

---

## **Caveat Auditor: Epistemic Trust and Conflicts of Interest**

Justin P. McBrayer


Department of Philosophy, Fort Lewis College, Durango, USA

### **ABSTRACT**

To place epistemic trust in someone is to take their word for something. Much of the existing literature on epistemic trust concerns epistemic authorities. But as important as authority is to epistemic trust, it pales in comparison to the epistemic importance of conflicts of interests. In economics, we say that buyers shouldn't take the word of sellers. *Caveat emptor*: let the buyer beware. I argue for a similar principle in epistemology. *Caveat auditor*: let the hearer beware. Others often have incentives to testify in ways that are odds with our epistemic goals. Given this, our epistemic trust in others should be calibrated to reflect the epistemically virtuous and perverse incentives of ourselves and others. This basic principle explains the need for epistemic caution in a wide range of applied topics from politics to product reviews to fake news.

### **KEYWORDS**

Testimony, caveat emptor, defeater, interest, trust

**CONTACT** Justin P. McBrayer  [jpmcbrayer@fortlewis.edu](mailto:jpmcbrayer@fortlewis.edu) Fort Lewis College, Department of Philosophy, 1000 Rim Drive, Durango, CO 81301, USA

© 2022 Informa UK Limited, trading as Taylor & Francis Group

## **Introduction**

Most of what we know is the result of what others have told us. I know that Ankara is the capital of Turkey, Julius Caesar was the first Roman emperor, and that salt is composed of sodium and chloride. In each case, I am relying on testimony. Doing so requires trust.<sup>1</sup> In particular, I have placed a kind of epistemic trust (as opposed to other species of trust) in the sources of that information. I trust Google maps on capitals, my 3<sup>rd</sup> grade teacher on Roman history, and the science textbook on the composition of salt. To place epistemic trust in someone or something is to take their word for it.

This raises one of the most significant questions in regulatory epistemology: who shall I trust?<sup>2</sup> If I have epistemic trust in no one but myself, my knowledge-scape will be barren. As Hardin (2009, 10) puts it, we trust others of necessity. But if I trust everyone, my worldview will be riddled with false beliefs. So, I need to know who to trust. Somehow, I should do what I can to sort the epistemically sound sources from the rest.

By way of answering the question of who we should trust, the contemporary literature focuses on the epistemic importance of expertise. The basic idea is that we should trust experts more (or perhaps exclusively). This advice raises many important questions. What makes someone an expert? How do we identify experts? What should we do when experts disagree with one another (the epistemic significance of disagreement debate)?<sup>3</sup> What should we do when experts disagree with us (the preemption vs. total evidence debate)?<sup>4</sup> In short, the contemporary literature about epistemic trust focuses largely on two issues: (a) expertise and (b) disagreement. No doubt each is important. But there is something that matters far more: incentives.

To that end, I'm going to set aside issues of epistemic competence to look at another dimension of testifiers. We want to know who to trust. I argue that we can't answer that question without looking at the world from an *economic* point of view. Clarifying the proper limits of our epistemic autonomy requires that

we ask ourselves what others stand to gain or lose by testifying. My thesis is that the incentives of testifiers are epistemically relevant, often more so than expertise. It's plausible that many readers will share the intuition that incentives matter for epistemology. But there's also nothing in the literature clearly defending the epistemic importance of incentives, describing the nature of two distinct sorts of incentives for testimony, modeling the effect of incentives on vindication and defeat, or arguing that the incentives of testifiers typically matter more for epistemology than features like expertise and peer disagreement. This paper does all four.

## The Opening Insight

Almost everyone knows you shouldn't trust a used car salesman. If you're in the market for a vehicle, perhaps the epistemically worst thing you can do is head to the used car lot and accept the testimony of the first person you meet. This is a paradigm case in which trust is not warranted. Why?

A bad explanation is that the salesman is not an expert. To the contrary, he is likely an expert in vehicles and their condition. If he weren't, he would likely go out of business. And even if he doesn't qualify as an expert in the absolute sense, he is almost certainly going to have more expertise than you. His epistemic position is better than yours; you are not epistemic peers. So, we can't explain why you shouldn't trust the salesman by citing facts about his expertise in either a relative or absolute sense.

Instead, a good explanation is that the salesman's economic interests are at odds with your epistemic interests. He has an incentive to mislead you. He makes progress on meeting his economic goals when you fail to meet your epistemic goals. This can happen in two ways. First, your *missing* the truth is sometimes a benefit to him (e.g. you don't realize that the brakes are shot). Second, your *believing falsely* is sometimes a benefit to him (e.g. you believe it's a 2018 when it's really a 2008).

On the other hand, your epistemic goals include not missing out on important truths and believing truly about the car. You do better on your goals when you realize the brakes are shot and accurately believe that the car is a 2008. There is a mismatch between your epistemic goals and the salesman's economic goals. That conflict of interest explains why you shouldn't place epistemic trust in him. The limit of trust is not a function of expertise or disagreement. It's about incentives.

## Incentives

The opening case shows us that incentives have epistemic implications. But what is an incentive?

An incentive is a background condition that motivates a sentient being to behave in one way rather than another. For example, holding out a treat provides an incentive for a dog to sit. The dog might have sat down anyway – an incentive isn't a necessary condition for an action – but he is more likely to sit when you hold out the treat. That's because the dog wants the treat. So, more carefully, incentives are background conditions that are tied to interests.

An interest is something in which we have a stake (Feinberg 1984). For example, I have an interest in being pain-free, financially comfortable, intellectually stimulated, and so forth. Incentives shape our pursuit of interests by making it more likely that we will meet our goals in some ways rather than others. Classic economics teaches that incentives come in three broad flavors: monetary, social, and moral (Levitt and Dubner 2005, 17). The basic idea is that your pursuit of your goals will be impacted by monetary, social, and moral incentives.

People's interests can obviously be crosswise in the sense of a zero-sum game: a gain for you is a loss to me. The used-car salesman has an interest in disposing of a low-quality car and doing so for the highest price possible. I have an interest in procuring a high-quality vehicle and paying as little as possible. When my interests are met, his are not, and vice-versa. It is common to talk about this sort of conflict of interest in

ethics. You don't want the grandson who stands to gain a fortune to decide when to pull the plug on his aged grandmother. But there can also be conflicts of interest in epistemology.

When interests are crosswise, our incentives follow suit. The salesman has an incentive to lie about the quality of the car, and I have an incentive to pretend that I can't pay that much. These incentives are background conditions that affect our pursuit of our goals, and in this case, they are running contrary to one another.

I don't want to give the impression that all incentives are rational or efficacious. Some incentives are perverse in the sense that they actually end up hindering the pursuit of a goal. The classic example is the cobra effect. When the British ruled India, they paid a bounty for dead cobras in Delhi. In the long run, this caused the number of venomous snakes in the area to skyrocket: people started breeding them for the money, and they often escaped. The bounty is an example of a perverse incentive.

And perverse incentives aren't restricted to snakes. For example, Nguyen ([forthcoming](#)) has recently argued that when it comes to experts, public accountability and transparency can function as perverse incentives. That's because when experts know that they are accountable to the public or that there needs to be a kind of transparency between their work and the view of non-experts, they have incentives to limit their conclusions or abandon certain types of evidence. In other words, when experts face pressure to explain their conclusions and share their evidence with the general public and when that's not easy to do, there's an incentive to repress such conclusions instead.

But some incentives are virtuous (for lack of a better term). They are effective at encouraging the right sort of behavior.<sup>5</sup> For example, the presence of police cars is an effective deterrent to speeding. Citizens who know about the police presence have an additional incentive not to speed, and that incentive changes their behavior in the desired direction. Peer review is a good incentive in scholarly research. Scholars submitting their work for the review process have strong incentives to make sure that their evidence is strong, their arguments clear, and their conclusions sound.

## Incentive Epistemology

Your epistemic trust in someone should take their incentives into account. That is not a claim about how to allocate *initial* trust in others. Suppose there's some general answer for how prior or initial trust should be distributed (e.g. an account of Bayesian priors for trust). For example, suppose it's true that you should allocate equal amounts of trust to epistemic peers. A general answer to the question of initial trust will explain how much you should trust the ordinary person on the street on the question of annual handgun deaths in the US.

Instead, this is a claim about a specific species of source-sensitive defeat or source-sensitive vindication. The idea is that awareness of testifier incentives should move your posterior trust levels up or down. Once you've established an initial level of trust for the ordinary person on the street, how should you revise that level of trust up or down once you realize that the person works for the NRA or the CDC?

Put in terms of epistemic virtues, the point is that the virtue of epistemic autonomy requires sensitivity to the incentives of our testifiers. King (2021) frames epistemic autonomy as a mean between the vices of servility and isolation. We don't want to simply accept everything we hear, but we don't want to be epistemic islands, either. Given that framing, those with a reliable disposition to trust the wrong people at the wrong times display an epistemic vice. We shouldn't trust others when there's a conflict of interests.

Put in terms of epistemic deontology, when testifiers have incentives that are at odds with your epistemic goals, your epistemic duties of due diligence increase. In business, the duty to perform due diligence is called *caveat emptor*: let the buyer beware. The seller has incentives to pass off cheap goods as high-quality and to do so at the highest price possible. The consumer shouldn't just take the company's word. That duty remains even in cases where the conflict of interest isn't readily apparent.

The same lesson applies here. *Caveat auditor*: let the hearer beware. When a testifier has incentives that are in-line with your epistemic goals, you have a reason to increase your trust in that source. But when a testifier has incentives that are at cross-purposes with your epistemic goals, your trust in that source should be reduced.

In fact, for many cases, you are better off relying on what little knowledge you have rather than trusting someone else who has a conflict of interest. In a case like that, independent thinking is justified even when your knowledge base is weak. If your choice is either to trust the used-car salesman about whether the car is in good shape or to trust yourself – however little you know about cars – you’re better off trusting yourself.

Here are all the usual disclaimers about my theoretical model:

Concern: What if our epistemic incentives are less than pure? For example, I’m trying to figure out how strong of a teacher I am, and I may be happier with a false belief. In a case like that, the fact that I have a conflict of interest with others won’t be a reason for me to discount their views—after all, it might be a problem on my side of the equation!

Reply: Our epistemic goals are almost always less than pure (McBrayer 2021, chapter 2). In particular, there’s good reason to think that we consume information recklessly when there’s little to no material cost of error. For example, when there is a high cost to locating accurate information with little practical value, most of us will settle for beliefs that are less accurate but provide consumptive benefits (Tebben and Waterman, [forthcoming](#)). But to see the role that conflicts of interest play in appropriate levels of trust, I’m assuming you want the truth, the whole truth, and nothing but the truth.

Concern: What if we don’t know our own incentives or goals? We almost never see them clearly (Wilson 2004). So, if I can’t tell what my epistemic goals are, it will be difficult for me to determine whether there’s a conflict of interest at play.

Reply: We should concede that, at best, we see through a glass darkly. But, again, for modeling purposes, I’m assuming that you know what you want. That assumption won’t affect what follows, and it’s similar to the sort of modeling assumptions we make in economics when we assume that consumers are perfectly knowledgeable about options in a marketplace.

Concern: Does your belief about the testifier’s incentives have to be justified? I mean, what if you thought he was out to dupe you, but in reality he wasn’t? What if you are just paranoid about the incentives of your testifiers and end up assuming all sorts of conflicts of interest that aren’t there?

Reply: Since the question for this paper is whether conflicts of interest provide a defeater for testimonial justification, mere belief about those conflicts is sufficient to do the job (Foley 1993). If you mistakenly believe that the person is out to trick you, you still have a reason to be wary of their testimony, even if you’re mistaken about your presupposition.

That assumption is a strong one: belief is sufficient for defeat. The conspiracy theorist has a reason to be wary of government testimony if she believes that the government has a conflict of interest. I think that’s the right analysis of the situation, and my diagnosis fingers an epistemic problem earlier in the scenario (the conspiracy belief itself). But there’s a case for weakening the assumption: perhaps it’s only true belief, justified belief, or knowledge about a conflict of interest that generates a defeater. Nothing in my analysis of how incentives matter hinges on this decision. Instead, the decision about which mental states are sufficient to count as awareness will affect the scope the problem I discuss below.

In sum, despite the fact that the real world is messy, it makes sense to build a simple model of conflicts of interest in order to trace out the implications of such conflicts for the epistemology of trust.

## Two Types of Testimonial Incentives

There are at least two ways in which a testifier could have incentives at odds with your epistemic goals. First, a testifier could benefit by signaling in various ways regardless of whether the signaling affects your belief states. Call those cases presentation incentives. Second, a testifier could benefit by getting you to believe in certain ways. Call those cases doxastic incentives. Each is an important genus of a conflict of epistemic interests.

In a case of presentation incentives, the testifier has an incentive to say things irrespective of whether they are true and independently of whether you believe them. In other words, the testifier's interests are advanced merely by the signaling, regardless of whether the signal is effective at altering the hearer's belief states. Saying things is good enough to meet the goals. For example, consider a stand-up comedian. She has an interest in telling a funny story even if the audience doesn't believe the story. That's not the point. Her interests aren't furthered or hindered based on what the audience believes in such a case. The advancement of her interests depends only on getting a laugh. In that case, the comedian has a presentation incentive rather than a doxastic one.

In a case of a doxastic incentive, the testifier has an incentive to say things only if the testimony alters your belief states in the right ways. In other words, the testifier's interests aren't advanced by the mere saying of things but by your believing them. What they say matters instrumentally – they don't get the goods merely by signaling. It must translate into belief on your part. For example, return to the used car salesman. He benefits not from expressing things but from getting you to believe (or not believe) certain things. His goals are not met by merely saying that it's a 2018 model. He does well only if he gets you to believe this. In that case, his incentive is aimed at your belief and not merely at his testimony.

And, of course, someone could have both presentation and doxastic incentives. The university president who makes a big speech about racial diversity on campus advances her goals by grandstanding but also by convincing her audience to agree with her about complex, politically charged topics. The president has an incentive to give the speech whether it changes people's minds or not. That's the presentation incentive of a virtue signal. And she also has an incentive to get her employees to agree with her. That's the doxastic incentive of convincing case. Both of these incentives remain in place even if the president's view on racial diversity is completely disconnected from the truth.

## Six Species of Testimonial Alignment

Given the difference between presentation and doxastic incentives, we need to keep each in mind. In any given case of testimony, either of these incentives might be perverse or at odds with our epistemic goals. In particular, these incentives can be aligned with our epistemic goals, misaligned with those goals, or neither. That gives us six categories of potential epistemic conflict to puzzle over:

1. Testifier has **presentation** incentives **aligned** with your epistemic goals.
2. Testifier has **presentation** incentives **neither** aligned nor misaligned with your epistemic goals.
3. Testifier has **presentation** incentives **misaligned** with your epistemic goals.
4. Testifier has **doxastic** incentives **aligned** with your epistemic goals.
5. Testifier has **doxastic** incentives **neither** aligned nor misaligned with your epistemic goals.
6. Testifier has **doxastic** incentives **misaligned** with your epistemic goals.

In this section, I'll outline the nature of each type of conflict. In the next, I'll discuss the epistemic implications of each. For simplicity's sake, I'll assume that you are engaged in an inquiry and understand the goals and incentives of both you and your testifier.

In case number 1, the testifier has a presentation incentive that is aligned with your epistemic goals. That means the testifier advances her interests when she testifies in ways that are truth-directed. Since your epistemic goals are also truth-directed, your incentives align. A good example of this case is a scholar producing peer-reviewed research. That person has a presentation incentive to testify in truth-related ways. Good research is aimed at the truth, well argued, well evidenced, etc. If a researcher testifies in ways that fall short of this standard, it hurts her interests (retraction, public shaming, job loss, etc.). But notice that the researcher doesn't have a doxastic incentive in a case like this. It doesn't matter much whether the people reading her papers change their minds – her interests are advanced by the testifying itself, regardless of how many people the testifying convinces.

In case number 2, the testifier has a presentation incentive that is neither aligned nor misaligned with your epistemic goals. That means that the testifier advances her interests when she testifies in certain ways, but these ways neither track what's true or what's false. She doesn't consistently testify in ways that are at odds with the truth, so it's not as if your incentives are misaligned. But she also doesn't share your epistemic goal of truth, and so her presentation incentives aren't aligned with yours, either. A good example of this case is a bullshitter. A bullshitter has incentives to say things and signal in certain ways regardless of whether those things are true (Frankfurt 2005). Bullshitters advance their interests by entertaining their listeners rather than guiding them to the truth. Bullshitters are unreliable testifiers, yet their unreliability goes in both directions. It's not as if they advance their interests purely by saying what's false, either. As Frankfurt (2005) puts it, 'although [bullshit] is produced without concern with the truth, it need not be false. The bullshitter is faking things. But this does not mean that he necessarily gets them wrong' (48–49).

In case number 3, the testifier has a presentation incentive that is misaligned with your epistemic goals. That means that the testifier does better with regard to advancing her interests when she testifies in ways that are at odds with the truth. It's good for her to say what's false. But your epistemic interest is the truth. And so, your incentives are misaligned. For example, consider the talking head on TV. Pundits and news guests have incentives to be intellectually uncharitable (instead of fair-minded), controversial (rather than humdrum), confrontational (rather than conciliatory), black-and-white (rather than nuanced), *et cetera*. Talking heads generate more buzz, gather larger audiences, and get shared on social media more, when they testify in ways that at odds with our best evidence. Unlike the bullshitter, the talking head will have strong incentives that are at odds with certain truths: conservatives have incentives to avoid truths that are inconvenient for their audiences, liberal anchors have incentives to avoid truths that are inconvenient for their audiences, etc. And the incentives are mere presentation incentives – they will be invited back to the show regardless of whether they manage to convince audiences.


The next three cases focus on doxastic incentives instead of presentation ones. In case number 4, the testifier has a doxastic incentive that is aligned with your epistemic goals. That means the testifier advances his goals only when you believe the truth. He testifies in order to convince you to alter your belief structure, and his interests are furthered when you get closer to the truth rather than further away. Your interests are aligned. A good example of this is a medical doctor whose pay structure is tied to patient outcomes. When you come in for a check-up, the MD will testify in all sorts of ways. He might tell you about vaccines, dieting, exercise, and so forth. But his interests aren't furthered by merely *telling* you what's true. He benefits economically only when you *believe* him (and then alter your lifestyle accordingly). We might think about people in case number 4 as our epistemic *interest* peers. No doubt those interests are not equally strong, of the same species, etc. Yet, they share the same goal, and the testifier's incentive structure mirrors our own.

In case number 5, the testifier has a doxastic incentive that is neither aligned nor misaligned with your epistemic goals. That means the testifier advances her goals when you come to believe certain things based on her testimony, but those propositions don't track either what's true or what's false. They are careless in the way that the bullshitter is careless, but in this case the negligent attitude is about what you believe. This is the cognitive equivalent of the bullshitter. This is perhaps the strangest of the six cases, but sometimes people have incentives of this sort. A good example is an attorney litigating a criminal case. The attorney is trying to get the judge or jury to see the facts of the case a certain way. When the prosecuting attorney and defense attorney disagree about some issue, they can't both be correct. And yet each has an incentive to get the audience to agree with them. This isn't a mere presentation incentive. The attorney's interests are advanced only when the presentation is effective at generating belief (or disbelief). And since those incentives sometimes track what's true and sometimes track what's false, in general they are neither aligned nor misaligned with your epistemic goals.

In the final case, number 6, the testifier has a doxastic incentive that is misaligned with your epistemic goals. The testifier advances his goals only when his testimony is enough to change your mind about something. The used car salesman with which we began is a paradigm case of doxastic incentives that are at cross-purposes with your own. You do best when you believe truly about the car on the lot, and he does best when you believe falsely (or at the least fail to believe truly). In that sense, your incentives are misaligned.

In sum, there are at least two different sorts of testimonial incentives and in each case, those incentives can be aligned with yours, misaligned with yours, or neither (Table 1).

**Table 1.** Six species of testimonial alignment (Table view)

Presentation Incentives	Doxastic Incentives
Testifier has presentation incentives <b>aligned</b> with your epistemic goals.	Testifier has doxastic incentives <b>aligned</b> with your epistemic goals.
Testifier has presentation incentives <b>neither</b> aligned nor misaligned with your epistemic goals.	Testifier has doxastic incentives <b>neither</b> aligned nor misaligned with your epistemic goals.
 Testifier has presentation incentives <b>misaligned</b> with your epistemic goals.	Testifier has doxastic incentives <b>misaligned</b> with your epistemic goals.

### The Sting of Defeat and the Boost of Vindication

Becoming aware of a testimonial incentive often, but not always, has an epistemic implication. In particular, our beliefs about incentives function as defeaters or vindicators for our initial epistemic trust. Upon finding out about these incentives, we should adjust the levels of our prior trust up or down. Put another way, testimonial conflicts of interests matter for epistemology.

When we believe that incentives are misaligned, our epistemic duties of diligence increase. Since the testifier advances her goals by speaking in ways that are at odds with your epistemic goals, you need to be cautious. In that case, discovery of a conflict of interest provides a defeater (or partial defeater) for the testimonial evidence. On the other hand, when incentives are aligned, our epistemic duties of due diligence

are relaxed. Since the testifier advances her goals only when yours are met, too, it's not caution but confidence that is warranted. In that case, discovery of an alignment of incentives vindicates the testimonial evidence – you should be even more confident in it than you were initially.

There are various accounts of how initial trust in strangers should be allocated. We can think of that as the justified level of prior trust. Suppose we have a good account in hand. The question for this paper isn't how that *initial* batch of trust should be allocated but instead how that level of trust should be *adjusted* given the testimonial incentives discussed previously. In other words, the issue developed in this paper isn't an account of prior trust but an account for how that prior trust should be revised in light of testimonial incentives. The question is about appropriate posterior levels of trust.

Here's a first pass. It's a matrix of the six species of testimonial alignment.<sup>6</sup> Given the emphasis on expertise in the philosophical literature, the cases are developed for both experts and non-experts. Stipulate that 1 is the least amount of trust you can have in someone else, 10 is the most trust, and that the correct account of initial justification implies that we should have prior confidence of level 5.<sup>7</sup> We can then make some idealizations that illustrate the appropriate adjustment of trust up or down from the initial trust levels.

Note that the absolute numbers plugged into the matrix aren't important (recall that I don't have a theory about how to assign initial trust). Instead, it's the relative difference that holds across instances that is important for my purposes. And since I need some kind of assignments to depict the epistemic boost or downgrade for various sorts of incentives, I've just made assumptions about initial trust. What matters is the relation that holds across each row (the difference between posterior trust in experts and non-experts) and the relation that holds across each column (the difference between posterior trust when someone's interests are aligned with yours misaligned, or neither) (Table 2).

**Table 2.** Posterior levels of trust (Table view)

<b>Presentation Incentives</b>	Expert	Non-expert
Aligned (pr article)	9	5
Neither (bullshitter)	6	4
Misaligned (talking head)	2	3
<b>Doxastic Incentives</b>	Expert	Non-expert
Aligned (medical doctor)	10	6
Neither (attorney)	6	4
Misaligned (car salesman)	1	2

As the table makes clear, finding out that someone is an expert and has incentives that align with yours provide a big boost to epistemic trust. In those cases, awareness of incentives and expertise provide a vindication of the testimony. On the other hand, finding out that someone is an expert and has incentives that are at odds with yours provides a significant decrease in appropriate levels of epistemic trust. In those cases, the testimony is handily defeated by awareness of expertise and conflicts of interests.

### Three Lessons About Trust

The table of appropriate posterior levels of trust is plausible because it gets a lot of things right. Here are its three most important features.

*Lesson Number One:* expertise is epistemically relevant for trust. As the contemporary philosophical literature on epistemic authority makes clear, the fact that someone is an expert is relevant to how much we should trust them. This importance is captured by the table of posterior trust levels: in each case, there is a



gap between the column on the left (for experts) and the column on the right (for non-experts). Expertise makes an epistemic difference. For example, look at the top row of the table. Even if two people have precisely the same presentation incentives, the trust you should place in the expert is nearly twice that of the non-expert. This explains why you have a much stronger reason to trust Dr. Fauci on vaccine efficacy than your well-intentioned neighbor.

However, even though expertise matters for trust, the table of posterior trust levels pinpoints two important points about how expertise matters. First, how much someone's expertise matters depends on the alignment of incentives. In other words, the fact that someone is an expert makes a great deal of difference in a case where both you and the expert have incentives that align. But in a case where there's a conflict of interest between the two of you, expertise matters much less.

This gradation of expert force is illustrated in the table. In the first row, expertise makes a difference of four. That's because your interests align with the expert. But in the second row, the difference is only two. And in the last row, the difference is only one. The punchline is that your epistemic trust should be calibrated first to incentives and only second to expertise. Someone's being an expert in a domain provides very strong reasons for trusting them when their interests align with yours but not otherwise.

The second interesting feature about expertise is that the directional value of expertise is reversed in cases where there's a conflict of interest. In other words, while the first interesting feature is about the size of the gap between experts and non-experts, the second feature is about the direction of the gap between experts and non-experts. As the table makes clear, when you and your testifier have interests that are aligned or at least not misaligned, expertise provides more of a reason to trust them. However, when you and your testifier have incentives that are misaligned, expertise provides a reason to distrust them. In other words, expertise boosts trust in cases of alignment and undermines trust in cases of a conflict of interest. For example, the third row indicates that in a case where the testifier has presentation incentives that are misaligned with your epistemic incentives, expertise makes things worse. You should trust an uninformed neighbor at level 3 but an expert who is out to deceive you at level 2.

The basic idea here is that experts have greater knowledge than non-experts (at least in their specialized domain). So, if an expert is out to deceive you, she'll do a better job than someone who is ignorant of the facts in that domain. That's why you're better off trusting your Uncle Frank at the used car lot rather than the salesman. The salesman is an expert whose incentives are misaligned with yours. At least Uncle Frank – for all of his ignorance – isn't out to deceive you about the little that he does know.

*Lesson Number Two:* incentives are epistemically relevant for trust. The fact that someone has incentives to testify or get us to believe a certain way is relevant to how much we should trust them. This importance is captured by the table of posterior trust levels: in each case, there is a difference between any given row and the row above or below it. Incentives make an epistemic difference.

The used car salesman case is a thought experiment that supports this lesson, but the table makes clear that the importance of incentives extend beyond that simple case. When you find out that someone's incentives are aligned with yours, your confidence in that source of information should increase. When you find out that someone's incentives are misaligned with yours, your confidence in that source of information should decrease.

One interesting implication of the importance of incentives concerns the difference between presentation and doxastic incentives. When it comes to experts whose incentives are aligned with yours, the alignment of doxastic incentives matters more to levels of trust than does the alignment of presentation incentives. In terms of the table, the top-left number in the doxastic incentives ranking is higher than the top-left number in the presentation incentives ranking. In other words, if you had to choose between trusting an expert whose cognitive incentives aligned with yours or one who merely had presentation incentives aligned with yours, choose the former.

*Lesson Number Three:* when it comes to trust, incentives are *more relevant* than expertise. The fact that someone has a conflict of interest is more relevant to whether you should trust them than whether they are an expert in the field. This importance is captured by the table of posterior trust levels: the bottom-left corner of the table for each sort of incentive is lower than the top-right number. Incentives matter more than expertise.

For example, if you believe that someone else's incentives are misaligned with your own, the table indicates that an appropriate level of trust in that person should be somewhere between a one and a three. On the other hand, if all you believed about a person was that she was an expert, the table indicates that an appropriate level of trust could range anywhere from a one to a ten. The punchline is that if you had to ask one question about your testifier, you're better off using that question to find out about conflicts of interests rather than academic degrees.

These three lessons capture the epistemic importance of conflicts of interests. In particular, the lessons are evidence of myopia in the contemporary professional literature: we have over-focused on the epistemic importance of expertise, particularly when other things like conflicts of interests matter more. In short, we've been thinking too much like philosophers and too little like economists. Going forward, we can apply these lessons – particularly those involving the epistemic trade-offs between experts and incentives – to think more clearly about a wide range of concrete issues in epistemology.

## The Application

Theorizing about the conditions under which awareness of incentives produce defeat or vindication matters for a whole host of applied issues. I sketch some of those in this closing section.

Some applications of the caution for conflicts of interest are obvious. Returning to our paradigm example, we should have a low level of trust in salespeople. They have powerful doxastic incentives that are at cross-purposes with virtuous epistemic goals. The same goes for politicians. They routinely have presentation incentives that are at odds with the truth, and in cases where their livelihood depends on votes from constituents, they have doxastic incentives, too. One truism in politics is that it's better to meet voters where they are than to try to change their minds. That's a perfect example of a conflict of interest.

Much of the COVID debate in the United States has wrongly focused on the question of expertise rather than incentives. We were quick to ask whether a testifier was an MD but far less quick to ask about the person's incentives. Anyone reading this paper likely received plenty of social media shares, forwarded emails, and newspaper clippings from putative experts in the field offering contrary advice on whether COVID was dangerous, whether vaccines were better than natural immunity, and whether alternative treatments were effective. The advice always came with notice of credentials.

But in most cases, determining who to trust wasn't about determining whether the person was an expert but about determining whether the person had a conflict of interest. The guy selling supplements or advertising dollars on a podcast has a conflict of interest. There are strong doxastic incentives at play. There's no rule that says doctors can't be bullshitters. And the political partisans who pushed masks or no masks have pretty obvious presentation incentives. Virtue signaling to your tribe is conducive to lots of social goods, and there's good reason to think that virtue signaling provides conformist incentives strong enough to override the epistemic virtues of crowd-sourcing information (Hill and Garner 2021).

A perhaps less obvious application is fake news (e.g. see Bernecker, Flowerree, and Grundmann 2021). Those producing the news have an incentive to build an audience. But you can build an audience in many ways, few of which are epistemically virtuous (McBrayer 2021). For example, anytime there is a marketplace for news, there will be incentives to tell readers and listeners what they want to hear, regardless of whether it's true (Tebben and Philip Waterman, [forthcoming](#)). That's a clear example of a presentation incentive. Further, news channels that want to build long-term audiences also have doxastic incentives at

play: they want their viewers to come to believe the message that they are selling so that they keep coming back for more.

This same lesson goes for the guests and speakers on news programming. Talking heads often have presentation incentives that are at odds with virtuous epistemic goals. It's good for them to be controversial, combative, entertaining, etc. That's what gets views and generates repeat invitations. Once we are aware of these incentives, we ought to lower our levels of trust in those individuals. That's not to say they don't often speak truly. Often, they do. But it's also true that they have perverse epistemic incentives, and we ought to take those into account when deciding who to trust. Given these incentives, we ought to lower our confidence in the news anytime we find ourselves in a news marketplace. Those on the news and those producing the news have conflicts of interest.

Online product reviews are often plagued by a conflict of interest. For example, I recently made a purchase on Amazon and received an email inviting me to write a product review. The company offered me a \$20 gift card to leave a positive review on the Amazon site, and it explicitly told me not to mention this offer in the review. In a case like this, I have a clear conflict of interest. My presentation incentives are at odds with the epistemic incentives of those reading the reviews. The conflict of interest in other reviews might be less obvious but still present. There are all kinds of presentation incentives: expressing rage or disappointment, boosting a reviewer profile, et cetera.

The lessons of conflicts of interest apply to us, too. For methodological purposes, I assumed that your epistemic goals were perfectly virtuous. I assumed, in other words, that you had an interest in getting to the truth, the whole truth, and nothing but the truth. That's often false. We often have perverse doxastic incentives. In that case, our own goals provide the conflict of interest.

For example, teachers have a strong incentive to think about themselves as good teachers. That incentive is often at odds with the evidence. There are a lot of bad teachers out there. In those cases, teachers have a conflict of interest. On the one hand, they have a doxastic incentive to believe that they are good teachers. On the other, they have an epistemic incentive to get to the truth. The two are at odds, indicating that our self-assessments are not likely to track the truth. The experimental data bears this out (Cross 1977). And so teachers ought to discount their own testimony about the quality of their teaching.

In closing, conflicts of interest are relevant for matters of trust, indeed more so than expertise. Finding out that a testifier has incentives that align with your epistemic incentives vindicates your trust and moves posterior levels of trust even higher. But finding out that a testifier lacks such incentives or even has incentives that are misaligned with your epistemic goals defeats your initial trust and moves your posterior levels of trust lower. Taking this epistemic lesson to heart will elucidate a number of concrete, epistemic issues.

## Notes

1. As McMyler (2011) puts it, testimonial knowledge is based on "taking another's word for things," "Taking things on another's authority," or "trusting another for the truth," (6).
2. For an introduction to the contemporary debate about the limits of epistemic authority, see Zagzebski (2012).
3. For the contours of this debate, see Christensen (2009) and Matheson (2015).
4. For the contours of this debate, see Constantin and Grundmann (2020), Grundmann (2021), Jäger (2016), and Wright (2016).
5. Incentives and interests might even be constitutive for understanding objectivity; see Willholt (2022) for a defense of this idea.
6. We could also make a three-dimensional version of this table that takes into account the importance of the topic. This would be a way to accommodate a kind of pragmatic encroachment for testimony. The basic idea is that a conflict of interest might matter more or less depending on the importance of believing truly in that case.
7. For simplicity's sake, I leave aside entirely the issue of distrust where that's a negative trust of some sort. In many cases, it's not just that your trust in someone could be low, but you could have a positive reason to distrust them.

## Acknowledgements

Dugald Owen and two anonymous referees provided helpful feedback on an early version of this paper. I'm also grateful to the audience at the Epistemic Autonomy Conference at the University of North Florida and to James Beebe and Jon Matheson in particular for the invitation. Michael Martin provided Latin expertise for the title. *Gratias tibi ago.*

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## References

- Bernecker, Sven, Amy Flowerree, and Thomas Grundmann. 2021. *The Epistemology of Fake News*. Oxford: Oxford University Press.
- Christensen, David. 2009. "Disagreement as Evidence: The Epistemology of Controversy." *Philosophy Compass* 4 (5): 756–767. .
- Constantin, Jan, and Thomas Grundmann. 2020. "Epistemic Authority: Preemption Through Source Sensitive Defeat." *Synthese* 197: 4109–4130. .
- Cross, K Patricia. 1977. "Not Can, but Will Teaching Be Improved?" *New Directions for Higher Education* 1977 (17): 1–15. .
- Feinberg, Joel. 1984. *Harm to Others: The Moral Limits of the Criminal Law*. Oxford: Oxford University Press.
- Foley, Richard. 1993. *Working Without a Net: A Study of Egocentric Epistemology*. Oxford: Oxford University Press.
- Frankfurt, Harry. 2005. *On Bullshit*. Princeton: Princeton University Press.
- Grundmann, Thomas. 2021. "Facing Epistemic Authorities: Where Democratic Ideals and Critical Thinking Mislead Cognition". In *The Epistemology of Fake News*, edited by Sven Bernecker, Amy Flowerree and Thomas Grundmann, 134–155. Oxford: Oxford University Press.
- Hardin, Russell. 2009. *How Do You Know?*. Princeton: Princeton University Press.
- Hill, Scott, and Renaud-Philippe Garner. 2021. "Virtue Signaling and the Condorcet Jury Theorem." *Synthese* 199: 14821–14841. .
- Jäger, Christoph. 2016. "Epistemic Authority, Preemptive Reasons, and Understanding." *Episteme* 13 (2): 167–185. .
- King, Nathan. 2021. *The Excellent Mind: Intellectual Virtues for Everyday Life*. Oxford: Oxford University Press.
- Levitt, Steven D., and Sephen J. Dubner. 2005. *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything*. New York: Harper Collins Publishers.
- Matheson, Jon. 2015. "Disagreement and Epistemic Peers." *Oxford Handbooks Online*. .
- McBrayer, Justin. 2021. *Beyond Fake News: Finding the Truth in a World of Misinformation*. New York: Routledge Press.
- McMyler, Benjamin. 2011. *Testimony, Trust, and Authority*. Oxford: Oxford University Press.
- Nguyen, C Thi. Forthcoming. "Transparency is Surveillance." *Philosophy and Phenomenological Research*. .
- Tebben, Nicholas, and John Philip Waterman. Forthcoming. "Expert Testimony and Practical Interests." *Inquiry*. .
- Wilholt, Torsten. 2022. "Epistemic Interests and the Objectivity of Inquiry." *Studies in History and Philosophy of Science* 91: 86–93. .
- Wilson, Timothy D. 2004. *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Cambridge, MA: Belknap Press.
- Wright, Sarah. 2016. "Epistemic Authority, Epistemic Preemption, and the Intellectual Virtues." *Episteme* 13 (4): 555–570. .
- Zagzebski, Linda. 2012. *Epistemic Authority*. Oxford: Oxford University Press.