The puzzle of the unmarked clock and the new rational reflection principle*

Adam Elga

August 9, 2012

1

Will it rain tomorrow? I'm not sure. My credence (subjective probability) that it will rain is 40%.

Is 40% the rational credence for me to have that it will rain? I'm not sure about that either. Properly taking all of one's evidence into account can be tricky. I'm not sure that I've done it exactly right.[1]

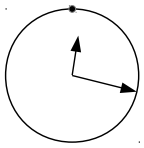
So: I am uncertain whether it will rain. And I am uncertain about the rational degree of belief for me to have that it will rain.

Is there a principle that links these two sorts of uncertainty? Or can any old beliefs about the weather be rationally combined with any old beliefs about what it is rational to believe about the weather?[2]

2

There does seem to be a principle that links the two sorts of uncertainty. Below I will motivate an improved version of just such a principle. But first, a puzzle:

Take a quick look at this picture of an "irritatingly austere" clock, whose minute hand moves in discrete 1-minute jumps:[3]



*Edited version to appear in *Philosophical Studies*. Please cite published version. Thanks to David Christensen, Paulina Sliwa, Sophie Horowitz, Maria Lasonen-Aarnio, Michael Titelbaum, Jenann Ismael, participants in the 2011 Brown Epistemology workshop, the Corridor Group, the Princeton Formal Epistemology reading group, and the 2012 Bellingham Summer Philosophy Conference, an audience at Stanford University, and especially Joshua Schechter.

[1]Compare to Christensen (2007, 2010).

[2]Here and elsewhere I use talk of "beliefs" as shorthand for talk of degrees of belief.

[3]The clock example is due to Williamson (2007), as adapted by Christensen (2010, 122-125), whose discussion I follow in this section. "Irritatingly austere" is from Williamson (2010, 13).

If your eyes are like mine, it won't be clear whether the clock reads 12:17 or some other nearby time. What should you believe about the time that the clock reads?

That, it seems, depends on what the clock really reads. If the clock really reads 12:17, then you should be highly confident that it reads a time near 12:17—99% confident, say, that the time is within a minute of 12:17. But you should be highly uncertain as between 12:16, 12:17, and 12:18. For definiteness, suppose that given your visual acuity, you should have roughly the same degree of belief in each of these three possibilities.

If the clock had instead indicated a different time—4:03, say—you should have instead been 99% confident that the time was within a minute of 4:03, but highly uncertain as between 4:02, 4:03, and 4:04. And the corresponding pattern holds for any other time, as well.

But now there is a problem. Suppose that you are 99% confident in H, the proposition that the time is within a minute of 12:17. You ask yourself: is 99% the rational level of confidence for you to have in H? You might reason as follows. Either the time (indicated on the clock) is 12:17 or not:

- If the time is 12:17, then 99% is the **correct** level of confidence for you to have in H.
- If the time is not 12:17, then 99% is an **irrationally high** level of confidence for you to have in H. For example, if the time is really 12:18, then you should have less than 99% confidence in H. For in that case, you should have more than 1% confidence that the time is 12:19, a possibility incompatible with H.

So on the one hand, you have 99% confidence in H. On the other hand, you think that 99% is a level of confidence that is definitely not too low, and is probably too high (since you think that the time is probably not exactly 12:17).[4] But that looks irrational.

Compare: Pangloss is 99% confident that the next round of Mideast peace talks will succeed. But he also thinks that he is often irrationally overconfident in good outcomes, and never irrationally underconfident in them. As a result, he thinks that 99% is a level of confidence that is definitely not too low, and is probably too high.[5]

Pangloss seems to have an irrational combination of attitudes. And, at least at first glance, your attitudes toward the unmarked clock look to be just

[4]Compare to Christensen (2010, 124): "it seems that [a subject looking at the clock] should think that .3 is probably too high a credence for her to have that [the clock reads a particular time], and certainly not too low."

[5]This example is based on the case of Brayden from Christensen (2010, 121-122).

as irrational. But given your imperfect ability to distinguish nearby times, your attitudes toward the clock seem to be perfectly *rational*. The puzzle is to resolve this conflict. What degrees of belief *should* you have about what time the clock displays?

3

A careful treatment of the puzzle would probe some of the assumptions made in the informal presentation above. Is the puzzle an artifact of the idealized way in which the setup was assumed to be completely rotationally symmetric? Or of a questionable assumption that the clock viewer is sure just how the position of the clock hand determines what it is rational for her to believe? Or of an undefended assumption that the viewer has perfect access to her exact degrees of belief? Or of the assumption that it is rational for the clock viewer to be uncertain what it is rational to believe?

These are all fair questions, but we needn't get caught up in the details. For the puzzle of the clock is an instance of a much more general conflict, a conflict that can't be avoided by tweaking the details of the clock setup. And laying out and resolving the more general conflict will resolve the puzzle as a side-effect.

4

To begin laying out the general conflict, recall the question from the end of section 1: can any old beliefs about the weather be rationally combined with any old beliefs about what it is rational to believe about the weather?

The answer is: no. For example:

> Joe is certain just what degrees of belief he ought to have. In particular, he is certain that he ought rationally have degree of belief 99% that it will rain. But despite this, his degree of belief that it will rain is only 1%.

I hope you agree that Joe's combination of attitudes is unreasonable. But if not, imagine chatting with Joe about the weather.

> "The evidence strongly supports that it will rain," he might say. "There are plenty of storm clouds nearby, and the barometric pressure is low. Furthermore, it has rained every day for the last month, and this is the rainy season. Yes, I'm quite certain of exactly what degrees of belief are rational for me, and that

> it is rational for me to be extremely confident that it will rain tomorrow."
>
> "So, will it rain tomorrow?" you ask.
>
> "No."

This dialogue makes dramatic that Joe's beliefs about the weather do not mesh properly with his beliefs about what he should believe about the weather.

Joe is unreasonable because he violates the following constraint[6]:

CERTAIN Whenever a possible rational agent is certain exactly what degrees of belief she ought rationally have, she has those degrees of belief.

This constraint is extremely plausible. But it covers only a very specific case—the case in which one is *certain* just what one should believe. Can it be generalized to cover cases in which one is uncertain about what degrees of belief one should have?

To see one natural way of generalizing the constraint, modify the case of Joe. Suppose that Joe is not certain that his degree of belief in rain should be *exactly* 99%. Instead, suppose that he is just certain that it should be quite high—say, greater than 90%. But despite this, Joe's degree of belief that it will rain is only 1%.

Again, Joe's combination of attitudes looks unreasonable.

If Joe is rational, it seems, his degree of belief that it will rain should be somewhere in the range of values that he thinks might be rational. Indeed, it seems that his degree of belief that it will rain should be some kind of average of those values.

This suggests a tempting way to generalize CERTAIN:[7]

RATIONAL REFLECTION $P(H|P' \text{ is ideal}) = P'(H)$ whenever P is the credence function of a possible rational subject S, H is a proposition, P' is a credence function, "ideal" means "perfectly rational for S to have in her current situation", and the conditional probability is well defined.[8]

[6]This constraint is a more cautious cousin of the "principle of non-akrasia" from Ross (2006, 277).

[7]RATIONAL REFLECTION is a variant of the principle "RatRef" from Christensen (2010, 122). A similar principle is introduced in Ross (2006, §10.3) under the name "The epistemic principal principle". Earlier work on related principles includes Goldstein (1983), Lewis (1986), van Fraassen (1984).

[8]To avoid complications, here and below I ignore self-locating beliefs, assume that

The statement above is a mouthful. But the guiding idea is simple: When one is rationally certain what credence function one should have, one should have that credence function. But when one is uncertain, then one should have as one's credence function a weighted average of the functions one thinks it might be rational to have.

For example, suppose that one is 50% confident that one should have credence function P_1 and 50% confident that one should have credence function P_2. Further suppose that $P_1(\text{rain}) = 70\%$ and $P_2(\text{rain}) = 90\%$. Then if one is rational, RATIONAL REFLECTION entails that one will have as one's degree of belief in rain the average of 70% and 90%—i.e., 80%.[9]

5

The story so far: I have presented the puzzle of the unmarked clock, and claimed that it is an instance of a more general conflict. As a first step in laying out the general conflict, I gave some motivation for RATIONAL REFLECTION. That principle connects one's beliefs about, say, the weather, with one's beliefs about what it is rational to believe about the weather.

To complete my explanation of the general conflict, I will need to address the question: is it ever rational to be uncertain about what it is rational to believe?

The answer is: yes. For example:[10]

HYPOXIA Bill the perfectly rational airline pilot gets a credible warning from ground control:

> Bill, there's an 99% chance that in a minute your air will have reduced oxygen levels. If it does, you will suffer from hypoxia (oxygen deprivation), which causes hard-to-detect minor cognitive impairment. In particular, your degrees of

every situation determines a unique ideally rational probability function, and assume that credences are countably additive and defined over a space of possibilities that is at most countably infinite.

[9]For readers familiar with the Reflection Principle from van Fraassen (1984), the Principal Principle from Lewis (1986), or RatRef from Christensen (2010), it may be helpful to note that RATIONAL REFLECTION entails

RATREF $P(H|I(H) = x) = x$
whenever P is the credence function of a possible rational subject S, H is a proposition, x is a real number, "$I(H) = x$" denotes the proposition that the ideal probability for S to have in H is x, and the conditional probability is well defined.

[10]Cf. Elga (2008), Christensen (2010).

> belief will be slightly irrational. But watch out—if this happens, everything will still seem fine. In fact, pilots suffering from hypoxia often insist that their reasoning is perfect—partly due to impairment caused by hypoxia!
>
> A few minutes later, ground control notices that Bill got lucky—his air stayed normal. They call Bill to tell him. Right before Bill receives the call, should he be uncertain whether his degrees of belief are perfectly rational?

The example invites us to answer "yes", for the following reason. Before Bill is told that he got lucky, he should be uncertain whether he is suffering from hypoxia, and so should be uncertain whether his degrees of belief are perfectly rational. And he should be uncertain about what degrees of belief it is rational for him to have.

One might reject this analysis. One might claim that Bill should be absolutely certain that he got lucky and avoided hypoxia. But that is implausible. Such certainty would be overconfidence on Bill's part—a failure to properly take into account ground control's credible warning.

The case of Bill shows that in some situations, rationality is compatible with uncertainty about what degrees of belief are rational. Indeed, Bill should think that he is in exactly such a situation. Let us record these conclusions:

MODESTY In some possible situations, it is rational to be uncertain about what degrees of belief it is rational for one to have. Furthermore, it can be rational to have positive degree of belief that one is in such a situation.

6

An independent argument supporting MODESTY goes by way of uncertainty about evidence. One way to be uncertain what one should believe is to be uncertain what evidence one has. The conclusions of anti-luminosity arguments from Williamson (2000, 2008) entail that in some situations, rationality is compatible with uncertainty about what evidence one has. And they entail that it can be rational to suspect that one may be in such a situation. So anti-luminosity arguments provide a route to MODESTY available even to those who reject the argument based on HYPOXIA.

7

So far I have argued for MODESTY, which entails that it is sometimes rational to be uncertain about what it is rational to believe. And I have given a motivation for RATIONAL REFLECTION, which is a constraint on how one's opinions of what is rational to believe ought to mesh with the rest of one's opinions. I hope I've convinced you that both of these claims are true.

It was a trap.

It turns out that MODESTY and RATIONAL REFLECTION are inconsistent with each other. That is the more general conflict I promised to explain. And it is the conflict at the root of the puzzle of the clock.

Here is a proof that MODESTY and RATIONAL REFLECTION are inconsistent with each other. (The proof may be skipped without loss of continuity.)

> **Proof**: Suppose that a particular subject has credence function P, and that P' is any credence function that the subject thinks might be ideal. Then if RATIONAL REFLECTION is true, for any proposition H, $P(H|P' \text{ is ideal}) = P'(H)$. In particular, when H is the proposition that P' is ideal:
>
> $$P(P' \text{ is ideal}|P' \text{ is ideal}) = P'(P' \text{ is ideal}).$$
>
> By the definition of conditional probability, the left hand side of this equation equals 1. So P' is *immodest*, in the sense that it assigns credence 1 to the claim that it itself is the ideal credence function for the subject to have. And the same is true for every credence function that the subject thinks might be ideal for her. So the subject is certain that rationality requires her to be certain about what credences it is rational to have. This conflicts with (the second sentence of) MODESTY.[11,12]

So we have an apparent paradox: we had initially plausible motivations for believing both MODESTY and RATIONAL REFLECTION, but now have seen that the two claims conflict.[13]

[11] Williamson (2010, Appendix) proves related results, delivering conditions for the necessary truth of RatRef (a reflection principle closely related to RATIONAL REFLECTION—see note 7).

[12] As was pointed out to me by Michael Titelbaum, the proof in the text is analogous to the proof that the "Old" Principal Principle is in tension with Humeanism about laws of nature (Hall 1994, Lewis 1994).

[13] One might try to avoid the conflict by advocating not RATIONAL REFLECTION but the very similar principle RatRef (described in footnote 7). This does not work because it is

The puzzle of the clock is an instance of this conflict. For recall that the puzzle depended on the assumption that the clock viewer should be uncertain about what it is rational for her to believe about the time. That assumption is an instance of MODESTY. And it depended on the assumption that it is unreasonable for the viewer to be 99% confident in a proposition, while thinking that 99% is a level of confidence that is certainly not too low and probably too high. That assumption derives from the same considerations that motivate RATIONAL REFLECTION.

How should the conflict between MODESTY and RATIONAL REFLECTION be resolved? There are a number of options:[14]

- We might reject MODESTY, and claim for example that rationality requires one to be certain just what degrees of belief are rational. This would require us to say that Bill the pilot should be certain that he has avoided hypoxia, and that the clock viewer should be certain exactly what it is rational for her to believe about the clock. That seems desperate.[15]

- We might reject RATIONAL REFLECTION and similar principles, insisting that beliefs about what it is rational to believe do not impose systematic constraints on one's other beliefs.[16] A defender of this line takes on the

unreasonable to accept both RatRef and MODESTY. That is because it is unreasonable to accept MODESTY without also accepting the following claim, which itself is incompatible with RatRef:

POSITIVE It can be rational to think something of the form "I'm not sure how confident I should be that P' is ideal. Maybe I should be 20% confident that it is, maybe 21%, or maybe another value." (Here 20% and 21% are placeholders for any two positive values, and P' can be any credence function.)

Proof that POSITIVE is incompatible with RatRef: For brevity, use "$V(H, v)$" to denote the proposition that the ideal credence to have in H is v, and use "$I(P')$" to denote the proposition that P' is the ideally rational credence function to have. Now suppose POSITIVE is true. Then for some possible rational person who has credence function P, there are distinct positive values x and x' such that $P(V(I(P'), x))$ and $P(V(I(P'), x'))$ are both greater than 0. But $I(P')$ is compatible with at most one of $V(I(P'), x)$ and $V(I(P'), x')$, since it settles what the ideal probability to have in $I(P')$ is. So at least one of $P(I(P')|V(I(P'), x))$ and $P(I(P')|V(I(P'), x'))$ equals zero. But RatRef entails that these two terms equal x and x' respectively, which are both positive. So POSITIVE contradicts RatRef.

[14]Compare to Christensen (2010, 124).

[15]Alternatively, we might reject the second sentence of MODESTY by saying that rationality requires one to be certain of the following falsehood: rationality requires one to be certain exactly what time the clock indicates. That seems just as desperate, since a rational viewer of the clock might well realize that she has an imperfect ability to discriminate nearby times. Thanks here to Kenny Easwaran.

[16]Williamson (2010) seems sympathetic to this approach.

burden of explaining away the initial appeal of such principles, and the seeming irrationality of the clock viewer's "99% confidence is probably too high and definitely not too low" stance.

- We might say that MODESTY and RATIONAL REFLECTION both express rational ideals, but admit that some rational ideals conflict with others.[17] This may be defensible in the end,[18] but there is a cost to admitting that the notion of perfect rationality is itself inconsistent. And this proposal seems not to give a clear answer to the question: what degrees of belief should the clock viewer have about the time?

This brief survey of options is not exhaustive, and the objections I have raised are not conclusive. But I hope to convince you that we can do better. We can resolve the conflict in a way that allows us to consistently hold on to MODESTY, and also to the considerations that motivate RATIONAL REFLECTION.[19]

All we need to do is amend RATIONAL REFLECTION. Let me explain.

8

Think back to how RATIONAL REFLECTION was motivated above (in section 4). The story started with this constraint:

CERTAIN Whenever a possible rational agent is certain exactly what degrees of belief she ought rationally have, she has those degrees of belief.

This constraint is extremely plausible and extremely cautious. It doesn't even rule out that one can be rationally certain that one is irrational. It just rules out that one can be rationally certain *exactly* what degrees of belief one should have, without having those degrees of belief.

The next step was to generalize this constraint to cover cases in which one is uncertain what degrees of belief one ought to have. It was suggested that when one is uncertain what credence function is rational, one should have a credence function that is a particular weighted average of the ones that one thinks might be rational. That is RATIONAL REFLECTION.

[17] This is proposed as a fallback position in Christensen (2010, 136).

[18] It may be defensible, for example, if we are forced to accept a similar conclusion about independent cases, as is argued in Christensen (2007).

[19] Compare: "the ideal outcome of thinking about [the puzzle of the clock] would be our finding a way of accommodating the intuitions behind [a principle similar to RATIONAL REFLECTION] while avoiding the difficulties we've been examining." (Christensen 2010, 136)

There was nothing wrong with the first step of the story: CERTAIN is correct. And there was nothing wrong with trying to generalize CERTAIN to cover more cases. But RATIONAL REFLECTION is the wrong way to generalize CERTAIN.

A better way of generalizing CERTAIN is brought out by the following line of reasoning.[20]

Suppose that you're considering what credence function it would be rational for you to have. Consider the candidate functions — the ones that you think might be ideally rational — as a kind of panel of purported experts. In the special case that you are sure what function is ideal, the panel contains just a single member, and you're sure that she is the true expert. In that case, you should just believe what the expert believes. That corresponds to CERTAIN.

But now suppose that you are uncertain which function is ideal. In that case, it is as if the panel contains a number of purported experts and you are uncertain which one is the true expert.

For concreteness, suppose that the panel consists of credence functions named Cassandra, Merlin, and Sherlock. Conditional on Sherlock being the true expert, what credences should you have?

It is tempting to answer: the ones that Sherlock has. That is how RATIONAL REFLECTION answers the question. But that answer is not in general correct. For Sherlock might himself be uncertain who is the true expert. And conditional on Sherlock being the true expert, you should *not* be uncertain who the true expert is.

So: conditional on Sherlock being the true expert, you shouldn't align your credences to Sherlock's. What should you do?

A warm-up question will point the way to the answer: What should be your credence that it will rain tomorrow, given that Sherlock is the true expert *and that many people will use umbrellas tomorrow*?

Answer: your credence should be rather high. And it should not in general equal Sherlock's unconditional credence that it will rain. For the information that many people will use umbrellas tomorrow provides strong evidence that it will rain tomorrow.

This suggests that your conditional credence should not equal Sherlock's credence that it will rain. Rather, it should equal Sherlock's credence that it will rain *conditional on (at least) the information that many people will use umbrellas tomorrow*.

[20]The line of reasoning, including the example involving the panel of experts, is due to Hall (1994, 510). Christensen (2010, 135) comes tantalizingly close to applying such reasoning to RATIONAL REFLECTION, but pulls back at the last moment.

More generally: your credences, conditional on Sherlock being the true expert, should equal Sherlock's credences conditional on Sherlock being the true expert. Further generalizing this thought yields the following principle:[21]

NEW RATIONAL REFLECTION $P(H|P' \text{ is ideal}) = P'(H|P' \text{ is ideal})$
whenever P is the credence function of a possible rational subject S, H is a proposition, P' is a credence function, "ideal" means "perfectly rational for S to have in her current situation", and the conditional probability is well defined.

An example will help illustrate the difference between the new principle and the old. Suppose that you are 50% confident that you should have credence function P_1 and 50% confident that you should have P_2. RATIONAL REFLECTION entails that if you are rational, your probability for a proposition H will be the average of $P_1(H)$ and $P_2(H)$. In contrast, NEW RATIONAL REFLECTION entails that if you are rational, your probability for H will be the average of $P_1(H|P_1 \text{ is ideal})$ and $P_2(H|P_2 \text{ is ideal})$.

(There is another strategy to motivate the new principle:[22] Suppose that a subject starts out wondering: what degrees of belief are rational for me? And suppose that she then learns the answer to that question. She will end up certain just what she ought to believe. And so CERTAIN will impose a constraint on her final state of mind. But that will indirectly impose a constraint on her initial state of mind—by way of an assumption about how the subject should update her beliefs when she gets new information. In other words, we can generalize CERTAIN by saying: Rational agents have states of mind that are consistent with CERTAIN, and *would remain consistent with* CERTAIN *were they to learn the truth about what they ought to believe*. This strategy yields a derivation of NEW RATIONAL REFLECTION in a special case, and so lends some credence to the truth of the principle in full generality.)[23]

[21]The constraint is named NEW RATIONAL REFLECTION because it stands to RATIONAL REFLECTION as the "New Principal Principle" stands to the "Principal Principle" (Hall 1994, Lewis 1994). (NEW RATIONAL REFLECTION also stands to RATIONAL REFLECTION as the "guru principle" from Elga (2007) stands to the "Reflection Principle" from van Fraassen (1995).) The derivation from footnote 23 of a special case of NEW RATIONAL REFLECTION can be adapted to yield a parallel derivation of a special case of the New Principal Principle.

[22]This second strategy adapts an idea from Ross (2006, 277-299). A similar argument was independently suggested to me by Boris Kment.

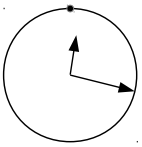
[23] The special case: a rational agent has probability function P at time 0, realizes that rational agents update their beliefs by conditionalization, and realizes that she is about to conditionalize on the truth of what probability function is ideal-for-her-at-time-0. Together with CERTAIN, these assumptions entail that the agent satisfies the instance of NEW

Moral: NEW RATIONAL REFLECTION is the right way to generalize CERTAIN. It expresses the manner in which a subject's opinions about what it is rational to believe constrain her other opinions. And it is perfectly consistent with MODESTY. So the conflict between MODESTY and RATIONAL REFLECTION has a satisfying resolution: drop RATIONAL REFLECTION and adopt NEW RATIONAL REFLECTION instead.

The end. Except for one final matter: it remains to address the puzzle of the clock.

9

Recall the setup:



You look at the clock, ending up 99% confident in H, the proposition that the time is either 12:16, 12:17, or 12:18. You are highly uncertain as between those three possibilities, assigning, say, 33% of your confidence to each of them.[24] That pattern of attitudes looks reasonable.

But as we saw in section 2, that pattern of attitudes also entails that you think that 99% is a level of confidence in H that is **definitely not too low** for you, and is **probably too high**. That makes your 99% confidence in H look unreasonable.

RATIONAL REFLECTION that applies to her.

Proof: Under the above conditions, suppose that agent conditionalizes on the information that probability function P' is ideal-for-her-at-time-0. As a result, at time 1 her new probability function is $P(-|P'$ is ideal-at-0$)$. By the assumption that rational agents conditionalize, this new probability function is ideal-for-her-at-time-1.

But at time 1, the agent is certain that P' was ideal-for-her-at-time-0. And she is certain that at time 0 she remained ideally rational by conditionalizing on the truth about what function was ideal-for-her-at-time-0. So she is at time 1 certain that the following probability function is ideal-for-her-at-time-1: $P'(-|P'$ is ideal-at-0$)$.

Now CERTAIN applies to the agent at time 1, and entails that the agent's probability function at time 1 is $P'(-|P'$ is ideal-at-0$)$. We have now derived two expressions for the agent's probability function at time 1. Equating them yields that for any proposition H:

$$P(H|P' \text{ is ideal-at-0}) = P'(H|P' \text{ is ideal-at-0}),$$

which is an instance of NEW RATIONAL REFLECTION.

[24]Nothing substantial would change if the 33%/33%/33% distribution were changed to one that favored 12:17 over the other two times.

So: the first line of reasoning concludes that your beliefs about the clock are reasonable. The second line concludes that those beliefs are unreasonable. What has gone wrong? That is the puzzle.

The answer is that the second line of reasoning is wrong. For what lies behind that reasoning is the thought that one's degree of confidence in H should always be a weighted average of the degrees of confidence that one thinks might be rational. That is an initially tempting thought. And it is a thought that follows from RATIONAL REFLECTION. But it is incorrect.

In contrast, this "averaging" thought *doesn't* follow from NEW RATIONAL REFLECTION. That is why the state of uncertainty about the clock described above is compatible with NEW RATIONAL REFLECTION. The bottom line is that there just isn't anything wrong with your state of uncertainty about the clock.

Then why does that state of uncertainty *seem* unreasonable?[25] Because it superficially resembles states of uncertainty that are genuinely unreasonable.

For instance, your state of mind superficially resembles the state of mind of Pangloss, the self-aware optimist from section 2:

> Pangloss is 99% confident that the next round of Mideast peace talks will succeed. But he also thinks that he is often irrationally overconfident in good outcomes, and never irrationally underconfident in them. As a result, he thinks that 99% is a level of confidence that is definitely not too low, and is probably too high.

Pangloss seems to exhibit the same pattern of uncertainty that you do. And Pangloss *is* unreasonable. That provides additional temptation to think that you, too, are unreasonable.

But the cases are different, and the reason Pangloss is unreasonable does not apply to you as a viewer of the clock. Let me explain what makes Pangloss unreasonable, and why no corresponding consideration applies to the viewer of the clock.

Let S be the proposition that the peace talks will succeed, and let P_G be Pangloss's credence function. For simplicity, suppose that Pangloss is sure that the ideal credence for him to have in S is either 99% or 66%, but has no idea which. It follows[26] that Pangloss has approximately 99% credence in S,

[25] For an independent and complementary diagnosis, see Horowitz and Sliwa (2011).

[26] It follows by the probability calculus that Pangloss's credence in S is a weighted average of $P_G(S|\text{99\% is ideal})$ and $P_G(S|\text{66\% is ideal})$. So this weighted average equals 99%. But since each of these terms is no greater than 100%, and since they are weighted roughly equally in the average, the only way for their average to be 99% is for each of them to be close to 99%.

conditional on the rational credence in S being 66%:

$$P_G(S|66\% \text{ is ideal}) \approx 99\%.$$

There looks to be a mismatch here. And there is: on any natural understanding of the case, such a conditional credence is totally unreasonable. Conditional on 66% being the ideal credence to have that the talks will succeed, Pangloss's credence that the talks succeed should *not* be approximately 99%. Rather, it should be approximately 66%.

Compare: in an ordinary case, one's credence that it will rain next week, conditional on the rational credence being 66%, should be approximately 66%. Only with a very special back-story would it make sense for that conditional credence to be anywhere near 99%. And the same is true for Pangloss's conditional credence above.

That is why Pangloss is unreasonable.

At first glance, the situation with the viewer of the clock looks similar. In particular, the viewer of the clock has 100% credence in H, conditional on the rational credence in H being 66%:

$$P(H|66\% \text{ is ideal}) = 100\%.$$

That conditional credence looks to involve the same sort of mismatch as Pangloss's. But in the clock case, the mismatch is only apparent. For the clock case has the following very special feature. Given the setup, the information "66% is the ideal credence to have in H" is strong evidence that H is true. Indeed, it is conclusive evidence that H is true, since the clock viewer is certain that

66% is the ideal credence for the viewer to have in H
only if
the time is either 12:16 or 12:18.

No corresponding claim holds for Pangloss.

The bottom line is that clock viewer's conditional credences exhibit the same apparent mismatch that Pangloss's do. The difference is that in the viewer's case, the mismatch is *only* apparent.

References

David Christensen. Does Murphy's Law apply in epistemology? Self-doubt and rational ideals. *Oxford Studies in Epistemology*, 2:3–31, 2007.

David Christensen. Rational reflection. *Philosophical Perspectives*, 24:121–140, 2010.

Adam Elga. Reflection and disagreement. *Noûs*, 41:478–502, 2007.

Adam Elga. Lucky to be rational. Paper presented at Bellingham Summer Philosophy Conference, June 2008. URL `http://www.princeton.edu/~adame/papers/bellingham-lucky.pdf`.

Michael Goldstein. The prevision of a prevision. *Journal of the American Statistical Association*, 78(384):817–819, 1983.

Ned Hall. Correcting the guide to objective chance. *Mind*, 103:505–518, 1994.

Sophie Horowitz and Paulina Sliwa. Level-bridging, uncertainty, and akrasia. Manuscript, June 2011.

David Lewis. A subjectivist's guide to objective chance. In *Philosophical Papers*, volume 2. Oxford University Press, Oxford, 1986.

David Lewis. Humean supervenience debugged. *Mind*, 103:473–490, 1994.

Jacob Ross. *Acceptance and practical reason*. PhD thesis, Rutgers University, October 2006.

Bas C. van Fraassen. Belief and the will. *Journal of Philosophy*, 81:235–256, 1984.

Bas C. van Fraassen. Belief and the problem of Ulysses and the Sirens. *Philosophical Studies*, 77:7–37, 1995.

Timothy Williamson. *Knowledge and its limits*. Oxford University Press, Oxford, 2000.

Timothy Williamson. Improbable knowing. Notes, 2007. URL `http://www.philosophy.ox.ac.uk/__data/assets/pdf_file/0014/1319/Orielho.pdf`.

Timothy Williamson. Why epistemology cannot be operationalized. In Quentin Smith, editor, *Epistemology, new essays*, chapter 11, pages 277–300. Oxford University Press, 2008.

Timothy Williamson. Very improbable knowing. Draft manuscript, 2010. URL `http://www.philosophy.ox.ac.uk/__data/assets/pdf_file/0015/19302/veryimprobable.pdf`.