Truth-tracking and the Problem of Reflective Knowledge

Joseph Salerno
Saint Louis University

In "Reliabilism Leveled" Jonathan Vogel (2000) provides a strong case against epistemic theories that stress the importance of tracking/sensitivity conditions. A tracking/sensitivity condition is to be understood as some version of the following counterfactual:

$$\text{(T)} \quad \sim p \;\Box\!\!\rightarrow\; \sim Bp$$

(T) says that s would not believe p, if p were false. Among other things, tracking is supposed to express the external relation that explains why some justified true beliefs are not knowledge. Champions of the condition include Robert Nozick (1981) and, more recently, Keith DeRose (1995). To my knowledge, the earliest formulation of the counterfactual condition is found in Fred Dretske's *conclusive reasons* condition (1971), which says, s would not have had the reason that she does for believing p, if p were false. Vogel contends that any such counterfactual condition on knowledge will render the theory of knowledge too strong. He believes that there is at least some possible reflective knowledge that cannot satisfy the counterfactual--viz., the possible knowledge that one does not believe falsely that p. The alleged impossibility of such reflective knowledge is taken by Vogel to be a decisive objection to the tracking theories advocated by Dretske, Nozick, DeRose[1] and others.

The criticism finds its roots in Vogel's earlier work (1987), and recurs in papers by Ernest Sosa (2002, 1996). Sosa suggests that the externalist idea behind tracking is on target, but that Nozick's counterfactual is a misbegotten regimentation of the idea. In its place Sosa offers his own counterfactual "safety" condition, which he feels properly captures the externalist idea. Sosa's counterfactual is not the topic of this paper. I mention it only to point out that the criticism that constitutes the subject of my investigation is meant to do a lot of work. In Sosa's case the criticism is meant to motivate his own counterfactual analysis, and in Vogel's case the criticism promises to be a silver bullet against a theory that has recently found renewed life in the work of Keith DeRose. It will be argued here that the criticism is misguided. My belief that I do not believe falsely that p can track the truth.

The Counterexample to Truth Tracking

Consider the proposition that I do not believe falsely that p. I should be able to know a proposition like that. Sometimes I know p and double-check my sources, thereby coming to know, additionally, that I do not believe falsely that p. Vogel's concern is that this kind of epistemic confidence cannot track the truth: if my belief that I am not mistaken were false, I would *still* believe that I'm not mistaken.

Vogel begins by translating the claim of epistemic confidence, "I do not believe falsely that p" as shown. Call it (EC):

$$\text{(EC)} \quad \sim(Bp \;\&\; \sim p),$$

Formally, (EC) reads, "it is not the case that both I believe p and p is false." The tracking condition, again, says, if p were false it would not be believed that p:

$$\text{(T)} \quad \sim p \;\Box\!\!\rightarrow\; \sim Bp.$$

Knowing (EC) then requires satisfying the following instance of (T):

$$\text{(T*)} \quad (Bp \;\&\; \sim p) \;\Box\!\!\rightarrow\; \sim B\sim(Bp \;\&\; \sim p).$$

It says this: if I were to believe p falsely, then I would not believe that I do not believe p falsely.

Evaluating this counterfactual is tricky business. Nevertheless, Vogel argues that clearly it is not satisfiable, because he thinks, "If you believe p, you believe that you do not falsely believe p." (2000, 611) Formally,

$$(*)\quad Bp \rightarrow B{\sim}(Bp \,\&\, {\sim}p).$$ [2]

Of course, if (*) is valid then (T*) is not satisfiable. In the relevantly close worlds where the antecedent of (T*) is true, you believe you are not mistaken. That is because, by (*), in *every* world where the antecedent of (*) is true, you believe you are not mistaken.

So if (*) is valid and (EC)-- i.e., '~(Bp & ~p)'--properly regiments the claim 'I do not believe falsely that p,' then it is not possible to track the truth of this claim. Therefore, if Vogel's logical resources are in order and tracking is a necessary condition on knowledge, then, absurdly, one cannot know that one does not believe falsely that p.

Objections

This is a good place to point out that both the assumption that (*) is valid and the assumption that (EC) best captures the expression of epistemic confidence are questionable. I address each of these worries in turn.

Vogel's criticism of tracking theories depends on the validity of (*). A number of concerns arise. First, it is not obvious that believing p entails having the higher-order belief that one is not mistaken in believing p. That implies that small children and other unreflective thinkers have beliefs about their own beliefs. More to the point, no contradiction flows from the assumption that there is a thinker who, for whatever reason, is able to form only first-order beliefs (i.e., beliefs that do not have the concept of belief as part of their content).[3] So we have some reason from the start to be suspicious of (*).

Second, Vogel cites as his defense of (*) the validity of a closure principle for knowledge. He writes, any proposition p itself entails that a belief that p is not false: p entails ~(Bp & ~p). So, to say that you know [p] but you fail to know that you do not believe falsely that [p] would be to reject the closure principle for knowledge. (2000, 610, note 15)

Vogel here refers to the principle stating that we know all the consequences of what we know. Call the principle (CP):

$$\text{(CP)} \quad (Kp \;\&\; p \rightarrow q) \rightarrow Kq.$$

There are several problems with this defense, not the least of which is that (CP) is invalid. (CP) notoriously implies that we know even the undiscovered consequences of what we know.[4]

One might weaken the closure principle by strengthening the antecedent in some way. To what extent we ought to strengthen the antecedent of (CP) is an open question.[5] That we ought to strengthen it, on the other hand, is the received view among those who believe that logical entailment can extend our knowledge. But strengthening the antecedent of (CP) will destroy Vogel's argument for (*). The thought here is that weakening (CP) will enjoin an analogous weakening of (*). But Vogel's central argument depends on something at least as strong as (*).

Another problem with Vogel's appeal to knowledge closure is that the principle he is defending--viz., (*) $Bp \rightarrow B{\sim}(Bp \;\&\; {\sim}p)$-- is not a consequence even of the strongest closure principle for knowledge--, viz., (CP) $(Kp \;\&\; p \rightarrow q) \rightarrow Kq$. It is at best a consequence of some closure principle for belief. Perhaps Vogel's strategy is this. If knowledge closure is valid, then the corresponding closure principle for belief is also valid. In other words, If $(Kp \;\&\; p \rightarrow q) \rightarrow Kp$ is valid, then so is $(Bp \;\&\; p \rightarrow q) \rightarrow Bq$. The reasoning here must be that if closure holds for knowledge, then it holds for every necessary condition on knowledge.

The important objection to this strategy is that it is simply false that closure holds of knowledge only if it holds of every necessary condition on knowledge. To think otherwise is to commit the fallacy of composition.[6] So I find no plausible argument in Vogel from the validity of closure for knowledge to the validity of (*).

If we hope to avoid closure principles altogether, then we might rehabilitate Vogel's criticism another way. All we need to do is to argue that

$$(T^*) \quad (Bp \,\&\, {\sim}p) \;\Box\!\!\rightarrow\; {\sim}B{\sim}(Bp \,\&\, {\sim}p)$$

is not satisfied in all those cases where, plausibly, one knows that one is not mistaken in believing p. But consider a case where s knows p and corroborates her belief with an independent source, thereby coming to believe that she is not mistaken. It might be argued that in the closest worlds where s is mistaken (i.e., where sBp & ~p), s continues to believe that she is not mistaken. The idea here is that in the closest worlds where she is mistaken, she continues to believe that she is not mistaken, even though in the actual world she checked her work by an independent method.[7]

There is an obvious objection to this outcome. For the case at hand, it is unclear that a world where {s is mistaken but believes she is not} is closer than a world where {s is mistaken but withholds judgement about whether she is mistaken}. After all, in the actual world s double-checked (by an independent method) whether p. The result of her double-checking was that p. She thereby concluded that she was not mistaken in believing p. Arguably, the closest worlds where she is mistaken, her double-checking would not have corroborated her belief that p, and so, she loses her reason for thinking she is not mistaken. To deny this is to suggest the following strange claim: a world where independent methods give us faulty results if any of our methods give us faulty results is a closer world than one where our independent methods do not go haywire in this way. My point here is that it is all but clear that (T*) is unsatisfiable or that some general argument avoiding closure principles will show that (T*) is unsatisfiable.

If Vogel can show in some other way that the closest worlds where s is mistaken in believing p are worlds where one would believe that one is not mistaken, then perhaps his argument may be revived. But this is yet to be done, and as we will presently see it would not help. The reason it would not help is that '~(Bp & ~p)' is not the best formal representation of the thought that I do not believe falsely that p.

I Do Not Believe Falsely that p

We are considering the statement, "I do not falsely believe p". One might be inclined to think that this is the negation of "My belief that p *is* false," in which case, with Vogel, we simply negate the claim "I believe that p and p is false", giving us [(EC)] ~(Bp & ~p). Whether a belief of this form tracks is actually not important, because the thought process behind this regimentation is subtly confused. The relevant formulation of "I do not falsely believe p" is not the negation of "My belief that p is false." These two statements are contraries, not contradictories. The point is developed here.

We begin by noticing that "I do not falsely believe p" (or "I am not mistaken in believing p") is ambiguous. On one reading, the claim that

(+) I do not falsely believe p

is equivalent to

(1) I believe p and p is not false.

Formally,

Bp & ~~p

or equivalently,

$$Bp \,\&\, p.$$

This reading has the trivial implication that I believe p. The second reading does not have this implication. Accordingly, it may reasonably be argued that I do not falsely believe p, if I do not believe p at all. On this reading, the claim that

(+) I do not falsely believe that p

is equivalent to

(2) it is not the case that both I believe that p and not-p.

Formally,

$$\sim(Bp \,\&\, \sim p).$$

Reading 2 does not have the implication that I believe that p. My failing to believe p is sufficient for the truth of 2.

The question is this. In the contexts in which, intuitively, I can know that I do not believe falsely that p, which reading best captures the proposition? In ordinary speech the reasonable claim that I do not believe falsely that p is an expression of epistemic confidence in one's belief that p. And it is this expression of confidence that Vogel is focusing on. For it is in the context of knowing p that it seems to him that one also knows that one is not mistaken in believing p. Here is his example,

> You see your long-time friend Omar, who is a perfectly decent and straightforward sort of person. Noticing his shiny white footwear, you say, "Nice shoes, Omar, are they new?" Omar replies, "Yes, I bought them yesterday". I think the following things are

> true: [a] You know Omar has a new pair of shoes. [b] You know that your belief that Omar has a new pair of shoes is true, or at least not false. (2000, 609-610)

My belief that p is not false occurs in a context where I believe p. And it is the belief that p that the thought "My belief that p is not false" is about. So arguably, "My belief that p is not false" is best represented by an expression that entails that I believe p. Reading 1 above, Bp & p, is the best reading.

Ernest Sosa offers the same criticism as does Vogel against tracking theories, though he makes no appeal to (*) or closure principles. He writes,

> Consider: (a) p, and (b) I do not believe incorrectly (falsely) that p. Surely no one minimally rational and attentive who believes both of these will normally know either without knowing the other. Yet even in the cases where one tracks the truth of (a), one could never track the truth of (b). After all, even if (b) were false, one would still believe it anyhow. (1996, 276)[8]

What is important here is that Sosa is considering contexts in which p is known, and so believed. The claim that I am not wrong in thinking p must then be saying something that may serve as an expression of epistemic confidence. And so, the claim that I am not mistaken is saying, among other things, that I believe p. If I am saying something that does not imply that I believe p, then I am not saying something that expresses my epistemic confidence in my belief that p. The reading that seems most appropriate then is reading 1, Bp & p. Reading 1 has the truth conditions that best suit it for ordinary speech, where one's claim that one is not mistaken is commonly used as an expression of epistemic confidence. At the very least reading 1 has the truth conditions that best suit it for the kind of case Sosa has in mind--viz., one in which a minimally rational and attentive agent believes both p and that her belief that p has not gone wrong.

Why then do Vogel and Sosa automatically interpret the claim that I am not mistaken with reading 2? I fear that they are confusing the claim that I am not mistaken with the negation of the claim that I am mistaken. But just as the claim that I am not mistaken in believing p ordinarily means something that implies that I do believe that p, so the statement that I *am* mistaken in believing p implies that I do believe p. And so, if I fail to believe p, then both claims are false. The two claims do not contradict, since the falsity of one does not preclude the falsity of the other. They are contraries, since (in addition to the joint satisfiability of their negations) they cannot both be true. The fact that they cannot both be true explains the confusion; it explains why Vogel and Sosa confuse "I do not believe falsely that p" for the negation of "I do believe falsely that p".

We may better represent the expression of confidence in one's belief as the claim that my belief is correct. This suggests that we regiment the thought as Reading (1), Bp & p. This regimentation gives us the right truth conditions. The proposition is false when and only when either p is not believed or p is false. Furthermore, Reading (1) properly formalizes "I am not mistaken" as a contrary of "I am mistaken" (formally, Bp & ~p).

Reading (1), Bp & p, is a stronger claim than Reading (2), ~(Bp & ~p). The former entails the latter, but not vice versa. The former better represents the claim that one is not mistaken, and the truth that it depicts can be tracked. I will show that in a moment. Before I do, it should be noted that Vogel considered Reading (1). He claims that he chose Reading 2 over Reading 1

> to avoid complications about counterfactuals with disjunctive antecedents. Such technical hazards will arise any time one tries to apply [thesis (T)] to knowledge of conjunctions. (2000, 611, note 17)

With all due respect, this is a terrible reason to opt for one reading over another. The correct logical characterization of a proposition (that is not a counterfactual) should be decided independently of any problems there may be with interpreting counterfactuals that embed that proposition. Otherwise, we should not interpret any apparent natural language conjunction as p & q, when evaluating a counterfactual that embeds its negation (in the antecedent place).

Back to Tracking

We seek to determine whether the belief that I am correct in believing p (i.e., Bp & p) can track the truth. The belief can track just in case the following counterfactual can be true:

$$\text{(T**)} \quad \neg(Bp \,\&\, p) \;\Box\!\!\rightarrow\; \neg B(Bp \,\&\, p).$$

This is the relevant instance of (T). Equivalently,

$$\text{(T**)} \quad (\neg Bp \vee \neg p) \;\Box\!\!\rightarrow\; \neg B(Bp \,\&\, p).$$

(T**) embeds the dreaded disjunctive antecedent that Vogel wished to avoid. But for the standard case a counterfactual with a disjunctive antecedent simplifies. That is, a counterfactual with a disjunctive antecedent, $(p \vee q) \;\Box\!\!\rightarrow\; r$, implies the conjunction of two counterfactuals, $(p \;\Box\!\!\rightarrow\; r) \,\&\, (q \;\Box\!\!\rightarrow\; r)$. For instance, “if it were to rain or frost, the game would be canceled” implies, “If it were to rain the game would be cancelled, and if it were to frost the came would be cancelled”. Of course there are cases that appear to invalidate the inference, but all of these are cases where one of the disjuncts is not taken to be a ‘real possibility’,[9] as in “If Spain were to fight on one side or the other [i.e., with the Allies or with the Axis in WWII], she would fight with the Axis” or in “If she were to run for Congress [i.e., the House or the Senate], she would run for the House.” There is no reason to think that (T**) exhibits the peculiarity of these latter examples. So we should allow the inference here.[10]

The relevant counterfactual is (T**) $(\neg Bp \vee \neg p) \Box\!\!\rightarrow \neg B(Bp \,\&\, p)$. So, if it is true, then the following two counterfactuals are true:

$$\text{(T**1)} \quad \neg p \Box\!\!\rightarrow \neg B(Bp \,\&\, p)$$

$$\text{(T**2)} \quad \neg Bp \Box\!\!\rightarrow \neg B(Bp \,\&\, p).$$

(T**2) is trivially satisfied, because any world where I do not believe p is a world where I do not believe Bp & p.[11] The significant point is that (T**1) is not impossible. In fact, it seems to be exactly that on which knowledge of one’s own correctness depends. It depends on whether I would have believed I am correct in thinking p, if p were false. For instance, if Omar did not have a new pair of shoes, I would not have thought that my belief that he does have a new pair shoes is correct. After all, Omar ‘is a perfectly decent and straightforward sort of person,’ and so would not have misled me. In the closest worlds where Omar does not have a new pair of shoes, I would not believe that he does (and so, I would not have thought that I have a true belief that he does). (T**1) is satisfied. Consider another case. I come to believe p and independently corroborate, thereby coming to believe that my initial belief is true, not false. In the relevantly close worlds where p is false, my independent source would not corroborate that p. And so, I would not have thought that my belief that p is true, not false. (T**1) is satisfied. The two counterfactuals, (T**1) and (T**2), are jointly satisfiable.[12] So, tracking condition (T) is satisfied in the relevant cases.

In conclusion, Vogel's proposal against truth-tracking rests on a subtle confusion about how to interpret an expression of epistemic confidence. The best interpretation allows for the possibility of sensitively believing that one is not mistaken. If Vogel's argument was intended to show that any renewed interest in tracking conditions is hopeless, then Vogel's argument falls short. If Sosa's criticism was intended to motivate his own counterfactual analysis, then Sosa's analysis is ill-motivated. Whatever the shortcomings of truth-tracking proposals, they do not include the impossibility of reflective knowledge that one is not mistaken.

References

David, M. and T. Warfield. forthcoming, "Knowledge-Closure and Skepticism." in Q. Smith, (ed.), *Epistemology: New Philosophical Essays*. Oxford: Oxford University Press.

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

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

Hawthorne, J. 2004. *Knowledge and Lotteries*, Oxford: Oxford University Press.

Lycan, W. 2001. *Real Conditionals*, Oxford: Oxford University Press.

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

Sosa, E. 1996. "Postscript to 'Proper Functionalism and Virtue Epistemology'" in J. Kvanvig, (ed.), *Warrant in Contemporary Epistemology: Essays in Honor of Plantinga's Theory of Knowledge*. Savage, MD: Rowman & Littlefield.

--------. 2002. "Tracking, Competence and Knowledge" in P. Moser, (ed.), *The Oxford*

Handbook of Epistemology. Oxford: Oxford University Press.

Vogel, J. forthcoming. “Varieties of Skepticism.” *Philosophy and Phenomenological Research*.

--------. 2000. “Reliabilism Leveled.” *The Journal of Philosophy* XCVII(11): 602-623.

--------. 1987. “Tracking, Closure, and Inductive Knowledge” in S. Luper(-Foy), (ed.), *The Possibility of Knowledge: Nozick and His Critics*. Lanham, MD: Rowman & Littlefield.

Warfield, T. 2004. “When Epistemic Closure Does and Does Not Fail: a Lesson from the History of Epistemology.” *Analysis* 64(1): 35-41.

Notes

The author wishes to thank Berit Brogaard, Susan Brower-Toland, Alicia Finch, Dan Haybron, Marion Ledwig, Steven Luper, Matt McGrath, Duncan Pritchard, Scott Ragland, Ernest Sosa, Kent Staley, Jim Stone and Adam Wager for useful discussion and comments. The majority of the paper was completed with the help of a Mellon Faculty Development Grant in the Summer of 2003.

1. For DeRose (1995) sensitivity is not a general requirement for knowledge, although the effect of denying that one has knowledge is to require that sensitivity be met thereafter. Thus Vogel’s argument threatens DeRose’s position. In an ordinary (non-philosophical) context where I think that I am not mistaken in believing p and you deny that I know this, it should be possible to rise to the challenge thereby satisfying the sensitivity condition. Vogel’s argument, in effect, denies that it is possible to rise to this challenge.

2. We are here interpreting Vogel's conditional (*) as a necessary material conditional, because only on this familiar interpretation is Vogal's argument valid. If (*) is treated as a material conditional, the argument is invalid. The material conditional, Bp → B~(Bp & ~p), is not sufficient to falsify (T*). The material conditional tells us only what is the case in the actual Bp-world, but to evaluate the counterfactual (T*) we need to know what is the case in the closest Bp&~p-worlds. Furthermore, if Vogel's conditional (*) is treated as a counterfactual, the argument is invalid. It is well known that strengthening the antecedent of a counterfactual does not preserve truth. In particular, Bp □→ B~(Bp & ~p) does not entail (Bp & ~p) □→ B~(Bp & ~p). And so, the truth of Bp □→ B~(Bp & ~p) does nothing to by way of undermining (T*), (Bp & ~p) □→ ~B~(Bp & ~p).

3. Thanks to Steven Luper for sharpening this point.

4. For instance, we know (the conjunction of) the axioms of arithmetic. But then by CP, we know that Goldbach's conjecture is false (if it is false). But clearly we do not know whether Goldbach's conjecture is false. We might in the familiar way weaken the closure principle to "we know all the *known* consequences of what is known," giving (WCP): (Kp & K(p→q)) → Kq. (WCP) is a more plausible closure principle. But it should be noted that Nozick and Dretske reject (WCP), and with it the stronger (CP). So any argument for rejecting the tracking condition that employs a closure principle at least as strong as (WCP) is not something by which Nozick or Dretske would be moved.

5. It has been emphasized repeatedly that even (WCP) [as defined in the previous note] is problematic as it stands. Perhaps one may know p and know that p entails q without believing (and so without knowing) q. The suggestion is that the antecedent of (WCP) be strengthened to include that the agent performs the deduction thereby coming to believe q. Other problems then arise. It might be that one knows p and knows that p entails q but in the time it takes to perform the inference from p to q the agent loses her knowledge that p. The suggestion here is that we strengthen the antecedent further to include that the agent retains her knowledge of the premise throughout. These concerns about the validity of (WCP) are developed in (Hawthorne 2004) and in (David and Warfield, forthcoming).

6. This point is the thesis of (Warfield 2004). Vogel recently acknowledged this lesson in his "Varieties of Skepticism" (forthcoming).

7. It may be noted that Nozick's treatment of these counterfactuals is not the familiar Lewis semantic. For reasons we need not go into here, Nozick requires, roughly, that the *sufficiently close* worlds (not just the closest worlds) are relevant for the proper evaluation of the counterfactual. This difference does not effect the above argument, since it is sufficient for showing that the Nozick counterfactual, $p \Box\!\!\rightarrow q$, is false that there is a closest p-world that is not a q-world. For if there is a closest p-world that is not a q-world, then there is a sufficiently close p-world that is not a q-world. The difference becomes crucial when demonstrating the *truth* of a counterfactual. On the Lewis approach the actual truth of p and q is sufficient for $p \Box\!\!\rightarrow q$, since the closest p-world (i.e., the actual world) is a q world. Not so on the Nozick

approach. The closest p-world may be a q-world, but for the counterfactual, $p \Box\!\!\rightarrow q$, to be true, it must be that all the sufficiently close p-worlds (not just the actual p-world) are q-worlds.

8. For another formulation of argument, see (Sosa 2002, 265).

9. See (Lycan 2001, 42-46).

10. We should not get hung up on the inference. It is apparent that even if we disallow the simplification, (T**) is satisfiable, and so, the bit of reflective knowledge in question is possible. See note 12.

11. Here I merely suppose that believing a conjunction entails believing each of the conjuncts.

12. Notice that if counterfactuals with disjunctive antecedents do not simplify, we need to evaluate (T**),

$$(\neg Bp \vee \neg p) \Box\!\!\rightarrow \neg B(Bp \,\&\, p),$$

as it stands, unsimplified. On the standard Lewis/Stalnaker semantic, this requires that we go to the closest (¬Bp v ¬p)-world and determine whether it's a ¬B(Bp & p)-world. The closest (¬Bp v ¬p)-world is either the closest ¬Bp-world or it is the closest ¬p-world. If the former, the satisfiability of (T**) hangs on the satisfiability of (T**2). If the latter, the satisfiability of (T**) hangs on the satisfiability of (T**1). As we have seen (T**1) is satisfiable, and so is (T**2). But then either way (T**) is satisfiable.