

ARTICLE

Manufacturing the Illusion of Epistemic Trustworthiness

Tyler Porter 

University of Colorado Boulder, Boulder, CO, USA
Email: tyler.porter@colorado.edu

(Received 4 April 2023; revised 24 February 2024; accepted 6 March 2024)

Abstract

There are epistemic manipulators in the world. These people are actively attempting to sacrifice epistemic goods for personal gain. In doing so, manipulators have led many competent epistemic agents into believing contrarian theories that go against well-established knowledge. In this paper, I explore one mechanism by which manipulators get epistemic agents to believe contrarian theories. I do so by looking at a prominent empirical model of trustworthiness. This model identifies three major factors that epistemic agents look for when trying to determine who is trustworthy. These are (i) ability, (ii) benevolence, and (iii) moral integrity. I then show how manipulators can manufacture the illusion that they possess these factors. This leads epistemic agents to view manipulators as trustworthy sources of information. Additionally, I argue that fact-checking will be an ineffective – or even harmful – practice when correcting the beliefs of epistemic agents who have been tricked by this illusion of epistemic trustworthiness. I suggest that in such cases we should use an alternative correction, which I call trust undercutting.

Keywords: Contrarian theories; epistemic manipulation; fact-checking; social epistemology; trust; trustworthiness

1. Introduction

Some people believe that vaccines can cause autism, even though “vaccinations are not associated with the development of autism or autism spectrum disorder” (Taylor *et al.* 2014: 1).¹ Here, we have two competing claims. The first is that vaccines do not cause autism. This claim represents what Coady calls the *official story* (2003). In short, the official story is the prevailing theory of the day. The second claim is that vaccines do cause autism. This is a *contrarian theory* – i.e., a theory that conflicts with the official story (*ibid.*).² Contrarian beliefs are widespread.³ In this paper, my goal is to explore one

¹Thanks to an anonymous reviewer for recommending this example.

²Alternatively, Cassam (2019: 28) suggests that what makes a theory contrarian is to be “contrary to [the] appearances or obvious explanation of events.” Either of these views works for the purposes of this paper.

³Bader *et al.* (2016) found that two-thirds of Americans held paranormal beliefs. Each of these paranormal beliefs is a contrarian theory. At that time, 46.6% of Americans thought that “places can be haunted by

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

explanation as to why people come to endorse contrarian theories. That is, there are epistemic manipulators (henceforth, manipulators) who trick epistemic agents into holding contrarian theories for personal gain.⁴ I will explain one mechanism that manipulators use to trick epistemic agents – and some consequences of that mechanism. I call this mechanism *the illusion of epistemic trustworthiness*. I explain this illusion by looking at an influential model of trust. The model identifies several important factors that people look for when building trust relations. I go on to explain how manipulators can fabricate the appearance of each of these factors, and thus build the illusion that they are trustworthy sources of information. Manipulators can use this illusion to get epistemic agents to believe contrarian theories for personal gain. Additionally, I will argue that once an epistemic agent is tricked into viewing a manipulator as epistemically trustworthy, standard practices such as fact-checking will become at best ineffective and at worst harmful. I will suggest that instead of fact-checking we should engage in the practice of trust undercutting.

2. Three possible explanations for contrarian belief

We all know someone who has come to believe a contrarian theory rather than the official story. But why do people believe contrarian theories? There are multiple possible explanations. The primary explanation given in the literature is that people who endorse contrarian theories possess epistemic vices. Epistemic vices can be understood as problematic character traits “such as close-mindedness, gullibility, active ignorance, and cynicism” that make it difficult to fruitfully engage in epistemic work (Nguyen 2021: 5). For example, Cassam (2016) gives the case of Oliver – an epistemic agent who believes that 9/11 was an inside job. Cassam says the best explanation for Oliver’s belief is that he has epistemic vices which lead him to endorse this contrarian view (for more on epistemic vice, see Kidd *et al.* 2020; Swank 2000).

I agree that sometimes epistemic vice explains why an epistemic agent comes to endorse a contrarian theory. However, there are other cases where it does not. Perhaps counterintuitively, it has been found that some contrarians have better epistemic practices and more true beliefs (related to the subject of their contrarian theory) than many people who believe the official story. For example, Lee *et al.* (2021) found that contrarian theorists hold workshops about how to gather and evaluate raw figures; Klein *et al.* (2019) found that contrarian theorists are highly interested in gathering and analyzing evidence; Harris (2018) observed that contrarian theorists seem to search out and evaluate evidence more often than people who accept the official story; and Kahan (2015) found that people who deny that climate change is a significant problem are often more aware of how climate change works than people who view it as a significant problem. In short, while we tend to think of contrarian theorists as crackpots living in bunkers wearing tinfoil hats, many of them are intelligent people with generally good

spirits”; 39.6% thought that “ancient, advanced civilizations, such as Atlantis, once existed”; 27% believed that “aliens visited Earth in our ancient past”; and so on. For more, see <https://blogs.chapman.edu/wilkinson/2016/10/11/paranormal-beliefs/>.

⁴I take this project to be about hostile epistemology. Hostile epistemology is “the study of how external forces might subvert the efforts of epistemic agents” (Nguyen 2021: 5; see also Nguyen 2023). It includes discussions of epistemic injustice (Fricker 2007), propaganda (Stanley 2015), echo chambers (Nguyen 2020), fake news (McBrayer 2020; Rini 2017), and so on. More specifically, this project is in combat epistemology – the study of how nefarious actors intentionally exploit the vulnerabilities of epistemic agents (Nguyen 2021: 4).

epistemic practices. They gather information (both good and bad) and try their best to think critically about it.⁵

An alternative explanation as to why people believe contrarian theories is that they make epistemic mistakes. After all, doing your own research is epistemically risky (Levy 2022). People may be bright and have access to good information, but fall short of the epistemic excellence required to solve some particular problem. It can take teams of trained experts to answer even one small part of a complex question. For example, the PBS documentary “King Arthur’s Lost Kingdom” explores what the Dark Ages looked like in the UK. Answering the question involved multiple teams of archeologists, DNA mapping, literary scholars, and high-energy physics machines. Plenty of intelligent people with generally good epistemic practices wouldn’t be able to gather and process all of that information without making any wrong turns – especially without proper expertise.⁶ These wrong turns can lead to endorsing contrarian theories. Thus, you can end up endorsing a contrarian theory simply because you made an epistemic mistake.

A third, somewhat darker explanation is that there are manipulators who guide epistemic agents into contrarianism for personal gain. In *The Merchants of Doubt*, Oreskes and Conway (2010) explored one way that manipulators can get people to reject the official story. That is, manipulators can manufacture evidence that functions to undercut the official story.⁷ In this paper, I examine an additional method of manipulation which I call the illusion of epistemic trustworthiness. Rather than undercutting existing evidence, this strategy is used by manipulators to become trusted sources of information. The idea that people can use trust as a tool for manipulation is not new. In *The Prince*, Machiavelli suggests that a good prince should display an appropriate amount of trust both to prevent the prince from becoming imprudent and also to not allow “excessive distrust to render him insufferable” (ibid.: 271). Thus, a prince can get others to like him, at least in part, by displaying the appropriate amount of trust toward others. More recently, empirical work has shown that trust can be abused for manipulative purposes (see Forster *et al.* 2016⁸; Williams and Muir 2019⁹). Additionally, Nguyen has also argued that trust is an important factor in the creation and maintenance of echo

⁵This explains why contrarian theorists are often more informed overall than people who don’t engage with the topic at all. After all, there is often no reason for people who believe the official story to interrogate their beliefs, collect more evidence, and so on. They simply trust experts to know what they are doing. Contrarian theorists, on the other hand, want to do the hard work themselves.

⁶Ballantyne (2019) says that expertise is built out of skills and evidence relative to a field. Anyone without the skills and evidence required to engage with a field’s question is not an expert, and thereby in a bad epistemic position to answer the field’s questions.

⁷In the book, examples primarily concerned companies and interest groups funding dubious scientific research. In one of the central case studies of the book, the chairman of R. J. Reynolds and his advisory board allocated 45 million dollars of biomedical research grants and succeeded in mass producing a body of scientific work that could be used to defend the tobacco industry. In a collection of internal documents, they explicitly said that the goal of the program was to develop “an extensive body of scientifically, well-grounded data useful in defending the industry against attacks” (*Merchants of Doubt*: 12: quoting an internal memo from Hobbs to Sticht).

⁸“Agents are subject to persuasion bias and repeatedly communicate with their neighbors in a social network. They can exert effort to manipulate trust in the opinions of others in their favor and update their opinions about some issue of common interest by taking weighted averages of neighbors’ opinions” (Forster *et al.* 2016: 1).

⁹Considers “how elements of communication, such as building rapport and the use of authenticity cues, may be used to invoke trust to effectively deceive others” (Williams and Muir 2019: 1).

chambers (2020); and elsewhere, that a sense of clarity acts as an epistemic litmus test for figuring out when we ought to terminate inquiry – a fact that manipulators can exploit to build trust (2021, 2023).¹⁰ I think each of these manipulation tactics exists.¹¹ They can be performed together or independently. But, importantly, if someone is manipulated by the illusion of epistemic trustworthiness it will impact viable strategies for convincing them to believe the official story.

I take epistemic vice, epistemic mistakes, and epistemic manipulation to each explain the existence of some set of contrarian beliefs. Additionally, I don't take these explanations to be mutually exclusive. Someone is more likely to make epistemic mistakes during research if they have epistemic vices. And a manipulator will likely have an easier time manipulating people who have epistemic vices. Thus these explanations can and will interplay in the real world. However, it is also true that you can make epistemic mistakes during research without epistemic vice playing a large role. Similarly, it is possible to be duped by manipulators without being particularly prone to epistemic vice. Thus, I find it important that the manipulation tactics I describe here don't require the presence of epistemic vice. For that reason, I will focus on how manipulators dupe what I will call *competent epistemic agents*. These agents need not be perfect reasoners (who among us is?). They need not even be particularly good reasoners. Rather, they merely need to lack epistemic flaws that are so dramatic we would call those flaws epistemic vices.¹² In that spirit, the remainder of this paper can be seen as explaining how manipulators can trick competent epistemic agents into endorsing contrarian theories by manufacturing the illusion of epistemic trustworthiness.

3. Epistemic litmus tests and trust

Epistemic agents are persons to whom (1) we can “ascribe knowledge and other epistemic states (such as justified or rational belief)” and (2) who play some role “in acquiring, processing, storing, transmitting, and assessing knowledge” (Goldberg 2021: 19). All actual epistemic agents are limited – both practically and cognitively. We are often not able to dedicate all of our time and attention fully to our epistemic goals. And even if we have that luxury in some cases, no one person could know everything there is to know. There is just too much information out there. In short, we face what Milligram calls *the problem of hyper-specialization* (2015: 2 and 27–44). That is, there is too much difficult epistemic work to do, so we must trust other epistemic agents to do some of that work for us (ibid.). Because of this, we often need to *outsource* some of our epistemic labor (see Levy 2022). We rely on others to gather and assess evidence, store and distribute that evidence, and so on. Thus, a proper assessment of many of our beliefs will include an epistemic assessment of the people we choose to become epistemically dependent upon (Goldberg 2021: 20).¹³

So, we cannot dedicate all of our time and resources to solving our epistemic problems. We need the help of others. However, we also cannot spend all of our time

¹⁰Frost-Arnold (2014) also talks about epistemic tricksters and imposters, and epistemic trustworthiness as an epistemic virtue.

¹¹I.e., manufacturing doubt, displaying the appropriate amount of trust, using clarity to get people to believe things, and the illusion of epistemic trustworthiness.

¹²Thanks to an anonymous reviewer for helping to clarify this point.

¹³Goldberg defines epistemic dependence as follows: “subject S2 is epistemically dependent on another subject S1, then, when an epistemic assessment of S2's belief... requires an epistemic assessment of the role S1 played in the process through which S2 acquired (or sustained) the belief” (2021: 20).

and attention figuring out which other people can help us out epistemically. Thus, we rely on heuristics to figure out who we can offload some of our epistemic labor to. In other words, we use an epistemic litmus test.¹⁴ Typically, epistemic litmus tests are easy and indicative tests for determining whether a belief is true or not. But these litmus tests can serve other epistemic functions as well. For example, according to empirical research, we are more likely to accept an idea as true if it is easy for us to understand (Kahneman 2011: chapter 5; Oppenheimer 2008). Drawing on this phenomenon, Nguyen has argued that a sense of clarity acts as an epistemic litmus test for figuring out when we ought to terminate inquiry (2021: 13). The question I am investigating here is related but importantly different. That is, Nguyen is exploring which things act as litmus tests for terminating inquiry (and how that litmus test can be exploited) (ibid.; 2023). I am discussing an epistemic litmus test for setting up epistemic dependency relations – and describing how this litmus test can be tricked by manipulators. Thus, here we need a test for whether we can responsibly outsource epistemic labor to someone else.

To figure out what a litmus test for epistemic dependency relations would look like, it will be helpful to look at how people set up dependency relations more generally. When people go about their lives they often need to rely on other people to do things for them. We rely on farmers to provide grocery stores with food; we rely on grocery employees to make that food accessible to us; and so on. These are dependency relations. Sometimes how society is set up dictates who we offload labor to (e.g., we don't typically seek out people we can rely on to stock our grocery shelves, we let companies do that for us). Other times, we need to figure out who we can responsibly offload labor to for ourselves (e.g., if I were a manager at a store, I would have the responsibility of finding people who could be relied on to stock shelves). To do this, we look for cues of trustworthiness.

Trust is usually described as a fundamentally three-place relation (Baier 1986; Hawley 2014; Hieronymi 2008; Holton 1994; Jones 1996). There is a trustor (the agent who is doing the trusting), a trustee (the agent who is being trusted), and something the trustor is entrusting to the trustee.¹⁵ To be trustworthy is a normative status of trustees in relation to trustors. Here, I am not interested in the normative question of what makes someone trustworthy.¹⁶ Instead, I am interested in the descriptive question of how people assess trustworthiness (and later, how that mechanism can be tricked). In an influential analysis of the empirical literature on trust, Mayer *et al.* found that a few key factors influence whether a trustor will decide to trust a trustee with something (1995: 717).¹⁷ First of all, as Deutsch (1958) made clear, risk is a central feature of trust. People only need to trust each other when there is some level of risk involved. To see why, imagine the following: you are trying to invest for retirement. Now, think of a world in which there is one clear best investment plan, and all investors always recommend that plan. In this world, there is no reason for you to figure out which investors are trustworthy. That is because there is no risk of getting bad investment advice. No matter who you go to you will get the best investment advice possible. In the real world, however, we need to figure out who is trustworthy because there are

¹⁴I am borrowing the term “epistemic litmus test” from Nguyen (2021).

¹⁵For an argument in favor of viewing trust as a two-place relation, see Domenicucci and Richard (2017).

¹⁶For more on what it takes to be trustworthy, see Hawley (2019).

¹⁷Thanks to an anonymous reviewer for bringing this study to my attention.

often risks associated with either trusting the wrong people or failing to trust at all.¹⁸ Thus, trust is only needed when risk is present.

Second, potential trustors have different inherent propensities to trust (Mayer *et al.* 1995: 715). “Propensity will influence how much trust [a trustor] has for a particular trustee before data on that particular party being available” (ibid.: 715). For example, consider the characters Ted Lasso (from the TV show, *Ted Lasso*) and Lord Voldemort (from the *Harry Potter* series):

High Propensity for Trust: Ted Lasso is an American football coach who gets hired to coach a European football (soccer) team. It is strange for someone who barely knows anything about European football to be given a coaching job at the highest level. But Ted doesn’t consider whether the job offer was given for devious reasons or not. He simply trusts that he was given the job for good reasons and moves to the UK.

Low Propensity for Trust: Lord Voldemort is a powerful evil villain. He is extremely paranoid that someone will try to take his power away. Because of this, he creates a complicated web of safeguards and never tells anyone the full extent of those safeguards. Additionally, he never fully trusts anyone. He always accompanies requests to complete tasks or keep secrets with threats of injury or death upon failure.

These examples show that – before a trustor has any information about a trustee – certain possible trustors will have a natural tendency to trust others. They are naturally trusting. Other possible trustors will only very reluctantly (if ever) place their trust in anyone. They are naturally suspicious. These are two extremes on a spectrum of propensity to trust.

And finally, there are three important factors that trustors look for when trying to determine whether a trustee is trustworthy. These are ability, benevolence, and integrity (ibid.: 717–24). Ability is understood as “that group of skills, competencies, and characteristics that enable a party to have influence within some specific domain” (ibid.: 717). As understood in the context of figuring out who to trust, trustors will look for evidence that the trustee can perform the task entrusted to them. For example, if you needed someone to watch your child for the evening, then you would want evidence that the babysitter could perform that task. Thus, you would likely look for references, experience, and so on. The second factor, benevolence, is understood as “the extent to which a trustee is believed to want to do good to the trustor” (ibid.: 718). Benevolence is important because it would be extremely risky to place your trust in someone who wished you harm. If you were holding onto a rope and needed to rely on someone to pull you up, it would be wise to trust the task to someone who wanted you to survive rather than someone who wanted you dead. Thus, when figuring out who to trust, trustors look for evidence that the trustee is benevolent toward them. Third, integrity can be understood as “the trustor’s perception that the trustee adheres to a set of principles that the trustor finds acceptable” (ibid.: 719). To better understand why integrity is important for developing trust, consider the following situation:

¹⁸Risk assessments will differ between people for a host of reasons. For example, risk assessments in contexts where both positive and negative outcomes are possible will be different than those in contexts where only positive outcomes are possible. For this paper, it is enough to say that an important condition for a trustor to place trust in a trustee is that the trustor thinks that some level of risk is present.

Water Case: You find yourself mildly thirsty. You would like a glass of water but have no way of getting it yourself. There are two people you could trust to get you water, Jenna and Steve. Jenna is willing to walk into town, spend some money, and bring you a bottle of water. Steve, on the other hand, will rob some nearby hapless tourists and bring you their water.

Presumably, you would likely trust Jenna to bring you the water rather than Steve. But why? Both can bring you water and both are benevolent toward you. You could successfully get water by trusting either party. Simply put, it matters to us that we offload labor to people who have moral sensibilities we are okay with. For one thing, we care that the task we are entrusting to someone else is completed morally. But more generally than that, it seems safer to trust people who we know will act according to certain moral norms. There is less chance of mishap.

How does this all connect to our epistemic litmus test? In short, when we need to offload epistemic labor the process will be similar to other cases of offloading. The presence of risk will kickstart our search for someone we can trust to complete the epistemic task for us. Some people (those with a high propensity for trust) will trust others before gathering much evidence of trustworthiness. In many cases, we may call this group of people gullible – and thus if these people come to believe contrarian theories, then the belief can be explained by an epistemic vice.¹⁹ Other people – those I am calling competent epistemic agents – will look for evidence of trustworthiness. That is, they will look for evidence that the trustee has epistemic ability, is benevolent toward them, and shares a sense of moral integrity with them.²⁰ This is where the manipulator will spring their trap.

4. The illusion of epistemic trustworthiness

Much like ordinary epistemic agents, manipulators trade in the acquisition, processing, storage, transmission, and assessment of information.²¹ But epistemic agents try to trade in justified beliefs and knowledge, however, manipulators don't care whether epistemic goods get promoted or not – as long as they profit from the results. In this section, I will explore the nature of one method manipulators use to build trust. I call it, the illusion of epistemic trustworthiness. It functions through two mechanisms: (1) guiding audience research in polluted epistemic environments to seemingly validate sensational or desirable assertions and (2) signaling the possession of certain character traits (related to intelligence, benevolence, and integrity) and credentials. This manipulation begins when a competent epistemic agent is looking for a trustworthy person to help them solve an epistemic problem.

¹⁹In some cases, people may trust without looking for evidence because they have no choice. For example, if your child is about to fall off the edge of a cliff and the only way to save them is to place your trust in a stranger, you may do so without looking for evidence of trustworthiness. In an epistemic case, such conditions may be sufficient to remove the charge of gullibility.

²⁰Perhaps other litmus tests can be employed in different circumstances. But we have good empirical reasons to think that these are the bits of evidence people look for when figuring out who to trust.

²¹Manipulators can either be individuals or groups. For example, the R. J. Reynolds Tobacco Company acted as a manipulative group (Oreskes and Conway 2010). However, individual people can also engage in epistemic manipulation on their own. Thus, I will allow both individuals and groups can be proper subjects of epistemic analysis. In doing so, I am following in the tradition of those who argue that epistemic communities (groups of epistemic agents) are proper units of analysis in and of themselves (Goldman 2011; Lackey 2020).

The first step for the manipulator will be to emphasize that it is risky to not listen to what they have to say. This can be done either implicitly or explicitly, and it can be positively framed or negatively framed. For example, cults will often promise potential members happiness, wealth, knowledge, eternal salvation, or some other set of goods. The risk of not joining, in such cases, is missing out on the promised goods. Other manipulators – like Alex Jones – claim that there are dangerous forces in the world. Manipulators will promise to keep you aware of the danger, or even give you knowledge that can save you from it.²² These are not mutually exclusive ways of indicating risk. But whichever way is employed will serve a similar function. That is, by suggesting that not taking some epistemic problem seriously is risky, the manipulator is giving competent epistemic agents a reason to look for an epistemic trustee. Here, not taking the problem seriously includes both disbelieving or suspending judgment about the risk. The manipulator will make their audience think that taking either of these routes is foolhardy.

Next, the manipulator will broadcast the notion that they are a potential epistemic trustee. This will likely involve the manipulator holding up truth-finding or rationality as their ultimate goal. Of course, genuine epistemic agents and communities do this as well. However, manipulators tend to mimic and over-emphasize this behavior. For example, conspiracy theorist Alex Jones' secondary news site NewsWars at one point had the tagline "Breaking News and Information: a strong bias for telling the truth." Similarly, cult leader Keith Raniere used a "tool" which he called "Rational Inquiry" to brainwash group members.²³ This emphasis on truth and rationality frames the manipulator's goals as being epistemic. In addition, the manipulator might explain why they – and (often) they alone – have the tools required to answer the epistemic questions at hand.²⁴ Surely this won't by itself be enough to dupe a competent epistemic agent. But it might coax them into viewing the manipulator as a possible epistemic trustee. This is when a competent epistemic agent would begin to look for evidence of trustworthiness – i.e., evidence of epistemic ability, benevolence, and moral integrity.

Let us begin with evidence of epistemic ability. Given that we are exploring what competent epistemic agents would do, I will assume that the epistemic agents involved will look for the evidence they ought to look for. Goldman and O'Connor (2021) point us toward four possible types of evidence that one ought to look for to assess epistemic ability. These include trying to see whether an agent's claims cohere with the claims of other trusted sources (coherence), directly verifying an agent's claims (verifiability),

²²For a more specific example, see <https://www.dailystar.co.uk/news/latest-news/mathematician-warns-world-risk-being-31042634>, where Eric Weinstein explains why people should be much more scared of the world ending in nuclear disaster than the currently are. Other examples might include knowing the word of God bringing you salvation, knowing about gold and crypto-currency helping you survive the collapse of civilization, and so on.

²³Other examples include: the Flat Earth Society's about page says that they are "standing with reason we offer a home to those wayward thinkers that march bravely on with REASON and TRUTH in recognizing the TRUE shape of the Earth – Flat" (capitalization original to the text); Bret Weinstein, a podcaster who has promoted COVID conspiracies and anti-vaccine rhetoric, says that on his podcast "we will explore questions that matter, with tools that work"; Donald Trump's social media site is called "Truth Social"; and Russel Brand created a news station on YouTube called "the Truws."

²⁴Recent empirical work has suggested that perceived access to secret knowledge (or secret knowledge-finding practices) is part of the draw in contrarian theorizing – particularly conspiracy theorizing and cultish beliefs (Imhoff and Lamberty 2017; Sternisko *et al.* 2020). For a philosophical analysis on the role of "fantasies of secret knowledge" in cults and conspiracist groups, see Munro (forthcoming).

identifying whether an agent seems generally knowledgeable (intelligence), and identifying relevant credentials (credentials).²⁵ Each of these counts as a type of evidence that someone has epistemic ability. Epistemic agents will assign different weights to the types of evidence that seem important in different cases.²⁶ For example, if you are trying to get directions to the nearest coffee shop then it would be sufficient to identify someone who seems knowledgeable about the surrounding area. If you are trying to determine whether climate change is happening, on the other hand, then you ought to seek out someone with the relevant credentials.

So, competent epistemic agents will look for verifiability, coherence, intelligence, and credentials as types of evidence pointing toward epistemic ability. Importantly, just because these agents look for the types of evidence they ought to doesn't mean they will evaluate that evidence perfectly. We are talking about competent agents, not ideal ones. And manipulators will seek to take advantage of that fact. To do so, manipulators will set themselves up in polluted epistemic environments.²⁷ Levy (2021) suggests that polluted epistemic environments are disadvantageous places to engage in epistemic work because they are full of misinformation. Misinformation is best understood simply as false information. As McBrayer puts it, "sometimes, misinformation is the result of bad actors (as in the case of propaganda), sometimes it's the result of negligence (like homemade coronavirus cures)" (2021: 3).²⁸ Setting themselves up in polluted epistemic environments gives manipulators two advantages. First, it will be harder for sincere epistemic agents to sort the good evidence from the bad evidence. Second, it will be easier for the manipulator to manufacture the illusion that they have epistemic ability. This is because the manipulators can guide audience research in those environments toward evidence – either planted or intentionally selected – that seemingly validates sensational or desirable assertions.

To see how, consider the example of Alex Jones. Jones has largely created his own polluted epistemic environment. He runs a radio show called InfoWars which invites questionable but (often) well-credentialed guests on to discuss controversial topics. He also runs other alternative news sites, such as [NewsWars.com](https://www.infowars.com/). Thus, he can always

²⁵Goldman and O'Connor (2021) frame these as possible methods for identifying experts. I would suggest that these are better understood as types of evidence that feed into our epistemic litmus test for who it is permissible to outsource epistemic labor too.

²⁶How much evidence we try to gather about the reliability of other people will often depend on the level of risk involved. For example, trying to find the nearest coffee shop is often not a very risky activity. We can afford to get the answer wrong. Therefore, if we are attempting to locate the nearest coffee shop we are likely to trust the testimony of someone who looks like they may know, without gathering very much evidence about the person. Deciding how to invest for retirement, on the other hand, is very risky. We typically cannot afford to arrive at an incorrect answer. So we will likely do a lot more work gathering evidence about whether we can trust the people giving us investing advice. In each case, we are still trying to find a reliable person to listen to. However, we are willing to accept less evidence of reliability in situations with low risk.

²⁷Generally, epistemic environments are environments in which epistemic communities and agents acquire, process, store, transmit, and assess knowledge (see Goldberg 2021). For example, if I as an epistemic agent wanted to find out "who won the World Series in 1963?" I could investigate that question in different environments. I could investigate in a library, or I could investigate on the internet. The environment in which I investigate the question will dictate the strategies I use to acquire the information. It would also provide different advantages and disadvantages to me as an epistemic agent.

²⁸One type of misinformation is fake news. Fake news "purports to describe events in the real world, typically by mimicking the conventions of traditional media reportage, yet is known by its creators to be significantly false, and is transmitted with the two goals of being widely re-transmitted and of deceiving at least some of its audience" (Rini 2017: E-45).

direct the audience to sources of information that he himself generates. In addition, Jones has surrounded himself with dubious sources run by people who support his project. These sources include [ThePeoplesVoice.tv](https://thepeoplesvoice.tv), [NewsPunch.com](https://news-punch.com), [NaturalNews.com](https://naturalnews.com), [NoMoreFakeNews.com](https://nomorefakenews.com), and so on. I have chosen two examples where Jones quite clearly displayed the technique of directing audience research within a polluted environment.

Alex Jones Gay Bombs: In 2015, Jones claimed that the U.S. government has used bombs that turn people gay “on our troops, in Vietnam... and in Iraq.” He also claimed that “[the U.S. government sprayed PCP on the troops] and that “they give the troops special vaccines that are really nano-tech that already reengineer their brains.” However, Jones put particular emphasis on the “gay bombs,” saying “**if you’re a new listener just type in ‘pentagon tested gay bomb’**” (bold added for emphasis). (InfoWars, October 16, 2015)

Alex Jones World Economic Forum: On August 19th, 2022 Jones claimed that the World Economic Forum had hired over 110,000 information warriors to control the online narrative and take down InfoWars. After introducing the story, Jones directed the listener to research the issue for themselves. In this case, he told people both to look at normally trusted sources, but also to check out a YouTube video by a channel called ThePeoplesVoice, and an article run by [NewsPunch.com](https://news-punch.com).²⁹

How do these stories make newcomers think Jones has epistemic ability? Well, someone may do as he recommends and follow up by fact-checking him. In other words, they will engage in what Levy calls *shallow research* (2022). “Shallow research consists in the consultation of sources we have good reason to regard as reliable and which are aimed at non-experts like us. We engage in shallow research by reading mainstream media, trade books and the like, attending public lectures and so on” (ibid.: 6). This will likely manifest as the new listener searching “Pentagon tested gay bomb” and “WEF information warriors” on a web browser. If they were to do so, they would see articles in the Guardian, the British Medical Journal, the New Scientist, and the BBC all confirming that the Pentagon did try to create a bomb that would turn enemy soldiers gay.³⁰ And they would see the World Economic Forum (WEF) did have an initiative to combat misinformation on the internet.³¹ These examples show a manipulator using a polluted epistemic environment to establish coherence. That is, Jones’ claims

²⁹The “WEF information warriors” story can be found at <https://thepeoplesvoice.tv/klaus-schwab-hires-millions-of-information-warriors-to-seize-control-of-the-internet/> and the listing of News Punch and The Peoples Voice as disinformation websites can be found at <https://www.factcheck.org/2017/07/websites-post-fake-satirical-stories/> under the heading News Punch. Acknowledgments to the podcast Knowledge Fight (episode #719) for covering this episode of InfoWars.

³⁰The Guardian wrote a piece titled “Air Force Looked at Spray to Turn Enemy Gay.” The BMJ wrote a piece titled “Gay Bomb and BMJ authors win prizes.” The New Scientist wrote a piece titled “Military Wins Ig Nobel Peace Prize for ‘Gay Bomb’.” And the BBC wrote a piece titled “the U.S. Military Pondered Love, Not War.”

Here, even a somewhat savvy internet user could be bamboozled. This is because, even though the trustworthy sites do not confirm Jones’ craziest claims – e.g., that the U.S. government has used bombs that turn people gay “on our troops, in Vietnam... and in Iraq” – they still confirm a pretty outlandish claim: that the U.S. government spent time and resources trying to develop a so-called “gay bomb.”

³¹<https://www.weforum.org/projects/combating-covid-misinformation>.

seemingly cohere with other independent sources turned up by their shallow research. I say “seemingly” because the exact claims – gay bombs were developed and used; the WEF hired misinformation warriors to take Jones down – do not cohere, but some nearby claims do. This seeming coherence might get a newcomer to InfoWars to walk away with some evidence that Jones has epistemic ability. Additionally, the second example also shows the manipulator beginning to expand their audience’s network of trusted sources to include those run by the manipulator himself (and his associates).

Coherence is just one type of evidence that people look for to establish epistemic ability. Another type of evidence is verifiability. To see how manipulators can use polluted epistemic environments to establish verifiability, consider the following example:

Flat Earth Society: The Flat Earth Society has forums dedicated to helping people work out various problems on their own. One of these forums includes some mathematical calculations meant to show that if the earth were round, then you would be unable to take a photograph of Chicago from the coast of Michigan. The forum explains how to work through the math, and then they show a photograph taken of Chicago from the coast of Michigan. Thus, they seem to show that the earth isn’t round.

This example shows how in polluted epistemic environments, manipulators can create puzzles for epistemic agents to work through and verify for themselves. In this case, the environment is polluted in ways that get people to endorse the following conditional: if the earth were round, then you would be unable to take a photograph of Chicago from the coast of Michigan. People are then asked (and sometimes taught how to in workshops) to work through the problem themselves. Upon verifying the math, the agent can then verify that the photograph is possible – either by accepting the photograph provided or going to take one themselves. This example shows how manipulators can create puzzles and plant evidence in ways that make people feel as though they are independently verifying answers on their own. This, in turn, acts as evidence for the epistemic agent that the manipulator has epistemically ability. Manipulators are often in a better position than traditional media to capitalize on verifiability as evidence of epistemic ability. This is because listening to the media is often not an interactive enterprise, but manipulation can be.

In both the Alex Jones and Flat Earth Society examples, the manipulators directed audience research in polluted epistemic environments to create evidence of epistemic ability (i.e., coherence and verifiability). An additional similarity between these cases is that the claims being made were sensational (the claims could also have been desirable). This might initially seem bad for the manipulator, as it has the potential to drive people away. But the sensational nature of these claims is features of the manipulator’s strategy, not bugs. Here are two reasons to think this. First, people tend to find novel, complex, and comprehensible claims interesting (Silvia 2005). Additionally, interest motivates people to engage further with the interesting claims (ibid.). The sensational claims made by manipulators are certainly novel, complex, and comprehensible. Thus, these claims will likely grab the interest of epistemic agents and thereby motivate them to engage with the manipulator more in the future. Second, people are more likely to remember when someone gets something shocking correct (e.g., learning the USA attempted to develop gay bombs) than when they get trivial things correct (e.g., the weather report from three days ago). Thus, getting shocking claims correct is more likely to get an epistemic agent to remember the manipulator as a possibly trustworthy

source of information. Thus, a manipulator can build evidence of epistemic ability by guiding audience research in polluted epistemic environments to seemingly validate sensational or desirable assertions. This strategy, deployed over and over, has resulted in people saying things like “we all know that [Jones has messed] some things up, right? But [Jones has] gotten so many things right” (Washington Post quoting Joe Rogan). In other words, some people have come to view Jones as an epistemic source worth listening to.

In addition to these techniques, manipulators will signal to their audience that they are intelligent and properly credentialed. Evidence of intelligence can be built in many ways. One way is intellectual virtue signaling (Levy 2023). Intellectual virtues are character traits that are helpful to possess when doing epistemic work (Roberts and Wood 2007). But, as Levy points out, intellectual virtue signaling is often not about actually possessing intellectual virtues (2023). Rather, it is about signaling “characteristics that other people will value” (ibid.: 311). For example, “it is conceivable that some individuals might attract attention by signaling intellectual virtues like empathy and humility,” but “in practice these virtues rarely do well” at getting others to view you as intelligent (ibid.: 315). The intellectual virtues that manipulators are likely to signal include quickness of mind (e.g., the ability to use language well, speak coherently about any topic, and so on), intellectual autonomy (e.g., the ability to think for oneself and not rely on others), and intellectual courage (e.g., the willingness to offer contrarian views) (ibid.: 316).³² Other techniques include manipulators promoting each other as intellectual exemplars³³; using traditional signals of intelligence like using classical music, dressing like stereotypical intellectuals³⁴; and so on. Competent epistemic agents will see these signals of intelligence as evidence of epistemic ability.

Finally, manipulators will also signal to their audience that they have the credentials necessary to answer some set of questions. Here, two strategies can be employed. The first is to inflate the credentials that the manipulator does possess. To see how this could be done, consider the following example:

Credential Inflation: Robert Malone is an M.D. who worked on mRNA vaccine technology during his early career. Malone has since claimed to have invented mRNA technology and has used that badge to promote vaccine skepticism on platforms like the Joe Rogan show. This claim is contrary to other experts in the field, such as Rein Verbeke. Verbeke spoke to the Atlantic about these claims, stating Malone and his co-authors “sparked for the first time the hope that mRNA could have potential as a new drug class” but “the achievement of the mRNA vaccines of today is the accomplishment of a lot of collaborative efforts.”³⁵

³²For empirical work verifying similar claims, see Williams and Muir (2019).

³³E.g., Eric Weinstein saying the Bret Weinstein and Heather Hays should have both earned Nobel Prizes.

³⁴E.g., Eric Weinstein said “you’ll notice that I almost always wear a jacket because I am the establishment in waiting not, you know, the sort of rebels living in the trees, enjoying terrorism and calling it freedom fighting. It requires an incredible amount of discipline to do this...” in an interview on Rebel Wisdom with David Fuller (acknowledgments to the “Decoding the Guru’s” Podcast, episode from August 27, 2021). Perhaps competent epistemic agents ought to merely find this funny. But it does show that Eric is attempting to signal to his audience that he is intelligent and ready to run the world.

³⁵See “The Vaccine Scientist Spreading Misinformation” in *The Atlantic* at <https://www.theatlantic.com/science/archive/2021/08/robert-malone-vaccine-inventor-vaccine-skeptic/619734/>.

In this example, Malone has inflated his already impressive credentials to seem better than the normal experts in academia and government.³⁶ Competent epistemic agents could look up Malone, see that he was involved in mRNA technology, and believe these inflated claims. This would count as evidence that Malone has the epistemic ability required to make definitive claims on COVID-19 vaccines.

The second strategy is to downplay the need for credentials to answer some set of questions. For example:

Credential Deflation: the first rule of the forum website called the Ornerly American – run by the author Orson Scott Card – states “we aren’t impressed by your credentials, Dr. This or Senator That. We aren’t going to take your word for it, we’re going to think it through for ourselves.”³⁷

This credential deflation signals to an audience that credentials won’t help with solving the important epistemic problems at hand. This may not work for people who place a high premium on expertise, but it could work perfectly well on merely competent epistemic agents. Thus, manipulators can get competent epistemic agents to accept that the manipulator has the proper credentials to answer some questions by either inflating their credentials or deflating the need for credentials at all. In either case, setting expectations around credentials and then meeting those expectations could look like evidence of epistemic ability to a competent epistemic agent.³⁸

In addition to evidence of epistemic ability, epistemic agents will also look for evidence of benevolence and integrity before placing their trust in a manipulator. Thus, manipulators will also employ techniques to make themselves look benevolent toward their audience. Displays of benevolence can take different forms. One common technique is called *love bombing*. Love bombing occurs when a manipulator makes exaggerated displays of attention and affection. These overt displays are used to make the victim feel a loving connection quickly. For example, Tourish and Vatcha (2005) say of love bombing (in the context of cults):

Love Bombing: “One of the most commonly cited cult recruitment techniques is generally known as ‘love bombing’ (Hassan, 1988). Prospective recruits are showered with attention, which expands to affection and then often grows into a plausible simulation of love. This is the courtship phase of the recruitment ritual. The leader wishes to seduce the new recruit into the organization’s embrace, slowly habituating them to its strange rituals and complex belief systems.” (Tourish and Vatcha 2005: 17)

Manipulators in our epistemic sense will use love bombing to gain and maintain an audience. Common tropes in the epistemic domain include manipulators saying that their audience is more intelligent than other people, or that they are the only ones

³⁶Another example is Eric Weinstein (PhD in mathematical physics, podcast host, and ex-managing director of Thiel Capital), who has claimed to have developed a theory of everything called “geometric unity.” He has said that this discovery makes him deserving of a Nobel Prize. While this example may not work well on people who know how academia works, it may work on people who don’t (i.e., merely competent epistemic agents). This is evidenced by the fact that Eric’s podcast has quite a substantial following.

³⁷See <http://www.ornery.org>.

³⁸Another way people can impact credential evaluations is by using expertise from one area to make themselves look like experts in another area. For more on this, see Ballantyne (2019).

who can see the truth. In problematic cases, this love bombing will then shift into abusive and controlling behavior. This can include any range of behaviors – e.g., financial or sexual exploitation, abuse, neglect, misleading for personal or political gain, and so on. Often in epistemic settings, the abuse involves financial exploitation without returning anything of genuine epistemic value. A second common technique for broadcasting benevolence involves the manipulator signaling to their audience that they (the manipulator) are actively working to protect them (the audience) from harm. This technique is used explicitly by Alex Jones, who often declares that he is fighting a war of information to protect his audience. By using techniques like these, manipulators can project an image of benevolence toward their audiences.

Finally, manipulators will employ techniques to make themselves appear to possess moral integrity. One strategy for doing this involves the manipulator broadcasting a set of moral sensibilities that they think their audience will agree with. Thus, you will see manipulators taking on populist interests and aligning themselves with political groups they think will appeal to an audience. For example, Alex Jones often makes partisan political statements such as emphasizing his love of the Second Amendment. An additional technique involves manipulators telling stories that paint themselves as possessing character traits that are associated with moral integrity. People are wired to identify moral character traits in others (see Uhlmann *et al.* 2015). We assess whether others have character traits by examining their actions (*ibid.*). Some actions are more communicative than others (*ibid.*).

Acts That Show Integrity: “More generally, acts that can be attributed to multiple plausible motives or causes (i.e., are high in attributional ambiguity; Snyder, Kleck, Strenta, & Mentzer, 1979) tend to be seen as low in informational value. In contrast, behaviors that are statistically rare or otherwise extreme are perceived as highly informative about character traits (Ditto & Jemmott, 1989; Fiske, 1980; Kelley, 1967; McKenzie & Mikkelsen, 2007). In addition, decisions that are taken quickly and easily (Critcher, Inbar, & Pizarro, 2013; Tetlock, Kristel, Elson, Green, & Lerner, 2000; Verplaetse, Vanneste, & Braeckman, 2007), that are accompanied by genuine emotions (Trivers, 1971), and that involve costs for the decision maker (Ohtsubo & Watanabe, 2008) are perceived as especially informative about character.” (*ibid.*: 74)

Thus, manipulators will likely tell stories involving themselves performing statistically rare or otherwise extreme acts that came at a personal cost. These stories will be accompanied by grand displays of emotion – e.g., warmth and compassion toward the vulnerable, great sadness for the existence of evil, anger, and vengefulness toward wrongdoing, and so on. Competent epistemic agents will take these stories as signals of moral integrity – and thus be more willing to place their trust in the manipulator.

So, because of a perceived risk competent epistemic agents will look to responsibly offload epistemic labor. Manipulators will set themselves up as possible sources of epistemic information, and manufacture evidence of epistemic ability, benevolence toward their audience, and moral integrity. Competent epistemic agents will look for evidence of these factors when trying to determine to whom they can responsibly offload epistemic labor. They will find the manufactured evidence and come to view the manipulator as a trustworthy source of information. This is how the illusion of epistemic trustworthiness is built. Importantly, in this story, competent epistemic agents are largely doing what they ought to do when seeking to offload epistemic labor. And

they can be duped in this way without the presence of significant epistemic vice or epistemic mistakes. Once a manipulator has duped an epistemic agent in this way, that agent will be willing to endorse other contrarian claims made by the manipulator. And the manipulator can exploit the agent more easily for personal gain. For example, by employing these tactics, Alex Jones can afford to spend nearly \$100k a month (presumably, largely from income from InfoWars) without returning any genuine epistemic services to his audience.³⁹ This is largely what makes the use of these tactics so manipulative – the willingness and active attempt to sacrifice epistemic goods for personal gain.

5. Fact-checking vs. undercutting epistemic trust

Epistemic institutions – such as universities and news outlets – have noticed the widespread nature of contrarian beliefs. Many of these institutions have begun trying to correct these beliefs. A common strategy for doing so is the practice of fact-checking. AP News has created a section of its website called “AP Fact Check.” The goal of this site is to evaluate and discredit misinformation (e.g., contrarian theories). Other news sites have implemented similar practices.⁴⁰ Intuitively – and from personal experience – fact-checking can work to combat contrarian beliefs. For a long time, I believed that vitamin C helped combat the common cold. It was only when I encountered someone fact-checking that claim that I changed my mind. As it turns out, there is “no consistent effect of vitamin C was seen on the duration or severity of colds in the therapeutic trials” (Hemilä and Chalker 2013). There is empirical evidence that backs up this personal anecdote. For example, Swire *et al.* (2017) showed that fact-checking can be effective at changing the strength of people’s contrarian beliefs.⁴¹ However, as stated at the beginning of this paper, belief in contrarian theories can be explained in (at least) a few different ways. These beliefs could be held because of epistemic vice, epistemic mistake, or epistemic manipulation. It seems plausible, *prima facie*, that how people come to hold contrarian theories will affect strategies for correcting those beliefs. For example, if you know that someone believes a contrarian theory because they made an epistemic mistake, correcting that belief might be as simple as doing some fact-checking. If someone believes a contrarian theory because of epistemic manipulation, the correction might not be that simple. In this section, I will argue that – while fact-checking can be a worthwhile endeavor – it will often be an ineffective strategy against combating contrarian beliefs held because of epistemic manipulation. Instead, I suggest we ought to try engaging in the practice of trust undercutting.

The notion that fact-checking is sometimes ineffective in correcting contrarian beliefs has been noted by both philosophers (Nguyen 2020, 2023; Novaes 2020) and scientists (Hart and Nisbet 2012; Nyhan and Reifler 2010; Nyhan *et al.* 2014). For example, Nyhan and Reifler (2010) showed that attempts to correct the beliefs of ideological groups often fail. Even worse, their results found that correction attempts could “backfire,” causing the ideological group to believe the corrected claims even more strongly (*ibid.*). The illusion of epistemic trustworthiness can give us an explanation

³⁹<https://abcnews.go.com/US/wireStory/alex-jones-spent-93000-july-sandy-hook-families-103186219>.

⁴⁰See the BBC’s “Reality Check,” CNN’s “Facts First,” and so on.

⁴¹In the study, participants were shown a series of both true and false statements that Donald Trump made during his campaign. People were asked to rate their confidence in each of the statements. Next, the experimenters let the participants know which statements were false. As a result, they found that all participants believed the false statements less after correction – including trump supporters. Thus, it seems that fact-checking can be an effective way to fight contrarian beliefs.

as to why this is – and suggest a possible alternative method of correction in cases of epistemic manipulation. In cases of epistemic manipulation, fact-checking is likely to fail to correct contrarian beliefs because the trustor's trust in the manipulator gives the trustor a reason to doubt the fact-checking source. To see why, consider the following two cases:

Fact-Checking the BBC: Assume that you have come to trust the BBC. You read a BBC article one morning about Taylor Swift breaking records during the Grammy nomination process.⁴² You come to believe that these events have transpired. Later, you see an article on Facebook claiming that Taylor Swift hadn't broken any records.

In this case, seeing an article on Facebook will not affect your belief that Taylor Swift broke records. This is because you have come to trust the BBC enough to form your beliefs based on their journalism. Additionally, however, this seems to give you a reason to actively distrust the random source you saw on Facebook. This source is telling you the opposite of something you know to be true. The same thing will happen to people who have come to trust manipulators. For example:

Fact-Checking InfoWars: Beth has come to trust InfoWars. She sees an article saying Grammy's lied about Taylor Swift breaking records to further the feminist movement. Later, Beth sees an article by the BBC saying Taylor Swift has broken records during the Grammy nomination process.

In this case, Beth seeing the BBC article will not affect her belief. This is because she has come to trust InfoWars enough to form her beliefs based on their journalism. Additionally, the BBC contradicting Alex Jones gives Beth a reason to distrust the BBC. They are, after all, telling her the opposite of something she believes very strongly. Again, this tracks with empirical work in this area. Perhaps unsurprisingly, "the persuasiveness of a message increases with the communicator's perceived credibility and expertise" (Lewandowsky *et al.* 2012). Additionally, Walter and Tukachinsky (2020) found that "corrections [of misinformation] are less effective if the misinformation was attributed to a credible source."⁴³ That is, if someone perceives a manipulator to be a credible source, their message is likely to be believed over and above competing sources.⁴⁴ Thus, the illusion of epistemic trustworthiness can explain – at least in some cases – why fact-checking fails to correct beliefs or backfires.

As we can see from the above examples, once someone trusts a source of information they will view sources offering conflicting views skeptically, or even come to distrust those alternative sources altogether. In cases like these, fact-checking likely won't be successful. Here we need an alternative practice for combatting contrarian beliefs. One possible strategy is to undercut the trustor's trust in the manipulator. If someone is a competent epistemic agent, then they will have formed their trust in InfoWars by looking for evidence of epistemic ability, benevolence toward their audience, and moral integrity. Thus, undercutting trust in a manipulator will involve undercutting evidence

⁴²<https://www.bbc.com/news/entertainment-arts-67381339>.

⁴³Thanks to an anonymous reviewer for recommending these two sources.

⁴⁴A similar point was also made by Novaes (2020).

of these factors. This strategy tracks well with some suggested corrections in the empirical literature. For example:

Credibility Corrections: “Corrections should criticize the credibility of the source of the misinformation. This serves two functions. First, source credibility is central to processing the initial (mis)information but not for the correction source. Thus, trying to undo the damage done by climate change deniers and vaccine skeptics with messages that rely, primarily, on the expertise of their sources is likely to be futile. Instead, the correction should focus on discrediting the sources of misinformation. For example, rather than emphasizing the knowledge of a climate science expert, messages should highlight the lack of expertise and relevant training of climate change skeptics. Second, questioning the credibility of the misinformation source can enhance the coherence of the corrective message. Put differently, discrediting the source as biased and lacking goodwill can explain the spread of the misinformation and make it easier for message consumers to maintain a coherent mental model that dismisses the misinformation.” (Walter and Tukachinsky 2020)

To make trust undercutting more concrete, consider the case of Beth once more. She trusts InfoWars – and this explains her contrarian belief. Fact-checking is likely to either not work or backfire. Instead, it may be prudent to undercut her trust in InfoWars. This could involve undercutting Beth’s perceptions of Jones’ epistemic ability – e.g., showing that Jones’ claims are incoherent, can’t be independently verified, or that he lacks intelligence and doesn’t have the proper credentials. However, these epistemic corrections are likely to be extremely difficult to implement. Perhaps a better starting point would be to show Beth that Jones is not benevolent toward his audience, or that he lacks moral integrity. For if Beth conforms to general trends, she would not trust someone she views as malevolent or lacking moral integrity. These are possibly easier points to challenge than unraveling Jones’ entire worldview (which Beth has likely adopted). Of course, this correction is likely to be very difficult. And it is not guaranteed to result in competent epistemic agents coming to believe the official story. Rather, successful implementation would leave Beth freed from her trust in a manipulator. Hopefully, this would allow her to find better sources to place her trust in. While the remaining difficulties are large, there was never likely to be a magic bullet for correcting contrarian beliefs formed by manipulative means. At the very least, trying to help people stop trusting epistemic manipulators seems like a good place to start.

6. Conclusion

One explanation for contrarian beliefs is epistemic manipulation. In this paper, I showed one mechanism by which manipulators can get (even competent) epistemic agents to endorse contrarian theories. I call this mechanism the illusion of epistemic trustworthiness. Manipulators build the illusion of epistemic trustworthiness by manufacturing evidence of epistemic ability, benevolence toward their audience, and moral integrity. By manufacturing this evidence, manipulators can get epistemic agents to view the manipulator as a trustworthy source of information. When an epistemic agent views a manipulator as trustworthy, fact-checking will be an ineffective way of combating the agent’s contrarian beliefs. Instead, we ought to engage in the practice of trust undercutting. This involves undercutting evidence of epistemic ability,

benevolence, and integrity. Hopefully, this strategy will allow the epistemic agent to be more open to finding better epistemic sources to place their trust in.⁴⁵

References

- Bader C., Day L.E. and Gordon A. (2016). Chapman Survey of American Fears, Wave 3. <https://doi.org/10.17605/OSF.IO/MJ4ET>
- Baier A. (1986). 'Trust and Antitrust.' *Ethics* 96(2), 231–60. <https://doi.org/10.1086/292745>.
- Ballantyne N. (2019). 'Epistemic Trespassing.' *Mind* 128(510), 367–95.
- Cassam Q. (2016). 'Vice Epistemology.' *The Monist* 99(2), 159–80.
- Cassam Q. (2019). *Conspiracy Theories*. Bristol: Polity Press.
- Critcher C.R., Inbar, Y. and Pizarro D.A. (2013). 'How Quick Decisions Illuminate Moral Character.' *Social Psychological and Personality Science* 4(3), 308–315. <https://doi.org/10.1177/1948550612457688>
- Coady D. (2003). 'Conspiracy Theories and Official Stories.' *International Journal of Applied Philosophy*, 17(2), 197–209.
- Deutsch M. (1958). 'Trust and suspicion.' *Journal of Conflict Resolution* 2(4), 265–279. <https://doi.org/10.1177/002200275800200401>
- Ditto P.H. and Jemmott J.B. (1989). 'From rarity to evaluative extremity: Effects of prevalence information on evaluations of positive and negative characteristics.' *Journal of Personality and Social Psychology* 57(1), 16–26. <https://doi.org/10.1037/0022-3514.57.1.16>
- Domenicucci J. and Richard H. (2017). In P. Faulkner and T. Simpson (eds), *The Philosophy of Trust*, pp. 150–161. Oxford: Oxford University Press.
- Fiske S.T. (1980). 'Attention and weight in person perception: The impact of negative and extreme behavior.' *Journal of Personality and Social Psychology* 38(6), 889–906. <https://doi.org/10.1037/0022-3514.38.6.889>
- Forster M., Mauleon A. and Vannetelbosch V. (2016). 'Trust and Manipulation in Social Networks.' *Network Science* 4(2), 216–43. <https://doi.org/10.1017/nws.2015.34>.
- Fricker M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.
- Frost-Arnold K. (2014). 'Imposters, Tricksters, & Trustworthiness as an Epistemic Virtue.' *Hypatia* 29(4), 790–807.
- Goldberg S. (2021). *Foundations and Applications of Social Epistemology*. Oxford: Oxford University Press.
- Goldman A. (2011). A guide to social epistemology. In Alvin I. Goldman & Dennis Whitcomb (eds.), *Social Epistemology: Essential Readings*. New York: Oxford University Press. pp. 11–37.
- Goldman A. and O'Connor C. (2021). 'Social Epistemology'. In E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. §3.6. Stanford: The Metaphysics Research Lab. <https://plato.stanford.edu/archives/win2021/entries/epistemology-social/>
- Harris K. (2018). 'What's Epistemically Wrong with Conspiracy Theorizing?' *Royal Institute of Philosophy Supplement* 84, 235–57. <https://doi.org/10.1017/s1358246118000619>.
- Hart P.S. and Nisbet E.C. (2012). 'Boomerang Effects in Science Communication: How Motivated Reasoning and Identity Cues Amplify Opinion Polarization about Climate Mitigation Policies.' *Communication Research* 39, 701–23.
- Hassan S. (1988). *Combating Cult Mind Control*. Rochester, Vermont: Park Street Press.
- Hawley K. (2014). 'Trust, Distrust, and Commitment.' *Noûs* 48(1), 1–20.
- Hawley K. (2019). *How To Be Trustworthy*. New York, NY: Oxford University Press.
- Hemilä H. and Chalker E. (2013). 'Vitamin C for Preventing and Treating the Common Cold.' *Cochrane Database of Systematic Reviews*, Issue 1, CD000980. <https://doi.org/10.1002/14651858.CD000980.pub4>.
- Hieronymi P. (2008). 'The Reasons of Trust.' *Australasian Journal of Philosophy* 86(2), 213–36. <https://doi.org/10.1080/00048400801886496>.
- Holton R. (1994). 'Deciding to Trust, Coming to Believe.' *Australasian Journal of Philosophy* 72(1), 63–76. <https://doi.org/10.1080/00048409412345881>.

⁴⁵I would like to thank Amir Ajalloeian, Bret Donnelly, Julia Staffel, Jonathyn Zapf, commentators and audience members at the Northwest Philosophy Conference, Notre Dame/Northwestern Graduate Epistemology Conference, and the European Congress for Analytic Philosophy, and an anonymous reviewer for all the comments and aid that made this paper possible.

- Imhoff R. and Lamberty P.K.** (2017). 'To Special to be Duped: Need for Uniqueness Motivates Conspiracy Beliefs.' *European Journal of Social Psychology* **47**(6), 724–34.
- Jones K.** (1996). 'Trust as an Affective Attitude.' *Ethics* **107**(1), 4–25.
- Kahan D.M.** (2015). 'Climate–Science Communication and the Measurement Problem.' *Political Psychology* **36**(s1), 1–43.
- Kahneman D.** (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus, and Giroux.
- Kelley H.H.** (1967). 'Attribution theory in social psychology.' *Nebraska Symposium on Motivation* **15**, 192–238.
- Kidd I.J., Cassam Q. and Battaly H.** (2020). *Vice Epistemology*. London: Routledge.
- Klein C., Clutton P. and Dunn A.G.** (2019). 'Pathways to Conspiracy: The Social and Linguistic Precursors of Involvement in Reddit's Conspiracy Theory Forum.' *PLoS ONE* **14**(11), e0225098. <https://doi.org/10.1371/journal.pone.0225098>.
- Lackey J.** (2020). *The Epistemology of Groups*. New York, NY: Oxford University Press.
- Lee C., Yang T., Inchoco G.D., Jones G.M. and Satyanarayan A.** (2021). 'Viral Visualizations: How Coronavirus Skeptics Use Orthodox Data Practices to Promote Unorthodox Science Online.' In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1–18). Association for Computing Machinery. Retrieved July 27, 2021, from <https://doi.org/10.1145/3411764.3445211>.
- Levy N.** (2021). *Bad Beliefs: Why They Happen to Good People*. Oxford: Oxford University Press.
- Levy N.** (2022). 'Do Your Own Research!.' *Synthese* **200**(5), 1–19.
- Levy N.** (2023). 'Intellectual Virtue Signaling.' *American Philosophical Quarterly* **60**(3), 311–24.
- Lewandowsky S., Ecker U.K.H., Seifert C.M., Schwarz N. and Cook J.** (2012). 'Misinformation and Its Correction: Continued Influence and Successful Debiasing.' *Psychological Science in the Public Interest* **13**(3), 106–31. <https://doi-org.colorado.idm.oclc.org/10.1177/1529100612451018>.
- Mayer R.C., Davis J.H. and Schoorman F.D.** (1995). 'An Integrative Model of Organizational Trust.' *Academy of Management Review* **20**(3), 709–34.
- McBrayer J.P.** (2020). *Beyond Fake News: Finding the Truth in a World of Misinformation*. New York, NY, USA: Routledge.
- McKenzie C.R.M. and Mikkelsen L.A.** (2007). 'A Bayesian view of covariation assessment.' *Cognitive Psychology* **54**(1), 33–61. <https://doi.org/10.1016/j.cogpsych.2006.04.004>
- Millgram E.** (2015). *The Great Endarkenment: Philosophy for An Age of Hyperspecialization*. Oxford: Oxford University Press.
- Munro D.** (forthcoming). 'Cults, Conspiracies, and Fantasies of Knowledge.' *Episteme*, 1–22.
- Nguyen C.T.** (2020). 'Echo Chambers and Epistemic Bubbles.' *Episteme* **17**(2), 141–61. <https://doi.org/10.1017/epi.2018.32>.
- Nguyen C.T.** (2021). 'The Seductions of Clarity.' *Royal Institute of Philosophy Supplement* **89**, 227–55.
- Nguyen C.T.** (2023). 'Hostile Epistemology.' *Social Philosophy Today* **39**, 9–32.
- Novaes C.D.** (2020). 'The Role of Trust in Argumentation.' *Informal Logic* **40**(2), 205–36.
- Nyhan B. and Reifler J.** (2010). 'When Corrections Fail: The Persistence of Political Misperceptions.' *Political Behavior* **32**, 303–30.
- Nyhan B., Reifler J., Richey S. and Freed G.L.** (2014). 'Effective Messages in Vaccine Promotion: A Randomized Trial.' *Pediatrics* **133**, e835–42. <http://doi.org/10.1542/peds.2013-2365>.
- Ohtsubo Y. and Watanabe E.** (2008). 'Do sincere apologies need to be costly? Test of a costly signaling model of apology.' *Evolution and Human Behavior* **30**, 114–123.
- Oppenheimer D.** (2008). 'The Secret Life of Fluency.' *Trends in Cognitive Sciences* **12**(6), 237–41.
- Oreskes N. and Conway E.M.** (2010). *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*. New York: Bloomsbury Press.
- Rini R.** (2017). 'Fake News and Partisan Epistemology.' *Kennedy Institute of Ethics Journal* **27**(S2), 43–64. <https://doi.org/10.1353/ken.2017.0025>.
- Roberts R.C. and Wood J.W.** (2007). In W. Jay Wood (ed.), *Intellectual Virtues: An Essay in Regulative Epistemology*. Oxford, GB: Oxford University Press UK.
- Silvia P.J.** (2005). 'What is Interesting? Exploring the Appraisal Structure of Interest.' *Emotion* **5**(1), 89–102.
- Snyder M.L., Kleck R.E., Strenta A. and Mentzer S.J.** (1979). 'Avoidance of the handicapped: an attributional ambiguity analysis.' *Journal of Personality and Social Psychology* **37**(12), 2297–306. <http://doi.org/10.1037/0022-3514.37.12.2297>. PMID: 160924.
- Stanley J.** (2015). *How Propaganda Works*. Princeton: Princeton University Press.

- Sternisko A., Cichocka A. and Van Bavel J.J.** (2020). 'The Dark Side of Social Movements: Social Identity, Non-Conformity, and the Lure of Conspiracy Theories.' *Current Opinion in Psychology* 35, 1–6.
- Swank C.** (2000). 'Epistemic Vice.' In G. Axtell (ed.), *Knowledge, Belief, and Character: Readings in Virtue Epistemology*, pp. 195–204. Lanham, MD: Rowman & Littlefield Publishers.
- Swire B., Berinsky A.J., Lewandowsky S. and Ecker U.K.H.** (2017). 'Processing Political Misinformation: Comprehending the Trump Phenomenon.' *Royal Society Open Science* 4, 160802. <http://doi.org/10.1098/rsos.160802>.
- Taylor L.E., Swerdfeger A.L. and Eslick G.D.** (2014). 'Vaccines are not Associated with Autism: An Evidence-Based Meta-Analysis of Case-Control and Cohort Studies.' *Vaccine* 32(29), 3623–29.
- Tetlock P.E., Kristel O.V., Elson S.B., Green M.C. and Lerner J.S.** (2000). 'The psychology of the unthinkable: taboo trade-offs, forbidden base rates, and heretical counterfactuals.' *Journal of Personality and Social Psychology* 78(5), 853–70. <http://doi.org/10.1037//0022-3514.78.5.853>. PMID: 10821194.
- Trivers R.L.** (1971). 'The Evolution of Reciprocal Altruism.' *Quarterly Review of Biology* 46(1), 35–57.
- Tourish D. and Vatcha N.** (2005). 'Charismatic Leadership and Corporate Cultism at Enron: The Elimination of Dissent, the Promotion of Conformity and Organizational Collapse.' *Leadership* 1(4), 455–80.
- Uhlmann E., Pizarro D. and Diermeier D.** (2015). 'A Person-Centered Approach to Moral Judgment.' *Perspectives on Psychological Science* 10(1), 72–81.
- Verplaetse J., Vanneste S. and Braeckman J.** (2007). 'You can judge a book by its cover: the sequel.: A kernel of truth in predictive cheating detection.' *Evolution and Human Behavior* 28(4), 260–271.
- Walter N. and Tukachinsky R.** (2020). 'A meta-analytic examination of the continued influence of misinformation in the face of correction: How powerful is it, why does it happen, and how to stop it?' *Communication Research* 47(2), 155–177. <https://doi.org/10.1177/0093650219854600>
- Williams E.J. and Muir K.** (2019). *Chapter 13: The Palgrave Handbook of Deceptive Communication*. London: Springer International Publishing.

Tyler Porter is currently a PhD student at the University of Colorado Boulder. His research is primarily centered around social epistemology, with a particular focus on social and environmental conditions that lead to problems for would-be-knowers.