory of epistemic vigilance, that the core motivation of human communication is the transmission of information, with an alternative notion, one that I term “Nietzsche’s Thesis.” Following this, in section six, I suggest that it would be better to replace the theory of epistemic vigilance with one that better fits the evidence, a theory that I term “Machiavellian vigilance.” Finally, I conclude by noting some implications of the failure of the theory of epistemic vigilance and the success of the theory of Machiavellian vigilance for social epistemology.

2. Epistemic Vigilance

P.T. Barnum criticizes the man who thinks “any statement from anybody is just as likely to be false as true and that the only way to decide which, is to consider whether truth or a lie was likely to have paid best in that particular case,” and seems confident that such a calculating approach to human communication is the exception, rather than the norm. According to the theory of epistemic vigilance, however, it is such calculation that is, in fact, the norm.

For example, in “Epistemic Vigilance,” the authors introduce their theory using the language of expected utility. They write that, “Neither [communicator nor addressee] is likely to invest ... effort without expecting some benefit in return. For the addressee, the normally expected benefit is to acquire some true and relevant information. For the communicator, it is to produce some intended effect in the addressee” (Sperber et al. 2010, 360).

Relatedly, in his 2020 summary of the theory of epistemic vigilance, Pascal Boyer describes human communication as an “arms race,” noting that, “In the case of human communication, the arms race consists in a competition between the capacity to make one’s utterances persuasive, on the one hand, and the ability to protect one’s beliefs from deception, on the other. Dan Sperber and colleagues called this latter capacity ‘epistemic vigilance,’ the motivation and capacity to detect and discard unreliable information, and to check arguments for their validity” (Boyer 2020, p. 95).

One difficulty in engaging with the theory of epistemic vigilance is that its proponents seldom—if ever—provide explicit arguments for it. In the inaugural paper “Epistemic Vigilance,” the authors describe their goals as to “present this claim” and “to integrate [it] into a coherent topic for further research . . . , rather than to present detailed arguments” (Sperber et al. 2010, 359). In subsequent publications, adherents of the theory have appealed to the “Epistemic Vigilance” paper as support for the thesis of epistemic vigilance, rather than attempting to advance new arguments in support of the thesis.

Despite this lack of an explicit argument in favor of the theory, it is possible to formulate an argument from the materials provided in “Epistemic Vigilance.” Stating this tacit argument explicitly will aid both in understanding the motivations behind the theory of epistemic vigilance and in evaluating its weaknesses.

Here is the argument that I take to underwrite the discussion in “Epistemic Vigilance” (cf. Sperber et al. 2010, 362):

1. We are, at least to some extent, vigilant.
2. If our vigilance didn't serve some purpose, it wouldn't be reasonable.
3. Our vigilance is reasonable.
4. Thus, our vigilance serves some purpose.
5. The only purpose our vigilance could serve is epistemic.
6. Thus, our vigilance is epistemic vigilance.

The bulk of the discussion in “Epistemic Vigilance” is devoted to establishing the first premise, that we are in fact vigilant with respect to our interlocutors. Where the proponents of epistemic vigilance go astray is in their failure adequately to assess the additional logical steps needed to convert that evidential support for the fact of vigilance into evidential support for the fact of *epistemic* vigilance. As I will demonstrate, the move from the fact that we are vigilant to the conclusion that we are *epistemically* vigilant is, in fact, not adequately motivated.

In evaluating the move from vigilance to *epistemic* vigilance, I will focus, in subsequent sections of this article, largely on premise 5, that the only purpose our vigilance could serve is epistemic. However, it will be worthwhile to review some of the evidence that proponents of the theory of epistemic vigilance adduce for premise 1: that we are, in fact, vigilant.

Proponents of epistemic vigilance appeal to two main sorts of evidence to support the claim that we are vigilant: evidence about vigilance in mature, adult interlocutors (cf. Sperber et al. 2010, 369-70), and evidence about the development of vigilance in children (cf. Sperber et al. 2010, 371-3).

Thus, for example, Sperber and his co-authors note that adult interlocutors very quickly form judgments about the trustworthiness of their conversation partners:

A striking illustration of the tendency to form general judgments of trustworthiness on the basis of very limited evidence is provided in a study by Willis and Todorov (2006). Participants were shown pictures of faces, for either a mere 100 milliseconds or with no time limit, and asked to evaluate the person's trustworthiness, competence, likeability, aggressiveness and attractiveness. Contrary to the authors' expectations, the correlation between judgments with and without time limit was not greater for attractiveness (.69) ... than for trustworthiness (.73), while the correlations for aggressiveness and competence were a relatively low .52. One might wonder if such split-second judgments of trustworthiness have any basis at all, but what this experiment strongly suggests is that looking for signs of trustworthiness is one of the first things we do when we see a new face (see also Ybarra et al., 2001). (Sperber et al. 2010, 369-70)

Relatedly, Boyer notes that “people are attentive to the sources of information and maintain an estimate of a source's reliability, which affects how they process information. Conversely, the more suspicious the information, the better we recall its source as unreliable” (Boyer 2020, p. 95).

Strikingly, as both of these examples demonstrate, both Sperber and his co-authors and Boyer elide the difference between evidence for vigilance and evidence for *epistemic* vigilance. This is particularly obvious in the case of Sperber and his colleagues, who recognize that split-second judgments fail to provide *good* evidence of trustworthiness (I'll return to this point in the next section), but who nevertheless appeal to the fact that we make such split-second judgments as evidence of *epistemic* vigilance.

However, even in the case of Boyer’s argument, a moment’s reflection should underscore the need for caution in interpreting his claims. The fact that people maintain estimates of a source’s reliability, for example, gives us no reason for concluding that those estimates of a source’s reliability are themselves *accurate*. Indeed, in section three, we will see that there is good reason to think that those estimates are in fact *inaccurate*. Similarly, the fact that people are more likely to recall a source as unreliable if it has provided suspicious information doesn’t provide reason for thinking that people’s suspicions about information are well-founded. Again, in section three, we will see that there is good reason for thinking that people’s suspicions about that source’s information often *aren’t* well-founded.

Perhaps the most significant body of evidence to which supporters of epistemic vigilance appeal concerns the development of vigilance in infants and children (cf. Sperber et al. 2010, 371-3). Here, for example, is how Boyer helpfully summarizes the extensive literature to which the proponents appeal:

Some rudiments of epistemic vigilance appear early in cognitive development. Infants, for instance, seem to be sensitive to the difference between expert and novice agents. Later, toddlers use cues of competence to judge different individuals’ utterances, and mistrust those who have been wrong in previous instances, or those who seem determined to exploit others, or more simply agents talking about something they cannot possibly know, such as objects they cannot perceive. (Boyer 2020, p. 96)

Dealing with this literature in detail would require a paper of its own (see, for example, the extensive discussion of these issues in Shieber 2015, 129-136). However, even a cursory consideration of the dialectic should suffice to establish that the evidence on vigilance in infants and children is insufficient to support the idea that what infants and children practice is *epistemic* vigilance.

The central reason for this is that—as the discussion in the following section will demonstrate—*adults* are abject failures at detecting honesty or competence in their interlocutors. If, as we will see, adults aren’t reliably sensitive to epistemically relevant qualities in their conversation partners, it would be surprising to expect infants and children to be more reliable at detecting honest or competent

informants. Rather, what the wealth of evidence gathered in the last few decades demonstrates is that infants and children are *vigilant*. Given the inaccuracy of human judgments of honesty and competence—discussed in the next section—it would be hasty to characterize this discrimination on the part of infants and children as *epistemic* vigilance.

3. People are Unreliable at Evaluating Honesty and Competence

The argument for the thesis of epistemic vigilance introduced in the previous section hinges on premise five, that the purpose our vigilance serves is *epistemic*—that is, that our vigilance involves a reliable sensitivity to epistemically relevant qualities of our conversation partners, such as their deceptiveness or competence. However, a great deal of evidence suggests that we’re *not*, in fact, reliably sensitive to either deceptiveness or competence (Cf. Shieber 2012; Shieber 2015).

Over the past five decades, social psychologists have investigated whether there are identifiable and reliable cues to deceptiveness, whether humans are at all attentive to those reliable cues, and whether humans are reliable at using those cues to detect deceptiveness. Theories investigating the cues to deception include work by Ekman (e.g., Ekman & Friesen 1969), factor theory (Zuckerman et al. 1981), interpersonal deception theory (Buller & Burgoon 1996), and Vrij’s attempts to elicit cues to deception by increasing cognitive load (Vrij & Granhag 2012).

Recent meta-analyses suggest that cue-based attempts to detect deception are doomed to fail. The most thorough meta-analysis (DePaulo et al. 2003) provides an exhaustive survey of a variety of potential cues to deception: non-verbal and paraverbal cues, as well as verbal and content-based cues. The researchers find that, although there *are* some cues associated with deceptiveness, there are very few valid behavioral cues of deceptiveness and even those cues only have small effect sizes (DePaulo et al. 2003; Hartwig & Bond 2011).

It is worth noting that researchers also find that people are not only bad at using cues to detect deception, but that they are also unaware of which cues they actually rely upon in attempting to detect deception (Hartwig & Bond 2011). As Sternglanz and his colleagues note, “For example, perceivers frequently say that they use lack of eye contact to determine that a sender is lying; however, in

actuality, lack of eye contact is only weakly related ($r = -.15$) to perceivers' judgments of deceptiveness. Consistent with classic findings that people are often misguided when reporting on their internal (often unconscious) cognitive processes (Nisbett & Wilson, 1977), people don't seem to know what cues they use when making deception judgments" (Sternglanz et al. 2019, 312).

In other words, despite many decades of research, the findings are remarkably consistent in demonstrating that humans are quite poor at deception detection. The intervening decades have done little to alter the assessment reached by Ekman and O'Sullivan (1991) that, "in every study reported, people have not been very accurate in judging when someone is lying. ... Average accuracy in detecting deceit has rarely been above 60% (with chance being 50%), and some groups have done worse than chance" (Ekman and O'Sullivan 1991, p. 913. For more recent confirmation of the Ekman and O'Sullivan claim, see the comprehensive meta-analysis by Bond & DePaulo 2006. Indeed, Sperber and his colleagues concede that human lie detection is unreliable; cf. Sperber et al. 2010, 370).

Could these results be an artifact of the laboratory setting in which most deception detection studies have taken place? The recent review paper by Sternglanz and colleagues suggests that this is not the case:

... Hartwig and Bond (2014) ... found that lies are equally detectable regardless of senders' degree of motivation, whether senders are students or non-students, whether senders are communicating about feelings versus facts, or the setting in which the senders' communication takes place. Hartwig and Bond interpret this finding as evidence that deception detection accuracy rates are not an artifact of laboratory settings, because the detectability of lies remains consistent across a variety of settings and situational variables. In other words, the low-accuracy rate of human lie detection without special training is stable and generalizable. (Sternglanz et al. 2019, 308)

Researchers have demonstrated the unreliability of human lie detection even for professionals whose jobs presumably depend on their ability to detect deception—customs officials (Kraut and Poe

1980), federal law enforcement officers (DePaulo and Pfeifer 1986), and police officers (Köhnken 1987). Furthermore, the ability to deceive without being detected is acquired early, as recent studies have shown that adults are no better at detecting deception in children than they are at detecting deception in adult subjects (Crossman and Lewis 2006).

The evidence, then, is quite strong that humans are unreliable at detecting deception. There is good reason to think that our judgments of competence are often equally unreliable. Recall the Willis and Todorov study discussed in the passage from Sperber et al., cited in the previous section. In that study, a mere 100-ms exposure to a photograph of a face was enough for respondents to form judgments about the trustworthiness of the person in the photograph—judgments that were highly correlated with judgments of trustworthiness formed without time limits. Of course, that study concerned *trustworthiness*, which seems more apt in calling into question human judgments of deceptiveness. Similar studies, however, also call into question judgments of competence.

For example, in a series of studies conducted in 2000, 2002, and 2004, Todorov and colleagues demonstrated that “inferences of competence, based solely on the facial appearance of political candidates and with no prior knowledge about the person, predict the outcomes of elections for the U.S. Congress” (Todorov et al. 2005, 1623). In other words, facial appearance seems highly correlated with judgments of competence, despite its lack of correlation with more objective measures of performance (cf. Mueller & Mazur 1996). This, in turn, is strong evidence that our judgments of competence are unreliable.

The study by Todorov and his colleagues is not the only reason to call into question the reliability of human competency judgments. In a similar study reported in *Science* in 2009, Antonakis and Dalgas hypothesized that naive competency judgments—based solely on facial appearance in photographs—would accurately predict election results. In a twist on the Todorov et al. study, Antonakis and Dalgas also compared children’s competency judgments against election results.

What they found was that naive adults’—and even children’s—competency judgments predicted actual election results with over 70% accuracy (Antonakis & Dalgas 2009). As Antonakis and Dalgas note, intelligence cannot be predicted on the basis of facial appearance alone (Zebrowitz et al. 2002). Furthermore, there is great variation in the competence of politicians (Simonton 2002). What

these results suggest, then, is that *actual* voters are choosing their politicians based on superficial qualities detectable in photographs of facial features alone—qualities that do not correlate with intelligence or competence.

Thus, we have now seen a great deal of evidence to suggest that humans are not in fact reliably sensitive to epistemically relevant qualities of their conversation partners. I have suggested that this calls into question premise five of the argument from the previous section, that the only purpose our vigilance could serve is epistemic. Before we proceed, however, it would be worth considering one potential response to this suggestion: perhaps we’re attempting to be vigilant about epistemically relevant qualities of our interlocutors, but we’re just failing.

The problem with this line of response is that it ignores premises two through four of the argument in section two, premises which give substance to the notion of “purpose” in premise five. In order to serve a purpose that makes vigilance reasonable, it would have to be the case that the cognitive resources we expend on vigilance are worth the cost. But this, in turn, means that it *wouldn’t* be reasonable to expend cognitive resources on vigilance if that vigilance wasn’t reliably attuned to significant properties of our interlocutors.

This is the point of the underlying evolutionary logic at the root of the arguments in favor of epistemic vigilance. For example, Sperber and colleagues suggest that, “The fact that communication is so pervasive despite [the] risk [of misinformation] suggests that people are able to calibrate their trust well enough to make it advantageous on average to both communicator and audience” (Sperber et al. 2010, 360). What the evidence we’ve reviewed here suggests, however, is that people *aren’t* able reliably to calibrate their trust on the basis of epistemically relevant properties. If that is correct, however, then this would suggest that our vigilance must be serving some purpose *other* than calibrating trust according to epistemically relevant qualities of our conversation partners.

4. Evolutionary Psychology Doesn’t Support Epistemic Vigilance

Since we are not reliably sensitive either to our interlocutors’ deceptiveness or to their competence, it would seem that the vigilance that we practice in response to our conversation partners is not *epistemic* vigilance. Before turning to a positive suggestion of what

we *are* vigilant towards, I want briefly to canvass some further reasons for thinking that it is a mistake to characterize our vigilance as *epistemic* vigilance, reasons that will motivate the move to the alternative notion of vigilance for which I will argue in sections five and six.

In the previous section, I appealed to the fact that we are not reliably sensitive to epistemically relevant qualities of our interlocutors in order to argue that our vigilance isn't properly characterizable as *epistemic*. In this section, I want to suggest two further reasons for questioning the characterization of our vigilance as epistemic. The first has to do with the phylogenetic origins of vigilance. The second has to do with the social elements of belief formation.

The first additional reason for questioning the theory of epistemic vigilance is that our vigilance is phylogenetically prior to the use of language. Many nonlinguistic, but social animals also practice vigilance with respect to their conspecifics. For example, monkeys will “pay” to see pictures of higher status monkeys (Deaner et al. 2005). Furthermore, monkeys will “play dumb” in the presence of higher status monkeys (Drea & Wallen 1999). Even baboons and ravens are sensitive to violations of social hierarchy (Sapolsky 2017, pp. 370 & 467). In other words, in the case of nonlinguistic social animals, it is clear that vigilance developed to allow animals to be sensitive to the status rankings of their conspecifics, in order to aid in coalition building. We will return to this point in sections five and six.

The second additional reason for questioning the theory of epistemic vigilance has to do with the nonepistemic aspects of vigilance stemming from the social dimension of belief acquisition (cf. Shieber 2013; Shieber 2019). Indeed, it would be surprising if an ability that evolved hundreds of thousands of years ago would help us navigate our contemporary information landscapes.

Social influences on belief formation are so manifold as to be difficult to trace. However, it would be worthwhile to survey a few such influences. I'll do so by citing observations made by none other than Sperber and his colleagues:

- “... it is quite common for a piece of information with no clearly identified source to be accepted and transmitted purely on the ground that it is widely accepted and transmitted—an obvious circularity” (Sperber et al. 2010, 380).

- “Csibra and Gergely (2009) have argued that people in general, and children in particular, are eager to acquire cultural information, and that this may bias them towards interpreting (and even over-interpreting) communicated information as having cultural relevance, and also towards accepting it” (Sperber et al. 2010, 381).
- “... many reputations are spread on a larger scale, by people with no knowledge relevant to their direct assessment. When an addressee has to decide whether or not to believe an unfamiliar source of information, she may have no other basis for her decision than her knowledge of the source’s reputation, which she is unable to assess herself, and which she is likely to accept for want of a better choice. All too often, reputations are examples of ideas which are accepted and transmitted purely on the ground that they are widely accepted and transmitted” (Sperber et al. 2010, 381).
- “... It may be that the content of the ideas matters less to you than who you share them with, since they may help define group identities. When what matters is the sharing, it may be that contents which are unproblematically open to epistemic evaluation would raise objections within the relevant social group, or would be too easily shared beyond that group. So, semi-propositional contents which can be unproblematically accepted by just the relevant group may have a cultural success which is negatively correlated with their epistemic value” (Sperber et al. 2010, 382).

I want to highlight two aspects of these observations of particular relevance to the discussion of vigilance and the epistemic: (1) we acquire much of what we believe by means of social mechanisms the workings of which are opaque to us, and (2) many of those social mechanisms are not responsive to considerations of truth or evidence. Since the workings of those social mechanisms are opaque to us, however, we are unable to recognize when our socially-derived beliefs are responsive to considerations of truth or evidence. This, however, is further reason for thinking that our vigilance cannot be characterized as epistemic.

There is an additional aspect of these observations worth emphasizing, as it reinforces one of the lessons from the previous section. If our vigilance were epistemic, then it would be responsive

to our interlocutors' reputations—*and* those reputations themselves would be based on epistemically relevant qualities. However, as Sperber and his colleagues note, “All too often, reputations are examples of ideas which are accepted and transmitted purely on the ground that they are widely accepted and transmitted.” Or, to cite Iago, that noted expert on human psychology, “Reputation is an idle and most false imposition; oft got without merit and lost without deserving” (*Othello* Act II, Scene iii; Shakespeare 1903, 112).

5. Nietzsche's Thesis

The evidence from the previous two sections suggests that we are not reliably sensitive to epistemically relevant qualities of our conversation partners. Thus, if our vigilance is somehow contributing to our evolutionary fitness, that vigilance cannot be epistemic.

This can seem puzzling. The standard view in philosophy—and in evolutionary psychology, for that matter—is that the purpose of conversation is the exchange of information. If that is the case, however, then it might seem obvious that any vigilance that we exercise in monitoring our conversation partners would have to do with the quality of the information that we receive, and would thus be epistemically relevant.

In this section and the next, I suggest that the standard view is mistaken. There is no single purpose to conversation. To the extent that the exchange of information is *a* purpose of conversation, it is only one among many. Furthermore, if we were to choose one purpose for conversation that took priority over the other purposes, it would in fact be a *mistake* to single out the exchange of information.

We began this discussion in section one by considering Melville and Barnum on trust and distrust. To appreciate that the exchange of information is not the primary purpose of conversation, it will help to turn to a third near-contemporary of both Melville and Barnum, Friedrich Nietzsche.

Nietzsche criticizes the traditional conception in his essay “On Truth and Lies in an Extra-Moral Sense,” an early work written in 1873, but not published until the late 1890s by Nietzsche's sister (cf. Clark 1990; Gemes 1992; Nehamas 2017). According to Nietzsche, far from serving the exchange of information, our conversation is instead a vehicle for self-presentation:

This art of dissimulation reaches its peak in man: here deception, flattery, lying and cheating, talking behind the backs of others, keeping up appearances, living in borrowed splendor, donning masks, the shroud of convention, playacting before others and before oneself—in short, the continual fluttering around the flame of vanity is so much the rule and the law that virtually nothing is as incomprehensible as how an honest and pure drive to truth could have arisen among men. (Nietzsche 2010, 11)

To be fair, Nietzsche shares with the proponents of the theory of epistemic vigilance the idea of conversation as an arms race. He notes that,

In the natural state of things, the individual, inasmuch as he wants to protect himself against other individuals, uses his intellect mostly for dissimulation. But because, out of both necessity and boredom, he wants to exist socially and in herds, man needs a peace treaty and strives at the least to rid his world of the crudest forms of *bellum omnium contra omnes*. (Nietzsche 2010, 12)

It is this, according to Nietzsche, that leads to the distinction between truth and falsehood: we agree to use words in a certain way so that we may be understood by others in our “herd.” Any intentional deviation from the agreed-upon use constitutes, for Nietzsche, a lie:

the liar uses valid designations—words—to make the unreal appear real; he says, for instance, “I am rich,” precisely when the proper designation for his condition would be “poor.” He misuses fixed conventions by various substitutions or even inversions of names. If he does this in self-serving or otherwise injurious ways, society will no longer trust him and will therefore exclude him from its ranks. (Nietzsche 2010, 12)

However, where Nietzsche parts ways with the proponents of the theory of epistemic vigilance is in the idea that what we value about

conversational exchanges is the possibility of acquiring accurate information. If that were the case, then we would always condemn untruthful communications and always welcome truthful ones. This, however, is not what we do:

... men flee not so much from being cheated as from being harmed by cheating. Even on this level, it is at bottom not deception they hate but the dire, inimical consequences of certain kinds of deception. So, too, only to a limited extent does man want truth. He desires the pleasant, life-preserving consequences of truth; to pure knowledge without consequences he is indifferent, to potentially harmful and destructive truths he is even hostile. (Nietzsche 2010, 12)

Nietzsche's point here seems to be twofold. First, he suggests—contra the proponents of epistemic vigilance—that our goal in conversation is *not* primarily to acquire truthful information. Indeed, we have no problem with deception that isn't harmful, and we are in fact *hostile* “to potentially harmful and destructive truths.” What, then, *is* our primary goal in conversation?

This is the second aspect of Nietzsche's point. If anything, it would seem that the primary goal of conversation is self-presentation: “deception, flattery, lying and cheating, talking behind the backs of others, keeping up appearances, living in borrowed splendor, donning masks, the shroud of convention, playacting before others and before oneself—in short, the continual fluttering around the flame of vanity.” Another way to say this is that we engage in conversation as a form of status game, to make ourselves look good and to monitor the status attempts of others.

This is the claim that I dub “Nietzsche's Thesis”:

Nietzsche's Thesis: In conversational interactions we are primarily concerned with our and our interlocutors' social status, rather than their truthfulness. How we evaluate those conversational exchanges has more to do with our concern for our own social status, rather than for truth.

At least some researchers working within the framework of evolutionary psychology have recognized the value of Nietzsche's Thesis. Here's how Simler and Hanson put a related observation:

Modeling the world accurately isn't the be-all and end-all of the human brain. Brains evolved to help our bodies, and ultimately our genes, get along and get ahead in the world—a world that includes not just rocks and squirrels and hurricanes, but also other human beings. And if we spend a significant fraction of our lives interacting with others (which we do), trying to convince them of certain things (which we do), why shouldn't our brains adopt socially useful beliefs as first-class citizens, alongside world-modeling beliefs? (Simler & Hanson 2017, 93)

In order to buttress their argument that communication doesn't primarily serve as a medium for informational exchange, Simler and Hanson pose four puzzles for information exchange theories of communication (cf. Simler & Hanson 2017, 164-7): (1) People don't keep track of conversational debts, (2) People are more eager to talk than to listen, (3) The criterion of relevance—from an information maximization perspective, conversations should bounce around willy-nilly, but they don't; and (4) Suboptimal exchanges—when people first meet, they seldom share the most important information they possess. Let's briefly consider these puzzles, the challenge they pose for information exchange theories, and how much better they comport with Nietzsche's Thesis.

The first puzzle involves the fact that people don't keep track of conversational debts. For example, if Bill tells Debbie some information—say, about trouble he's having with a colleague at work—he doesn't then refrain from telling Debbie further information until she has shared an equal amount of information with him. If information was the commodity of value in communication, however, then we would be much more rigorous about keeping account of who is in our informational debt—which is not, in fact, something we do at all.

The second puzzle involves the fact that people are much more eager to talk than listen. It is a frequent occurrence that people can't wait for a turn in a conversational exchange. In fact, in multi-person conversations, different speakers often jostle for a chance to speak. This would not make sense if the point of communication was to receive valuable information. If that were the case, then the wise course of action would be to speak as little as possible, while

acquiring as much information from others as possible. Again, such behavior is, however, far from the rule.

The third puzzle involves the criterion of relevance. This is the criterion that suggests that conversational exchanges ought to be relevant. If Jane is talking about her sick uncle, then Sidney can't follow up on Jane's contribution by talking about rising semiconductor prices in Singapore. If the point of communication was information exchange, however, then this is actually the opposite of what we would expect. The most informative contributions wouldn't be constrained by relevance; rather, conversations would bounce around willy-nilly from one topic to the next.

The fourth puzzle involves the fact of suboptimal informational exchanges. When two people meet for the first time, they often exchange pleasantries or share information, at most, about what their job is or their familial situation. If the purpose of communication was information exchange, however, people would take advantage of meeting new people to find out the most significant information they possess—who knows if they will have another opportunity to acquire information from these strangers a second time!? Again, such behavior is not, in fact, what we observe in actual conversational exchanges.

Notice that Nietzsche's Thesis provides a solution to all of these puzzles. According to Nietzsche's Thesis, conversation is an opportunity for speakers to enhance their status—through “deception, flattery, lying and cheating, talking behind the backs of others, keeping up appearances, living in borrowed splendor, donning masks, the shroud of convention, playacting before others and before oneself.” This explains both the first and second puzzles: we are happy to share more information with others, and to speak more and listen less. It also explains the third and fourth puzzles. Both the criterion of relevance and suboptimal exchanges serve to function as displays of concern for and interest in our conversation partners. Displaying concern for and interest in our conversation partners itself serves to maintain our status with those conversation partners.

Here's how Simler and Hanson put the point, referring to the work of the psychologist Geoffrey Miller and the cognitive scientist Jean-Louis Dessalles:

To resolve these puzzles, both Miller and Dessalles suggest that we stop looking at conversation as an exchange of information, and instead try to see the

benefits of speaking as something other than receiving more information later down the road. Specifically, both thinkers argue that speaking functions in part as an act of showing off. Speakers strive to impress their audience by consistently delivering impressive remarks. This explains how speakers foot the bill for the costs of speaking we discussed earlier: they're compensated not in-kind, by receiving information reciprocally, but rather by raising their social value in the eyes (and ears) of their listeners. (Simler & Hanson 2017, 167)

It is important to deal with the fact that some might read Dessalles in particular as an ally of the proponents of epistemic vigilance. This is because Dessalles thinks that the specific way in which we enhance our status is by proving ourselves to be good informants (Dessalles 2007). This aspect of Dessalles's view, however, is a weakness rather than a strength.

What Dessalles gets right is that conversational exchanges serve status functions—enhancing or maintaining status. What Dessalles gets wrong is that he is too limited in his understanding of what sorts of behavior might maintain or enhance status. That is, Dessalles fails to recognize that what strategies will enhance—or, at the very least, maintain—status varies depending on the local status group. Some status groups—at a scientific conference, for example—may well put a high premium on significant and novel information. Many more status groups—including most of the ones in which we are embedded in our day-to-day conversational interactions—prioritize different conversational goals (cf. the points by Machery in Machery et al. 2010, 865-6).

The key point to note in the context of the present discussion is that the pursuits that achieve status value within a given status group in which a conversation takes place determine the goals for that particular conversation. That is, status considerations determine the relevant conversational goals. This explains how it is that, although the particular conversational goals may vary depending on the status group, the underlying goal is always the same: the maintenance or enhancement of status.

There is perhaps no better example of this than that of the character of the barber in Melville's *Confidence Man*. Over the course of two chapters, Melville skillfully illustrates how appeals to status can cause even the most stalwart cynic to abandon his distrust.

Challenged by a confidence man, in the guise of “FRANK GOODMAN, Philanthropist, and Citizen of the World” (Melville 2009, 319), the barber defends his “No Trust” sign by pointing to his experience in dealing with the prevalence of human deception. He asks, “can one be for ever dealing in macassar oil, hair dyes, cosmetics, false moustaches, wigs, and toupees, and still believe that men are wholly what they look to be?” The barber continues that speech, with its echoes of Nietzsche’s catalog of “continual fluttering around the flame of vanity”:

“To contrast the shamefaced air behind the curtain, the fearful looking forward to being possibly discovered there by a prying acquaintance, with the cheerful assurance and challenging pride with which the same man steps forth again, a gay deception, into the street, while some honest, shock-headed fellow humbly gives him the wall. Ah, sir, they may talk of the courage of truth, but my trade teaches me that truth sometimes is sheepish. Lies, lies, sir, brave lies are the lions!” (Melville 2009, 315)

The confidence man Goodman, however, wears down the barber’s resistance in a series of exchanges that echo Barnum’s appeals to humanity. “To say that strangers are not to be trusted,” Goodman cajoles,

“does not that imply something like saying that mankind is not to be trusted; for the mass of mankind, are they not necessarily strangers to each individual man? Come, come, my friend,” winningly, “you are no Timon to hold the mass of mankind untrustworthy. Take down your notification; it is misanthropical; much the same sign that Timon traced with charcoal on the forehead of a skull stuck over his cave. Take it down, barber; take it down to-night. Trust men. Just try the experiment of trusting men for this one little trip. Come now, I’m a philanthropist, and will insure you against losing a cent.” (Melville 2009, 312)

These appeals eventually have the desired effect. The barber removes the sign, takes Goodman's assurance that he will be reimbursed for any lost wages suffered by rendering his services on credit ... and lets Goodman be the first to walk out of the shop without paying for his shave.

6. Machiavellian Vigilance

Recall the argument attributed to Sperber and colleagues in section two:

1. We are, at least to some extent, vigilant.
2. If our vigilance didn't serve some purpose, it wouldn't be reasonable.
3. Our vigilance is reasonable.
4. Thus, our vigilance serves some purpose.
5. The only purpose our vigilance could serve is epistemic.
6. Thus, our vigilance is epistemic vigilance.

Though we have seen good evidence for premise 1 and have, for the sake of argument, accepted premises 2-4, we have acquired excellent reasons to doubt premise 5. Indeed, the evidence that we've canvassed provides better support for the idea that we pay attention to social status rather than truth.

In other words, the theory of epistemic vigilance is false; we haven't evolved to keep track of our conversation partners for signs of honesty and competence. Rather, what is true is the theory of what I dub "Machiavellian vigilance":

Machiavellian vigilance: We have evolved to keep track of our conversation partners for signs of their relative social status and in the service of maintaining or enhancing our own social status.

In other words, I suggest replacing premise 5 of the epistemic vigilance argument with:

- 5*. The most likely purpose for our vigilance is Machiavellian.

Leading to the conclusion, 6*. Thus, our vigilance is Machiavellian vigilance.

We have already seen that appealing to social status considerations makes better sense of our communicative goals than appealing to information exchange. In what remains of this section, I will review evidence to suggest that our vigilance is better understood as Machiavellian vigilance than epistemic vigilance.

First, recall Boyer's summary that "people are attentive to the sources of information and maintain an estimate of a source's reliability, which affects how they process information. Conversely, the more suspicious the information, the better we recall its source as unreliable" (Boyer 2020, 95). What we saw in section two was that Boyer offers no reason for thinking that people's estimates are reliable, or that people's suspicions are based on accurate credibility judgments. In fact, as we saw in section three, there is a great deal of evidence that people are unreliable at estimating source reliability and that people don't make accurate credibility judgments.

People *are*, however, accurate at making *status* judgments. In other words, the evidence shows that people aren't accurate at remembering reliability, but rather coalitional and status information.

For example, Boyer notes that "... people do not recall information about out-groups and in-groups in the same way. They are much more distressed by disagreements with in-groups than with out-groups" (Boyer 2020, p. 53). Furthermore, Boyer cites "... memory experiments [that] demonstrate that people presented with conversations between unknown third parties automatically attend to who is allied to whom.... This information is then retrieved from memory more easily than other features of the interaction. That is, people may have fuzzy memories of exactly what was said, but they recall quite clearly who opposed whom" (Boyer 2020, p. 59).

Indeed, Boyer actually makes a claim that is quite close to Nietzsche's Thesis when he notes that "... the motivation to recruit is an important factor in people's processing of their beliefs. That is to say, beliefs are preselected in an intuitive manner, and those that could not trigger recruitment are simply not considered intuitive and compelling" (Boyer 2020, p. 108). Again, this provides evidence in favor of the Machiavellian vigilance thesis and *against* the epistemic vigilance thesis.

Additionally, in section four we already noted that many of the cognitive structures recruited in keeping track of our conversation partners are phylogenetically prior to human development and were evolved to subserve status tracking. For example, we noted that monkeys will "pay" to see pictures of higher status monkeys

(Deaner et al. 2005) and will “play dumb” in the presence of higher status monkeys (Drea & Wallen 1999). We also noted that baboons—as well as ravens—are sensitive to violations of social hierarchy (Sapolsky 2017, pp. 370 & 467).

The evolutionary origins of our focus on social status are reflected in our neurobiology. Going against group consensus is associated with increased amygdala activation in the brain, linked to feelings of aggressiveness or fear (Sapolsky 2017, p. 44). Furthermore, the brain can group faces by social status in as little as 50 milliseconds—as quickly as it processes gender or race (Sapolsky 2017, p. 423). Finally, brain studies demonstrate that people become more vigilant and conscious of others’ emotions when interacting with peers demonstrating obvious higher status cues (Kraus & Mendes 2014; Mattan et al. 2017).

Not only are we hardwired to make rapid status judgments, but those judgments—in contrast to our judgments of deceptiveness or competence—are remarkably accurate. In a review paper, Smith and Galinsky write that:

For example, dominant and submissive facial postures differentially affect brain activity (in the mid-superior temporal sulcus, lingual gyrus, and fusiform gyrus) in as little as 200 ms (Chiao et al., 2008). In a series of studies, ... Mast and Hall (2004) asked participants to look at candid photographs of pairs of employees and indicate the status difference within each pair (i.e., Employee A is higher status, Employee B is higher status, they are of equal status). Regardless of whether they saw the entire photograph at once or only one employee at a time, participants were very accurate in assessing status differences (mean r 's = 0.71 and 0.60, respectively). In fact, ... Mast and Hall point out that ‘in comparison to other domains of interpersonal sensitivity, accuracy of judging status seems high’ (p. 157). Participants were equally accurate in judging same-sex dyads as opposite sex dyads and in judging targets of the same sex as of the opposite sex. (Smith & Galinsky 2010, 922)

It is also important to note that we perform these rapid and reliable status judgments without attending to them—or even being aware of them.

We can see how effortlessly and nonconsciously we modulate our behavior to account for status differences by considering the visual dominance ratio. The visual dominance ratio is calculated by dividing the proportion of time one spends looking at an interlocutor while speaking to them by the proportion of time one spends looking at an interlocutor while listening to them. A visual dominance ratio close to or above 1 indicates relatively similar or higher status, while a visual dominance ratio well below 1 indicates relatively lower status.

Here is how Mlodinow (2013) describes the data regarding visual dominance ratio:

What is so striking about the data is not just that we subliminally adjust our gazing behavior to match our place on the hierarchy but that we do it so consistently, and with numerical precision. Here is a sample of the data: when speaking to each other, ROTC officers exhibited ratios of 1.06, while ROTC cadets speaking to officers had ratios of 0.61; undergraduates in an introductory psychology course scored 0.92 when talking to a person they believed to be a high school senior who did not plan to go to college but 0.59 when talking to a person they believed to be a college chemistry honor student accepted into a prestigious medical school; expert men speaking to women about a subject in their own field scored 0.98, while men talking to expert women about the women's field, 0.61; expert women speaking to nonexpert men scored 1.04, and nonexpert women speaking to expert men scored 0.54. These studies were all performed on Americans. The numbers probably vary among cultures, but the phenomenon probably doesn't. (Mlodinow 2013, 143)

Overall, what the evidence suggests is that, although we are unreliable with respect to our sensitivity to the deceptiveness or competence of our interlocutors, we are in fact highly reliable with respect to our sensitivity to our own and others' social status. This suggests

that the theory of epistemic vigilance is false, and that we should supplant that theory with the much better supported theory of Machiavellian vigilance.

7. Conclusion

The wealth of evidence we've considered is damning for the theory of epistemic vigilance. The theory of epistemic vigilance fails to account for our unreliability in deception detection and our inability accurately to detect competence. It stands in tension with the phylogenetic origins of our vigilance for status and cannot deal with the multiple social mechanisms of belief formation. Finally, the theory is unable to deal with the many puzzles plaguing information exchange theories of communication.

Contrast the failure of the theory of epistemic vigilance with the success of the theory of Machiavellian vigilance. The theory of Machiavellian vigilance explains our unreliability in deception detection and our inability to detect competence: we are vigilant with regard to our own and others' social status, rather than their epistemically relevant qualities. Furthermore, the theory of Machiavellian vigilance comports well with the phylogenetic origins of our vigilance for status and provides promising avenues for better explaining the social mechanisms of belief formation. Finally, the theory of Machiavellian vigilance—via the related Nietzsche's Thesis—provides a solution to the puzzles plaguing information exchange theories of communication.

In other words, we should abandon the theory of epistemic vigilance in favor of the theory of Machiavellian vigilance. Given the widespread influence of the theory of epistemic vigilance—the inaugural paper, “Epistemic Vigilance” has been cited more than 1,600 times—this result alone is significant. However, this finding is not only of independent interest, but also of great interest within the field of social epistemology. A large number of theories in social epistemology appeal to the notion of epistemic vigilance to support the idea that justified or warranted belief on the basis of testimony requires epistemic vigilance (e.g., Kappel 2014; Tebben & Waterman 2016; Dutilh Novaes 2020; Fricker 2021; Grodniewicz 2022).

So, to cite merely one recent example, Levy takes the theory of epistemic vigilance to be well-established and suggests that it shows that “we filter testimony by reference to cues that correlate with

reliability ... Filtering testimony in these ways is the behavior of a rational animal, not an unthinking conformist” (Levy 2021, 47).

The failure of the theory of epistemic vigilance removes the perhaps strongest remaining support for the idea that rational belief on the basis of testimony requires reliable monitoring of one’s informants for trustworthiness and competence (cf. Shieber 2022). Replacing that theory with the theory of Machiavellian vigilance will enable us to refocus our attention on more promising directions in the epistemology of testimony—in particular attempts to explain testimonial justification, warrant and knowledge in terms of the socially distributed cognitive networks in which we are embedded (cf. Shieber 2013; Shieber 2015; Shieber 2019).

References

- Antonakis, J. and Dalgas, O., 2009. Predicting elections: Child's play!. *Science*, 323(5918), pp.1183-1183.
- Barnum, P.T. 1866. *Humbugs of the World*. London: John Camden Hotten.
- Bergmann, J.D., 1969. The original confidence man. *American Quarterly*, 21(3), pp.560-577.
- Blumenthal, S.L. 2019. Counterfeiting Confidence: The Problem of Trust in the Age of Contract. In: A.L. LaCroix, S. Levmore, M.C. Nussbaum, eds. *Power, Prose, and Purse: Law, Literature, and Economic Transformations*. Oxford: Oxford University Press, pp. 15-50.
- Boyer, P. 2018. *Minds Make Societies*. New Haven: Yale University Press.
- Buller, D.B. and Burgoon, J.K., 1996. Interpersonal deception theory. *Communication theory*, 6(3), pp.203-242.
- Chase, R., 1949. Melville's Confidence Man. *The Kenyon Review*, 11(1), pp.122-140.
- Chiao, J.Y., Iidaka, T., Gordon, H.L., Nogawa, J., Bar, M., Aminoff, E., Sadato, N. and Ambady, N., 2008. Cultural specificity in amygdala response to fear faces. *Journal of cognitive neuroscience*, 20(12), pp.2167-2174.
- Clark, M., 1990. *Nietzsche on truth and philosophy*. Cambridge: Cambridge University Press.
- Crossman, A.M. and Lewis, M., 2006. Adults' ability to detect children's lying. *Behavioral Sciences & the Law*, 24(5), pp.703-715.
- Csibra, G. and Gergely, G. 2009. Natural pedagogy. *Trends in cognitive sciences*, 13(4), pp.148-153.

- Deaner, R.O., Khera, A.V. and Platt, M.L., 2005. Monkeys pay per view: adaptive valuation of social images by rhesus macaques. *Current Biology*, 15(6), pp.543-548.
- DePaulo, B.M., Lindsay, J.J., Malone, B.E., Muhlenbruck, L., Charlton, K. and Cooper, H., 2003. Cues to deception. *Psychological bulletin*, 129(1), 74-
- DePaulo, B.M. and Pfeifer, R.L., 1986. On-the-Job Experience and Skill at Detecting Deception 1. *Journal of Applied Social Psychology*, 16(3), pp.249-267.
- Dessalles, J.L., 2007. *Why we talk: The evolutionary origins of language*. Oxford: Oxford University Press.
- Drea, C.M. and Wallen, K., 1999. Low-status monkeys “play dumb” when learning in mixed social groups. *Proceedings of the National Academy of Sciences*, 96(22), pp.12965-12969.
- Dutilh Novaes, C., 2020. The role of trust in argumentation. *Informal Logic*, 40(2), pp.205-236.
- Ekman, P. and Friesen, W.V., 1969. Nonverbal leakage and clues to deception. *Psychiatry*, 32(1), 88-106.
- Ekman, P. and O'Sullivan, M., 1991. Who can catch a liar?. *American psychologist*, 46(9), 913-
- Fricker, E., 2021. Epistemic Self-Governance and Trusting the Word of Others: Is There a Conflict?. In J. Matheson and K. Loughheed, eds. *Epistemic Autonomy*. New York: Routledge, pp. 323-342.
- Gemes, K., 1992. Nietzsche's Critique of Truth. *Philosophy & Phenomenological Research*, 52(1), pp.47-65.
- Grodziewicz, J.P., 2022. Effective filtering: Language comprehension and testimonial entitlement. *The Philosophical Quarterly*, published early online: <https://doi.org/10.1093/pq/pqac064>.
- Hartwig, M. and Bond Jr, C.F., 2011. Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological bulletin*, 137(4), 643
- Hartwig, M. and Bond Jr, C.F., 2014. Lie detection from multiple cues: A meta-analysis. *Applied Cognitive Psychology*, 28(5), pp. 661-676.
- Kappel, K., 2014. Believing on trust. *Synthese*, 191, pp.2009-2028.
- Keltner, D., Van Kleef, G.A., Chen, S. and Kraus, M.W., 2008. A reciprocal influence model of social power: Emerging principles and lines of inquiry. *Advances in experimental social psychology*, 40, pp.151-192.

- Knight, P., 2016. Reading the market: genres of financial capitalism in gilded age America. Baltimore: Johns Hopkins University Press.
- Köhnken, G., 1987. Training police officers to detect deceptive eyewitness statements: Does it work?. *Social Behaviour*.
- Kornblith, H., 2021. *Scientific Epistemology: An Introduction*. Oxford University Press.
- Kraus, M.W. and Mendes, W.B., 2014. Sartorial symbols of social class elicit class-consistent behavioral and physiological responses: a dyadic approach. *Journal of Experimental Psychology: General*, 143(6), p.2330.
- Kraut, R.E. and Poe, D.B., 1980. Behavioral roots of person perception: The deception judgments of customs inspectors and laymen. *Journal of Personality and Social Psychology*, 39(5), p.784.
- Levy, N. 2021. *Bad Beliefs: Why They Happen to Good People*. Oxford: Oxford University Press.
- Machery, E., Dessalles, J. L., Cowie, F., & Alexander, J. 2010. Symposium on J.-L. Dessalles's *Why we Talk* (OUP, 2007): Precis by J.-L. Dessalles, commentaries by E. Machery, F. Cowie, and J. Alexander, Replies by J.-L. Dessalles. *Biology & Philosophy*, 25: 851-901.
- Mast, M.S. and Hall, J.A., 2004. Who is the boss and who is not? Accuracy of judging status. *Journal of Nonverbal Behavior*, 28, pp.145-165.
- Mattan, B.D., Kubota, J.T. and Cloutier, J., 2017. How social status shapes person perception and evaluation: A social neuroscience perspective. *Perspectives on Psychological Science*, 12(3), pp.468-507.
- Melville, H. 2009. *The Confidence-Man: His Masquerade*. Stephen Materson, ed. London: Penguin Books.
- Mercier, H. 2020. *Not Born Yesterday: The Science of Who We Trust and What We Believe*. Princeton: Princeton University Press.
- Mercier, H., Sperber, D. 2011. Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2), 57–74.
- Mercier, H., Sperber, D., 2017. *The Enigma of Reason*. Cambridge, MA: Harvard University Press.
- Mlodinow, L., 2013. *Subliminal: How your unconscious mind rules your behavior*. New York: Vintage.

- Mueller, U. and Mazur, A., 1996. Facial dominance of West Point cadets as a predictor of later military rank. *Social forces*, 74(3), pp.823-850.
- Nehamas, A., 2017. Nietzsche on Truth and the Value of Falseness. *Journal of Nietzsche Studies*, 48(3), pp.319-346.
- Nietzsche, F. 2010. *On Truth and Untruth: Selected Writings*. T. Carman, ed. and trans. New York: Harper Perennial Modern Classics.
- Nisbett, R.E. and Wilson, T.D., 1977. Telling more than we can know: Verbal reports on mental processes. *Psychological review*, 84(3), 231-
- Sapolsky, R.M., 2017. *Behave: The biology of humans at our best and worst*. New York: Penguin.
- Shakespeare, W., 1903. *The tragedy of Othello*. London: Methuen.
- Shieber, J., 2012. Against credibility. *Australasian Journal of Philosophy*, 90(1), pp.1-18.
- Shieber, J., 2013. Toward a truly social epistemology: Babbage, the division of mental labor, and the possibility of socially distributed warrant. *Philosophy and Phenomenological Research*, 86(2), pp.266-294.
- Shieber, J., 2015. *Testimony: A philosophical introduction*. New York: Routledge.
- Shieber, J., 2019. Socially Distributed Cognition and the Epistemology of Testimony. In Fricker, M., Graham, P.J., Henderson, D. and Pedersen, N.J., eds. *The Routledge Handbook of Social Epistemology*. New York: Routledge, pp. 87-95.
- Shieber, J., 2022. Challenging the ability intuition: From personal to extended to distributed belief-forming processes. *Philosophical Issues*, 32(1), pp.351-366.
- Simler, K. and Hanson, R., 2017. *The elephant in the brain: Hidden motives in everyday life*. Oxford: Oxford University Press.
- Smith, P.K. and Galinsky, A.D., 2010. The nonconscious nature of power: Cues and consequences. *Social and Personality Psychology Compass*, 4(10), pp. 918-938.
- Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origg, G. and Wilson, D., 2010. Epistemic vigilance. *Mind & Language*, 25(4), pp. 359-393.
- Sterelny, K., 2021. *The Pleistocene social contract: Culture and cooperation in human evolution*. Oxford: Oxford University Press.
- Sternglanz, R.W., Morris, W.L., Morrow, M. and Braverman, J., 2019. A review of meta-analyses about deception detection.

- The Palgrave handbook of deceptive communication, pp.303-326.
- Tebben, N. and Waterman, J.P., 2016. Counterfeit testimony: lies, trust, and the exchange of information. *Philosophical studies*, 173, pp.3101-3117.
- Thurman, J. 2017. Philip Roth E-Mails on Trump. *New Yorker*. Jan. 22, pp. 18-19.
- Todorov, A., Mandisodza, A.N., Goren, A. and Hall, C.C., 2005. Inferences of competence from faces predict election outcomes. *Science*, 308(5728), pp.1623-1626.
- Vrij, A. and Granhag, P.A., 2012. Eliciting cues to deception and truth: What matters are the questions asked. *Journal of Applied Research in Memory and Cognition*, 1(2), pp.110-117.
- Willis, J. and Todorov, A., 2006. First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological science*, 17(7), pp.592-598.
- Zuckerman, M., DePaulo, B.M. and Rosenthal, R., 1981. Verbal and nonverbal communication of deception. In *Advances in experimental social psychology* (Vol. 14), pp. 1-59.