Epistemic Autonomy, Trust, and Conflicts of Interest: A Reply to McBrayer

Joshua Brecka, University of Toronto, j.brecka@mail.utoronto.ca

_____

Abstract

Whether we ought to trust the testimony of another person is often thought to primarily depend on features like their evidence, knowledge, or level of expertise. No doubt these are epistemically relevant features. However, a recent paper by Justin P. McBrayer (2024) convincingly argues that a testifier's interests may be more important than their expertise when it comes to how we should allocate our epistemic trust in them. Just think of the proverbial used car salesman. You should not trust what he tells you, and realizing that he is an expert only seems to make things worse. In this reply, I consider McBrayer's basic insight, constructively criticize his view, and tentatively explore further territory.

S*ocial Epistemology* recently released a special issue—edited by Jonathan Matheson—entitled "Epistemic Autonomy" (2024). For some, the notion of epistemic autonomy may conjure images of rugged epistemic individuals and trailblazers marching to the beat of their own free-thinking drums. However, as this excellent and wide-ranging edition to the extant literature on epistemic autonomy makes clear, to be epistemically autonomous is (in part) to think well for oneself (Battaly 2021; Matheson 2021; 2023; King 2021). And, as Matheson (268) puts the point in his introduction to the special issue, thinking *for* yourself does not require thinking *by* yourself (original emphasis).[1] So, epistemic autonomy—i.e., thinking well for yourself—should involve properly marshaling the epistemic resources of others, paradigmatically via trusting them and the things they tell you.

Unfortunately, not everyone is trustworthy. There are hucksters and liars, fools and frauds who may wish to harm or deceive us. We must rationally allocate our trust. So, epistemic autonomy involves discerning knowledgeable speakers from ignorant ones and experts from non-experts. But there is more to it than that. In his thought-provoking contribution to this special issue, Justin P. McBrayer (2024) convincingly argues that the *interests* and *incentives* of testifiers are highly relevant to rational epistemic trust allocation. McBrayer argues that when a testifier's interests conflict with your epistemic goals, you should lower your epistemic trust in what they tell you. In short: conflicts of interest are relevant to proper trust in testimony (299).

McBrayer's fundamental point is underappreciated, and the connections between epistemic autonomy, rational trust, and incentives are, to me, well worth thinking about. Still, we might have further questions. First, we might wonder what the ultimate aim of an epistemology of incentives ought to be. Second, we might wonder what kind of psychological connection, if any, to potential facts about a testifier's interests and incentives is needed for testimonial defeat. Lastly, we might wonder about the moral limits on judgments about how a speaker's incentives influence their testimony. Perhaps lowering our trust in a speaker whose interests conflict with our own can, in certain situations, wrong them. In this reply, I will summarize McBrayer's view and consider the impact of these questions. While I will critically engage

---

[1] Matheson here echoes King (2021, 88).

with parts of McBrayer's view, what I write is meant as an invitation to hear more rather than an objection.

Incentive Epistemology

To begin to see his point, consider McBrayer's preferred example. Suppose you are looking to buy a used car and you go to a local used car lot. After meeting the used car salesman at the lot, what level of trust should you place in what he tells you? According to McBrayer (291), you should view the used car salesman skeptically—i.e., you should lower your epistemic trust in him and his testimony. According to McBrayer, the salesman's interests and incentives, rather than his expertise, best explain why you ought to lower your trust. After all, the used car salesman has a strong interest in selling a car to you at the highest possible price, while you (presumably) have a strong interest in gaining knowledge about the car's condition. The salesman makes progress towards his goal when you fail at yours. Thus, your epistemic goals conflict with the salesman's economic interests. McBrayer argues that one should lower their epistemic trust in a testifier when the testifier's interests conflict with one's truth-directed goals. That is, conflicts of interest represent "a specific species of source-sensitive defeat" for testimony (292).

For McBrayer, incentives are background conditions tied to an agent's interests that motivate agents to act in some way rather than in others (291). Incentives shape an agent's pursuit of their interests because incentives are by their nature motivating.[2] In part, this is because *interests*—that to which incentives are tied—are things in which we have a stake. Thus, when interests conflict, so do incentives. Since the salesperson has an interest in—i.e., has a stake in—making money, they have an incentive to lie to you or mislead you about the car to get you to buy it. Assuming your goals with respect to the car are truth-directed, your interests and incentives are crosswise with those of the salesman. Importantly, McBrayer's is a view about *posterior* levels of epistemic trust. This is not a view about how we ought to initially allocate trust (292).

As we saw earlier, epistemic autonomy requires one to deal with testimony properly. The proposal here extends this insight: epistemic autonomy requires being properly sensitive to the incentives of testifiers. When a testifier has a conflict of interest this is the wrong time for trust. And trusting the wrong people at the wrong time displays an epistemic vice. We should instead increase our epistemic due diligence (292).

In slogan form: "*Caveat auditor:* let the hearer beware" (293). When a speaker's interests conflict with your own truth-directed interests, this is a defeater for testimonial belief (293). For McBrayer, merely believing that one's source has a conflict of interest is enough to generate this defeater (293). Thus, we should adjust our level of trust in a source once we acquire beliefs about their testimonial incentives.[3] If their incentives align with our goals, we

---

[2] However, McBrayer (291–292) is careful to point out that incentives are neither necessary nor sufficient for action. They merely make acting in a certain way more likely.

[3] McBrayer distinguishes between what he calls *presentation incentives*—incentives to signal in various ways regardless of hearer uptake—and *doxastic incentives*, incentives to get a hearer to believe something (294). Each

should upgrade our trust. If their incentives are neither aligned nor misaligned with our goals, we should have a middling level of trust. If their incentives are misaligned with our goals, we should decrease trust. Indeed, incentives are more relevant than other epistemically significant features of testifiers like expertise (298).[4]

While I have ignored some of the intricacies of McBrayer's view, having the basics on the table should be enough for now.

What Should We Want from Incentive Epistemology?

We can interpret the idea that conflicts of interest are relevant to rational epistemic trust modestly or strongly. Modestly, it might be that conflicts of interest are relevant to rational epistemic trust in the sense that (our beliefs that) our source's interests conflict with our epistemic goals *can* be a reason to lower our trust in our source. Or, more strongly, it might be that conflicts of interest are relevant to rational epistemic trust in the sense that we are *rationally required* to lower our epistemic trust in response to (our beliefs that) our source's interests conflict with our epistemic goals. These two alternatives strike me as importantly different. The first suggests that conflicts of interest are one, albeit important, factor among many that together determine the rationality of one's degree of trust in a source's testimony. The second seems to suggest that conflicts of interest *just are* an overring reason to lower one's trust.

I take McBrayer to be advocating something closer to the first interpretation. But distinguishing between the two interpretations prompts the question: what should we want from an incentive epistemology? How strong do we want an incentive epistemology to be?

At first, we might think the modest interpretation is all we could want. We all have various interests. We have a plethora of incentives pulling us in different directions, at various intensities nearly all the time. So, the stronger interpretation seems too strong—it misses the complexity of real-world interests and incentives. However, some adjusted version of the strong view seems desirable to me. A more powerful view might help us cut through the messiness of the real world. Ultimately, I think that an incentive epistemology should provide an account of when a testifier's interests we should want from an incentive epistemology is an account of when a testifier's interests and incentives rationally require us to lower our epistemic trust in their testimony.

Conflicts of Interest and Testimonial Defeaters

One route to this stronger version of the view starts by considering how (and whether) one must be aware of a purported conflict of interest to have a reason to lower one's trust.

---

type of incentive on the part of a speaker can either align or misalign with a hearer's epistemic goals, or neither (294–296). While this distinction is useful, it should not affect my later discussion and criticisms. So, I will gloss over it.

[4] Indeed, McBrayer argues that expertise *undermines* trust when interests conflict because experts are better equipped to deceive you (298).

McBrayer holds that believing that a testifier has a conflict of interest provides one with a defeater for testimonial belief (293).[5] Belief also seems necessary for such defeaters on his account. I wish to flag two issues here.

First, some of McBrayer's other remarks might suggest a different view. For instance, he writes that "[t]he fact that someone has incentives to testify or get us to believe a certain way is relevant to how much we should trust them" (298). Depending on how we interpret "relevant to", this comment seems to suggest that the *fact* itself is a defeater. McBrayer also writes that "[w]hen you find out that someone's incentives are misaligned with yours, your confidence in that source of information should decrease" (298). "Finding out" strikes me as closest to "learning" which suggests that the relevant psychological connection is knowledge. I do not mean to be pedantic.[6] Rather, I think these remarks open space for a different, more externalist, approach.

Second, note that couching things in terms of defeaters fits better with the modest approach. To see why, we need to talk briefly about defeaters. A defeater is just some piece of evidence or a reason that counts against some belief that p. We can further distinguish between *psychological* and *normative* defeaters.[7] A psychological defeater for the belief that p is a reason had by the agent that suggests that p is false or otherwise unacceptable. A normative defeater for the belief that p is a reason that an agent *ought to* have that suggests that p is false or otherwise unacceptable. Even if the agent fails to notice this reason, it can still defeat their initial belief.

Moreover, defeaters can themselves be defeated. For example, suppose that, after glancing out the window of my third-floor apartment into the street below, I come to believe that there is a Ferrari SF90 parked outside. But suppose that my friend later tells me that the car is actually a Lamborghini Aventador. My friend's testimony gives me a psychological defeater for my belief. But suppose further that my friend is very obviously unreliable at identifying cars, a fact I really ought to know. My previous psychological defeater is thus normatively defeated. This chain of defeated defeaters can go on and on. An *undefeated defeater* is what we are left with when we are presented with a defeater that is not itself defeated.

Now, one thing we might want from a stronger incentive epistemology is an account of when conflicts of interest generate undefeated defeaters, not merely defeaters. Distinguishing between psychological and normative defeaters helps us do this. Perhaps, as McBrayer holds, merely believing that a testifier has a conflict of interest can defeat testimonial belief. But this

---

[5] Indeed, McBrayer (293) writes that "[t]he conspiracy theorist has a reason to be wary of government testimony if she believes that the government has a conflict of interest". He identifies an upstream epistemic failing (like the conspiracy belief) as the real culprit.

[6] McBrayer (293) writes that little of his overall case hinges on his choice here. I agree that his fundamental insight will not be affected much. But the view I will later gesture at looks quite different in the end.

[7] I here set aside the undercutting and rebutting distinction. However, I take it that in most cases a conflict of interest constitutes an undercutting defeater because it undercuts the reliability of one's source. For more on the distinction between psychological and normative defeat in the context of testimony, see Lackey (2011, 73–75).

defeater can itself be easily defeated. That is, failing to do one's epistemic due diligence when forming beliefs about a testifier's interests and incentives might constitute a defeater-defeater. Alternatively, the fact that a testifier very obviously has a conflict of interest seems to generate a normative defeater regardless of one's belief.

But when will one have an undefeated defeater for testimonial belief stemming from a conflict of interest on the part of their source? There might be too many factors here to give a cohesive answer. Still, I will tentatively suggest one way forward.[8] One is rationally required to lower one's epistemic trust in a testifier because the testifier's interests conflict with one's epistemic goals if and only if one is in a position to know this about the testifier. Thus, sometimes the mere fact that a speaker has a conflict of interest is enough to defeat testimonial beliefs. Moreover, falsely believing that a speaker has a conflict of interest is improper grounds for downgrading trust in their testimony.

What about cases in which all of our evidence strongly yet misleadingly points to our source having a conflict of interest? Some might find my view implausible because it holds that we are not rationally required to lower our trust in such cases. I do not think this is so implausible. We can fail to properly track facts about another's interests and intentions. But our failure can sometimes be *excused*. In such cases our natural human capacities for detecting the interests and intentions of others may function normally but, perhaps due to one's body of evidence or features of one's epistemic environment, they still lead us astray.

To see how my (tentative) proposal might work, consider the following cases.

(Not a Salesperson): You are walking through a used car lot and are looking to buy. Interested in hearing someone else's opinion you see a man heading towards you. You see the man wearing a hat and shirt regularly worn by employees at the lot, you remember the man discussing car sales with other customers, and the man introduces himself as "something of a salesman". The man tells you that the car in front of you has high-quality brakes. On the basis of your evidence, you judge that he has a conflict of interest. However, the man is not a salesperson and has no conflict of interest.

(Not a Salesperson)*: All is as it is in (Not a Salesperson) except this time you do not have good evidence to think that the man has a conflict of interest. He is not dressed like a salesperson at the lot, he does not identify himself as such, and you have no background evidence supporting the judgment that he is a salesperson at the lot. However, you judge that he has a conflict of interest totally reflexively and not on the basis of evidence. The man tells you that the car has high-quality brakes. Again, the man is not a salesperson and has no conflict of interest.

---

[8] The following is exploratory and meant more as a sketch than a worked-out view. Indeed, I make some controversial assumptions about the nature of defeaters. For a picture of defeaters closer to the one I will assume, see Dutant and Littlejohn (2021).

(Missed Salesperson): Similarly to (Not a Salesperson), you have excellent evidence to support the judgment that the man has a conflict of interest. However, out of misplaced love of neighbor and the desire to see the best in others, you ignore your evidence and judge that the man is not a salesperson and does not have a conflict of interest. You judge that he is just a helpful passerby. The man tells you that the car in front of you has high-quality brakes. In this case, the man is a salesperson, and he has a severe conflict of interest.

(Missed Salesperson)*: Things are very similar here to (Not a Salesperson)*. That is, you have no evidence to support the judgments that the man is a salesperson, or that he has a conflict of interest. In light of your lack of evidence, you judge that the man does not have a conflict of interest. The man tells you that the car in front of you has high-quality brakes. Again, the man is a salesperson, and he has a severe conflict of interest.

In (Not a Salesperson) I think you are not rationally required to lower your epistemic trust in the man on the purported grounds that he has a conflict of interest. Since being in a position to know is factive, you are not in a position to know that he has a conflict of interest. Some may worry that this verdict is unintuitive. Let me try to assuage the worry.

First, note that in cases like (Not a Salesperson), one would be epistemically blameless for lowering their trust. Thus, while they have not complied with an epistemic norm, they are excused for their failure. Second, note that the norm I am considering is specifically about conflicts of interest. That is, even if one is not rationally required to lower their trust according to the norm we are considering, it might be that their body of evidence is strong enough that they are required to lower their trust for some other reason. That is, the same evidence that might mislead one to falsely think that a testifier has a conflict of interest might be good evidence for truly thinking that the testifier is a liar, unreliable, or simply mistaken. Relatedly, (Not a Salesperson)* strikes me as a case in which you are not rationally required to lower your epistemic trust. Moreover, you would not be excused were you to lower your trust because you would be doing so irrationally.

The first two cases are meant to show that believing that a testifier has a conflict of interest is not sufficient for being rationally required to lower one's epistemic trust on the purported grounds of a conflict of interest. (Missed Salesperson) and (Missed Salesperson)*, on the other hand, are meant to show that belief is not necessary. I take both to be cases in which you are rationally required to lower your epistemic trust in the man because you are in a position to know that he has a conflict of interest. In (Missed Salesperson) you would not be excused for failing to lower your trust. However, in (Missed Salesperson)* your misleading evidence excuses your rational failure.[9]

While it would require far more argument, I think this picture fits better with the view of epistemic autonomy with which we started—that of thinking well for oneself. Lowering

---

[9] We might instead want to say that you are not in a position to know here, and thus are not rationally required to lower your trust. This outcome still fits with my view. The details will depend on what we think it is to be in a position to know something.

one's epistemic trust in another because one mistakenly and inexcusably attributes a conflict of interest to the other does not, to me, seem like a way of thinking well for oneself. Rather, it seems like a failure to do one's epistemic due diligence. Likewise, failing to lower one's epistemic trust in another who is obviously in a conflict of interest seems again like a failure of (full) epistemic autonomy. My above proposal captures these intuitions. Moreover, we would ideally want our powers of trust to perfectly track facts about speakers' interests. The epistemically ideal truster would skillfully attune their trust to the *fact* that a speaker has a conflict of interest, not merely to potentially misleading signs of such facts. My above proposal incorporates this thought too. However, like skillful action, being virtuous depends on both *whether* one meets/fails to meet the relevant evaluative norm and *how* one meets/fails to meet it. Thus, in the above examples, we can be excused for improperly trusting by falsely attributing a conflict of interest to a testifier when our evidence is misleading. I think this feature of my proposal helps us see: (i) how proper trusting is important for the virtue of epistemic autonomy, and (ii) that proper trusting is a virtue worth cultivating for its own sake.[10]

The Moral Limits of Incentive Epistemology

Offering testimony has clear moral limits. For instance, plausibly one should not lie (perhaps with certain exceptions). But there are also moral considerations when it comes to receiving testimony. For instance, G. E. M. Anscombe (1979, 150) wrote that: "It is an insult and may be an injury not to be believed. At least it is an insult if one is oneself made aware of the refusal, and it may be an injury if others are."

Anscombe's thought, I take it, is that when an honest testifier offers you (what they take to be) good reasons, you can wrong them by not duly considering their reasons. Perhaps similar moral concerns arise when one disregards another's testimony because one attributes to them a conflict of interest.

We have been mostly dealing with cases in which a conflict of interest might consciously lead a speaker to lie or mislead. For instance, the used car salesman is fully aware that he stands to gain monetarily by lying to you about the breaks. However, interests and incentives can also influence us unconsciously.[11] For example, the corrupt politician might mislead her constituents by saying positive yet false things about a corporation. The politician may be strongly influenced to spread these falsehoods because the corporation in question makes large donations to her campaign. Her financial/professional interests conflict with the epistemic interests of her constituents. But the politician need not consciously decide to spew her talking points because of her financial interests for us to see her as being motivated by them. Indeed, she may even sincerely claim that she says these things because she believes them. Still, we want there to be room here to say that the politician is ultimately acting for a

---

[10] See Carter (2022) for one account of trust as a skill or virtue.
[11] This is not meant as a psychological claim. I just mean that one is not always consciously aware of how incentives impact the things they do and say.

different reason. We should want our account to say that it can be rational to lower one's trust in the politician because of her conflict of interest.

Cases like this are cases of *bad faith*. When one is in bad faith one avows a reason for their speech, actions, or attitudes that is not their real motivating reason.

Now, attributing a conflict of interest to a testifier is one thing. Claiming that this conflict is a reason, or perhaps *the* motivating reason, that they issued their testimony is another. We do the second, at least implicitly, when we lower our trust in a testifier because we attribute to them a conflict of interest. This seems to follow straightforwardly from McBrayer's view. Presumably, we care about conflicts of interest because they are situations in which a source is more likely to lie to us. In such cases, a speaker says what they do, at least in part, *because* of their interests. To lower one's trust in a speaker because one attributes to them a conflict of interest is to, at least implicitly, accuse them of being motivated by their interests or incentives to lie or mislead. So, attributing a conflict of interest to a speaker and subsequently lowering one's trust in them is to see them as being in bad faith. When we do this, it is like accusing the speaker of only saying what they do because of their interests and incentives, not for epistemically rational reasons.

To say that someone believes that p for psychological reasons—that is, to say something like "you only say that because you are paid to say that"—is to engage in what A.K. Flowerree (2023) calls a *psychologizing explanation*. Thus, when we lower our trust in a testifier because we attribute to them a conflict of interest, we are (at least implicitly) psychologizing them. But, as Flowerree points out, psychologizing can be harmful. It can wrong others by damaging interpersonal relationships, insulting those who offer sincere testimony, and shutting down honest conversation (Flowerree 2023, 972). This is because psychologizing does not give due course to another's reasons. Indeed, privately psychologizing another to oneself could be disrespectful because we treat them 'as if' they were sincere while silently disregarding them (973). So, psychologizing others has moral implications. I think incentive epistemology does as well.

Yet, it is sometimes morally permissible to psychologize a speaker. It is also surely sometimes morally permissible to lower our trust in response to a conflict of interest. After all, sometimes people really are in bad faith. Sometimes salespeople really do lie to scam you. So, according to Flowerree, what are the moral limits on psychologizing? Flowerree (978) argues that it is permissible to psychologize another person "when you are in a position to know the other person is in bad faith." I suggest that we follow Flowerree here. We risk harming a testifier if we lower our trust in them because we attribute to them a conflict of interest. We should only do so when we are in a position to know this.

The moral limits of incentive epistemology fit nicely with the epistemic limits discussed in the last section. Together they form a nice joint package. Furthermore, we might think that putting moral limits on incentive epistemology can help us think well for ourselves in a more robust sense. Indeed, like other kinds of autonomy, epistemic autonomy comes with moral limits.

Conclusion

Justin P. McBrayer's "*Caveat Auditor:* Epistemic Trust and Conflicts of Interest" (2024) is an excellent addition to a rich and philosophically rewarding special issue on epistemic autonomy. In this reply to McBrayer, I distinguished between a stronger and weaker reading of his view. Then I sketched one way of moving to the stronger view and explored whether there might be moral limits on epistemic trust. Future work on the connections between practical interests, rational trust in testimony, and epistemic autonomy will no doubt prove fruitful.

References

Anscombe, G. E. M. 1979. "What is it to Believe Someone?" In *Rationality and Religious Belief*, edited by C. F. Delaney, 141–151. South Bend: University of Notre Dame Press.

Battaly, Heather. 2021. "Intellectual Autonomy and Intellectual Interdependence." In *Epistemic Autonomy*, edited by Jonathan Matheson and Kirk Lougheed, 153–172. Routledge.

Carter, J. Adam. 2022. "Trust as Performance." *Philosophical Issues* 32 (1): 120–147.

Dutant, Julien and Clayton Littlejohn. 202. "Defeaters as Indicators of Ignorance." In *Reasons, Justification, and Defeat*, edited by Brown, Jessica and Mona Simion, 223–246. Oxford: Oxford University Press Oxford.

Flowerree, A. K. 2023. "When to Psychologize." *Australasian Journal of Philosophy* 101 (4): 968–982.

King, Nathan L. 2021. *The Excellent Mind: Intellectual Virtues for Everyday Life*. Oxford University Press.

Lackey, Jennifer. 2011. "Acquiring Knowledge from Others." In *Social Epistemology: Essential Readings*, edited by Goldman, Alvin and Dennis Whitcomb, 71–91: Oxford University Press Oxford.

Matheson, Jonathan. 2024. "The Philosophy of Epistemic Autonomy: Introduction to Special Issue." *Social Epistemology* 38 (3): 267–273.

Matheson, Jonathan. 2023. *Why it's OK Not to Think for Yourself*. Taylor & Francis.

Matheson, Jonathan. 2021. "The Virtue of Epistemic Autonomy." In *Epistemic Autonomy*, edited by Matheson, Jonathan and Kirk Lougheed, 173-194. Routledge.

McBrayer, Justin P. 2024. "*Caveat Auditor:* Epistemic Trust and Conflicts of Interest." *Social Epistemology* 38 (3): 290–301