Self-Fulfilling Beliefs: A Defense

Paul Silva Jr.
University of Cologne

Self-fulfilling beliefs are, in at least some cases, a kind of belief that is rational to form and hold in the absence of evidence. The rationality of such beliefs have significant implications for a range of debates in epistemology. Most startlingly, it undermines the idea that having strong evidence for the truth of *p* is necessary for it to be rational to believe *p*. The rationality of self-fulfilling beliefs is here defended against the idea that their rationality is incompatible with a compelling closure principle.

Key Words: rationality, justification, self-fulfilling beliefs, closure, evidentialism

1. Rationality and Self-Fulfillment

Background conditions can be such that merely believing *p* ensures that *p* is highly likely to be true on one's evidence. These are the kind of *self-fulfilling beliefs* at issue here. Many epistemologists have thought self-fulfilling beliefs can be rational to form even if one lacks *prior* evidence for their truth, and they have put this fact to work in epistemology (Foley 1991; Velleman 1989; Reisner 2015; Peels 2015; Sharadin 2016; Raleigh 2017; Anthill 2020; Silva and Tal 2021; Dahlback forthcoming). Perhaps most startlingly, self-fulfilling beliefs undermine the simple evidentialist idea that having evidence that strongly supports *p*'s truth is necessary for it to be rational to believe *p*.[1]

Here's an illustrative case:

> **Mindreading Machine**. There is an infallible brain scanner that displays recently formed beliefs. It will scan you shortly. As you wait you remember that, for any number *n*, if you now come to believe *that the scanner will display the number n* then it will display *n* (since the machine displays your numerical belief about *n* by displaying the number *n*). So if you believe that the scanner will display 1 then it will display 1, and if you believe that it will display 2 then it will display 2, and so on. You have no significant evidence to think you will form a belief about one number rather than any other. Knowing all this, *is it rational for you to believe that the scanner will display any particular number?*

Despite your current lack of evidence about what number the scanner will display, it seems rational for you to believe that it will display a 1, and rational to believe that it

[1] Self-fulfilling beliefs are naturally paired off with *self-defeating beliefs*: propositions one's evidence conclusively supports iff one does not believe them. These undermine the converse evidentialist thesis. See Silva (2018, 2023) and Silva and Bernecker (forthcoming) for discussion of the import of self-defeating beliefs when it comes to thinking about reasons, rationality, knowledge, and awareness. See Silva and Tal (2021) for discussion of how knowledge-first forms of evidentialism can survive cases of self-fulfilling and self-defeating belief.

will display a 2, and so on. For upon forming any such belief it will be highly likely on your evidence that your belief is true assuming: that you know that the scanner is infallible and that you're able to know by introspection what numerical belief you've formed.

Marxen points out that such intuitions about the rationality of self-fulfilling beliefs can be explained with:

> *True Belief and Ex Ante Rationality* (TEAR): Necessarily, if *S*'s believing that *p* in conditions *c* ensures that *p* is sufficiently likely on *S*'s evidence to be true upon believing that *p*, then it's rational for *S* to believe that *p*.[2]

TEAR is a principle about *ex ante* rationality (propositional justification) and not *ex post* rationality (doxastic justification). For it concerns when it's rational *to believe* that *p*, whether or not one actually believes that *p* and whether or not one believes that *p* for the 'right reasons' or in the 'right way' (Silva and Oliveira 2022).

2. The Closure Argument Against TEAR

Marxen (2021) objected that principles along the lines of TEAR are inconsistent with:

> *Addition Closure*. Necessarily, if *p* is rational for *S* to believe, then (*p* or *q*) is rational for *S* to believe.

To demonstrate this inconsistency we are given the following case:

> **Logic Class**. Friedrich is attempting again to pass logic. Given Friedrich's poor performance in his philosophy classes, the proposition that his philosophical abilities are not prodigious is rational for him to believe. He's narcissistic, though, and so he irrationally believes (P) *that his philosophical abilities are prodigious*. Applying the Addition inference rule to what he believes, he comes to believe (TvP) *that his philosophical abilities are prodigious or today will be terrible*. While he has come to believe this proposition, he hasn't come to believe that today will be terrible. He knows that the chance (T) *that today will be terrible* is minuscule since—in keeping with his self-absorbed ways—he meticulously records how his days go. While the chance is quite low, there have been some terrible days for him over the past few years. And, almost always, the sole reason why these days are terrible is that he firmly believed that the day would be terrible: whenever he does, this belief makes him depressed, which in turn makes the day terrible [Hence, the belief that T is self-fulfilling]. Thankfully, Friedrich is generally cheerful, and has strong reason to think that he will, as usual, avoid the pessimistic belief that today will be terrible. [So Friedrich is most likely to have a *middling* day: one that is neither terrific (since he's very unskilled at logic) nor terrible (since he's not likely to believe that today will be terrible).] (Marxen 2021: 3)

[2] TEAR differs immaterially from the principle that Marxen criticizes: (RSTB) if *p* is likely [on your evidence] if you believe it, then *p* is rational for you to believe.

Consider the general features Logic Class is supposed to possess:

A. At *t1*, Friedrich's total evidence ensures that neither T nor P are likely to be true. Rather, both T and P are *very* likely false and so is their disjunction (T∨P). (Supposition)
B. At *t1*, Friedrich is in a situation where believing that T would ensure that T is likely to be true upon believing that T, but Friedrich has not (at *t1*) come to believe that T. (Supposition)
C. At *t1*, Friedrich believes that P, but this belief is irrational. (From condition A)
D. At *t1*, through a competent deduction Friedrich formed the further disjunctive belief that (T∨P), but this disjunctive belief is not supported by Friedrich's total evidence at *t1*. (From condition A)
E. At *t1*, it's rational for Friedrich to believe the disjunction (T∨P). (From B, TEAR, and Addition Closure)

E says (T∨P) is rational to believe at *t1*. But, Marxen argues, D indicates the opposite:

> But is (T∨P) rational for Friedrich to believe? It is not. (T∨P) is *unlikely*, even if he believes (T∨P). … As T and P are each *unlikely*, even after Friedrich believes (T∨P), it's *unlikely* if he believes it. And, since this is the case, it's plausible that (T∨P) is *irrational* for Friedrich to believe [contra E]. So, TEAR and Addition Closure are incompatible. (Marxen 2021: 3).

If Marxen is right that it's irrational at *t1* for Friedrich to believe (T∨P), then Addition Closure, TEAR, and A-E are jointly inconsistent. Closure principles as weak as Addition Closure have a high degree of antecedent plausibility (Marxen 2021: 7-9), and Marxen argues the only plausible solution is to give up TEAR.

3. Disambiguations

We have to clarify the content of TEAR. TEAR lacks temporal parameters and can be interpreted either *diachronically* or *synchronically*. It can also be interpreted as applying to belief *states* or belief-forming *processes*. Here is a synchronic interpretation:

> *TEAR: Stative-Synchronic* (TEAR-SS): Necessarily, if *S*'s believing that *p* **at *t*** in conditions *c* ensures that *p* is sufficiently likely on *S*'s evidence to be true upon believing that *p* **at *t***, then it's *ex ante* rational for *S* to believe that *p* **at *t***.

Mindreading Machine supports this reading because upon forming the belief that, say, the machine will display the number 1, one's evidence will support the truth of that claim. So even if one lacked prior evidence for that claim, once the belief is formed one's evidence is sure to support it at that time.

Mindreading Machine also supports a diachronic claim that concerns not belief *states*, but belief-forming *processes*:

> *TEAR: Procedural-Diachronic* (TEAR-PD): Other things being equal, if *S*'s believing that *p* in conditions *c* at *t* ensures that *p* is sufficiently likely on *S*'s evidence to be true upon believing that *p* at *t*, then **at some time prior to** *t* it's rational for *S* **to enact a process of belief revision** that yields a belief that *p* at *t*. (cf. Velleman 1989: 63; Raleigh 2017: 333)

Mindreading Machine supports this because it not only seems rational to believe *p* at *t* in that case, but also rational *to revise* one's attitudes so that one comes to hold that belief at *t*. TEAR-PD's 'other things being equal' qualifier is to set aside discussion of the potential range of cases where revising one's beliefs about whether *p* involves epistemic trade-offs. There are delicate issues surrounding the rationality of belief-revision processes (Podgorski 2017).

For now note that neither TEAR-SS nor TEAR-PD imply the idea that the rationality of holding a belief at some time *t* secures the rationality of holding that same belief *at an earlier time*. To get this result we need a principle like the following:

> *TEAR: Stative-Diachronic* (TEAR-SD): Necessarily, if *S*'s believing that *p* in conditions *c* **at some time *t*** ensures that *p* is sufficiently likely on *S*'s evidence to be true upon believing that *p* **at *t***, then it's *ex ante* rational for *S* to believe that *p* **at some earlier time *t*-**. (cf. Dahlback forthcoming)

As we will see, TEAR-SD is the principle that Marxen implicitly exploits.

4. TEAR-SD Without Defeaters

TEAR-SS, TEAR-PD, and TEAR-SD each say something substantially different about *what* self-fulfilling beliefs are sufficient for. TEAR-SS says that they're sufficient for *ex ante* rational belief *at the time* the evidentially self-fulfilling belief is formed; TEAR-PD says that they're typically sufficient for the rationality of enacting a belief revision *process*; and TEAR-SD says that they're sufficient for *ex ante* rational belief *at some earlier time*. In this way, TEAR-SD has a retrospective aspect that TEAR-SS and TEAR-PD lack.

This retrospective aspect of TEAR-SD makes it vulnerable to problems with defeaters that exist at earlier times. Take Mindreading Machine. Coming to believe at time *t* (I) *that the machine will display the number 12,345* is a self-fulfilling belief for you at *t*. Importantly, this is true whether or not you actually come to believe (I) at *t* or at any other time. Thus, TEAR-SD implies that prior to forming any numerical belief at some earlier time, *t*-, it is *actually* rational for you to believe (I) at *t*-. But suppose you knew at *t*- (II) *that the machine will not display the number 12,345*. (Assume God or a time-traveler gave you testimonial knowledge at *t*- that you would only form some other numerical belief.) Knowledge entails rationality, so (II) is also rational to believe at *t*-. Thus, it follows that (I) and (II) are both rational for you to believe at *t*-. But (I) and (II) are obvious contradictories and we don't want to say that it's *ex ante* rational to believe such contradictions.

Problems with defeaters are remedied with 'no-defeater' clauses and we should add such to TEAR-SD:

> *TEAR: Stative-Diachronic-No-Defeaters* (TEAR-SD-ND): Necessarily, if *S*'s believing that *p* in conditions *c* at some time *t* ensures that *p* is sufficiently likely on *S*'s evidence to be true upon believing that *p* at *t*, then it's *ex ante* rational for *S* to believe that *p* at some earlier time *t-*, **provided at *t- S* lacks sufficient reason to believe that *p* is false.**[3]

Knowing that *p* is false ensures that one has sufficient reason to believe that *p* is false. So, given your knowledge of (II), TEAR-SD-ND does not imply that it's rational to believe (I) at *t-*.

5. Against the Closure Argument

Recall that Marxen sought to highlight the logical incompatibility of TEAR and Addition Closure by drawing our attention to how D and E conflict: D implies the irrationality of believing (T∨P) at *t1*, while E affirms the rationality of believing (T∨P) at *t1*.

But to reach E we need a disambiguated version of TEAR. Will any disambiguation of TEAR yield E? Yes, we can reach E with TEAR-SD. For it implies that Friedrich can rationally believe T–and hence (T∨P)–at the earlier time *t1* because there is a later time, *t2*, at which the self-fulfilling belief in T can be held. So E can be derived as Marxen intended with TEAR-SD.

But we've already seen that TEAR-SD is false for reasons unconnected to Addition Closure. This is corrected for with TEAR-SD-ND. *But TEAR-SD-ND will not allow us to derive E*. For Marxen's case indicates that Friedrich has sufficient reason to believe (T∨P) is false at t1. This is due in part to the high likelihood that (T∨P) is false (see condition A). But, in my view, it is also in part due to the fact that in Marxen's case Friedrich knows that it would be unusual (abnormal) for him to have a terrible day (cf. Smith 2016, Silva forth). So the most plausible stative-diachronic disambiguation of TEAR does not allow us to reach E. Contradiction avoided.

Could TEAR's other disambiguations lead to contradiction? TEAR-PD cannot as it's not about belief states, but belief-forming processes. Addition Closure has nothing to say about belief-forming processes. Contradiction avoided.

Could TEAR-SS allow us to derive E? No. The synchronic character of TEAR-SS prevents it from being used to derive E. For, according to B, the antecedent of TEAR-SS is not satisfied due to the fact that the self-fulfilling belief in T is not formed at *t1*. In order to use TEAR-SS to produce a contradiction we have to look to some future time, *t2*, at which Friedrich actually holds the self-fulfilling belief that T. Then it will follow from TEAR-SS that T is rational to believe at *t2*. From this and Addition Closure we get:

---

[3] The no-defeater condition is limited to rebutting defeaters, and omits undercutting and higher-order defeaters. It's difficult to show that *ex ante* rationality (as opposed to ex post rationality and knowledge) is inconsistent with these kinds of defeaters. Those who think so can add a more robust no-defeater clause to TEAR-SD-ND. This will not be relevant in answering Marxen's challenge.

F: At *t2,* it's rational for Friedrich to believe the disjunction (TvP).

Can we derive a contradiction from F and D? No. D is just a claim about what Friedrich's *total evidence fails to support at t1*, not a claim about what *it's (ir)rational for him to believe at some future time t2*, e.g. the time at which he hosts the self-fulfilling belief that T. So Addition Closure, TEAR-SS, A-D, and F form a jointly consistent set.

To get a contradiction with TEAR-SS we must assume that one's evidence at earlier times constrains what is rational to believe at future times. That is:

> *Primacy of Present Evidence-Diachronic* (PPE-D) Necessarily, it's *ex ante* rational for *S* to believe that *p* at **some future time** *t* only if *S*'s total evidence at **some earlier time** *t-* ensured that *p* is sufficiently likely to be true.

With PPE-D we can produce a contradiction involving A-D, F, TEAR-SS, and Addition Closure. But without PPE-D, or some diachronic principle in its neighborhood, one cannot conclude from D that it's not rational for Freidrich to believe (TvP) at *t2*.

We cannot now fully assess PPE-D. But we can quickly observe that it's inconsistent with a wide range of theories of rationality. For example, it's inconsistent with many externalist theories of rationality. Take reliabilism. Reliabilists have held that there are reliable processes that are not evidence-dependent, i.e. they don't take new evidence as input in the reliable belief-forming process (Goldman 2012). This implies the existence of cases where agents form new rational beliefs although they lacked prior evidence that supported those new beliefs. This is inconsistent with PPE-D.

PPE-D is also inconsistent with standard synchronic evidentialist views on which one's (*ex ante*) rational beliefs *at t* are constrained just by one's total evidence *at t*:

> *Primacy of Present Evidence-Synchronic* (PPE-S). Necessarily, it's rational for *S* to believe that *p* at some time *t* if and only if *p* is sufficiently likely to be true on one's total evidence at *t*.

According to this, if one's beliefs are supported by one's total evidence at the time one hosts the belief, that belief is at least *ex ante* rational. Thus, it does not matter *how* one ended up in that belief *state*; it could have been a completely irrational *process* of belief revision. According to advocates of PPE-S, the irrationality of a process of belief-revision impacts the *ex post* rationality of one's belief, not its *ex ante* rationality (Feldman and Conee 2004; Feldman 2014; Hedden 2015).

PPE-D is also inconsistent with the intuitive judgments many make about self-fulfilling beliefs in cases like Mindreading Machine. In such cases many find it intuitive to treat one as ending up with a rational belief despite the lack of prior evidence for the target belief. So PPE-D's problems are many and varied.

Could we use PPE-S to threaten TEAR-SS? Not clearly. For if one's total evidence fails to sufficiently support one's self-fulfilling belief at *t2*, then the antecedent of TEAR-SS will not be satisfied, as it requires that one's total evidence makes *p* sufficiently likely at *t2*.

In conclusion: there is no plausible disambiguation of TEAR that is inconsistent with Addition Closure.

Acknowledgements. I'm grateful to the AJP editors and referees whose kind attention significantly improved this project. This project was written while under funding from the Alexander von Humboldt Foundation.

References

Antill, Gregory (2020) 'Epistemic freedom revisited', *Synthese* 197(2): 793–815. doi:10.1007/s11229-018-1735-6.

Dahlback, Morten (forthcoming) 'Infinitely Permissive', *Erkenntnis*.

Feldman, Richard and Earl Conee (2004) *Evidentialism: Essays in epistemology*. Oxford University Press.

Feldman, Richard (2014) 'Justification is Internal', in Matthias Steup, John Turri, and Ernest Sosa eds., *Contemporary Debates in Epistemology*: 337–350. Blackwell.

Foley, Richard (1991) 'Evidence and Reasons for Belief', *Analysis* 51(2): 98–102. doi:10.1093/analys/51.2.98.

Goldman, Alvin (2012) *Reliabilism and Contemporary Epistemology*. Oxford University Press.

Hedden, Brian (2015) 'Time-Slice Rationality', *Mind* 124(494): 449–491. doi:10.1093/mind/fzu181.

Marxen, Chad (2021) 'Closing the Case on Self-Fulfilling Beliefs', *Australasian Journal of Philosophy*. On-line first. doi.10.1080/00048402.2021.1967416.

Peels, Rik (2015) 'Believing at Will is Possible', *Australasian Journal of Philosophy* 93(3): 524–41. doi:10.1080/00048402.2014.974631.

Podgorski, Abelard (2017) 'Rational Delay', *Philosophers' Imprint* 17(2): 1-19. Hyperlink.

Raleigh, Thomas (2017) 'Another Argument Against Uniqueness', *The Philosophical Quarterly* 67(267): 327–346. doi.10.1093/pq/pqw058.

Reisner, Andrew (2015) 'A Short Refutation of Strict Normative Evidentialism', *Inquiry* 58(5): 477–85. doi.10.1080/0020174X.2014.932303.

Sharadin, Nathaniel (2016) 'Nothing but the Evidential Considerations?', *Australasian Journal of Philosophy* 94(2): 343–361. doi.10.1080/00048402.2015.1068348.

Silva, Paul (2018) 'Explaining Enkratic Asymmetries: Knowledge-First Style', *Philosophical Studies* 175 (11): 2907-2930. doi.10.1007/s11098-017-0987-1.

Silva, Paul (2023). *Awareness and the Substructure of Knowledge*. Oxford University Press.

Silva, Paul and Sven Bernecker (forth) 'Evidence, Reasons, and Knowledge in the Reasons-First Program', *Philosophical Studies*.

Silva, Paul and Luis R. G. Oliveira eds. (2022) *Propositional and Doxastic Justification: New Essays on their Nature and Significance*. Routledge.

Silva, Paul and Eyal Tal (2021) 'Knowledge-First Evidentialism and the Dilemmas of Self-Impact.' In Kevin McCain, Scott Stapleford, and Matthias Steup, eds., *Epistemic Dilemmas*: Chapter 11. Routledge.

Velleman, David (1989). *Practical Reflection*. Princeton University Press.