FUMERTON'S PUZZLE FOR THEORIES OF RATIONALITY

Ru Ye

Richard Foley has presented a puzzle purporting to show that all attempts in trying to find a sufficient condition of rationality are doomed. The puzzle rests on two plausible assumptions. The first is a level-connecting principle: if one rationally believes that one's belief p is irrational, then one's belief p is irrational. The second is a claim about a structural feature shared by all promising sufficient conditions of rationality: for any such condition, it is possible that one's belief satisfies it and yet one rationally believes that it doesn't. With the two assumptions, Foley argues that a sufficient condition of rationality is impossible. I explain how exactly the puzzle goes and try to offer a solution. If my solution works, all theorists of rationality who accept certain level-connecting principles will need to add an extra condition to their favorite rationality-making condition.

Keywords: rationality, Fumerton's puzzle, level-connecting principle

1. Introduction

A theory of rationality answers the question 'what is it for a belief to be rational' by saying

> For any person S and any proposition p, S's belief that p is rational if and only if it satisfies condition C.

Here 'C' refers to different properties according to different theories. For example, it refers to 'fitting one's evidence' according to evidentialism, and it refers to 'being reliably formed' according to (a crude form of) reliabilism.

Each theory has its own problems. However, Richard Foley [1990] has presented a puzzle purporting to show that all such theories would fail. The puzzle rests on two assumptions. The first is a level-connecting principle, which says that if one rationally believes that one's belief p is irrational, one's belief p is irrational. The second is a claim about a structural feature shared by all promising candidates of rationality C: for any C, it is possible that one's belief p satisfies C and yet one rationally believes that it doesn't. With the two assumptions, Foley argues that any interesting candidate C is not sufficient for rationality.

Foley calls the puzzle 'Fumerton's puzzle', since it originates from Richard Fumerton [1990: 117–28]. This puzzle, as Foley observes, has largely gone unnoticed.[1]

[1] Lasonen-Aarnio [2014] discusses a problem similar to the puzzle and argues that the problem can be solved only if we reject NEGATIVE (which I explain later). But I think this principle is very plausible and I want to explore solutions without giving it up.

I try to offer a solution to Fumerton's puzzle in this paper. In section 2, I explain how the puzzle goes. In section 3, I argue that Foley doesn't get the real source of the puzzle and bring out what I think is the real source. Then I offer a solution in section 4, refine it in section 5 and address a potential worry about the solution in section 6. If my solution works, all theorists of rationality (who take certain level-connecting principles seriously) will need to add an extra condition to their favourite rationality-making conditions.

2. How the Puzzle Goes

Fumerton's puzzle relies on what are called 'level-connecting principles', which connect *what is rational to believe* to *what is rational to believe what is rational to believe*. Various versions of level-connecting principles are discussed in the literature (see Smithies [2012]). The one relevant to the puzzle is

> NEGATIVE
> If it is rational for S to believe that her belief p is irrational, then it is irrational for her to believe p.

This principle is widely accepted[2], and the most commonly seen reason is summarised by Declan Smithies [2012: 285] as follows. Suppose for reductio that NEGATIVE is false. Then there would be cases in which it is rational for a person S to believe that her belief p is irrational, and yet it is still rational for her to believe p. In at least some of those cases, it would be rational for S to believe 'p, but it is irrational for me to believe p'. Then it looks like she could assert 'p, but it is irrational for me to believe that p'. But such assertions seem irrational. How could you rationally assert 'it is going to rain tomorrow, but it is irrational for me to believe that it is going to rain tomorrow'? If you should believe that it is irrational for you to believe that it is going to rain tomorrow, it seems that you should just give up that belief.

Given a level-connecting principle like NEGATIVE, we can see how exactly Fumerton's puzzle goes. It starts with the claim

> POSSIBLE-1
> All interesting candidates C of rationality share this feature: there exists some proposition p such that it is possible that one's belief that p satisfies C and yet one rationally believes that it doesn't.[3]

[2] For example, Lasonen-Aarnio [2014], Bergmann [1997], Christensen [2007, 2010], Feldman [2005], Smithies [2012] and Vogel [2006] all endorse NEGATIVE. Besides, as Bergmann [1997: 405] has noticed, Goldman, Nozick and Plantinga all include a 'no defeater' condition (namely, one doesn't believe that one's belief is defeated) into their accounts of justification or warrant, which suggests that they would also endorse NEGATIVE. That said, this principle is not uncontroversial; it is denied, for example, by Williamson [2011].

[3] I restrict POSSIBLE-1 to all *interesting* or *promising* candidates C of rationality, because there are some trivial sufficient conditions that do not have the feature mentioned in POSSIBLE-1. For example, 'being rational' is sufficient for rationality, and yet if NEGATIVE is true, it is not possible that one's belief has the property of being rational but one rationally believes that it doesn't. But we can all agree that 'being rational' is not an interesting candidate of rationality. It is just not the kind of properties epistemologists have in mind when they look for rationality-making properties.

To see this, suppose C is 'fitting one's evidence'. It seems possible that my belief that p fits my evidence and yet it is rational for me to believe that it doesn't. For example, it is possible that I have a perfect proof for a math theorem, but I hear from my professor that the proof is problematic. Then it seems that I would have a rational but false belief that believing the theorem doesn't fit the evidence I have.[4]

Or suppose C is 'being reliably formed'. It seems possible that my belief p is in fact reliably formed and yet I have overwhelming evidence that it is not. Then it seems that I would have a rational but false belief that my belief p is not reliably formed.

So, POSSIBLE-1 is very plausible. Now suppose that I am in this situation: I rationally but falsely believe that my belief p doesn't satisfy C. Suppose also that I have a rational background belief that C is necessary for rationality. It follows that it is rational for me to believe that my belief p is irrational. By NEGATIVE, this means that my belief p is *irrational*. However, we have supposed that my belief p does satisfy C. So, C is not sufficient for rationality.

To sum up, Foley's argument is that given NEGATIVE and the common feature of C mentioned in POSSIBLE-1, any interesting candidate C of rationality is not sufficient for rationality. Therefore, we have a puzzle: a plausible level-connecting principle, combined with an observation about the structural feature of interesting candidates of rationality, implies that an interesting sufficient condition of rationality is impossible.[5]

---

[4] Thanks to a referee for this suggestion: perhaps the evidentialist could deny POSSIBLE-1 by claiming that when one acquires evidence that believing p doesn't fit one's evidence, it automatically follows that believing p doesn't fit one's evidence. So in the present case, once I hear from the professor that my proof is problematic, believing the theorem no longer fits my evidence.

However, this reply seems implausible. Before I hear from my professor, believing the theorem fits my evidence because it is entailed by the math axioms I know and I can 'see' the entailment through the perfect proof I have. When I hear from the professor, I have to believe that I must have made a mistake in the proof, but the theorem is still entailed by the axioms and in some sense I can still 'see' how it follows from the axioms. Therefore, it seems that when I hear from the professor, my overall evidence still supports the theorem.

Several authors make the same point in defending claims similar to POSSIBLE-1. For example, in discussing the peculiarity of higher-order evidence (evidence to the effect that one's belief is irrational), Christensen [2010: 192–3] argues that taking such evidence seriously would force one into a state of cognitive imperfection, because one would have to give up the attitude supported by one's overall evidence. Also see Lasonen-Aarnio [2014: 321–6] for her defence of POSSIBLE-1.

[5] It is worth noting that by 'Fumerton's puzzle' Foley means two separate puzzles. The first is the puzzle I describe here, which purports to show that any C is not sufficient for rationality. The second is a puzzle purporting to show that any C is not *necessary* for rationality. This second puzzle is quite similar to the first one, except that the level-connecting principle it relies on is

> POSITIVE
> If it is rational for S to believe that her belief p is rational, then it is rational for her to believe p.

And the assumption about the common feature of C is

> POSSIBLE-1*
> All interesting candidates C of rationality share this feature: there exists some proposition p such that it is possible that one's belief p doesn't satisfy C and yet one rationally believes that it does.

The puzzle about necessity is that if S is in a situation in which her belief p doesn't satisfy C and yet she rationally believes that it does, and if S has the background belief that C is sufficient for rationality, S would

3. What is the Real Source of the Puzzle?

As we have seen, Foley thinks that the sources of the puzzle are NEGATIVE and POSSIBLE-1. In this section, I argue that POSSIBLE-1 is *not* really the source of the puzzle. To see why, it is useful to discuss a solution that Foley offers but ultimately rejects.[6]

Since a source of the puzzle, according to Foley, is that for any alleged sufficient and necessary condition C of rationality, it is possible that one's belief satisfies C and yet one rationally believes that it doesn't, perhaps a solution is to avoid this possibility by revising C into C+:

C+(p) = C(p) and it is irrational to believe not-C(p)[7]

Foley rejects this revision because 1) the same puzzle would arise at the level of C+ just like it does at the level of C, because it is possible that one's belief satisfies C+ and yet one rationally believes that it doesn't; and 2) it would give us a viciously circular account of rationality. The account would say *One's belief p is rational if and only if 1) p satisfies C and 2) it is irrational for one to believe that it doesn't*, which is viciously circular given the role of the term 'irrational' played in the condition.

There is a third problem with the revision, which Foley doesn't address but is worth noting because it motivates what I think is the correct solution. The problem is that the revision doesn't get the source of the puzzle right. Although the puzzle starts with the

---

rationally believe that her belief p is rational. Her belief p would be rational according to POSITIVE, even though it doesn't satisfy C by supposition. Therefore, C is not necessary for rationality.

I set aside this puzzle about necessity, because I think that POSITIVE and POSSIBLE-1* are much less plausible than NEGATIVE and POSSIBLE-1 respectively. Smithies's main argument for POSITIVE is that to deny it would license as rational Moorean assertions of the form 'not-p, but it is rational for me to believe p' or 'it is an open question whether p, but it is rational for me to believe p'. But I don't think this fact lends strong support for POSITIVE, because such assertions don't sound so awkward. At the very least, they are much less awkward than the assertions permitted by the denial of NEGATIVE—assertions of the form 'p, but it is irrational for me to believe p'. Especially, if rationality could be understood as permission instead of requirement, it seems that I could refrain from believing p even though I think that believing p is permissible. For example, it is not so awkward for me to say 'it is an open question whether it is going to rain tomorrow, although it is permissible for me to think that it is going to rain tomorrow'.

POSSIBLE-1* is not so plausible because some theorists of rationality are going to deny that for their favourite condition C of rationality, it is possible that one's belief doesn't satisfy C and yet one rationally believes that it does. For example, the evidentialist Richard Feldman has the famous slogan 'evidence of evidence is evidence'. So for Feldman, it is impossible that you do not have evidence for p, and yet you have evidence (and are rational in believing) that you have evidence for p. In contrast, the same argument could not be used to deny POSSIBLE-1. There is no slogan 'evidence of non-evidence is non-evidence'.

Therefore, I set aside the puzzle about necessity given that the two assumptions it relies on are not very plausible. Of course, if you are not convinced and think that the puzzle is as serious as the one about sufficiency, you will have to view my solution to be offered later as only a partial solution to Fumerton's puzzle.

[6] Foley's final solution is to reject NEGATIVE. But I think Smithies's argument for NEGATIVE is convincing, and the principle is also too intuitive to give up. Therefore, I want to explore solutions without rejecting it.

[7] Foley's revised condition is a little different from C+. He revises C(p) into

Either C(p) *or you rationally believe that C(p)*, and it is irrational for you to believe not-C(p)

As you can see, this revision is motivated both by the puzzle about necessity resulted from POSITIVE (see my footnote 5) and by the puzzle about sufficiency resulted from NEGATIVE. Since I am only concerned with the latter puzzle, I drop the disjunct 'or you rationally believe that C(p)'.

possibility that one's belief satisfies C and yet it is rational for one to believe that it doesn't, this possibility merely nicely explains the puzzle, but it is not really the source of it. What generates the puzzle is not POSSIBLE-1 but POSSIBLE-2:

> POSSIBLE-1
> For any C, there is some proposition p such that it is possible that one's belief p satisfies C and yet *it is rational for one to believe that it doesn't.*
>
> POSSIBLE-2
> For any C, there is some proposition p such that it is possible that one's belief p satisfies C and yet *it is rational for one to believe that one's belief p is irrational.*

As I explained in section 1, Foley thinks that the common feature of C that leads to the puzzle is POSSIBLE-1. He thinks that C's sufficiency for rationality is endangered because it is possible that one's belief satisfies C and yet one rationally believes that it doesn't. But what endangers C's sufficiency for rationality is POSSIBLE-2, namely, the possibility that one's belief satisfies C and yet one rationally believes that it is irrational. This is so for two reasons.

The first is that one's rational belief 'my belief p doesn't satisfy C' doesn't have defeating power by itself, even though C is in fact necessary for rationality. That belief has defeating power only when one has a rational background belief to the effect that C is necessary for rationality.

To see this, suppose reliability is in fact necessary for rationality. I believe that it is raining outside through my visual perception, which is in fact reliable, but I hear from my physician that it is not. This testimony enables me to rationally (but falsely) believe that my raining belief is not reliably formed. Suppose, however, I have no reason at all to think that this means my belief is irrational, or unreasonable, or shouldn't be held in any way. I might (rationally) think that my visual experience is enough for my raining belief to be rational. So, from my rational point of view, my belief is still rational even though it is not reliably formed. Then it is hard to see why believing that my raining belief is not reliably formed has defeating power for the raining belief. After all, my raining belief *is* reliably formed, and from my rational point of view, this belief is still rational.[8]

Therefore, one's rational belief 'my belief p doesn't satisfy C' has defeating power only in conjunction with the rational background belief that somehow not satisfying C means that one's belief p is irrational. This shows that the feature of C that endangers its sufficiency for rationality is not the possibility that one's belief p satisfies C and yet it is rational for one

---

[8] To further support this point, notice that Smithies's argument for NEGATIVE does not apply to the following claim:

> RELIABILITY
> If one rationally believes that one's belief p is not reliably formed, one's belief p is irrational.

Denying RELIABILITY would license as rational some assertions of the form 'p, but my belief p is not reliably formed'. But this doesn't support RELIABILITY even if we suppose that reliability is in fact necessary for rationality, because such assertions can sometimes be rational. For example, if I don't believe that unreliability implies irrationality, my assertion 'p, but my belief p is not reliably produced' is not awkward, just like my assertion 'p, but my belief p is not liked by my parents' is not awkward.

to believe that it doesn't, but the possibility that one's belief p satisfies C and yet it is rational for one to believe that it is irrational. After all, the role of the background belief is just to make sure that the latter possibility is realized when the former possibility is. This is the first reason why it is POSSIBLE-2, not POSSIBLE-1, that is the source of the puzzle.

The second reason is that the possibility that one's belief p satisfies C and yet it is rational for one to believe that it is irrational could be realized in many ways—it doesn't need to go through the possibility that one's belief p satisfies C and yet one rationally believes that it doesn't. Other ways of realizing that possibility could also endanger C's sufficiency. This is because one's rational belief 'my belief p is irrational' could be arrived at through ways that have nothing to do with one's belief about C, without any impact on its defeating power.

To see this, choose your favourite C. Let it be 'fitting total evidence' or 'being reliably formed'. Suppose my belief p satisfies condition C. However, I rationally believe that a belief is rational only if my professor likes it. Then I learn that my professor doesn't like belief p and thereby form a rational belief that my belief p is irrational. Clearly, this rational belief is not arrived at through my rationally believing that p doesn't satisfy C in conjunction with relevant background belief. But it still has defeating power: when I form the rational belief that my belief p is irrational, my belief p is no longer rational. Therefore, C is not sufficient for rationality. So the possibility that my belief satisfies C and yet I rationally believe that it *doesn't* is not required to threaten C's sufficiency. Realizing this possibility is only one way to realize the possibility that my belief p satisfies C and yet I rationally believe that my belief p is *irrational*. When this latter possibility is realized through other ways, C's sufficiency would also be endangered.

In conclusion, it is POSSIBLE-2, not POSSIBLE-1, that is the real source of the puzzle. It means that revising C into C+ by adding the condition 'it is irrational for one to believe that C is not satisfied' is heading towards a wrong direction. That revision attempts to avoid the possibility mentioned in POSSIBLE-1, whereas what we should try to avoid is the possibility mentioned in POSSIBLE-2. This means that the extra condition we need is 'it is irrational for one to believe that one's belief p is irrational'. We should revise C into

C++(p) = C(p) and it is irrational for one to believe that one's belief p is irrational.

It seems that only this revision could solve the puzzle.[9]

[9] Foley mentions another solution, which is suggested by Fumerton himself. It denies a single sense of rationality and introduces indefinitely many senses:

S's belief that p is Rational-1 iff C is met for p.
S's belief that p is Rational-2 iff it is Rational-1 and it is irrational for her to believe that belief p is not Rational-1.
. . .
S's belief that p is Rational-n iff it is Rational-(n−1) and it is irrational for her to believe that her belief p is not Rational-(n−1).

The idea is that when you rationally believe 'my belief p does not satisfy C', your belief p is irrational in some higher order sense of rationality (Rational-2), but it could remain rational in the basic sense (Rational-1).

Foley calls the solution 'ingenious', although he doesn't think it is the best solution. However, it faces some serious problems. First, it seems self-contradictory. On the one hand, it denies a single, unifying sense of

However, this revision would make the circularity problem worse. C+ has only one occurrence of 'irrational' in it, while C++ has two.

Later on, I will solve the circularity problem in two steps. First, I will replace the first occurrence of 'irrational' in C++ with other terms. The replacement will preserve the intuition behind C++ and still give us a sufficient condition of rationality. Second, I will argue that although the second occurrence of 'irrational' cannot be eliminated because as I have argued, it is the rational belief that one's belief is *irrational* that is the source of the defeating power, this occurrence would not render the account of rationality *viciously* circular.

4. Eliminating Vicious Circularity

Now, how to eliminate the first occurrence of 'irrational' in C++? Recall that we have supposed all along that C is a necessary condition of rationality.[10] If C is necessary for any proposition to be believed rationally, then a good way to ensure *it is irrational to believe (ir) it is irrational to believe p* is to make sure *ir doesn't satisfy C*. Therefore I suggest we revise C++ into C*:

> C*(p) = C(p) and *ir* doesn't satisfy C
> or simply: C*(p) = C(p) + not-C(*ir*)[11]

And the corresponding proposal of rationality becomes:

> NEW PROPOSAL
> One's belief that p is rational if and only if 1) p satisfies C and 2) the proposition *ir* doesn't satisfy C.

So, if C is 'being reliably formed', NEW PROPOSAL says

> One's belief that p is rational if and only if the belief p is reliably formed, and there is no reliable process in one such that if the process were operated, it would produce the belief 'it is irrational for me to believe p'.

---

rationality. On the other hand, it uses the term 'irrational' in the introduction of all higher-order senses of rationality, which suggests that there is a unifying sense of rationality after all.

Second, just like C+, it doesn't get the real source of the puzzle. Notice that all the introduced senses of rationality are parasitic on Rational-1: Rational-2 is parasitic on Rational-1; Rational-3 is parasitic on Rational-2 and thus parasitic on Rational-1 . . . So the solution still assumes that all rational beliefs of the form 'my belief p is irrational' have to be arrived at through the belief 'my belief p doesn't satisfy C' or the belief 'it is rational for me to believe my belief p doesn't satisfy C' . . . This assumption is false because, as I have argued, my rational belief 'my belief p is irrational' could be arrived at through beliefs that have nothing to do with C.

[10] This supposition is safe because, as I have explained in footnote 5, Foley's argument for the puzzle about necessity is problematic, which means that we can safely assume that C's necessity for rationality is not endangered, and what we are concerned with is just its sufficiency.

[11] Notice that C* is very different from Foley's extra condition in C+. The extra condition in C+ is *It is irrational to believe 'p does not satisfy C'*, whereas the extra condition in C* is *the proposition 'It is irrational to believe p' does not satisfy C*.

And if C is 'fitting one's evidence', NEW PROPOSAL says

> One's belief that p is rational if and only if believing p fits one's evidence, and believing 'it is irrational for me to believe p' doesn't fit one's evidence.

This revision solves the puzzle, because it doesn't have the feature mentioned in POSSIBLE-2, which is the real source of the puzzle as I have argued. That is, it is impossible that C* is satisfied for p and yet it is rational to believe *ir*:

> For C* to be satisfied for my belief p, it must be the case that
> (1) My belief p satisfies C and *ir* doesn't satisfy C.
> But for me to rationally believe *ir*, it must be the case that
> (2) *ir* satisfies C.
> This is because C is a necessary condition of rationality.

Since (2) contradicts the second conjunct of (1), it is impossible for both (1) and (2) to be true. Therefore, it is impossible that C* is satisfied for my belief p and yet it is rational for me to believe *ir*. Then revising C into C* would solve the puzzle.[12] Now there is never going to be a situation in which my belief p satisfies C* and yet it is rational for me to believe 'my belief p is irrational'. So NEGATIVE would not imply that C* is not sufficient for rationality.

For example, suppose C is 'fitting one's evidence'. This condition is not sufficient for rationality because, as is shown in the math proof example in section 1, it is possible that one's belief in the theorem fits one's evidence, and yet it is rational for one to believe that the belief is irrational because it doesn't fit one's evidence. Now, this possibility is ruled out if we add the extra condition that believing 'it is irrational to believe p' doesn't fit one's evidence. In the math proof example, this extra condition is not satisfied, because believing 'it is irrational to believe the theorem' does fit one's evidence. Therefore, this is not a case in which the evidentialist version of C* is satisfied and yet it is rational to believe that one's belief p is irrational. That is, this case doesn't show that the evidentialist version of C* is not sufficient for rationality.[13]

[12] Notice that C* still has the feature mentioned in POSSIBLE-1: It is possible that one's belief p satisfies C* but one rationally believes that it doesn't. This doesn't endanger the sufficiency of C* because, as I have argued, that a condition has the feature mentioned in POSSIBLE-1 is not enough (and also not necessary) to endanger its sufficiency for rationality.

[13] Thanks to a referee for pointing out to me that this is a case in which *no* doxastic attitude is rational: believing the theorem is not rational because it doesn't satisfy the extra condition, and disbelieving or suspending judgment is not rational because it doesn't satisfy the original condition 'fitting one's evidence'. So we have an epistemic dilemma. I am not worried about this result because the dilemma arises not because of the move from C to C*: it seems that any theorist who takes seriously a level-connecting principle like NEGATIVE would face an epistemic dilemma (see Christensen [2007] and Lasonen-Aarnio [2014]). Therefore, it's no wonder that the move from C to C*, which is motivated by solving a puzzle given rise by NEGATIVE, could not avoid epistemic dilemmas.

Besides, it's my suspicion that a variety of theories of rationality cannot rule out epistemic dilemmas, regardless of whether they endorse level-connecting principles—there is no guarantee that for every person and every proposition, there must be a doxastic attitude that satisfies the condition specified by the theories. This means that we could always come up with some weird examples in which no attitude satisfies the condition.

5. A Complicated Version of C*

Before I address the remaining circularity problem with C*, there is a worry I need to deal with: The original condition C is supposed to be sufficient and necessary for rationality. Its sufficiency is put into question due to the puzzle, but its necessity is still intact. Now, I have revised C into C* so that we have a sufficient condition of rationality, but the problem is that C* seems too strong to be necessary. My solution to the puzzle would not be successful if in coming up with a sufficient condition of rationality, I make the condition too strong to be necessary. Here is how the worry goes.

First, recall how we get to C*. I have argued that the source of the puzzle is the possibility that my belief p satisfies C and yet I rationally believe *ir*. To eliminate this possibility, we revise C into C++ by adding the condition 'it is irrational to believe *ir*'. But C++ would give us a viciously circular account of rationality because of the first occurrence of 'irrational' in it. To eliminate the vicious circularity, we revise C++ into C* by replacing the condition 'it is irrational to believe *ir*' with the condition '*ir* doesn't satisfy C'. That is how we get to C*. That is, the sufficient condition we really want is C++, and we only revise it into C* to eliminate the first occurrence of 'irrational'.

However, to eliminate 'irrational' by replacing 'it is irrational to believe *ir*' with '*ir* doesn't satisfy C' is problematic, because the two conditions are not equivalent. Although to ensure '*ir* doesn't satisfy C' is a way to ensure 'it is irrational to believe *ir*' given that C is a necessary condition of rationality, it is not the only way. The reason is as follows:

> Just like it is possible that
>   C(p) but it is rational to believe *ir*,
> it is possible that
>   C(*ir*) but it is rational to believe (ir^2) it is irrational to believe *ir*.
> But since NEGATIVE says that for *any* proposition q, if it is rational to believe that it is irrational to believe q, it is irrational to believe q, the above possibility would give us the following possibility when we let q = *ir*:
>   C(*ir*) but ir^2.

That is, it is possible that *ir* satisfies C and yet it is irrational to believe *ir*. This means that to ensure '*ir* doesn't satisfy C' is only one among many ways to ensure 'it is irrational to believe *ir*'. Therefore, although C*(p) is a way to get C++(p), it is not the only way. Revising C++ into C* has yielded a condition stronger than the one we want.

This worry is reasonable. However, the revision is still along the right track. We just need to weaken C* a little. For convenience, I use the following abbreviations:

> For any proposition p, R(p) = It is rational for me to believe p;
> As before, *ir* = It is irrational for me to believe p, and ir^2 = It is irrational for me to believe *ir*;
> ir^3 = It is irrational for me to believe ir^2;
> . . .
> ir^n = It is irrational for me to believe ir^{n-1}.

Using the above abbreviations, we get

> C++(p) = C(p) and ir^2
> C*(p) = C(p) and not-C(*ir*)

The worry, as I have explained, is that C* is too strong for C++ because 'not-C(*ir*)' is only one way to ensure 'ir^2'. It is possible that

> C(*ir*) but R(ir^2)

'R(ir^2)' would also give us 'ir^2' according to NEGATIVE, even in the presence of C(*ir*).

This suggests that the extra condition we use to replace 'ir^2' in C++ should not be 'not-C(*ir*)'. Rather, it should be

> 1) not-C(*ir*), or C(*ir*) but R(ir^2)

However, 1) cannot be the appropriate sufficient condition of rationality, because the occurrence of the term 'rational' in it would still make the relevant proposal viciously circular. To eliminate the vicious circularity, we need to use the same trick as the one used in section 4. We replace 1) with

> 1)* not-C(ir), or C(*ir*) but [C(ir^2) and not-C(ir^3)]

However, the worry comes back again: 'C(ir^2) and not-C(ir^3)' is only one way to get 'R(ir^2)'. We would also get 'R(ir^2)' if we have 'C(ir^2) and [C(ir^3) but R(ir^4)]'.[14] This means that 1)* should be revised into

> 2) not-C(*ir*), or C(*ir*) but [C(ir^2) and not-C(ir^3)], or C(*ir*) and C(ir^2) and [C(ir^3) but R(ir^4)]

But again, 2) cannot be the sufficient condition we are looking for, because the occurrence of the term 'rational' in it would make the relevant proposal viciously circular. To eliminate the vicious circularity, we have to use the same old trick and replace 2) with

> 2)* not-C(*ir*), or C(*ir*) but [C(ir^2) and not-C(ir^3)], or C(*ir*) and C(ir^2) and C(ir^3) but [C(ir^4) and not-C(ir^5)]

[14] This is because C(ir^2) and [C(ir^3) but R(ir^4)] = C(ir^3) and [C(ir^2) and R(ir^4)]
= C(ir^3) and [C(ir^2) and R(It is irrational to believe ir^3)]
⇒ C(ir^3) and [C(ir^2) and it is irrational to believe ir^3]
= C(ir^3) and [C(ir^2) and it is irrational to believe it is irrational to believe ir^2]
⇒ C(ir^3) and C++(ir^2) ⇒ C++(ir^2) ⇒ R(ir^2)

But then the worry would arise again: 'C(ir^4) and not-C(ir^5)' is only one way to get 'R(ir^4)'. Then we have to go one step further . . .

The pattern is clear. It shows that the condition we use to replace the extra condition 'ir^2' in C++ should be

n)* not-C(ir), or [C(ir) and C(ir^2) and not-C(ir^3)], or [C(ir) and C(ir^2) and C(ir^3) and C(ir^4) and not-C(ir^5)] or, . . . , or [C(ir) and C(ir^2) and . . . and C(ir^{2n}) and not-C(ir^{2n+1})]

Therefore, C++ should be revised into

Grand-C* = C(p) and n)*

It can be proved that Grand-C* is equivalent to C++, the real sufficient and necessary condition of rationality which is dismissed only because of vicious circularity. (Please see Appendix for the proof.)

Now, you might think that Grand-C* is too complicated to be a useful candidate of rationality. To explain why a belief p is irrational even when it satisfies C, we have to show that each disjunct in n)* is not satisfied, and this is too difficult, if not impossible. This problem, however, needs not bother us. The actual human cognitive capacity and our actual cognitive situation are limited such that our beliefs normally don't satisfy even the second disjunct in n)*. For example, take C to be 'being supported by sufficient evidence'. How many of us find ourselves in a situation (and how often) in which we not only have sufficient evidence for p, but also sufficient evidence for 'it is irrational to believe p' and sufficient evidence for 'it is irrational to believe it is irrational to believe p'? I would say very few, very rare. So, unless the situation is extremely unusual, we can safely assume that the second and further disjuncts are not satisfied. This means that we can just use C* as a handy substitute for Grand-C*. To check whether someone's belief p is rational, we normally just need to check whether C is satisfied for p and whether it is satisfied for 'it is irrational to believe p'. We don't need to go on to check whether it is satisfied for 'it is irrational to believe it is irrational to believe . . . it is irrational to believe p', because it is normally not satisfied for such a complicated proposition.

Therefore, I suggest we use C* as a candidate for rationality while keeping in mind that a complete version is Grand-C*. This strategy should be good enough to assuage the worry that Grand-C* is too complicated to be a useful account of rationality.

This completes my attempt to eliminate the first occurrence of 'irrationality' in C++. It leaves us with the second occurrence. Does C* make NEW PROPOSAL, the corresponding account of rationality, viciously circular given that there is still an occurrence of 'irrational' in it? I now argue that the answer is negative.

6. The Remaining Circularity is not Vicious

A conceptual account is circular when the condition used to account for the concept involves the concept itself. Whether circularity in a conceptual account is vicious depends on what we expect from the account. For example, circularity would be vicious if we expect a conceptual account to explain the meaning of the concept to those who do not have it, because the account would not enable them to understand the meaning of the concept.

However, if we expect a conceptual account to give, not meaning-explanation, but grounding properties of the property referred to by the concept in question, then circularity is sometimes vicious and sometimes not. Consider

> SAFETY
> It is rational for S to believe p iff it is rational for her to believe 'p is safe'.

The circularity here seems vicious.[15]

Here is Another example. Suppose our best theory of evidence has it that whether one has sufficient evidence for p is grounded in whether the evidence makes believing p rational. Then this theory of rationality

> RATIONAL-EVIDENCE
> S's belief p is rational iff S has sufficient evidence for it.

is viciously circular. However, consider this theory of rationality:

> DISPOSITION
> S's belief p is rational if and only if S is disposed to judge it to be rational.

The circularity in DISPOSITION doesn't seem vicious.[16]

What distinguishes vicious circularity from non-vicious ones if we expect an account to give the grounding properties of the property in question? Although a complete answer is beyond the scope of this paper, I think the above examples indicate two marks of vicious circularity.[17]

---

[15] Here I disagree with a common suggestion made by McGinn [1983] among others, which claims that an account is not viciously circular as long as it is 'not trivial'. One problem among others is that trivialness is not necessary for vicious circularity. For example, SAFETY is not trivial—the claim that the rationality of belief p is determined by the rationality of the belief 'p is safe' is quite substantive, and yet the circularity in it still seems vicious.

[16] For example, Humberstone [1997] agrees with Johnston [1989] and McGinn [1983] that the circularity in DISPOSITION is not vicious.

[17] Perhaps the most comprehensive treatment of vicious circularity is given by Humberstone [1997]. Roughly, his proposal is that circularity in a conceptual account is not vicious just in case it occurs in a 'compositionally independent' context. A predicate ϕ is compositionally independent when there exists some proposition p such that the truth-value of $\phi[p]$ is independent of the truth-value of p. For example, an account of the concept 'person' might say

> For any object x, x is a person if and only if x believes that he is a person.

According to Humberstone, this account is not viciously circular because the predicate 'believing' is compositionally independent given that belief is non-factive. It would become viciously circular if we replace 'believes' with 'knows', given that knowledge is factive.

This proposal is certainly interesting, and it would give me what I want—it implies that the circularity in NEW PROPOSAL is non-vicious, because the circularity occurs in a context given by the predicate 'not-C', The first is that because of circularity, the account in question violates some constraints on the grounding relation. For example, since RATIONAL-EVIDENCE grounds rationality in evidence, whereas our best theory of evidence grounds evidence in rationality, RATIONAL-EVIDENCE violates the asymmetry constraint on grounding, which says that if a grounds b, then b doesn't ground a.

The second mark of vicious circularity is that, due to the circularity, the account in question fails to give the grounding property. SAFETY is viciously circular for this reason. We expect SAFETY to give us the grounding properties of rationality, while what it does is to ground the rationality of one belief on the rationality of another.

In conclusion, circularity is vicious in an account if the account is expected to give meaning-explanation, and it needs not be vicious if the account is expected to give grounding properties.

Now, the question is: what do we expect from a conceptual account of rationality? It is safe to say that it is not offering an explanation of the meaning of 'rationality' to those who have no idea what 'rationality' means. We offer conceptual accounts to *philosophers*, who have some pre-theoretical understanding of rationality. This pre-theoretical understanding is what we philosophers rely on when we offer or evaluate a theory of rationality.

I believe that the conceptual accounts given by philosophers in general aim to give grounding properties. An analysis of knowledge aims to explain what it is in virtue of which a proposition is knowledge. An analysis of rationality aims to explain what it is in virtue of which a belief is rational. If so, NEW PROPOSAL is not viciously circular because it doesn't have the two marks of vicious circularity I have discussed. Consider the reliablist version of NEW PROPOSAL:

> S's belief that p is rational if and only if the belief p is reliably formed, and there is no reliable process in S such that if that process is operated, it would produce a belief of the form 'it is irrational for me to believe p'.

The circularity here doesn't violate the asymmetrical constraint on grounding, and I cannot see any other constraint that it violates. It also doesn't make NEW PROPOSAL fail to give grounding properties of rationality. For NEW PROPOSAL says that the rationality of a belief is grounded in facts of reliability about the belief. Therefore, the circularity in NEW PROPOSAL is not vicious. It is much more similar to the circularity in DISPOSITION than it is to the circularity in SAFETY and RATIONAL-EVIDENCE.

7. Conclusion

Fumerton's puzzle for theories of rationality is the puzzle that all attempts in finding a sufficient condition of rationality are doomed given a level-connecting principle like NEGATIVE and given that one can always rationally but falsely believe whether one's belief satisfies the condition in question. I have explained the real source of the puzzle and have

---

which is compositionally independent for all promising candidates C. However, I think Humberstone's proposal is problematic and therefore I don't want to rely on it.

offered a solution to it by adding a condition to prominent theories of rationality. I have argued that this revision solves the puzzle without making the corresponding account of rationality viciously circular. If my solution works, all theorists of rationality (if they accept NEGATIVE) need to add this condition to their favourite rationality-making property.

Appendix: Grand-C* is equivalent to C++.

Recall that for any proposition p,

$$\text{Grand-C*(p)} = \text{C(p) and n)*}$$
$$= [\text{C(p) and not-C}(ir)] \text{ or } [\text{C(p) and C}(ir) \text{ and C}(ir^2) \text{ and not-C}(ir^3)]$$
$$\text{or . . . or } [(\text{C(p) and C}(ir) \text{ and . . . and C}(ir^{2n}) \text{ and not-C}(ir^{2n+1})]$$
$$\text{C++(p)} = \text{C(p) and } ir^2$$
$$\text{R(p)} = \text{it is rational to believe p}$$

Besides, we can stipulate

$$ir^0 = p$$

I first prove (I) Grand-C*(p) entails C++(p).

Let i be a natural number and Di be the i-th disjunct in Grand-C*(p). We can show that Di entails C++(p) for each i. Recall that I have shown in section 4 that

(*) For any proposition q,
C(q) and not-C(it's irrational to believe q) ⇒ C(q) and it is irrational to believe it is irrational to believe q ⇒ C++(q) ⇒ R(q).

The last step obtains because, as I have argued, C++ is the real sufficient and necessary condition of rationality, which is revised into C* only because of vicious circularity.

Then notice

$$\text{Di} = \text{C(p) and C}(ir) \text{ and . . .and C}(ir^{2i-2}) \text{ and C}(ir^{2i-1}) \text{ and } [\text{C}(ir^{2i}) \text{ and not-C}(ir^{2i+1})]$$
$$\Rightarrow \text{C(p) and C}(ir) \text{ and. . . and C}(ir^{2i-2}) \text{ and C}(ir^{2i-1}) \text{ and R}(ir^{2i}) \qquad ((*))$$
$$\Rightarrow \text{C(p) and C}(ir) \text{ and. . . and C}(ir^{2i-2}) \text{ and C}(ir^{2i-1}) \text{ and } ir^{2i} \qquad \text{(NEGATIVE)}$$
$$= \text{C(p) and C}(ir) \text{ and. . . and C}(ir^{2i-1}) \text{ and } [\text{C}(ir^{2i-2}) \text{ and } ir^{2i}]$$
$$\Rightarrow \text{C(p) and C}(ir) \text{ and. . . and C}(ir^{2i-1}) \text{ and C++}(ir^{2i-2})$$
$$\Rightarrow \text{C(p) and C}(ir) \text{ and. . . and C}(ir^{2i-1}) \text{ and R}(ir^{2i-2})$$
$$\Rightarrow \text{C(p) and C}(ir) \text{ and. . . and C}(ir^{2i-1}) \text{ and } ir^{2i-2} \qquad \text{(NEGATIVE)}$$

Repeating the last five steps in the above reasoning, we have

$$Di \Rightarrow C(p) \text{ and } C(ir) \text{ and}\ldots \text{ and } C(ir^{2i-1}) \text{ and } ir^{2i-2}$$
$$\Rightarrow C(p) \text{ and } C(ir) \text{ and}\ldots \text{ and } ir^{2i-4} \Rightarrow \ldots \Rightarrow C(p) \text{ and } C(ir) \text{ and}\ldots \text{ and } ir^{2i-(2i-2)}$$
$$\Rightarrow C(p) \text{ and } ir^{2}$$
$$\Rightarrow C{+}{+}(p)$$

This shows that each disjunct in Grand-C*(p) entails C++(p). It follows that Grand-C*(p) entails C++(p) and the proof for (I) is complete.

Next, I show that (II) The falsity of Grand-C*(p) entails the falsity of C++(p). To prove this, I need the following assumption, which is very plausible:

LIMIT
For any plausible candidate of rationality-making feature C, this is true:
For any person S and any proposition p, there exists a natural number m such that not-C(ir^{m}) is true for S.

For example, suppose C is 'S's belief p is reliably formed'. Then LIMIT says that there exists a natural number m such that S has no reliable process to form the belief 'it's irrational to believe it's irrational to believe . . . it's irrational to believe p' with m times of occurrences of 'irrational'. This proposition must have become so complicated when m is so big that it is beyond S's cognitive capacity to even understand what this means.

We can now prove (II) with this assumption. Suppose Grand-C*(p) is false. We now prove that C++(p) is false.

Given LIMIT, we can suppose that for the given proposition p, u is the *minimal* natural number such that not-C(ir^{u}). That is, for any natural number u* < u, we have C(ir^{u*}). u must be even. This is because if u were odd, some disjunct in the following disjunction would be true:

[C(p) and not-C(ir)] or [C(p) and C(ir) and C(ir^{2}) and not-C(ir^{3})] or . . . or
[(C(p) and C(ir) and C(ir^{2}) and C(ir^{3}) and . . . and C(ir^{2n}) and not-C(ir^{2n+1})]

For example, if u = 1, the first disjunct is true; if u = 3, the second disjunct is true. So if u were odd, some disjunct must be true, which means that the whole disjunction would be true. But this disjunction is exactly Grand-C*(p). Its truth contradicts our supposition that Grand-C* is false. Therefore, u must be even. This means that the following disjunction J must be true:

J = [not-C(p)] or [C(p) and C(ir) and not-C(ir^{2})] or [C(p) and C(ir) and C(ir^{2}) and C(ir^{3}) and not-C(ir^{4})] or . . . or [(C(p) and C(ir) and . . . and C(ir^{2n-1}) and not-C(ir^{2n})]

For example, if u = 0, the first disjunct is true; if u = 2, the second disjunct is true.

Let i be a natural number and Ji be the i-th disjunct of J. We can show that Ji entails the falsity of C++(p) for each i. Notice

$$
\begin{aligned}
\text{Ji} &= \text{C(p) and C}(ir) \text{ and} \ldots \text{and C}(ir^{2i-3}) \text{ and C}(ir^{2i-2}) \text{ and [C}(ir^{2i-1}) \text{ and not-C}(ir^{2i})] \\
&\Rightarrow \text{C(p) and C}(ir) \text{ and} \ldots \text{and C}(ir^{2i-3}) \text{ and C}(ir^{2i-2}) \text{ and R}(ir^{2i-1}) \qquad ((*)) \\
&\Rightarrow \text{C(p) and C}(ir) \text{ and} \ldots \text{and C}(ir^{2i-3}) \text{ and C}(ir^{2i-2}) \text{ and } ir^{2i-1} \qquad \text{(NEGATIVE)} \\
&= \text{C(p) and C}(ir) \text{ and} \ldots \text{and C}(ir^{2i-2}) \text{ and [C}(ir^{2i-3}) \text{ and } ir^{2i-1}] \\
&\Rightarrow \text{C(p) and C}(ir) \text{ and} \ldots \text{and C}(ir^{2i-2}) \text{ and C++}(ir^{2i-3}) \\
&\Rightarrow \text{C(p) and C}(ir) \text{ and} \ldots \text{and C}(ir^{2i-2}) \text{ and R}(ir^{2i-3})
\end{aligned}
$$

Repeating the last five steps in the above reasoning, we have

$$
\begin{aligned}
\text{Ji} &\Rightarrow \text{C(p) and C}(ir) \text{ and} \ldots \text{and R}(ir^{2i-3}) \Rightarrow \text{C(p) and C}(ir) \text{ and} \ldots \text{and R}(ir^{2i-5}) \\
&\Rightarrow \ldots \Rightarrow \text{C(p) and C}(ir) \text{ and} \ldots \text{and R}(ir^{2i-(2i-1)}) \\
&= \text{C(p) and C}(ir) \text{ and} \ldots \text{and R}(ir) \\
&\Rightarrow \text{R}(ir)
\end{aligned}
$$

Since R(*ir*) contradicts the second conjunct ir^2 in C++(p), Ji entails the falsity of C++(p). So, J entails the falsity of C++(p). Since J must be true given the supposition that Grand-C*(p) is false, we conclude that the falsity of Grand-C*(p) entails the falsity of C++(p). This completes the proof for (II). (I) and (II) together show that Grand-C* is equivalent to C++.[18]

Cornell University

REFERENCES

Bergmann, M. 1997. Internalism, Externalism and the No-Defeater Condition, *Synthese* 110/3: 399–417.

Christensen, D. 2007. Does Murphy's Law Apply in Epistemology? Self-doubt and Rational Ideals, *Oxford Studies in Epistemology* 2: 3–31.

Christensen, D. 2010. Higher-Order Evidence, *Philosophy and Phenomenological* Research 81/1: 185–215.

Feldman, R. 2005. Respecting the Evidence, *Philosophical Perspectives* 19/1: 95–119.

Foley, R. 1990. Fumerton's Puzzle, *Journal of Philosophical Research* 15: 109–113.

Fumerton, R. A. 1990. *Reason and Morality: A Defense of the Egocentric Perspective*, Ithaca: Cornell University Press: 117–28.

Humberstone, I. L. 1997. Two Types of Circularity, *Philosophy and Phenomenological Research* 57/2: 249–280.

Lasonen-Aarnio, M. 2014. Higher-Order Evidence and the Limits of Defeat, *Philosophy* and Phenomenological *Research* 88/2: 314–45.

McGinn, C. 1983. *The Subjective View: Secondary Qualities and Indexical Thoughts*, Oxford: Clarendon Press.

Smithies, D. 2012. Moore's Paradox and the Accessibility of Justification, *Philosophy and Phenomenological Research* 85/2: 273–300.

Smith, M., Lewis, D. and Johnston, M. 1989. Dispositional Theories of Value, *Proceedings of the Aristotelian Society* 63: 89–174.

Vogel, J. 2006. Externalism Resisted, *Philosophical Studies* 131/3: 729–42.

Williamson, T. 2011. Improbable Knowing, in *Evidentialism and its Discontents*, ed. T. Dougherty, New York: Oxford University Press: 147–64.

[18] This article originates from a presentation I gave to Sage School of Philosophy at Cornell University in 2013. Many thanks to Nicolas Silins, to anonymous referees and to the audiences at the presentation.