Yes, Safety is in Danger

Tomás Bogardus and Chad Marxen

What's a *safe* belief?

* A subject's belief is safe just in case the method she employed to arrive at that belief did not put her in serious epistemic danger, that is, serious danger of arriving at a false belief thereby.

* To say that a subject's belief that p is safe is to say that, were she to believe that p via this method, p would be true.

* Not easily would she have believed falsely via that method.

Are all known beliefs safe?

* If the safety condition is true, then a given belief is known *only if* it is *safe.*
* Here's the general recipe for whipping up such a counterexample: first, pick the most virtuous belief-forming method you can imagine, and have a subject form a belief via that method.
* Second, add a twist of fate: put the method in danger of malfunctioning, but let the danger remain purely counterfactual.

Atomic Clock

* Suppose that the world's most accurate clock hangs in Smith's office, and Smith knows this.

* This radiation sensor is very sensitive, however, and could easily malfunction if a radioactive isotope were to decay in the vicinity. This morning, against the odds, someone did in fact leave a small amount of a radioactive isotope near the world's most accurate clock in Smith's office.

* The clock is running normally at 8:22a.m. when Smith enters her office. Smith takes a good hard look at the world's most accurate clock—what she knows is an extremely well-designed clock that has never been tampered with—and forms the true belief that it is 8:22a.m.

Preliminary Diagnosis

* Since that danger remains purely counterfactual—since the clock could have malfunctioned but in fact remained the world's most accurate clock; since things *could* have gone less well epistemically but *didn't*—it's quite tempting to allow that Smith knows the time on the basis of the clock.

* Very easily could she have believed falsely via that clock: had the isotope decayed, which it easily might have, Smith easily might have believed falsely on the basis of the clock.

* So, it seems that *Atomic Clock* is a case of a *known* belief that isn't *safe*.

Is Smith's belief *safe?*

* As any timeshare owner can tell you, some things aren't how they seem.

* Fernando Broncano-Berrocal attempts to argue that **Atomic Clock** is not all it looks to be. He concedes that Smith does *know* the time on the basis of the clock. However, Broncano-Berrocal argues that, contrary to appearances, Smith's belief is formed *safely* in **Atomic Clock**.

What method did Smith use?

* Evaluating whether the safety condition holds requires that we examine the method used by Smith in the actual world, and then compare the results of *that very method* in nearby possible worlds.

* Checking a broken clock is a different way of forming beliefs about the time than is checking a working clock. The method that Smith actually uses—checking the world's most accurate clock—is different from the method that would easily have led her astray, namely checking a clock that was stopped by that radioactive isotope.

Broncano-Berrocal Proposal

* Even though they feel the same "from the inside," forming a belief about the time *via this clock while it's working* is different from forming such a belief *via this clock while it's broken*.

* (R4) For any type of method of belief-formation m_1 and for any type of method of belief-formation m_2, $m_1 = m_2$ if and only if

    * (i) m_1 and m_2 are globally reliable to the same degree with respect to the same field of propositions and the same range of circumstances,

    * (ii) they are both based on vision or olfaction or audition or taction or gustation or testimony or deduction or induction or memory etc., and

    * (iii) the circumstances in which the target belief is formed via m_2 are in the set of circumstances with respect to which m_1 is globally reliable.

If Broncano-Berrocal is right, is Smith's belief safe?

* Broncano-Berrocal claims that "[i]n the circumstances in which Smith uses m_2, a radioactive isotope has decayed disrupting the clock's sensor and stopping the clock.

* Those circumstances are not in the set of circumstances with respect to which m_1, Smith's actual method, is globally reliable. Therefore, condition (iii) does not hold."

* And therefore, **Atomic Clock** does not, contrary to appearances, feature a subject who knows via a method that could have gone awry. It features, rather, a subject who knows via a method (a working clock), but who easily might have failed to know via a *different* method that would have been indistinguishable to her "from the inside" (a broken clock). But that doesn't show that the method the subject *actually* used was unsafe.

Broncano-Berrocal's Argument

* 1. If (R4) is true, then Smith's belief is *safe* in Atomic Clock.

* 2. (R4) is true.

* 3. Therefore, Smith's belief is *safe* in Atomic Clock.

Our response

* (R4) is false.

Why is (R4) false?

* First, (R4) entails that there are no unreliable methods.

* Second, (R4) entails that there are no reliable methods that have failure conditions.

Are there unreliable methods?

* Are there any unreliable ways of forming beliefs? No need to answer; of course there are.

* Take tyromancy: divination based on observing cheese, especially when it is coagulating. Call that method "m_1," and imagine employing it to arrive at some belief: you consult your coagulating cheese, and interpret its pattern as predicting that the Democrats will gain control of the United States House of Representatives in 2014.

Are there unreliable methods?

* Now imagine using that method again, at some later time, in similar circumstances, to answer the same question. On the second occasion, call it "m_2." Is $m_1 = m_2$? By hypothesis, yes, it's the same terrible method: tyromancy. So far, so good; this all seems perfectly possible.

* (R4) For any type of method of belief-formation $m1$ and for any type of method of belief-formation m_2, $m_1 = m_2$ if and only if

  * (iii) the circumstances in which the target belief is formed via m_2 are in the set of circumstances with respect to which m_1 is globally reliable.

Can a reliable method ever fail?

* Suppose, for example, that you form a belief about the product of a complicated multiplication problem on the basis of your electronic calculator here, dry, at sea-level. name that method "m_1."

* Now suppose you use your calculator again, but this time under water, where calculators like that are unreliable. On that second occasion, baptize the method you used "m_2." Is $m_1 = m_2$?

* Intuitively, we'd say, yes of course: you've used the same method in two different circumstances (the first auspicious, the second not).

Can a reliable method ever fail?

* (R4) For any type of method of belief-formation $m1$ and for any type of method of belief-formation m_2, $m_1 = m_2$ if and only if
    * (iii) the circumstances in which the target belief is formed via m_2 are in the set of circumstances with respect to which m_1 is globally reliable.
* In the case we described, condition (iii) is not satisfied. The circumstances in which the target belief is formed via m_2 are *not* in the set of circumstances with respect to which m_1 is globally reliable. By hypothesis, m_1 is unreliable under water, and m_2 was used under water. So, (R4) entails that $m1 \neq m2$.

Should one accept (R4)?

* Broncano-Berrocal offers the safety theorist a victory so ruinous it's tantamount to defeat. Instead, let's cleave to the conviction that methods can fail.

* In **Atomic Clock,** Smith's clock is the same method of telling the time whether the clock is running or stopped (though we suggest you use that method only when it's running).

Conclusion

* But if there really is only one method in play in **Atomic Clock,** then Smith is indeed awash in a sea of nearby possible worlds where she forms false beliefs using the same method that delivered a true belief in the actual world.

* And so her belief really was formed unsafely, despite the fact that—as Broncano-Berrocal admits—Smith's belief is genuine knowledge.

* But then knowledge need not be safe, and the safety condition is false. And therefore safety really is in serious danger.