THE BACKWARD CLOCK, TRUTH-TRACKING, AND SAFETY*

We present *Backward Clock*, an original counterexample to Robert Nozick’s truth-tracking analysis of propositional knowledge, which works differently from other putative counterexamples and avoids objections to which they are vulnerable. We then argue that ways of analyzing knowledge in terms of safety, as suggested by Duncan Pritchard, cannot withstand *Backward Clock* either.

I. THE BACKWARD CLOCK AND THE TRUTH-TRACKING ANALYSIS OF KNOWLEDGE

Nozick’s analysis is that

\[ S \text{ knows that } p, \text{ using method } M \text{ of arriving at a belief whether } p, \text{ just in case} \]
(1) \( p \)
(2) \( S \) believes, using \( M \), that \( p \)
(3) In the closest (that is, most similar) worlds to the actual world in which not-\( p \) (and in which \( S \) uses \( M \)), \( S \) does not believe that \( p \)
(4) In the closest (that is, most similar) worlds to the actual world in which \( p \) (and in which \( S \) uses \( M \)), \( S \) believes that \( p \).\(^1\)

Nozick’s sensitivity condition (3) remains a live option in the literature.\(^2\) It seems to explain why you may gain knowledge in:

*Normal Clock*

You habitually nap between 4 pm and 5 pm. Your method of ascertaining the time you wake is to look at your clock, one you know has always worked perfectly reliably. This clock is analogue so its hands sweep its face continuously. However, it has no second hand.\(^3\) Awaking

---

\(^*\) The authors contributed equally to this article. We are grateful for useful discussion with Mark McBride, Lee Min Ying, and Bernadette Chin.


\(^3\) We stipulate this to ensure parity with two other examples, *Stopped Clock* and *Backward Clock*.
at 4:30 pm, you see that its hands point to 4:30 pm. Accordingly, you form the belief that it is 4:30 pm. And it is indeed 4:30 pm because the clock has continued to work perfectly reliably.

Your true belief that it is 4:30 pm is sensitive to falsehood. Had it been any time other than 4:30 pm when you looked at the clock, then you would not believe that it is 4:30 pm. In other words, in the closest worlds to the actual world in which it is not 4:30 pm (and in which you look at the clock to tell the time), you do not form the false belief that it is 4:30 pm. Instead you form some other true belief about what time it is. For example, had it been 4:31 pm when you looked at the clock, then you would not form the false belief that it is 4:30 pm. Instead you would form the true belief that it is 4:31 pm.

Your true belief also satisfies Nozick’s truth-adherence condition (4). If you had looked at the clock at 4:30 pm while being slightly closer to it, then you would still believe that it is 4:30 pm. In other words, in worlds close to the actual world in which it is 4:30 pm (and in which you look at your clock in order to ascertain the time), you believe that it is 4:30 pm.

Nozick’s sensitivity condition deals nicely with Gettier-type apparent counterexamples to the JTB analysis of knowledge as justified true belief, such as:

**Stopped Clock**

You habitually nap between 4 pm and 5 pm. Your method of ascertaining the time you wake is to look at your clock, one you know has always worked perfectly reliably. Like Normal Clock, it has an analogue design so its hands are supposed to sweep its face continuously. However, it has no second hand. Awaking at 4:30 pm, you see that its hands point to 4:30 pm. Accordingly, you form the belief that it is 4:30 pm. And it is indeed 4:30 pm because exactly twenty-four hours ago a stray fleck of dust chanced to enter the clock’s mechanism, stopping it.

We would all agree that you do not know that it is 4:30 pm. If any intuition is unassailable, this is one. One very plausible explanation of your ignorance is that your belief is luckily true. You were lucky to look at the clock exactly twenty-four hours after it stopped working, at the only instant during the hour when you nap at which its hands could have pointed to the correct time.

---

4 This description avoids the claim that the clock shows 4:30 pm. John Biro (“Showing the Time,” *Analysis*, lxxiii, 1 (January 2013): 57–62, at p. 58) argues that a stopped clock does not show any particular time.

5 An alternative explanation of your ignorance is that you are not justified (in the sense required by knowledge) in believing that it is 4:30 pm. See Adrian Heathcote,
Nozick’s analysis explains your ignorance in *Stopped Clock*; your belief that it is 4:30 pm is insensitive to falsehood. If it were not 4:30 pm but some other time, then by looking at the clock you would still believe that it is 4:30 pm. In other words, in possible worlds close to the actual world in which it is not 4:30 pm but, say, 4:31 pm (and in which you look at the clock to tell the time), you form the *false* belief that it is 4:30 pm. Now consider:

**Backward Clock**

You habitually nap between 4 pm and 5 pm. Your method of ascertaining the time you wake is to look at your clock, one you know has always worked perfectly reliably. Unbeknownst to you, your clock is a special model designed by a cult that regards the hour starting from 4 pm today as cursed, and wants clocks not to run forwards during that hour. So your clock is designed to run perfectly reliably backwards during that hour. At 4 pm the hands of the clock jumped to 5 pm, and it has been running reliably backwards since then. This clock is analogue so its hands sweep its face continuously, but it has no second hand so you cannot tell that it is running backwards from a quick glance. Awakening, you look at the clock at exactly 4:30 pm and observe that its hands point to 4:30 pm. Accordingly, you form the belief that it is 4:30 pm. As in *Stopped Clock*, you have a true belief that it is 4:30 pm. And again, you do not know that it is 4:30 pm because your belief is luckily true. For you were lucky to look at it at exactly 4:30 pm, at the only instant during the hour when you nap at which its hands could have pointed to the correct time.\(^6\) Since you do not know that it is 4:30 pm in *Stopped Clock* then surely you do not know that it is 4:30 pm in *Backward Clock* either, for the salient commonality is that your belief that it is 4:30 pm is luckily true.

But—and here is our point—unlike in *Stopped Clock*, your belief that it is 4:30 pm is *sensitive to falsehood*. If it were not 4:30 pm but some other time, then by looking at the clock you would *not* believe that it is 4:30 pm. Instead you would form *some other false belief* about what time it is. For example, if you had looked at it at 4:31 pm, then you would not form the false belief that it is 4:30 pm. Instead you

---

\(^6\) *Backward Clock* cannot be blocked by the fact that the hands point to the correct time only for a mere instant during the hour that you nap. Probability theorists might say that this is a “zero probability event,” but will admit that although events that are impossible have zero probability, the converse need not hold. As in *Stopped Clock*, you can still look at the clock at the only instant at which its hands point to the correct time.
would form the *false* belief that it is 4:29 pm. To satisfy the sensitivity condition, you need not form a true belief about what time it is in the counterfactual situation (as you do in *Normal Clock*). You only need to fail to form a particular false belief—perhaps, by forming a different false belief about what time it is instead (as you do in *Backward Clock*). Thus for Nozick, whether you know that it is 4:30 pm depends on whether your belief is truth-adherent.

It is indeed truth-adherent. Had you looked at the clock at 4:30 pm while being slightly closer to it, then you would still believe that it is 4:30 pm. In other words, in worlds close to the actual world in which it is 4:30 pm (and in which you look at your clock to tell the time), you believe that it is 4:30 pm.

Of course the closeness of possible worlds to actuality is vague, but close possible worlds cannot include those in which the *mechanism* of the clock differs from its *actual* mechanism. This is because the truth-adherence of your belief that it is 4:30 pm in *Normal Clock* resides in the fact that you would still have that belief in slightly changed circumstances in which the mechanism of the clock continues to work perfectly reliably. Likewise, the worlds close to the actual circumstances of *Stopped Clock* surely include those in which the mechanism of the clock is stopped.

Thus the truth-tracking analysis is too weak; that is, it treats non-knowledge as knowledge. *Backward Clock* also shows to be too weak a broadly Nozickian analysis of knowledge proposed by Rachael Briggs and Daniel Nolan. This is that you know that \( p \) just in case you have a true belief that \( p \) that you are not disposed to have in the circumstance in which \( \neg p \) and that you are disposed to have in the circumstance in which \( p \).

In *Backward Clock* you have a true belief that it is 4:30 pm. However, you are not disposed to believe that it is 4:30 pm by looking at the clock in the circumstance in which it is not 4:30 pm but 4:31 pm. Instead you are disposed to believe that it is 4:29 pm. Moreover, you are disposed to believe that it is 4:30 pm by looking at the clock in the circumstance in which it is 4:30 pm. Yet you do not know that it is 4:30 pm.

We claim that *Backward Clock* is not a case of knowledge because (as in *Stopped Clock*) your belief that it is 4:30 pm is luckily true. While we could instead appeal to the unreliability of your method

---

7 Rachael Briggs and Daniel Nolan, “Mad, Bad and Dangerous to Know,” *Analysis*, lxxii, 2 (April 2012): 314–16. Unlike *Stopped Clock, Mad* and *Bad* aim to show that Nozick’s analysis is too strong, that is, incorrectly excludes knowledge, while *Dangerous to Know* is supposedly a case of a safe true belief not constituting knowledge.
of ascertaining the time, we are hesitant to explain why your belief
is not knowledge in these terms. If this were our reason for claim-
ing ignorance in *Backward Clock*, it might be thought that we are
committed to an analysis of knowledge in terms of generally reli-
able methods.8 Such analyses face difficulties, including the gener-
ality problem.9 Fortunately we are uncommitted to such analyses, as
our appeal is to luck, not unreliability.

It might also be claimed that in *Backward Clock* you are not jus-
tified in forming any belief about what time it is by looking at the
clock during its backward-running hour, because to be so justified
you would have to check that its hands are still moving forwards.10
In that case you would realize that it is running backwards. But even
if this claim is true, that will help neither Nozick nor Briggs and
Nolan, since their analyses do not mention justification.

II. THE ORIGINALITY OF BACKWARD CLOCK

Plausible counterexamples to Nozick’s analysis include Ray Martin’s
*Racettrack*, George Pappas and Marshall Swain’s *Generator*, and Laurence
Bonjour’s *Clairvoyant*,11 as well as Saul Kripke’s *Red Barn*, his *Deceased
Dictator*, and his *Sloppy Scientist*.12 Fred Adams and Murray Clarke argue
persuasively that these are not counterexamples.13 Unlike *Backward
Clock*, which shows Nozick’s analysis to be too weak, *Generator* aims to
show that it is too strong, that is, incorrectly excludes knowledge.14 Nor

8 For example, Steven Luper-Foy, “The Causal Indicator Analysis of Knowledge,”
9 See Earl Conee and Richard Feldman, “The Generality Problem for Reliabilism,”
10 As Heathcote formulates this claim, “Knowing the time by looking at a clock is a
matter of confirming that it is still running” (op. cit., p. 312, our italics). We think that this
puts the bar too high in most cases, although it is plausible that to be justified in believ-
ing that the time is what the hands of the clock point to, you need to be justified in believing
that the clock is still working reliably. In fact we only occasionally confirm that
our clocks are still working reliably. These occasions sustain our justified confidence
that they have continued to work reliably on the much more frequent occasions on
which we only glance at them and hence normally tell the time, in other words, gain
both knowledge and justified belief of what time it is.
11 See Raymond Martin, “Empirically Conclusive Reasons and Scepticism,” *Philosophical
Method,” *Analysis*, xlii, 1 (January 1983): 28–33; George S. Pappas and Marshall Swain,
“Some Conclusive Reasons against ‘Conclusive Reasons’,” *Australasian Journal of Phi-
losophy*, li, 1 (1973): 72–76; and Laurence Bonjour, “Externalist Theories of Empiri-
12 Kripke gave these widely known examples at a session of the American Philosophical
Association in the 1980s.
13 Fred Adams and Murray Clarke, “Resurrecting the Tracking Theories,” *Australasian
14 Pappas and Swain, op. cit., p. 66.
do Red Barn or Deceased Dictator aim to show that it is too weak. Unlike Backward Clock, Clairvoyant appeals crucially to the notion of evidence and is arguably a case of knowledge, as is Sloppy Scientist. Arguably, Racetrack involves a failure of sensitivity.

Other examples that aim to show that Nozick’s analysis is too weak include Garrett’s Philosophical Father and as discussed by Lars Gundersen, Hologram Vase. Robert Gordon and Graeme Forbes argue respectively that each is a case of knowledge. Finally, consider Alvin Plantinga’s

Brain Lesion
You have a brain lesion that causes you to believe that you have a brain lesion. Since the cause of your belief is your brain lesion, your belief is sensitive. If you had no brain lesion then you would not believe that you have one. Pritchard thinks that it is clear that you do not know that you have a brain lesion, because your belief is true despite your cognitive abilities. But unlike in Backward Clock, your belief does not seem to be truth-adherent. The neural encoding of the content of belief is a subtle thing, and only a very specific type of lesion would

15 Red Barn aims to show that one knows that there is a red barn before one without knowing (because of a failure of sensitivity) that there is a barn before one. See Adams and Clark, op. cit., p. 215. Deceased Dictator aims to show that one knows the conjunction that a dictator is dead and that one has read an uncontradicted report that he is dead, but does not know (because of a failure of adherence) that he is dead. See ibid., p. 216.

16 Specifically, that the clairvoyant has a truth-tracking method of arriving at a true (believed) prediction but also has overwhelming evidence against it.

17 See Adams and Clark, op. cit., p. 220.

18 See ibid., p. 218.


20 You decide whether Tom’s father is a philosopher by believing that he is one just in case Tim’s father is a philosopher. You accept testimony that Tim’s father is a philosopher, and so believe that Tom’s father is a philosopher. Unbeknownst to you, Tim is Tom’s brother. B. J. Garrett, “Nozick on Knowledge,” Analysis, xliii, 4 (October 1983): 181–84.

21 You believe that there is a vase in a box upon seeing a hologram of a vase being displayed in the box. Unbeknownst to you, the hologram is switched on when and only when a vase in the box depresses a lever at the bottom of the box. Lars Gundersen, “Tracking, Epistemic Dispositions and the Conditional Analysis,” Erkenntnis, lxix, 3 (May 2010): 353–64.


generate a belief with the content that one has a lesion. With a slightly different lesion, different beliefs might be generated, or none at all. As Kelly Becker notes, “the brain lesion is not truth-conducive in nearby worlds.”

_Backward Clock_ stands apart from previous counterexamples in two ways. First, no example of truth-tracking belief not constituting knowledge is as incontestable. Second, and as previously noted, one’s belief that \( p \) is _sensitive in a special way_; were it false that \( p \), then one would not form the false belief that \( p \) but _would instead form some other false belief._

**III. THE BACKWARD CLOCK AND SAFETY**

We now discuss views on which safety is a necessary condition for knowledge. While Nozick’s analysis of knowledge does not include a safety condition, more recent analyses do. Consider Pritchard’s formulation:

"Content-fixed Safety"

If \( S \) knows that \( p \), then \( S \)’s true belief that \( p \) could not have easily been false.26

This formulation entails that you do not know that it is 4:30 pm in _Stopped Clock_. You could have easily falsely believed that it was 4:30 pm if you had looked at the clock at another time during the hour that you nap. This makes content-fixed safety look like a useful way of excluding cases of non-knowledge. But as we will show, it fails to exclude non-knowledge in _Backward Clock_.

On one straightforward reading, content-fixed safety allows _Backward Clock_ to be a case of knowledge. Your true belief that it is 4:30 pm could not have easily been false, because the only time you would form a belief with that content in nearby worlds is at 4:30 pm. Looking at the clock at other times would produce false beliefs with different contents. But as written, content-fixed safety only seems to concern whether beliefs with the _same content_ could have easily been false, incorrectly allowing your belief that it is 4:30 pm to be knowledge. Pritchard recognizes this problem, writing that

…what we are interested in is rather how the agent forms her beliefs in similar circumstances and in response to the same stimuli. These beliefs may be beliefs _that_ \( p \), but equally they may be beliefs in distinct propositions.27

26 Pritchard, _op. cit._, p. 253.
A belief can be unsafe if, in nearby worlds, the agent “forms her belief on the same basis and yet ends up with a false belief.”\textsuperscript{28} Modifying content-fixed safety in line with this emphasis on fixing the \textit{basis} rather than the propositional content of the belief produces the following:

\textit{Basis-fixed Safety}

If $S$ knows that $p$ on basis $B$, then $S$ could not have easily formed a false belief on basis $B$.

This better expresses what Pritchard intends.\textsuperscript{29}

To apply basis-fixed safety to \textit{Backward Clock}, we need to know what counts as the basis of your belief that it is 4:30 pm. But here we confront the disquieting fact that subtly different construals of the basis lead to different results as to whether your belief is safe. If $B$ is that the hands point to 4:30 pm, then we get the unsatisfactory result that it is indeed safe. For in \textit{Backward Clock}, you would not easily form a false belief on the basis that the hands point to 4:30 pm, since that only happens when it is 4:30 pm and your belief is true. To get the desired result that your belief is unsafe, the basis must be construed more loosely, perhaps so that $B$ is that you look at the clock. You can easily form a false belief on the basis that you look at the clock—indeed, you would form a false belief on that basis at all times except 4:30 pm during the hour that you nap. Which of these is the basis?

It must be that the hands point to 4:30 pm. That you look at the clock is not a sufficient basis for believing that it is 4:30 pm, as this leaves open where the hands are pointing. You need to see that the hands point to 4:30 pm to have grounds for believing that it is 4:30 pm.\textsuperscript{30} But since this basis does not allow your belief that it is

\textsuperscript{28} Ibid., p. 257.

\textsuperscript{29} He writes when discussing sensitivity, “Strictly speaking, we should be relativizing this principle—and the safety principle offered below—to the actual basis for the agent’s belief in $p$...” (\textit{ibid.}, p. 250 note 6). We see this relativization as worth making explicit, as it affects the structure and consequences of the principle.

\textsuperscript{30} Pritchard introduces the notion of a basis as “grounds for believing” (\textit{ibid.}, p. 251). In \textit{Mathema}, the title character of which believes that $12 \times 13 = 156$ because her calculator displays this result (yet unbeknownst to her, is generating results randomly), he suggests a weaker account of a basis:

While there is indeed no similar circumstance in which we can imagine Mathema forming a belief that $12 \times 13 = 156$ on the same basis and yet believing falsely, we can certainly imagine lots of similar circumstances in which Mathema forms her belief on the same basis and yet ends up with a false belief, such as the similar situation in which the calculator generates a different result. Mathema’s belief is thus unsafe... (\textit{ibid.}, p. 257).
4:30 pm to be easily false in *Backward Clock*, substituting basis-fixed safety for sensitivity incorrectly allows it to count as a case of knowledge. *Backward Clock* is distinctive in posing this problem. In *Stopped Clock*, you could have easily formed a false belief on the basis that the hands point to 4:30 pm, since they do so at all times during the hour that you nap, so basis-fixed safety correctly excludes your belief that it is 4:30 pm from being knowledge. In contrast, *Backward Clock* involves a basis that appears only when the belief it supports is true, so that even sophisticated safety conditions incorrectly predict knowledge. Perhaps there is an account of what a basis is that avoids this unhappy result. If so, it deserves to be clearly formulated.  

This problem seems to afflict Pritchard’s own analysis of knowledge, which combines a safety condition with an “ability condition,” namely that

\[ S \text{ knows that } p \text{ if and only if } S \text{’s safe true belief that } p \text{ is the product of her relevant cognitive abilities.} \]

By “product of her relevant cognitive abilities,” Pritchard means that although S’s cognitive success is not fully explained by her cognitive abilities, it is “to a significant degree creditable to her cognitive agency.” But while your cognitive abilities do not fully explain why your belief that it is 4:30 pm is true, it is certainly produced by them, with the result that your true belief is significantly creditable to them. In both *Normal Clock* and *Backward Clock*, your true belief that it is 4:30 pm is the product of the same cognitive abilities, namely your ability to accurately observe the position of the hands of the clock, your ability to understand how such positions represent time, your ability to recall that the clock has always worked perfectly reliably, and your ability to induce from all this that it is 4:30 pm. The exercise of these cognitive abilities allows you to know that it is 4:30 pm in

---

31 Generality problems emerge for looser construals of the basis, as noted by David Manley, “Safety, Content, Apriority, Self-Knowledge,” *this journal*, civ, 8 (August 2007): 403–23, at p. 408. Is \( B \) that you look at the clock, that you look around at that moment, or that you use visual perception?


Normal Clock, so your exercising them cannot be what prevents you from knowing that it is 4:30 pm in Backward Clock. Pritchard’s analysis is thus too weak, treating Backward Clock as a case of knowledge.

A third way of formulating the safety condition is to fix both the content and the basis of the belief. Ernest Sosa proposes this, as

\[
\text{Sosa-safety} \\
\text{If } S \text{ knows that } p \text{ on basis } B, \text{ then } S \text{ could not have easily formed the false belief that } p \text{ on basis } B. \]  

Sosa-safety cannot withstand Backward Clock however one construes B. This is because possible worlds close to the actual world (those in which the mechanism of the clock is perfectly reliable in the way that it runs backwards during the hour when you nap) in which you believe that it is 4:30 pm on the basis of looking at your clock (or by seeing where its hands point), are indeed worlds in which it is 4:30 pm. So substituting Sosa-safety for sensitivity again yields an analysis of knowledge that is too weak.

IV. CONCLUDING REMARKS

Backward Clock shows that Nozick’s analysis of knowledge is uncontestably too weak, and cannot be remedied by Briggs and Nolan’s dispositional formulation. It uniquely exemplifies a true belief not constituting knowledge, despite being non-trivially sensitive, non-trivially adherent, content-fixed safe, basis-fixed safe, Sosa-safe, and satisfying Pritchard’s ability condition. This suggests that it will be a useful test case for other analyses of knowledge.

Singapore Management University

National University of Singapore

56 Pritchard later refines his characterization of a cognitive ability as a disposition to reliably form a knowledge-conducive belief that is suitably integrated with the agent’s other belief-forming dispositions (ibid., pp. 261–62). In Normal Clock you have the disposition to reliably form the knowledge-conducive belief that it is 4:30 pm that is suitably integrated with your dispositions to form true beliefs about the position of the hands of the clock, about how such positions represent time, about the past reliability of the clock, and about what time it is. Exactly the same holds of Backward Clock. Pritchard adds that “while we do ordinarily relativize abilities to environments, we do not tend to do this in a very fine-grained way unless there is a specific reason to do so” (ibid., p. 268). There appears to be no salient difference between the environments of Normal Clock and Backward Clock that could motivate relativizing your cognitive abilities to them.


58 Beliefs in necessary truths are trivially adherent and trivially sensitive, as they are true in all possible worlds.