It’s OK to Make Mistakes: Against the Fixed Point Thesis¹

Forthcoming in Episteme.

Can we make mistakes about what rationality requires? A natural answer is that we can, since it is a platitude that rational belief does not require truth; it is possible for a belief to be rational and mistaken, and this holds for any subject matter at all. However, the platitude causes trouble when applied to rationality itself. The possibility of rational mistakes about what rationality requires generates a puzzle. When combined with two further plausible claims – the enkratic principle, and the claim that rational requirements apply universally – we get the result that rationality generates inconsistent requirements. One popular and attractive solution to the puzzle denies that it is possible to make rational mistakes about what rationality requires. I show why (contra Titelbaum (2015b), and Littlejohn (2015) this solution is doomed to fail. Consequently, we are left with the surprising result that solving the puzzle will require pursuing one of three highly unintuitive solutions that have so far not proved popular – we must accept that rationality sometimes generates dilemmas, reject the enkratic principle, or defend a conception of rationality for which the requirements of rationality do not apply universally.

Section 1 outlines the puzzle. Section 2 motivates what can seem initially like the most attractive solution to the puzzle: denying that it is possible to make rational mistakes about what rationality requires. I then outline two ways one might argue for the rational impermissibility of mistakes about what rationality requires. The first is via the claim that justification requires knowledge, and the second is via what Titelbaum calls “the fixed point thesis”. I suggest some reasons to be suspicious of the first route (section 2) before moving on to the main focus of the paper, arguing against the fixed point thesis (section 3). I conclude that we should reject the fixed point thesis, and consequently give up on solving the puzzle by

¹I thank the following people for helping me to think through this puzzle on numerous occasions over the past few years, in some cases being kind enough to read drafts and point out (sometimes less than rational) mistakes. In particular I am grateful to: Maria Alvarez, Marvin Backes, Sebastian Becker, Corine Besson, Jessica Brown, Charles Côté-Bouchard, Mike Coxhead, Philip Ebert, Miguel Egler, Giada Fratantonio, Patrick Greenough, Josh Habgood-Coote, John Heron, Ole Hjortland, Nick Hughes, Clayton Littlejohn, Jessica Leech, Matt McGrath, Rosanna O’Keefe, Quentin Pharr, Graham Priest, Lewis Ross, Justin Snedegar, Sophie Stammers, Kurt Sylvan, Fenner Tanswell, Ravi Thakral, Mark Textor, Brian Weatherson, Caspar Wilson, and Jake Wojtowicz. I thank the AHRC for supporting this research.
denying the possibility of making rational mistakes about what rationality requires.

1 The Puzzle

The puzzle can be stated as an inconsistent triad. The following three claims all have a good deal of initial plausibility, but if held together they generate inconsistent requirements of rationality:

(1) Requirements of rationality apply universally, to all agents regardless of their situation.
(2) The Enkratic Principle is one of the requirements of rationality.
    
    *Enkratic Principle:* Do not believe that you are rationally required to believe P without also believing $P^2$.
(3) It is rationally permissible to make mistakes about any subject matter whatsoever, including what rationality requires.

Suppose we accept (1). By (1), the requirements of rationality apply universally and in all cases. Then suppose that you have a mistaken belief about what rationality requires, and that this mistake is rationally supported, as (3) permits. If we also accept (2), the enkratic principle, as one of the universally applicable requirements of rationality then you are also rationally required to have consistent first and higher order beliefs. This means that you are rationally required to either adopt the first order beliefs recommended by your mistaken higher order beliefs about what rationality requires, or give up the mistaken higher order beliefs and obey the actual requirements of rationality at the first order. Given that the mistaken higher order beliefs are rationally supported, and the requirements of rationality that they by stipulation conform to apply in all cases, there seems to be little reason to give them up. So, if we accept all of (1) - (3) then in cases where agents make a rational mistake about rationality, they are required both to adopt a first order belief that contravenes the genuine requirements (by (2)), and not to adopt beliefs that contravene the genuine requirements (by (1)). In other words, accepting all of (1) - (3) means that rationality generates inconsistent requirements.

To respond to the puzzle, we can either simply accept the result that rationality does generate inconsistent requirements (see Christensen, 2004), or give up one of (1)-(3). In the following

---

2 Here I state the enkratic principle in its commissive form, but it equally prohibits one from believing that one is rationally required to refrain from believing P while failing to refrain from believing P. I will be using the ‘wide-scope’ reading of the enkratic requirement (Broome, 1999; Kolodny, 2005).
section I first discuss why giving up (3) can initially seem to be the most attractive option of the alternatives before going on to assess the arguments in its favour. I ultimately conclude that giving up (3) is, despite appearances, not a good option at all.

2 No rational mistakes about rationality

At first glance, denying (3) can seem like the most attractive option when compared to the alternatives.

Accepting that rationality generates dilemmas is at least a surprising result, since a natural thought about rationality is that it provides recommendations for how agents should believe. In this case, it's recommendations are inconsistent and so at least one must be ignored. Another natural thought is that rationality is a coherent system, applicable in all situations. This also must be given up if rationality in fact generates inconsistent requirements in some situations.

Giving up on (1) would involve a radical overhaul in our understanding of what rationality involves. According to a traditional understanding of rational requirements, they are universally applicable, and hold equally for all agents. They are traditionally of the form 'if conditions C obtain, you are rationally permitted to believe P'. As Lasonen-Aarnio (2014) and Littlejohn (2015) argue, giving up on this traditional understanding of rational requirements would involve serious costs. For example, it would no longer be possible to list traditional requirements such as 'do not believe contradictions', instead we would need to consider individuals and their belief states separately. It may turn out that this is in fact the correct way to proceed (e.g. see Gibbons, 2013; Kvanvig, 2014), but the radical nature of the overhaul required gives some indication why this option can seem less attractive at the outset.

Giving up on (2) does not seem any better. The enkratic principle has a great deal of intuitive plausibility and finds wide-ranging support: Smithies (2012) claims that denying the enkratic principle constitutes Moorean absurdity, Greco (2014) that epistemic akrasia could only be rational in cases where our minds are fragmented, and Titelbaum (2015b) is so confident in the enkratic principle's plausibility that he asserts it as a premise of his argument. As Horowitz (2014) and Littlejohn (2015) argue, if the enkratic principle is not a requirement of rationality, then some very bizarre and intuitively irrational reasoning patterns would be allowed to count as rational.
This leaves rejecting (3) as the only remaining option. This option allows us to preserve both the enkratic requirement and the commitment to there being universally applicable requirements of rationality. Rejecting (3) as a way to solve the puzzle has received a recent wave of support (Littlejohn 2015, Titelbaum 2015b).

Rejecting (3) will nevertheless require some argument. There are two main ways to do this: by rejecting the idea that it is possible to have rational false beliefs in general, or by defending what Titelbaum (2015b) calls “the fixed point thesis” – the claim that although rational false beliefs are possible, they are not possible in the domain of rationality.

Of these two strategies I will focus on the fixed point thesis, since denying the possibility of rational false belief faces various difficulties. Intuitively we experience many cases of false belief that we are inclined to evaluate in some positive sense - justified false beliefs have traditionally not been thought of as being particularly problematic. Those who deny their possibility usually explain away cases of apparently epistemically good false belief by appeal to excuses (Littlejohn, forthcoming; Sutton, 2005, 2007; Williamson, forthcoming). Those who employ good epistemic methods, but fail to form true beliefs through no fault of their own are not justified, but they are excused for violating the knowledge norm of justification. As others have pointed out³, the problem with this is that it runs together an important normative distinction between good epistemic conduct and excusable false belief that does not involve good epistemic conduct. The same notion – excuse – is applied to cases that intuitively demand different accounts, for example: a) unfortunate souls raised in a cult to ignore the dictates of reason, and b) unfortunate victims of Gettier cases.

One way knowledge first epistemologists might attempt to mark the difference between the victims of cults and the victims of Gettier cases is by using the notion of rationality – perhaps Gettiered agents are rational because they employ usually good methods of reasoning that unfortunately do not in this case result in knowledge, while the victims of cults are irrational in virtue of employing bad reasoning methods, but excused for being so. This option would not be available if rational false beliefs, like justified false beliefs, were prohibited. As such, I take the option of rejecting (3) by prohibiting rational false beliefs to be a hard sell. The rest of this paper will be focussed on considering arguments for the more plausible option for rejecting

³ See (Brown, forthcoming; Cohen & Comesaña, 2013; Kelp, 2016).
(3), via a defence of the fixed point thesis.

3 The Fixed Point Thesis

Littlejohn (2015) and Titelbaum (2015b) solve the puzzle by rejecting (3), and they do so by defending the fixed point thesis. The fixed point thesis says that mistakes about what rationality requires are not rationally permissible. As Titelbaum puts it: ‘mistakes about the requirements of rationality are mistakes of rationality’ (2015b: 253).

The fixed point thesis (henceforth FPT) is surprising. Given that rationality does not normally require correctness, and we might think that it should be possible for agents to make rational mistakes about any topic at all. The FPT denies this, claiming that rationality is different from other topics in this respect – agents are not rationally permitted to make the same kinds of mistakes about it as in other domains. If one takes a view at all about what rationality requires, then it must be true in order to be rationally permissible. I will discuss both Littlejohn’s and Titelbaum’s arguments in favour of the FPT and conclude that they both fail. The upshot of this is that the puzzle is not to be solved by rejecting (3), and we must instead either reject (1) or (2), or accept that rationality sometimes generates inconsistent requirements.

3.1 Argument from Indefeasible Justification

Titelbaum (2015b) argues that the explanation for why we should think that rational mistakes about rationality are impossible is that contrary to (3), all agents in fact have reason to comply with the rational requirements that in all cases overrides any other putative rational support they might have for false beliefs about what rationality requires. This is because, as a matter of fact, 'every agent possesses a priori, propositional justification for true beliefs about the requirements of rationality in her current situation' and this justification is 'ultimately empirically indefeasible' (2015b: 276). Titelbaum goes on to argue that given this, there could 'never be a situation in which empirical considerations outweigh a priori justification for rational requirements' (2015b: 276 fn.48).

If Titelbaum is successful in his appeal to indefeasible apriori justification for the rational requirements, then this kind of defence of objectivism looks promising. In the following sections I will suggest some reasons why I think it fails.
Titelbaum’s defence of objectivism involves an appeal to universal possession of particular kinds of justificatory assets. It can be summarised as follows:

**Assets:** All agents, in all possible situations, possess a priori propositional justification for the rational requirements that is indefeasible.

This claim requires some explanation. It can be read either as a factual claim about which justificatory assets agents in fact have, or a conceptual claim about the requirements of rationality, that they are such that they generate these justificatory assets concerning them. Either way, **Assets** requires some explanation. Titelbaum provides no such explanation when he introduces the claim, and as it stands it is somewhat surprising. More recent advocates of a priori justification have typically thought of it as defeasible at best (see BonJour, 1998).

If **Assets** is to be read as making a factual claim, then it does not seem like something that could be established via a priori reasoning. The claim that, in fact, all agents in all situations possess indefeasible a priori justification seems like something that would require at least minimal consideration of agents' actual situations. **Assets** looks more plausible if it is read as a conceptual claim about the requirements of rationality.

One possible explanation is that the rational requirements are obvious, certain, or impossible to doubt. If the rational requirements were similar or closely linked to basic principles of logic, then this might be thought to have some plausibility. It is possible that this is Titelbaum’s thought, since he spends a good deal of the early sections of the paper dismissing traditional problems associated with logical omniscience - the worry that if rationality requires conformity to logic, then this will generate an overly demanding set of rational requirements, requiring agents to, for example, believe all of the logical consequences of their beliefs. However, it is not clear that requirements of rationality for belief can be mapped neatly on to basic logical

---

4 Ichikawa and Jarvis (2013) and Smithies (2012) make similar points about our having propositional justification for the requirements of rationality.

5 Field (1998) is one exception.

6 A related motivation might be an appeal to something like ‘rational intuition’. I will not consider this in detail, primarily because even if intuition can provide a reliable guide to truths about the rational requirements, it is unclear how subjects could distinguish genuine intuitions from subjective feelings of obviousness, and so an intuition-based motivation seems to face the same problems as one based on subjective obviousness.
principles. The difficulties in specifying the exact relationship between logical principles and requirements or prescriptions for belief are well documented, and not limited to problems of logical omniscience\(^7\). So, we cannot assume that logical principles automatically generate requirements of rationality for belief.

Then again, Titelbaum also emphasises the importance of considering agents' situations in determining their rationality. He thinks that the presence or absence of 'rational flaws' can be determined by evaluating an agent's 'state against her situation' (2015b: 259). This suggests that the kind of requirements he has in mind might be situation-specific. If this is correct, then these requirements will be more subtle and less obvious than basic logical principles. The fact that they are obvious, then, cannot be used as an explanation for why we should think we have indefeasible justification for the requirements of rationality\(^8\).

An argument based on the *Assets* claim would require considerable further defence. I have explored the more obvious avenues above, and they do not seem fruitful. As such, I will move on to the other available argument for the FPT, Littlejohn's (2015) argument from liability.

### 3.2 Argument from Liability

Littlejohn (2015) provides another argument for the FPT, one which appeals to the idea that the rational requirements generate liabilities: just as citizens of a country are liable to pay taxes simply in virtue of being citizens subject to tax laws, agents are ‘liable’ for their beliefs simply in virtue of being agents subject to the requirements of rationality.

Littlejohn sums up his argument from liability as follows: "the fixed-point thesis isn't true because we all happen to have evidence for the right list of rational requirements; rather, it's true because the grounds for saying that someone's attitudes are irrational is that those attitudes reveal a kind of incompetence with respect to handling reasons and their demands. As it happens, mistaken beliefs about what rationality requires will manifest that kind of

---

\(^7\) See (Grice, 2001; Harman, 1986; Hawthorne & Bovens, 1999; Macfarlane, ms.)

\(^8\) Entitlement (see Wright, 2004a, 2004b, 2014) might be thought to be a possible explanation for the assets claim, since this generates justification without requiring any particular attitudes. However, this is also usually thought of as defeasible, so any argument making use of this would still require explanation for the indefeasibility of the justificatory assets involved.
incompetence". Earlier on the same page he also says that "rationality requires an understanding of what's required when reasons apply to you" (2015: 14).

We can reconstruct his argument for the FPT in the following way:

**The Liability Argument:**
1) Rationality is competence in handling the reasons that apply to you.
2) Competence in handling reasons that apply to you requires understanding what is required when reasons apply to you.
3) Having false beliefs about the rational requirements involves failing to understand what is rationally required of you in your particular situation.
4) If you fail to understand what is rationally required of you in your particular situation then you manifest an incompetence with respect to rationality.

Conclusion: mistakes about what rationality requires are not rationally permissible (FPT).

The liability argument takes rationality to be a matter of competence in handling reasons. Agents manifest this ‘competence’ by fulfilling the rational requirements. So, believing rationally is a matter of manifesting competence, and manifesting competence is a matter of fulfilling the rational requirements. The argument says that mistakes about what rationality requires differ from mistakes about other topics because false beliefs about rationality manifest an incompetence with respect to rationality, and this incompetence means that those beliefs cannot be rational.

It is the fourth premise of the argument that causes trouble for the liability argument. There are two ways of reading ‘fail to understand’, but both lead to undesirable consequences that any defender of the FPT should want to avoid. I will outline the two possible ways of reading ‘fail to understand’ in the fourth premise, and show how each leads to problems for defenders of the FPT.

3.2.1 The Strong Reading
One way to read ‘fail to understand’ is to read it as ‘lack true beliefs about’. I will call this the ‘strong reading’.
**Strong Reading:** to fail to understand P is to lack true beliefs about P

Taking the strong reading, the fourth premise of the liability argument says that anyone lacking true beliefs about what rationality requires is manifesting rational incompetence. The problem with this is that it is too demanding – for many candidate rational requirements it seems to be the case that agents can fulfil the requirements without holding any beliefs at all about what the requirements are. It is possible to fulfil the non-contradiction requirement by refraining from believing contradictions, and it is possible to refrain from believing contradictions while also suspending belief about what rationality requires of you. Consider the following agents:

**Innocent Agents:** lack some (and perhaps all) true beliefs about what rationality requires. They suspend on some or all questions of what is required by rationality. Despite this, their beliefs are completely in line with what is in fact required by rationality.

According to the strong reading, innocent agents fail to understand what is rationally required of them, and so count as incompetent with respect to rationality. There are at least two reasons to resist this result. Firstly, as already noted, it is far too demanding. Given the various disagreement over what the rational requirements are – whether, for example, one is rationally required to believe lottery propositions, conciliate in the face of peer disagreement, or believe the logical consequences of one’s beliefs, it seems that all but a few enlightened epistemologists will count as rationally competent.

Secondly, on the strong reading, the fourth premise of the liability argument introduces further requirements on rationality, over and above those covered in (1). In order to count as competent with respect to rationality, agents must not only obey the universally applicable requirements set out in (1), but they must also have second order beliefs about what they are rationally required to believe. This is a somewhat awkward addition, reminiscent of Carroll’s tortoise⁹ - it means that defending the FPT in this way has the result that it is not enough for agents to obey the requirements at the first order, they are also required to have second order beliefs about what is required at the first order. We might wonder how far up this demanding requirement goes. Must agents also believe correctly about their second order beliefs – that is,

---

⁹ See Carroll, 1895).
must they hold the correct beliefs about what rational competence requires (i.e. that it requires correct belief about what rationality requires at the first order)? If so, it seems to generate a needlessly complex picture of the requirements of rationality.

This awkward result might push a defender of the FPT to adopt what I will call the ‘weak reading’ of fail to understand instead. Unfortunately, this is no better.

3.2.2 The Weak Reading
The weak reading takes ‘fail to understand’ to mean ‘hold mistaken beliefs about’.

*Weak Reading*: to fail to understand P is to hold mistaken beliefs about P

This allows innocent agents to count as rational, but accuses those who explicitly believe falsehoods about what rationality requires of incompetence with respect to rationality. This is an improvement on the strong reading since it seems to say the right thing about innocent agents – they count as rational in virtue of believing in line with the requirements at the first order and lacking mistaken beliefs at the second order. However, it seems to say the wrong thing about other kinds of agents. Consider the following agents:

*Misguided Akratic Agents*: These agents hold false beliefs about rationality but for whatever reason contravene these false beliefs and end up believing in accordance with rationality.

Both the weak and the strong readings of ‘fail to understand’ mean that the liability argument takes these agents to be manifesting incompetence with respect to rationality, in virtue of their holding mistaken beliefs about rationality. However, it is not immediately clear why misguided akratic agents should be treated differently to innocent agents. Like innocent agents, misguided akratic agents obey the requirements of rationality at the first order, it is only their higher order beliefs that are sub-par. The weak reading means that innocent and misguided akratic agents are treated differently by the liability argument – whereas innocent agents were forgiven for their lack of true beliefs, misguided akratic agents are deemed incompetent with respect to rationality, despite the fact that both kinds of agent obey the requirements of rationality at the first order. As such, defenders of the liability argument who take the weak reading – the only
plausible reading remaining once we reject the strong reading – require some explanation for this difference in treatment.

One explanation that might be offered is that the mismatch between higher and first order beliefs in cases of akrasia simply exhibit an obvious kind of irrationality, and this should be explanation enough for difference in treatment. However, this somewhat blunt response risks undermining some of the simple and intuitive appeal of views that make use of requirements of rationality; views committed to (1) have the at least surface level advantage of being straightforward to apply – they say simply that to be rational is to comply with the requirements of rationality. However, the weak reading’s assessment of the misguided akratic agent as irrational, despite his complying with all the requirements at the first order, shows that views that preserve (1) by adopting the FPT are not quite as straightforward as they might otherwise appear. Having false beliefs about rationality renders misguided akratic agents rationally incompetent, despite the fact that all their first order beliefs are held rationally. This means that rationality requires more than simply fulfilling the rational requirements mentioned in (1), according to the liability argument it also generates requirements at the higher order. On the weak reading, this is the requirement to avoid false beliefs about what rationality requires. However, assuming it is possible to fulfil the first order requirements while holding false beliefs at the higher order, it is not clear why this extra requirement to avoid false beliefs is necessary. It seems to me an open question whether the putative irrationality of akrasia is sufficient to justify this now rather complicated view of rationality, and it requires some argument.

Further support for the implausibility of the idea that rationally competent agents must avoid beliefs about what is rationally required of them can be drawn from the literature on skill. It is well documented that skilled agents – that is, agents competent at various tasks – do not always have true beliefs, in fact they often have false beliefs, about what they are required to do in order to perform the task successfully. Some might object to the possibility of misguided akratic agents on the grounds that it cannot be possible for all of a misguided akratic agent’s beliefs to fulfil the requirements of

---

10 For more on this see (Brownstein, 2014; Montero, 2016; Reed, McLeod, & Dienes, 2010).
rationality. The thought here is that rationality must be such that if you do everything right, you cannot end up believing falsehoods – falsehoods in general must be due to some epistemic failing, so this must apply equally to falsehoods about rationality. Falsehoods about rationality must be the result of some rational failing on your part. This objection must be mistaken if we accept the non-factivity of rationality. Furthermore, it is an objection that goes further than the FPT and prohibits all rational false belief. It does not explain why false beliefs about rationality are particularly problematic.

The liability argument is caught between a rock and a hard place. On the one hand, taking the strong reading of ‘fail to understand’ and requiring agents to have beliefs, true ones, about what rationality requires of them in order to count as manifesting rational competence seems absurdly strong. It seems that the liability argument needs to make an exception for innocent agents to avoid absurdity. On the other hand, taking the weak reading in order to make this exception invites the question as to why misguided akratic agents should not also be granted an exception – on the grounds that their first order beliefs obey the requirements of rationality.

4 The best of a bad bunch?

Even friends of the FPT admit it is counterintuitive. Littlejohn introduces his solution to the puzzle, which is committed to the FPT, as ‘the best of a bad bunch’ (2015: 11). What he means by this is that if one wants to avoid giving up (1), and one is uncomfortable with the idea that rationality sometimes generates dilemmas, then one faces a choice between giving up the enkratic principle (2) or accepting the fixed point thesis (and so giving up (3)). Given this set of options, and the intuitive appeal of the enkratic principle, accepting the FPT can look like the lesser of two evils. However, in order to make this line of reasoning work, friends of the FPT must do two things. First, they must say why neither giving up (1), nor accepting that rationality sometimes generates dilemmas are options worth pursuing. Second, they must explain why the costs of the FPT that I have outlined in the previous sections are less problematic than giving up the enkratic principle. I am not convinced that appeal to its intuitive plausibility is sufficient to do this job.

11 Titelbaum (2015b) seems to defend something like this. Another way to reject the possibility of misguided akratic agents is by denying the possibility of epistemic akrasia completely, as Adler (2002) and Owens (2002) do.
12 Horowitz (2014) has some important arguments against rejecting the enkratic principle. The further claim that the costs associated with rejecting the enkratic principle that she indicates are worse than the costs associated with accepting the FPT remains to be argued for.
It is worth noting that there are reasons to think that the FPT, a narrow-scope requirement, is a consequence of the wide-scope enkratic principle, as Titelbaum has argued (Titelbaum, 2015a, 2015b). If Titelbaum is right about this, and if we have independent reasons for accepting the enkratic principle, then we must also accept the FPT. Titelbaum’s result, however, can be read in two ways. We can either take it to show that we are committed to the FPT, or we can take it to show that since the enkratic principle commits us to the FPT, we ought to give up the enkratic principle. I will not adjudicate this here, but I do not think we need to read Titelbaum’s result as decisive in favour of the FPT.

Indeed, what I have shown here is that solving the puzzle by rejecting (3) is a route one should be at the very least suspicious of. The best way to motivate a denial of (3) is via the FPT, but the FPT is a bold, surprising thesis and the arguments in its defence have been found wanting. Since neither the argument from indefeasible justification, nor the argument from liability can be made to work, rejecting (3) is not the best way to solve the puzzle, and it is worth exploring some of the other options.
References

Macfarlane, J. (ms). In What Sense (if any) is Logic Normative for Thought?


