WHERE TRACKING LOSES TRACTION

Mitchell Barrington

Dianoia Institute of Philosophy

ABSTRACT[1]

Tracking theories see knowledge as a relation between a subject's belief and the truth, where the former is responsive to the latter. This relationship involves causation in virtue of a sensitivity condition, which is constrained by an adherence condition. The result is what I call a *stable causal relationship* between a fact and a subject's belief in that fact. I argue that when we apprehend the precise role of causation in the theory, previously obscured problems pour out.

This paper presents thirteen distinct and original counterexamples to Nozick's tracking theory—many of which also constitute problems for more recent tracking theories. I begin by discussing how tracking relates to causation: Nozick invokes causation through conditions similar to those of Lewisian causal dependence. As a result, when causal dependence is not necessary for causation, Nozick fails to identify knowledge. I then address the inability of causation to capture epistemically important concepts, such as justification and truth. I conclude by discussing the underlying asymmetries between causation and knowledge that undermine any attempt to reduce knowledge to a purely metaphysical relation.

INTRODUCTION

Tracking theories grant knowledge to beliefs that track the truth: when the belief is true, it is believed; when it is not true, it is not believed. The appeal of these theories comes from seeing knowledge as a responsiveness of a subject to their environment where "the world's state is the independent variable and your beliefs are the dependent variables" (Roush 2009: 218). This paper explicitly addresses Robert Nozick's (1981) theory—as it stays truest to the idea of tracking and is by far the most widely discussed—though much of the following discussion is relevant to other tracking theories (e.g. Dretske 1971; Armstrong 1973; DeRose 1995; Roush 2005; Adams, Barker, and Clarke 2017).

The conditions for Nozick's theory are as follows, where p is any proposition believed by a subject S:

1.   p
2.   Sb(p)
3.   ~p → ~Sb(p)
4.   p → Sb(p)

The conditions state the following: 1) p is true; 2) S believes that p; 3) if p weren't true, S would not believe that p; 4) if p were true, S would believe that p (Nozick 1981: 176). Nozick uses possible world semantics to explicate the third and fourth conditions. The third condition (sensitivity) is understood as: in the closest possible worlds where p is false, S does not believe that p. The fourth condition (adherence) is understood as: in nearby possible worlds where p is true, S believes that p (Nozick 1981: 173). Nozick himself recognised that his account of knowledge via subjunctive conditionals was "no great leap" from Alvin Goldman's (1967) causal theory of knowledge (Nozick 1981: 689). However, once we understand the precise role that causation plays in Nozick's theory, counterexamples become easy to tease out.

Nozick's "tracking" conditions closely align with the conditions of David Lewis's (1973) counterfactual theory of causation. Lewis defines causal dependence (the concept he uses to analyse causation) as consisting of the truth of two conditionals. E causally depends on C iff:

1) In the closest possible world where C occurs, E occurs.

2) In the closest possible world where C doesn't occur, E doesn't occur (Lewis 1973: 563).

As a consequence of the alignment between Lewis's and Nozick's theories, beliefs recognised as knowledge by Nozick involve a relation of Lewisian causal dependence holding between p and b(p).

Nozick's subjunctive conditionals do not align exactly with Lewis's, however: the truth of Nozick's conditionals depend on multiple possible worlds (what Nozick calls the closest *neighbourhoods* of p and not-p) (1981: 681), whereas Lewis merely requires us to look at the closest *one* world (unless two or more are tied for closeness). This slight dissimilarity leads Nozick to require something slightly stronger than a causal relationship: the causal relationship must be strong enough to connect p and b(p), and not-p and not-b(p), respectively, despite minor changes in circumstances. This property of causal relationships I will call *stability*. As long as a stable causal relation holds between p and Sb(p), S will be granted knowledge of p by Nozick's account.

TRACKING FAILS TO MAP ONTO CAUSATION

Nozick's requirement for stable causation works well most of the time: most cases of knowledge seem to involve stable causation, and vice versa. However, there are instances where his theory fails to identify causation and thus fails to recognise knowledge. Both of the following cases concern problems for Lewis's counterfactual theory of causation, which translate into problems for Nozick's theory of knowledge. And while Lewis has the resources to deal with these problems, Nozick does not. Both of these issues would be solved if his conditions explicitly invoked causation, rather than operating through an imperfect picture of causation.

1. Preemption

As a result of the tracking theory's alignment with Lewisian causal dependence, cases where causal dependence is not necessary for causation can be translated into cases where Nozick's theory fails to recognise knowledge. This phenomenon (causation without causal dependence) arises in cases of *preemption* (Lewis 1973: 567). Suppose I throw a rock that breaks a window. The counterfactual theory of causation correctly identifies my throwing of the rock as the cause

of the window's breaking: if I had not thrown the rock, the window would not have broken. Though now suppose that, had I not thrown the rock, a bird would have flown into the window and broken it anyway. Now, the theory fails to recognise the causation because if I had not thrown it, the window would still have broken (on account of the bird).

Nozick considers a case of preemption (though without recognising it as such), prompting him to alter his theory. In this case, a grandmother is watching her grandson running around outside. Grandmother believes that Grandson is well, and she is right.[2] Although if he were sick, he would have stayed at home, and the family would have lied to her to spare her upset. Clearly, Grandmother knows that Grandson is well. But if he were *not* well, she would still believe he was (Nozick 1981: 179). Grandmother's belief that Grandson is well is caused by Grandson's being well. Yet if this cause did not occur, another cause of her belief would come into play: the family's false testimony. There is a second, potential cause that is preempted by the actual cause.

Nozick's solution is to stipulate that the method used to form the belief must remain the same in the other possible worlds we consider. He revises his tracking conditions as follows:

3. If p weren't true *and S were to use M to arrive at a belief whether (or not) p*, then S wouldn't believe, *via M*, that p.

4. If p were true *and S were to use M to arrive at a belief whether (or not) p*, then S would believe, *via M*, that p (1981: 179; my emphasis).

In the grandmother case, if Grandson were sick and Grandmother were to use the same method (presumably, looking at him) to determine if he is well, then she would see that he is not well, and would not believe that he is. Her belief, via this method, tracks the truth.

There is a generality problem involved with determining which method was used to form any belief. The problem is that any instance of a belief-forming method being is a token; in saying that the method must remain the same across possible worlds is to say each token method must be of the same type. But what is the type? When she looked at Grandson, did Grandmother use the

---

[2] Nozick's (1981: 179) case concerns the belief that Grandson is "well (or at least ambulatory)." For the sake of clarity, my case involves Grandmother simply believing that "Grandson is well." I thank the reviewer for pointing out that my argument would not work if I were to use Nozick's case (where Grandmother believes the disjunction).

4

method of "eyesight," "looking at Grandson," "looking at objects that are roughly twenty feet away," or "looking at family members at 11:20 am on a Thursday"? Nozick does not offer a principled way of typing the method.[3] But moreover, there are cases of preemption where the tracking theory blunders no matter *how* we type the method.

Nozick's requirement of holding the method fixed amounts to a constraint on the content of the causal chain. The modification is somewhat analogous to altering the counterfactual theory of causation to require that "if A had not occurred, C would not have occurred *as a result of an A-type event*." In the example of my rock-throwing causing the window to break, we would specify that only *rock-throwing* (and not birds) can cause the window to break. It solves the problem, but it is not hard to think of a new example that it cannot solve: suppose that, were I not to throw the rock, then someone else would have thrown a rock at the window. Now, if I had not thrown my rock, the window would still have broken *on account of rock-throwing*, and so the theory fails to recognise my rock-throwing as the cause.

Nozick only gets away with this solution because looking at Grandson playing is clearly a different method from trusting the family's testimony. Nonetheless, similarly to the another-rock-thrower example, we can make the issue arise again by supposing the preempted method is of the same type as the original method. For example, we can construct a case where if Grandson were not well and Grandmother uses the method of looking at him, he would be reading a book (instead of running around); and Grandmother would still think he is well, just interested in the book he is reading. Once again, her belief is not sensitive. We can avoid *this* error by cutting the method along even finer margins: instead of "looking at Grandson," we could suppose the method used was "watching Grandson playing outside." But even then, we merely have to suppose that if Grandson were unwell, he would be playing lazily in the sandpit, and Grandmother would think he was just tired. If we cut it even finer, as "watching Grandson run around," we could suppose that if he were not well, he would still be running around, but sluggishly, and Grandmother would think he is just getting fat. In each case, we still want to say she knows he is well when she sees him zip around. The preemption problem for the tracking theory amounts to this: however we type

---

[3] This point has been discussed in (Williamson 2000; and Roush 2005: 39). In view of these issues, throughout the rest of the paper I will explicitly state how I think how I think the method should most plausibly and charitably typed in each instance.

the belief-forming method, we merely need to suppose that using this method preempts using another method *of the same type* (whatever that may be).

Even if we were to type the method so finely that there could be no pre-empted method that would fall under the same type, we would lose sensitivity for a different reason. For instance, suppose we type the method as "watching Grandson zipping around." In considering the third condition, we are now taken to a possible world where not only is Grandson unwell, but he still zips around. The method is typed so finely that it restricts the evidence (Grandson's zipping around) from changing. Consequently, there is no way Grandmother could discriminate between him being well and unwell, and so she fails the sensitivity condition. Typing methods so finely results in no belief ever being sensitive, because our evidence is so restricted that it is incapable of discriminating between p and not-p. Grandson's zipping around is good evidence that he is well, but if the method is cut so finely, he will zip around even if he is not well.

David Lewis's (1973) counterfactual theory of causation solves problems of preemption by analysing causation in terms of *chains* of causal dependence, rather than causal dependence *simpliciter*. So, even though Grandmother's belief is not causally dependent on Grandson's being well (due to the preemption), they are connected by a chain of causal dependence via her looking at him. In merely requiring causal dependence, Nozick's theory fails to latch onto causation properly.

2. Similarity and Backtracking Counterfactuals

Consider a causal chain where an event causes some fact, and also causes a subject to form a belief in that fact. For instance, if I see a nearby lightning strike and form the belief "I am about to hear thunder," the lightning causes me to believe I will hear thunder, and also causes me to hear the thunder. Whether or not this chain is knowledge-producing depends on what kind of counterfactual we employ.

We should notice that *if* we are to make p not occur, then either we must make E not occur (making p not occur by removing its cause), or we must sever the causal relationship between E and p. For example, if we are to make it the case that there is no impending thunder, then we must

make it the case that either 1) the lightning never struck, or that 2) the lightning did not cause the thunder. The former utilises what Lewis calls a *backtracking* counterfactual; the latter utilises a *non-backtracking* counterfactual (Lewis 1979).

Lewis argued for the use of non-backtracking counterfactuals in his theory (Lewis 1973: 566), but Nozick *must* permit backtracking counterfactuals, or else the above structure of belief will never produce knowledge. Consider my belief that "I am about to hear thunder," formed on the basis that I have just seen lightning. If I weren't about to hear thunder, and we use the non-backtracking counterfactual, then I would still have seen lightning and would thus still believe that I am about to hear thunder; my belief is not sensitive. When using the backtracking counterfactual, however, the chain is rightly recognised as knowledge-producing: if I weren't about to hear thunder, I would not have seen lightning nearby, and thus would not believe I will hear thunder. So, unlike Lewis, Nozick must use backtracking counterfactuals. The reason Nozick is tied to backtracking counterfactuals is that, in his view, knowledge arises when the fact is causally dependent on the evidence, and so severing the causal relationship between the evidence and the fact will fail to produce knowledge.

The problem for Nozick is that sometimes we have no choice but to use non-backtracking counterfactuals. To create such cases, we merely construct the world so that changing the truth value of p via changing the causal relationship between E and p is less of a departure from reality than backtracking to prevent E from occurring in the first place.

Let us take an example. Working under a government with an aggressive foreign policy, I am given the order to authorise the dropping of a massive nuclear bomb on a hostile country. As I assess the severity of the impact, I confirm that the effects were as expected: the entire country has effectively been wiped from the earth. Since I know my friend Frank was in this country, I form the belief that Frank is dead.

I will add one complicating detail to the scenario. There is a way for an individual to survive this particular blast, though this method is not known to anyone. For whatever reason, one will survive if, at the moment of impact, they are covered in tin foil, standing on their head with their mouth full of tomato juice and humming the birthday song. Given my evidence both of the power of the bomb and the resulting conditions, it seems uncontroversial to posit that I *know* that Frank

is dead—despite the fact that there is one obscure and unknown way to survive the blast. Nozick disagrees.

In the closest possible world where Frank is not dead, do I still believe he is dead? It seems we have a couple of kinds of possible world where Frank is alive. In the backtracking counterfactual, Frank survives because the bomb was never dropped in the first place. In the non-backtracking counterfactual, the bomb was dropped, but it does not kill Frank because he happened to be doing the survival method. The question we are to ask is, which scenario is closer to reality? The backtracking counterfactual would involve preserving just about every person, animal, and object in the country. The non-backtracking counterfactual would involve a relatively minor alteration to the life of Frank. Clearly, changing Frank's life (to make him be performing the survival method) would be far less significant than the changes we would have to make to erase the dropping of the bomb entirely. So, the closest possible worlds where Frank survives are those where he was doing the survival method. And if Frank *were* performing the survival method, I would still believe that he is dead: my belief is not sensitive. Thus, Nozick's theory gives us the implausible result that I don't know that Frank is dead.

The reason the case of the nuclear bomb has given rise to an insensitive belief is that the case has been constructed so that the cause (the bomb being dropped) is so significant that it is less of a departure to sever the causal relationship between the cause and the event than it is to backtrack and remove the cause entirely. Consequently, we are forced to use the non-backtracking counterfactual. Most cases escape this problem because, if E gives us good evidence for p, then if p were false, then E would not have occurred. However, if we manipulate the world so that removing E requires substantial enough changes, then we become inclined to keep E and make it not cause p.[4]

---

[4] The problem does not arise for Lewis's theory of causation because he is not obliged to recognise causation between p and b(p). In fact, he goes to great lengths to avoid a causal relationship arising in such a case, addressed as "the problem of epiphenomena" (Lewis 1973: 566).

CAUSATION FAILS TO CAPTURE EPISTEMIC CONCEPTS

Before we resolve that Nozick should have explicitly required (stable) causation—rather than trying to invoke it through causal dependence across multiple worlds—we will see that there are problems with having causation play such a central role in epistemology. Causation fails to recognise epistemic concepts that are important to the analysis of knowledge: justification and truth.

Justification

Whether evidential justification is necessary for knowledge is controversial. Externalists, like Nozick, instead hash out their justification conditions in terms of things outside the subject's cognitive access (BonJour 2002). And while I take no stance on the issue here, we ought to take as a minimal constraint of externalist theories that they do not grant knowledge to beliefs formed in stark contrast to the available evidence. The first example of this sort concerns cases where a belief tracks the truth but is formed by ignoring one's evidence; here, tracking is not sufficient for knowledge. The second example concerns cases where a belief is formed on exceedingly good evidence but fails to track the truth; here, tracking is not necessary for knowledge. From these cases, we find that when any externalist theory so starkly opposes one's propositional justification, it is difficult to side with the theory.

3. Propositional Justification: Insufficiency

Suppose I go to my doctor to get tested for a disease. The doctor gives me a choice. She explains that Test A, which was used in the past, has recently seen overwhelming amounts of evidence against its reliability. In fact, it has been discredited so much that it will soon no longer be used. But since its use is still legal, I can choose to use it if I want. Alternatively, I can use another test, which is considered to be highly reliable. Of course, there is a twist: it turns out that Test A is actually exceedingly accurate. The discrediting evidence contains an unbelievable number of coincidences. It is overwhelmingly unlikely, given my evidence—yet nonetheless true—that Test A is reliable.

Suppose I choose to take Test A, have it come back positive, and form the belief that I have the disease. Surely I did not *know* that I had the disease: it is extremely lucky that a test with such a copious amount of evidence against its reliability turned out to be accurate. Still, if I did not have the disease, and I use the method of trusting Test A to determine whether or not I have it, the test would come back negative, and I would not believe it. If I were to have the disease and use this method, the test would come back positive, and I would believe it. Here, my having the disease caused the test to come up positive, and the test result caused me to believe I have the disease in a stable causal chain. Nevertheless, what no relation of causal dependence can take account of is whether I am propositionally justified in trusting that such a causal chain exists. In this case, I was not.[5]

4. Propositional Justification: Non-Necessity

Plato's early dialogues paint a picture of Socrates facing his impending death with admirable nobility. Indeed, we can only wonder what he was thinking at the moment when he was made to drink the hemlock. It seems we could never know. That said, I know he was not thinking about the new *PlayStation*. Look, I am pretty certain Socrates was not thinking about the new *PlayStation* when he was handed the hemlock; there are very few things I would more confidently claim to know. Nevertheless, in the closest possible worlds where Socrates *is* thinking about the new *PlayStation*, I would still believe that he was not. In this case, I have exceedingly strong evidence for my belief, but if we reach out to the (absurdly far-away) possible worlds where Socrates *was* thinking about the new *PlayStation*, my belief will not change accordingly. It is difficult to see why my beliefs in such far-off possibilities determine what I can be said to know in this world. Strong, knowledge-producing evidence does not need to be a part of the causal chain that connects the fact to our belief.

---

[5] I should note that this counterexample hinges, in some degree, on the importance of evidence for knowledge. Nozick, an explicit externalist, would deny this importance, and those who sympathise with this view may not buy the premise of the counterexample. I thank the reviewer for pointing this dependence out.

5. Doxastic Justification: Typing Generality

While it is not difficult to strip tracking conditions away from propositional justification, the tracking theory does better with aligning with doxastic justification. Propositional justification concerns whether the information we have justifies us in holding a belief; doxastic justification, on the other hand, concerns the specific information on which we actually form our belief. And since Nozick's theory demands that our belief-forming method produce truth-tracking beliefs, these belief-forming processes are almost always doxastically justified. However, when we attribute doxastic justification to a belief-forming method, we "type" it more broadly than Nozick's method stipulation does. Consequently, we can create cases where our method is unreliable typed broadly, and thus doxastically unjustified, but reliable typed narrowly, and thus truth-tracking. This unjustified belief is ascribed knowledge on Nozick's theory.

Suppose there is a president of a country who has done a great job governing for the past twenty years. Although, last week he hit his head and went into a coma. He has just woken up with amnesia and has forgotten everything about his presidency. While in hospital, he reads an article that documents his time in power. The article accurately reports that he was a great president. Despite not having the resources to assess the article's veracity, it strokes his ego. And out of arrogance and pride, he forms the belief that the article is accurate. Of course, he does not *know*, but he tracks the truth.[6]

Sensitivity asks whether, in the closest possible worlds where the article is *not* accurate, and the president uses the "did it stroke my ego?" method, he would still believe that it is accurate. We have two types of worlds to consider here. First, there are the worlds where the article is inaccurate by virtue of saying he was a good president when he was, in fact, poor. Second, there are the worlds where the article is inaccurate by saying he was a poor president when he was, in fact, good. The first kind of world is far from actuality because it would require changing his governance throughout his entire presidency so that he did a mediocre job; the article remains the same, saying that he was a good president. The second kind of world merely requires us to change the article, so that it does *not* say he was good, when in fact he was. Clearly, the second kind of

---

[6] The core of this example was given by (Roush 2005: 39) as an example of where Nozick's theory works well. With my changes, however, it constitutes a counterexample to both Roush's and Nozick's tracking theories.

world is closer than the first. In these worlds, the article would not stroke his ego, and he would thus not believe it is accurate. His belief, therefore, is sensitive.

Adherence asks whether, in the closest possible worlds where the article *is* accurate, and the president uses the "did it stroke my ego?" method, he would still believe that it is accurate. Once again, we *could* change his entire presidency, so that the article accurately describes his poor presidency, but the closer worlds are clearly those where almost everything is left the same so that the article accurately describes his great presidency. In these worlds, the president's ego is stroked, and so he believes that the article is accurate.

The president tracks the truth here because there is a stable causal chain connecting his great presidency to the article reporting that his presidency was great, to his believing that the article is accurate (in virtue of his conceit). The reason his belief-forming method tracks the truth but is nevertheless not doxastically justified is that judgments of truth-tracking type the method more narrowly than those of doxastic justification. Nozick's method requirement asks whether the method—as used by this particular person, for this particular belief—is reliable. When we assess doxastic justification, however, we ask whether the method used—when used by people *generally*—is reliable. For instance, *the president's* use of the "did it stroke my ego" method when forming a belief about *his presidency* is reliable and will consistently give him true beliefs—since he was, in fact, a good president. However, when we assess the belief's doxastic justification, we ask whether it is a good belief-forming method for people *generally*. And whether a statement strokes someone's ego says nothing about whether it is true.

6. Doxastic Justification: Inverse Relationships to the Truth

We can also create a mismatch between a method's tracking the truth and doxastic justification by having the causal chain take on a negative relationship to the truth. In these cases, the chain rewards poor belief-forming practices. For example, suppose I am having dinner with my husband's friend, a doctor, with whom he previously had a relationship. Naturally, I do not like this man. I dislike him so much that whatever he says, I think to myself, "I bet he's wrong about that"—blindly forming beliefs in contradiction to whatever he says. When he tells me that antibiotics treat influenza, I form the belief that "antibiotics do *not* treat influenza." As it turns out, this man is

almost as petty as me: he is nefariously feeding me misinformation out of jealousy of my husband's and my relationship. So, when he told me that antibiotics treat influenza, he told me something he knew to be false. Of course, believing the opposite of what someone says because you dislike them is an irresponsible belief-forming practice, and I certainly do not *know* that antibiotics do not treat influenza. (My only evidence is that I just heard an expert say that they *do*!) Nevertheless, I track the truth.

If antibiotics did treat influenza, and I used the method of contradicting the doctor's testimony, then the doctor would know they do and, in order to deceive me, tell me that they *don't* treat influenza (or tell me nothing at all). Consequently, I would not believe that "antibiotics treat influenza." My belief is sensitive. In nearby worlds where antibiotics do not treat influenza, and I use the method of blindly contradicting the doctor's testimony, the doctor will tell me that they do, and I will believe that they don't. My belief is adherent.

In this example, we create a causal chain that takes on a negative relationship to the truth. Good belief-forming habits (such as believing the expert) will, in this case, form wrong beliefs. This phenomenon occurs on account of the doctor being reliable regarding this fact, but also lying. The result is that his testimony is reliably *wrong*. Consequently, beliefs formed in blind contradiction to his testimony will be reliably correct, but nevertheless doxastically unjustified.

Truth

As we have seen, Nozick's tracking theory requires that one's belief be stably connected via a causal sequence to the fact (or the truthmaker of the fact) it concerns. Other beliefs that the subject holds can constitute members of this sequence. As long as the chain stably connects p to b(p), Nozick's theory will count it as knowledge. The problem is that a belief inferred from a false belief should not be knowledge, but Nozick's theory has no way of excluding false beliefs from the causal chain. The following two cases concern two ways in which a stable causal chain can use false beliefs to transmit tracking properties onto true beliefs—which Nozick misapprehends as knowledge. In both cases, the subject tracks the truth but does not know.

7. Inference Through Falsity

Nozick's complicated relationship with inferential knowledge and closure is well documented:[7] Nozick's theory fails to attribute knowledge to propositions that have been inferred from a known proposition. However, there is also a problem on the contrary: false beliefs can occupy a place in a stable causal chain. Of course, these beliefs will not be considered knowledge by Nozick because they fail the truth condition. However, if this belief goes on to cause a true belief, Nozick will attribute knowledge to the true belief.

Suppose I have a friend named John Smith. John, unbeknown to me, has a brother: Jack Smith. One day I am walking along the street and walk past Jack—whom I mistake for John. After the encounter, having believed I saw John, I form the belief that I have just seen a member of the Smith family. Now, I do not *know* that I have seen a member of the Smith family: I do not even know the person I saw, or that he even existed! Nevertheless, I track the truth.

If I had not seen a member of the Smith family, I would not have seen Jack, mistaken him for John, formed the belief that I had seen John, and consequently formed the belief that I had seen a Smith. In slightly changed circumstances where I *do* see a Smith, I will see Jack, think he is John, form the belief that I saw John, and then form the belief that I have seen a Smith. (Alternatively, if I saw another Smith, we can assume I would have correctly identified them and believed that I saw a Smith.) Nozick's account correctly excludes my belief that I saw John from being a case of knowledge. However, this false belief causes me to form the more general belief that I saw *a Smith*, which is stably connected by a causal sequence to my actually seeing a Smith. If I have a false belief in p, I should not (in ordinary cases) be able to come to know q from inferring it from p. Nevertheless, Nozick's theory allows tracking to be retained through a causal sequence containing a false belief.

8. Counteracting Errors

Suppose I am on vacation in Australia; and though I love the beach, I am afraid of sharks. As I am about to go in the water, I remind myself that shark fins are hooked, whereas dolphin fins are

---

[7] E.g., Kripke's (2011) treatment is quite comprehensive.

straight. However, I am mistaken: it is the other way around. Nonetheless, I see a hooked fin, think it is a shark, and get out of the water with haste. Despite my ignorance, my belief that I have seen a shark is correct: for a bizarre reason, this hooked fin belongs to a shark. You see, the Australian government noticed that tourists have stopped coming to Australia due to fear of sharks. So, they implemented a secret program to cut all sharks' fins to look like dolphin fins. Soon thereafter, environmental groups noticed the apparent decrease in the number of sharks. To satisfy the environmentalists, the government cut the fins of dolphins to look like shark fins. (Perhaps cutting the sharks' fins again would threaten their health.) All that is to say, now shark fins look like dolphin fins, and vice versa.

As a consequence of this strange series of events, my belief that I saw a shark is not only correct, but it tracks the truth. In the closest possible worlds where I did not see a shark and used the fin-identifying method to form a belief about whether I have seen a shark, I would have seen a dolphin with a shark-looking fin and believed it was a dolphin, or I would have seen nothing. Either way, I would not believe I have seen a shark. In nearby possible worlds where I *have* seen a shark, and I use the method of identifying its fin, I will see a dolphin-looking fin and believe it is a shark.

Here, there is a stable causal chain connecting my seeing a shark to my believing I have seen a shark. My belief is caused by the event of me seeing the shark, in conjunction with my external belief about the shape of shark fins. This external belief is false, and it predisposes me to transmit falsehood onto the belief in question. However, if we add another error to the causal sequence (in this case, my ignorance of the government's shark-cutting program), which *also* predisposes me to err, I will end up with a true belief; the errors counteract. Further, since the true belief is connected to the fact via this stable causal chain, the tracking properties are retained. Nozick suffers because a causal chain cannot distinguish true beliefs from false beliefs as members of the causal chain.

THE CAUSATION RELATION AND THE KNOWLEDGE RELATION

While the preceding section concerned problems with causation capturing *aspects* of knowledge, the following section concerns broader asymmetries between causation and knowledge. In these

cases, there is an asymmetry between the causal relationship holding and the subject occupying a positive epistemic position. That is, a causal relationship between a fact and a subject's belief does not map onto the *knowledge* relation between a belief and the fact. The first case concerns how a subject can unwittingly be the cause of their own belief's truth. The next three discuss ways in which the subject can misunderstand the causal chain, yet remain stably causally connected to the fact. I conclude by identifying what I believe is the fundamental asymmetry between the metaphysical relation of causation and the epistemic property of knowledge—which, I argue, prevents us from reducing knowledge to a purely metaphysical relation, such as causation.

9. Subject Causality

While it is an advantage that Nozick does not require the fact to cause the belief (thus allowing for knowledge of the future), his theory is vulnerable to cases of reversed causation—cases where one's holding of a belief causes itself to be true. Nozick recognised this problem and admitted to not being able to solve it without *ad hocery* (1981: 195-96). Despite recognising part of the problem, Nozick did not recognise its extent.

Suppose that I wake up one morning after a virus has spread, leaving me the only person alive. As I live by myself in the countryside, I am not aware of what has happened. Because of my religious upbringing, I have believed in God my whole life, and have been so sheltered from society that I think everyone believes in God. Clearly, my belief that everyone believes in God is irrational. However, since I am the only one alive, and since I believe in God, my belief that *everyone* believes in God turns out to be true. I obviously do not *know* that everyone believes in God, but this belief is caused by my own belief in God, and my belief in God also causes it to be the case that everyone believes in God (since I am the only person alive). There is thus a causal chain connecting my belief in God to my belief that everyone believes in God, and this belief's truth. Consequently, I track the truth. However, I certainly do not *know*.

In the closest possible worlds where *not* everyone believes in God, and I use the method of attributing my own beliefs to everyone, do I still believe that everyone believes in God? Well, there are two contenders for the closest possible worlds. Firstly, there are worlds where someone else survives the virus, and they do not believe in God. Secondly, there are worlds where I do not

hold my belief in God. The second kind would require us to erase one belief of mine; the first kind would require us to preserve the life of a whole other human—along with their myriad other beliefs. Clearly, it is closer to actuality to merely change my belief. And in such worlds where I do not believe in God, were I to use the method of attributing my belief to everyone, I would not form the belief that everyone believes in God. So, my belief is sensitive. In the nearby possible worlds where everyone believes in God, and I use the method of attributing my belief to everyone, would I believe that everyone believes in God? By necessity, I would. So, my belief is also adherent.

These kinds of counterexamples do not only concern beliefs. We might suppose, in the same scenario, that I go on daily walks, assuming that everyone does (via the method of attributing my *habits* to everyone). If it were not the case that everyone went on daily walks and I use the method of ascribing my habits to others, I would not go on daily walks, and thus wouldn't believe that everyone does. If everyone went on daily walks, and I use the method of ascribing my habits to others, I will believe everyone goes on daily walks.

These cases concern subjects who unwittingly cause their own beliefs to be true. Providing this causal connection is stable, Nozick calls it knowledge.

10. Skipped Links in the Causal Chain

Suppose I am sitting on the couch alongside my cat, Whiskers, as a thunderstorm is approaching. When thunderstorms are nearby, I get scared. Though it is not the thunder that scares me; it is the lightning. As a matter of fact, I have been deaf my whole life, so I do not even hear the thunder. Instead, my lack of hearing has heightened my sensitivity to light, making nearby flashes of lightning overwhelming. I do not realise that it is my unique condition that makes me scared of lighting, and I assume that everyone gets scared by lightning. Now, as lightning strikes nearby, I get scared. I also form the belief that Whiskers will get scared. Whiskers is not fazed by the nearby lightning, but the loud clap of thunder does scare him. As I see Whiskers jump, I think to myself "I was right; he *did* get scared."

I did not *know* Whiskers would get scared: I thought he was scared by the lightning, whereas he was actually scared by the thunder—a phenomenon of which I was not even aware! If

I were to try to prevent him from getting scared, I would draw the curtains, which would do nothing to stop him from getting scared by the thunder. Still, since the lightning caused my belief and also caused the thunder, which caused my belief to be true (it caused the cat to be scared), there is a stable causal chain connecting my belief to the fact.[8]

So, if Whiskers weren't about to get scared (and I were to use the method of reasoning based on my own fright), then there would have been no thunder to scare Whiskers, and consequently no lightning to scare me; and so I would not have believed Whiskers would get scared. Conversely, in slightly changed circumstances where Whiskers *is* about to be scared (and I use the method of attributing my fear to Whiskers), the lightning would have been close enough to scare me, causing me to believe that Whiskers would be scared. My belief tracks the truth. Nevertheless, since I have missed a crucial link in the causal chain—the thunder, which is the entire reason *why* Whiskers gets scared—I can not claim to have knowledge. In these cases, being connected to the fact by a stable causal chain does not produce knowledge if I am not aware of *how* I am so connected.

11. Superfluous Links in the Causal Chain

Similar to how knowledge is lost when we fail to apprehend important links in the causal chain, convoluting the causal chain with superfluous links can cost us knowledge, yet retain tracking. Suppose I am about to leave to take an exam at university when I see that it is forecast to hail. Since my roommate will be in the room above my exam room, I ask him to warn me that it is hailing by stomping on the floor, so I will hear and be able to move my new car to shelter. As I am in the exam, I hear pounding coming down on the ceiling. I immediately form the belief that my roommate is warning me, and consequently form the belief that it is hailing. Sure enough, I run outside, and it is. Excited by the genius of my plan, I engage in some epistemic flattery, claiming to myself, "I knew it!" When I get to the parking lot, however, I see my roommate arriving on

---

[8] Just to make sure, we might suppose that any lightning nearby enough to scare me, and make me think Whiskers would be scared, is also nearby enough to cause loud enough thunder to scare Whiskers. And conversely, any thunder loud enough to scare Whiskers is close enough to be preceded by a bright enough lightning strike to scare me.

campus. Confused about how he warned me about the hail if he is only just arriving, I look back at the building I was in. I realise my error: there is not a room above the exam room; it has its own roof. The pounding I heard was the hail coming down on the roof—and could not possibly have been caused by stomping in the room above!

I do not *know* it was hailing: my rationale, that I heard my roommate stomping on the floor, was entirely mistaken. Still, my belief tracks the truth. In the closest possible worlds where it is not hailing, and I use the method of listening for pounding on the ceiling, I will not hear anything and thus will not believe that it is hailing. In slightly changed circumstances where it *is* hailing, and I use the method of listening for pounding, I will hear the hail coming down on the roof, believe it is my roommate warning me, and believe that it is hailing. So, my belief tracks the truth.

12. Distinct Causal Chains

Not only can we arrive at a truth-tracking belief by erroneously adding or omitting links in the causal chain, but we can also create an entirely new causal chain in our heads that—as long as one link connects it to the real chain—will produce a truth-tracking belief.

Suppose we are at war with the island next to ours. We bomb them, then watch for an indication of their surrender—under the impression that our bombing did severe damage. They fire a flare into the sky, which we take to be a sign that they have surrendered. As we start celebrating, I form the belief that they have surrendered. However, the flare was not a surrender signal. They accidentally fired it while hastily preparing to strike back. In actuality, our bombing did virtually no damage to them: they have the resources to deal comfortably with our attacks, and there was no chance of them surrendering on account of our bombing. Nevertheless, they only had one flare, and they were only engaging in the war under the assumption that *if* they were struggling, they could call for support via the flare. The fact that they accidentally used the flare means that they have no backup and will have to surrender. So, in the absence of a flare, they wave a white flag, signalling their surrender—which we, being busy celebrating, do not even see.

I hold the true belief that the enemy has surrendered. However, I certainly don't *know* that they have surrendered: I formed the belief under the impression that our bombing did so much damage that they had no choice but to fire a flare to signify their surrender. In reality, our bombing

did nothing to them, and the flare did not signify their surrender. Nevertheless, if they had not surrendered, they would not have fired the flare, and so I would not have seen the flare and formed the belief that they were surrendering. Further, in slightly changed circumstances where they *had* surrendered, were I to use the method of looking to the sky for the flare, I would have seen the flare, and formed the belief that they had surrendered. I track the truth because, while the causal chain operating and the causal chain I *believe* to be operating are distinct, they are nonetheless connected by one event—the firing of the flare.

13. Causation in Epistemology

As we have seen, whether or not a belief tracks the truth is a matter of there being a stable causal chain connecting p to b(p). However, the existence of such a chain does not speak to the epistemic position of the subject. It might be noted that Alvin Goldman's (1967: 363) causal theory solves the preceding four cases by requiring that the subject be able to reconstruct the causal chain that produced their belief adequately. This criterion is an attempt to align the metaphysical notion of causation with the epistemic notion of knowledge. It accounts for cases where the subject fails to properly apprehend the causal chain connecting p to b(p). However, this alignment is precarious, and underlying problems can be teased out, as the following example illustrates. In this case, even a stable causal chain that has been understood perfectly can fail to produce knowledge.

Suppose I am sitting in my house when the power goes out briefly. Being inclined to supernatural explanations, I form the belief that a ghost walked through the wall, interrupting the circuit. Given what I know, it is far more likely that there was a brief supply shortage in the power grid. Nevertheless, let us suppose my unlikely explanation is the true one. Nozick rightly denies this belief is knowledge because the causal chain is not stable: there are close possible worlds where a power shortage (not a ghost) causes the power to go out. However, we can make this belief track the truth by making changes to the way the world is, in order to make worlds where other explanations hold further from actuality. Despite these worlds being further from actuality, my epistemic position does not change.

Take the same example concerning the power outage, but now suppose that, unbeknown to me, my house is hooked up to thousands of backup power supplies. The probability of all these

failing at once is approximately zero. Since I am not aware of these power supplies' existence (if I were, the power-supply-failure hypothesis would be discredited), it is no *more* reasonable to believe a ghost walked through my wall in this case than in the original case. However, the worlds where a power supply failure occurs *are* further from actuality now. This change, while epistemically irrelevant, does affect the truth of the theory's two subjunctive conditionals (the third and fourth conditions).

We can bring this point into full force by reconstructing the example. Suppose that aliens have been in control of human affairs since their origin and have spent decades monitoring my house to ensure the power does not go out. The aliens have the technology to make sure my power does not go out for any reason at all, with one exception: ghosts. Any time a ghost walks through the wall, the power will go out. Now, the world is so constructed that the *only* nearby worlds where the power goes out are ones where ghosts walk through the wall. All possible worlds where my power goes out for any reason *except* ghosts are extremely far away from actuality.

The causal chain between a ghost walking through my wall and my belief that a ghost walked through my wall has been made stable by these changes to our world. If a ghost had not walked through the wall and I used the method of looking for a power outage, the power would not have gone out (since nothing else could have caused the outage in any nearby world), I would not have attributed it to ghosts, and thus would not have believed that a ghost walked through the wall. In slightly changed circumstances where a ghost does walk through the wall, and I use the method of checking for a power outage, the power will have gone out, and I will believe that a ghost walked through the wall. Not only is the causal chain stable, but I understand it perfectly. Nevertheless, my epistemic position is no different than in the original example, and I certainly do not *know* that a ghost walked through the wall.

These alterations change which possible worlds are nearby, and thus which ones we look at when assessing whether I have knowledge, on Nozick's account. However, the question of whether the ghost hypothesis is the only one that could have occurred without making drastic changes to the world, and the question of whether it is worthy of epistemic praise, are two very different questions: one concerns the way the world is, and the other concerns the evidence I have for believing how the world is. It is the latter with which the analysis of knowledge concerns itself.

References

Adams, F., Barker, J. A., and Clarke, M. 2017. 'Knowledge as Fact-Tracking True Belief.' *Manuscrito,* 40: 1-30.

Armstrong, D. M. 1973. *Belief, Truth and Knowledge*. Cambridge: Cambridge University Press.

BonJour, L. 2002. Internalism and Externalism. In *The Oxford Handbook of Epistemology*, ed. P. K. Moser, 234-64: Oxford University Press.

DeRose, K. 1995. 'Solving the Skeptical Problem.' *Philosophical Review,* 104: 1-52.

Dretske, F. 1971. 'Conclusive Reasons.' *Australasian Journal of Philosophy,* 49: 1-22.

Goldman, A. 1967. 'A Causal Theory of Knowing.' *Journal of Philosophy,* 64: 357-72.

Kripke, S. 2011. Nozick on Knowledge. In *Philosophical Troubles: Collected Papers*, ed. S. Kripke, 162-224. Oxford: Oxford University Press.

Lewis, D. 1973. 'Causation.' *Journal of Philosophy,* 70: 556-67.

———. 1979. 'Counterfactual Dependence and Time's Arrow.' *Noûs,* 13: 455-76.

Nozick, R. 1981. *Philosophical Explanations*. Vol. 92. Cambridge, MA: Harvard University Press.

Roush, S. 2005. *Tracking Truth: Knowledge, Evidence, and Science*. Oxford: Oxford University Press.

———. 2009. 'Précis of Tracking Truth.' *Philosophy and Phenomenological Research,* 79: 213-22.

Williamson, T. 2000. *Knowledge and Its Limits*. Oxford: Oxford University Press.