

A paradox of rejection

Thomas N. P. A. Brouwer

Received: 21 March 2014 / Accepted: 12 August 2014 / Published online: 20 September 2014
© Springer Science+Business Media Dordrecht 2014

Abstract Given any proposition, is it possible to have rationally acceptable attitudes towards it? Absent reasons to the contrary, one would probably think that this should be possible. In this paper I provide a reason to the contrary. There is a proposition such that, if one has any opinions about it at all, one will have a rationally unacceptable set of propositional attitudes—or if one doesn't, one will end up being cognitively imperfect in some other manner. The proposition I am concerned with is a self-referential propositional attitude ascription involving the propositional attitude of rejection. Given a basic assumption about what constitutes irrationality, and a few assumptions about the nature of cognitively ideal agents, a paradox results. This paradox is superficially like the Liar, but it is importantly different in that no alethic notions are involved at all. As such, it stands independent of the Liar and is not a 'revenge' version of it. After setting out the paradox I discuss possible responses. After considering several I argue that one is best off simply accepting that the paradox shows us something surprising and interesting about rationality: that some cognitive shortfall is unavoidable even for ideal agents. I argue that nothing disastrous follows from accepting this conclusion.

Keywords Rejection · Bilateralism · Propositional attitudes · Paradox · Rationality · Dialetheism · Indeterminacy

T. N. P. A. Brouwer (✉)
Department of Philosophy, School of Divinity, History and Philosophy,
Northern Institute of Philosophy, University of Aberdeen,
Old Brewery, High Street, Aberdeen AB24 3UB, Scotland, UK
e-mail: tnpabrouwer@gmail.com

1 Rejection

In recent years there has been some enthusiasm among philosophers of logic for taking as primitive not only a propositional attitude of acceptance (and a speech act of assertion that expresses it) but also a propositional attitude of rejection (and a speech act of denial that expresses it). If assertion is akin to saying ‘yes’ to a question, denial is akin to saying ‘no’: they are equally strong but opposed ways of taking an opinion on something. As such both are distinct from and incompatible with the suspension of judgement. Taking denial/rejection as primitive—a view known as ‘bilateralism’—is a departure from the Fregean orthodoxy which identifies denying something with asserting its negation. But many have thought that this ideological enrichment earns its keep.¹ Here are some of the uses it has been put to:

- To explain the normative import of logical consequence. It has been proposed that if $A_1, A_2, \dots, A_n \vdash B$ holds then one should not accept all of A_1, A_2, \dots, A_n without also accepting B. But this seems too demanding, because the things one accepts will generally entail multitudes. To make the requirement more reasonable, one option is to restrict it to known entailment (Field 2009). But alternatively one could say that if $A_1, A_2, \dots, A_n \vdash B$ holds, one should not accept all of A_1, A_2, \dots, A_n while *rejecting* B (Restall 2005).
- To account for the significance and meaning of negation. It’s been proposed, for instance, that negation amounts to something like an embeddable rejection sign (Bendall 1979; Price 1990).
- To account for the possibility of disagreement in a way consonant with dialetheism: the view that there are true contradictions. On such a view, accepting a proposition and its negation is sometimes the rational thing to do, and so disagreement can’t in general be conceived as one person accepting a proposition and another accepting its negation. After all, if the proposition in question is a truth of the form $A \ \& \ \neg A$, both persons may accept A and accept $\neg A$ and not be in disagreement. Instead, the dialetheist conceives of disagreement as one person accepting a proposition and another rejecting it—attitudes which are always incompatible (Priest 2005).
- To characterise in neutral terms the differences between different (classical, intuitionist, paraconsistent) logics of negation (Restall 2005, 2013).
- To solve the so-called ‘categoricity problem’ for propositional logic (Smiley 1996; Rumfitt 2000).

Which general principles one takes to constrain the attitudes of acceptance and denial will depend to some extent on the logical principles one endorses. Intuitionists and paracompleatists may hold that sometimes one should reject a proposition and reject its negation, though this is a matter of disagreement (cf. Wright 2001; Caie 2012). Dialetheists generally hold that in some cases one should accept a proposition and also its negation. Classicists might not allow either of those options. Everyone seems to agree, however, on the following:

[PSA] One cannot rationally accept and reject the same proposition.

¹ Frege (1919). For a general introduction to the notion and a vindication of it *pace* Frege, see Smiley (1996).

This principle, which we might call the Principle of Self-Agreement or PSA for short, is held to be a basic claim about rationality.² But more than that, it is held to be constitutive of the notion of rejection, and crucial for the uses that notion is put to. The principle, in any case, seems eminently plausible to me.³

The Principle of Self-Agreement should be clearly distinguished from the claim that accepting and rejecting the same proposition is impossible, either psychologically or metaphysically.⁴ That is a very different claim, and less plausible to my mind. It is certainly implausible that anyone would, with full presence of mind, resolve to accept a proposition and also to reject that proposition, but it seems perfectly conceivable that someone sufficiently confused could be led by one train of thought to accept P and by another to reject P, and not notice that they had done so.

Note also that the plausibility of the principle is independent of whether one is a ‘bilateralist’, i.e. a primitivist about rejection, or whether one takes the Fregean line of identifying it with the acceptance of a negation. For those philosophers taking the Fregean view—who will not be dialetheists, for reasons already stated—the Principle of Self-Agreement would rule out accepting a proposition and accepting its negation, and it would thus be something like a pragmatic ancillary of the Law of Non-Contradiction. Not being dialetheists, they will not have a problem with that. The emergence of the paradox similarly does not depend on whether one is a primitivist about rejection or not. Although interest in the notion of rejection is most commonly found among bilateralists, the paradox I present is there for everyone to enjoy.

To run the paradox we also require three further assumptions. The first one is a broadly metaphysical one concerning propositional attitudes:

[A] Accepting that one bears an attitude of acceptance or rejection to a proposition suffices for bearing the attitude in question towards that proposition.

I think this is a plausible assumption, but it is open to some interpretation. On a strong reading, accepting that I have a certain attitude *constitutes* having that attitude (in either a metaphysical or conceptual manner) and so suffices for it. On a weaker reading, it simply happens to be the case (perhaps as a matter of psychological regularity) that whenever I accept that I have some attitude X, I have that attitude. We need nothing

² Throughout this paper, the notion of rationality I discuss is a *synchronic* one, that is, a notion of rationality that governs a subject’s propositional attitudes as they are at some given moment. This is to be contrasted with a *diachronic* notion of rationality, which governs transitions from one set of propositional attitudes to another. So, to illustrate, it is a requirement of synchronic rationality that one’s beliefs be consistent (let’s say) but it is a requirement of diachronic rationality that when one’s beliefs are inconsistent, one give up those inconsistency-generating beliefs that are least justified (let’s say). There is a debate to be had about whether one of these notions of rationality is to be cashed out in terms of the other, but that debate doesn’t matter for anything I’ll say in this paper.

³ Though I would like to be able to provide an argument for the principle, I would not know where to look for premises—it seems too basic for such. If one took the notion of rejection to be analysable in terms of assertion and negation, as per the Fregean view (above), one could argue for it from claims about negation, unless one were a dialetheist. But the dialetheist, as we have seen, already has reason to reject the Fregean view. If one takes the notion of rejection as primitive, one has a good reason for accepting the principle, for without it, it becomes hard to say what the opposition between acceptance and rejection consists in.

⁴ This is a view Graham Priest defended at one point, as evinced by his (2006), pp. 98–99. He has since backed off from it, though not in print, as far as I know.

more than the weak reading for the paradox. It need not even hold without exception, as long as we can presume it to hold in any given case, absent evidence to the contrary.⁵

Then, finally, we need two assumptions concerning ideal cognitive agents:

[B] For any attitude one bears to a proposition, one accepts that one bears this attitude to it.

[C] For any attitude one does not bear to a proposition, one rejects that one bears this attitude to it.

These assumptions are fairly demanding. It is plausible that few if any actual agents will satisfy them in full generality. Nevertheless, they are not so strong that we cannot expect actual agents to satisfy them for a good many propositions. Assumption B will presumably fail in cases where subjects simply haven't reflected on their own attitudes. These cases will presumably be numerous in actual agents, but an ideal agent will (I submit) display a full awareness of their own attitudes. Assumption C will most obviously fail in cases where an agent hasn't given any thought to some proposition at all, let alone taken an attitude towards it. Such cases will again be numerous in actual agents, but an ideal agent can be expected to have considered what attitudes to take (or not take) on any given topic.

There is, of course, ample room for disagreeing about what an ideal cognitive agent looks like. I do not mean to legislate. I think there is probably a range of notions of ideality, suitable for a range of different theoretical purposes. But I do think that there is a recognisable notion of cognitive ideality out there which involves satisfying assumptions like B and C, and indeed A. One might call an agent that is ideal in that way a *transparent* agent: they are fully and accurately opinionated about their own attitudes on any given matter, both about the ground-level ones and about the higher-order ones (i.e. those that pertain to their own attitudes). It is my concern in this paper to say something about the limits of agents of that sort.

All that being said, it is worth noting that if we were to take a non-ideal but reasonably clear-headed agent, expose them to some particular proposition P and the evidence concerning P, and ask them to try and make up their mind about it, then we may reasonably expect them to end up satisfying A, B and C *with regard to P*. I think this will also hold for the paradoxical proposition to be discussed shortly, and as a result the paradox also has some bearing on the nature of non-ideal agents. But the case is clearer with ideal agents, so that is what I will focus on.

In the next section I examine a self-referential proposition which is such that one cannot fail to have irrational attitudes towards it—irrational in the sense of violating the Principle of Self-Agreement—if one satisfies assumptions A, B and C. It is broadly analogous to a Liar sentence, but truth and falsity play no role in generating this puzzle, and in addition the most obvious analogues of gappy or gluttony approaches do not stop the paradox.

⁵ Exceptions might be cases in which one is under- or misinformed about the contents of a proposition which one endorses or denies. Maybe in such cases one can accept that one accepts something while failing to actually accept it. But the case is not clear. One might hold, for instance, that if a lack of comprehension of P stops one from having attitudes of acceptance or rejection to it, that same lack of comprehension would stop one from having second-order propositional attitudes involving P.

After setting out the paradox, I will discuss a number of responses to it, some more promising than others. I will ultimately suggest that we are best off accepting that there are cases of unavoidable irrationality. In the final section I'll investigate what the upshot of that would be. I argue that nothing unacceptable follows from it: we have simply learned something new and surprising about ideal cognitive agents.

2 Paradox

Consider the following self-referential proposition, which I'll call "R", as short for "Rejecter":⁶

[R] I reject R.

What stance can one rationally take towards R, if any? There are *prima facie* four options: (i) accept it, (ii) reject it, (iii) do neither or (iv) do both. Options (iii) and (iv) seem to be the closest analogues of, respectively, gappy⁷ and glutty solutions to a Liar paradox, though nothing turns on that analogy. Of these four options, we can strike off (iv) immediately, since it runs headlong into the Principle of Self-Agreement.

Let's take a closer look at the other options. What about (i)? If one accepts the proposition, then one accepts that one rejects R. Can one rationally do this? As it turns out, no. Accepting R entails accepting that one rejects it, for that is the content of R. By assumption A, accepting that one rejects R entails that one rejects R. Thus one will both accept and reject R, which by PSA one cannot rationally do.

So much for option (i), then. What about (ii)? If one rejects R, one rejects that one rejects R, for that is the content of R. Is that problematic? As it turns out, yes. On assumption B, rejecting R entails accepting that one rejects it. So, one ends up both accepting and rejecting that one rejects R. And that is equivalent to accepting and rejecting R.

So much then for option (ii). The only option left to consider is (iii): neither rejecting nor accepting R. This seems to me to be the most straightforward analogue of a gappy response to an alethic paradox. This option initially seems promising: after all, not accepting R does not entail that one rejects it, so one doesn't straightforwardly get the problems of option (ii). And likewise, not rejecting R does not entail that one accepts it, so one doesn't straightforwardly get the problems of option (i). But problems do emerge if we take into account assumption C.

On assumption C, not accepting R entails rejecting that one accepts R. This does not lead to any problem. But on assumption C not rejecting R also entails rejecting

⁶ The form of self-reference that goes on here is a bit less straight-forward than in the case of the Liar sentence, though I do not think it in any way problematic. While the Liar says of itself that it is false, the Rejecter says of the speaker that the speaker rejects it. Self-reference is still achieved, but it runs by way of the speaker's attitudes. Although less popular now, in the past various recipes for combating paradoxes have involved banning self-reference from the language, most famously the recipe given in Tarski (1933). Whether the type of anti-paradox legislation that rules out Liar-like self-reference will also automatically rule out R is not obvious: it would depend on the fine print.

⁷ There are other responses to R that are also reminiscent of gappy solutions. I discuss these in the next section.

that one rejects R. And rejecting that one rejects R is problematic, as we saw in the case of option (ii): via assumption B one again ends up accepting and rejecting that one rejects R.

It seems, then, that one cannot take any one of options (i), (ii), (iii) or (iv) without violating the Principle of Self-Agreement. Insofar as that principle is constitutive of rationality, an ideal cognitive agent cannot avoid being irrational when it comes to R.

3 Some initial responses

Although R is a bit like the liar paradox, it is not an alethic paradox.⁸ Since it involves propositional attitude ascriptions, we might call it a pragmatic paradox. Paradoxes are the kind of thing we usually try to get rid of. Are there any promising ways to rid ourselves of the Rejecter? There are certainly some options worth considering. One quite promising solution to a very similar paradox is defended in [Caie \(2012\)](#). I think Caie's solution ultimately fails, however, and I will explain why. But before I do that I first want to discuss more briefly two other options which, though less promising, are still worth touching on.

Infinite-Order Agnosticism. To avoid the irrationality threatened by the Rejecter an agent might try not to have any attitudes towards R, but also no attitudes about one's attitudes toward R, and no attitudes about *those*, etc. If one were in that position, the assumptions A through C could not be deployed to get one into trouble. This option could be seen as another (perhaps more thorough-going) analogue to a gappy response to the Liar. It seems psychologically unlikely that we ordinary mortals can put oneself in this state through a process of reflection. But nevertheless, it's a state we can be in, for it is the state one would be in if one has never considered R. So that would be a basic way of avoiding the looming irrationality: avoid the proposition R. But, first off, this won't be an option for anyone who has already had the misfortune of encountering R. Second, and more to the point, the kind of ideal cognitive agent we are considering,

⁸ Terence Parsons discusses a similar sophism in his (1984), formulated using denial rather than rejection: 'I deny this very sentence'. He notes one cannot coherently deny it, but doesn't take the matter much further than that. [Caie \(2012\)](#) discusses a close relative of the Rejecter (formulated in terms of disbelief) and proposes a response to it. I'll discuss that response in Sect. 4.

The Rejecter is also worth comparing to a paradox discussed by Graham Priest in his (2005), chapter 6. It involves a proposition p: "it is irrational to believe that p" (he regards it as a variant of the 'irrationalist's paradox' due to Greg Littman). To believe it would be to believe something which one considers irrational, and since that is irrational, one ought not believe it. But if it's indeed irrational to believe it, then that shows that p is true, and that one ought to believe it. One ends up in a situation where one ought and ought not believe p.

If we reformulate p as 'one ought to reject p', a resemblance between Priest's paradox and the Rejecter emerges. The two involve some of the same issues. For instance, its (reformulated) derivation arguably involves something like a normative analogue of our assumption A: one ought to reject what one accepts one ought to reject. As such, we can see Priest's discussion as an anticipation of the Rejecter.

However, Priest's discussion is brief, effectively restricted to a discussion of option (ii), above. It's not clear how Priest's paradox would fare on the other options: at least *prima facie*, it seems possible to reject p (i.e. take option iii) without landing in irrationality; one could reject something without believing that one ought to. Option (iv) also seems *prima facie* safe. Furthermore, Priest's bullet-biting solution to his paradox is to accept that there are dialetheias concerning what's rational; while my proposed response (below) will also be of a bullet-biting nature, it does not involve buying a contradiction, and so is far less logically committal.

one that satisfies assumption C, cannot evade any proposition entirely: even if they have no opinions about some proposition, they will still be aware of their lack of opinions about that proposition, and that is enough to get the paradox going. So this response does not pan out in the case of the ideal cognitive agents we're interested in.

Incoherentism. One might resort to denying the Principle of Self-Agreement. If that principle is dropped, the problem does indeed go away. But note the cost: we would lose pretty much any grip we had on the notion of rejection. It is essential to the notion of rejection that it is somehow opposed to acceptance—it's hard to see how one would capture this opposition, if not through the principle in question. Furthermore, the notion of rejection could no longer play most of the roles it has been called upon to play (see Sect. 1). For example, dialetheists would not be able to account for disagreement in terms of acceptance and rejection, any more than they can do so in terms of acceptance and negation.⁹ And, last but not least, this looks like a fairly *ad hoc* move.

These, then are the options I consider less promising. More mileage is to be gotten out of a response proposed by Michael Caie in his (2012), concerning a paradox very similar to the Rejecter. According to that proposal, one ought to have (and ideal agents would have) *indeterminate* attitudes towards R.

4 Indeterminism about the rejecter

Perhaps the proper stance toward R is not one of (i)–(iv) but rather another one in which it is (v) indeterminate whether or not one accepts R and indeterminate whether or not one rejects it, and hence indeterminate whether one does both, and whether one does neither. The thought would be that since one isn't determinately in any one of the problematic states, no irrationality results.

For this to work, one would also have to reject some instances of the law of excluded middle. One should deny that one either accepts or doesn't accept R, and that one either rejects or doesn't reject R. If one doesn't, then one can still derive the paradoxical conclusion, by arguing by cases from the disjunction of (i)–(iv). So this solution looks to be a non-classical and more specifically paracomplete one.¹⁰ In any case, that is the approach that Caie (2012) takes.

⁹ This makes our paradox an interesting challenge for dialetheists. While arguably no-one has it easier than them when it comes to alethic paradoxes, the very ideology they need in order to account for disagreement generates a paradox that they cannot respond to in their patented manner.

¹⁰ There may be *some* mileage in classical solutions, but they don't quite seem to get the job done. Suppose one were an epistemicist about indeterminacy. Then one might suggest that though there is a fact of the matter about which stance one takes to R, we are in principle ignorant of which one it is. In that case, the assumptions (b) and (c) seem to fail; possibly also (a). So in that case, if the stance one is in fact taking is one of (i) through (iii), irrationality does not ensue. But since one may in fact be taking stance (iv), it is not determinately the case (by epistemicist lights) that one is not in an irrational state. The best result the epistemicist can secure here is that one also isn't determinately in an irrational state. Supervaluationism about indeterminacy hits the same snag. Of course, one could claim that though it is indeterminate whether one is taking stance (i), (ii) or (iii), it is determinate that one isn't taking stance (iv). One way of guaranteeing this is to say that taking stance (iv) is simply (psychologically or metaphysically) impossible, as Priest does. But as noted above, it is unclear what the motivation would be for saying this. Saying that (iv) would be irrational would be to beg the question: we want to know whether or not we are in fact rational when it comes to R, and we don't take it as given that we are.

I think the indeterminist approach will not work, and I will try and show that in a bit. But let's just suppose, for the sake of argument, that it does get the job done: if one's attitudes to R are indeterminate, one will not end up violating the Principle of Self-Agreement. That still leaves some philosophical questions unanswered. As in the case of infinite-order agnosticism, it seems unlikely that one can, by deliberating, put oneself in a state of attitudinal indeterminacy. A more promising line to take, perhaps, is that the natural result of reflecting upon R is confusion about R, and that this confusion should be diagnosed as a state of attitudinal indeterminacy of the above sort. If that is so, then the surprising lesson of the Rejecter paradox is not that ideal cognitive agents are inevitably irrational, but that they are inevitably confused.

But is it obvious that one ought to diagnose confusion about R that way? In particular, is that diagnosis more plausible than saying that to be confused about R is to end up, *nolens volens*, taking stance (iii), i.e. neither accepting nor rejecting R? The indeterminate state may have the advantage of staving off irrationality, but that of course doesn't mean that it's the state we in fact naturally end up in. To reason in that way would be to reason *ad consequentiam*. We thus end up with a question of philosophical psychology—what exactly does confusion consist in?—one that I have no idea how to settle properly.¹¹

A proponent of the indeterminacy view might fairly complain at this point, though. Given that we are talking about ideal agents, is it really relevant to ask how they manage to get themselves into the indeterministic state? We already know that ideal agents are unrealistic; we can just assume them to be capable of the required feats of mental agility, even if those are unrealistic. That is a fair point. The main issue, ultimately, is whether the indeterminist approach gets the job done. But alas it doesn't: although it may handle the straight Rejecter, there is a revenge paradox lurking just around the corner. Take the following proposition:

(R*) I determinately reject R*.

This proposition—call it the Determined Rejecter—lands us in irrationality in all of the ways that R did, and more. Given that determinacy is factive (i.e. if one determinately rejects R*, one rejects R*) one could use assumptions A–C to show that one gets into the same trouble with R* as with R. However, to make the case clear it may help to rejig A–C so that they pertain explicitly to determinate attitudes:

[A*] Determinately accepting that one determinately bears an attitude of acceptance or rejection to a proposition suffices for determinately bearing the attitude in question towards that proposition.

[B*] For any attitude one determinately bears to a proposition, one determinately accepts that one determinately bears this attitude to it.

¹¹ However, the diagnosis of confusion in terms of neither accepting nor rejecting may have something going for it. When we first come to consider proposition R, we presumably don't yet have any attitudes towards it. In other words, we start out with stance (iii). It's hard to see how, in the process of considering R, we would naturally end up in a state of indeterminate attitudes towards it, instead just remaining in the unopinionated state that we started out with. Since all options look equally bad, there doesn't seem to be anything that would motivate us to change our set of attitudes, irrationality-inducing as that set may be.

[C*] For any attitude one determinately does not bear to a proposition, one determinately rejects that one determinately bears this attitude to it.

These assumptions are slightly different from the original ones, in that any attitude ascription is replaced with an explicitly determinate attitude ascription. In certain small ways, these assumptions are both weaker and stronger than the original three.¹² But bearing in mind that these assumptions are supposed to hold of ideal cognitive agents (of the ‘transparent’ sort) I think they are no less plausible than the original three.

With these reformulated assumptions in mind, let’s again consider options (i)–(iv), imagining in each case that the attitudes in question are borne (or not borne) with determinacy. Option (iv), determinately accepting and determinately rejecting R^* , is still a non-starter. Nothing changes here by the addition of a determinacy operator.

What about option (i), just determinately accepting R^* ? Well, given the content of R^* , this amounts to determinately accepting that one determinately rejects R^* . By A^* this entails that one determinately rejects R^* . So one determinately accepts R^* and determinately rejects it*, same as before.

What about option (ii), just determinately rejecting R^* ? Well, by B^* , one will then determinately accept that one determinately rejects R^* . And given the content of R^* , that entails that one determinately accepts R^* itself. And so one determinately accepts R^* and determinately rejects it, as before.

What about option (iii), determinately doing neither, i.e. determinately not accepting R^* and determinately not rejecting it? Well, if one determinately does not reject R^* , then by C^* one determinately rejects that one determinately rejects R^* . But given the content of R^* , that entails rejecting R^* . As we can see above, that gets us back into trouble.

That leaves the indeterminist’s preferred option (v): having indeterminate attitudes to R^* . What happens in that case? Given that the assumptions above explicitly only rule on determinate cases, they will not serve to get one into trouble. But now consider this further assumption:

[C**] For any attitude one does not determinately bear to a proposition, one determinately rejects that one determinately bears this attitude to it.

This assumption is not to be confused with C^* above: the negation and the determinacy operator are reversed here in the antecedent. But C^{**} isn’t any more objectionable than C^* . Consider: if one’s attitudes to a proposition happen to be indeterminate, then it will determinately not be the case that one determinately rejects the proposition in question. And given that this is determinately not the case, it seems entirely appropriate to determinately reject it.¹³

¹² The assumptions are of broadly conditional form, which means that one would strengthen them by either weakening the antecedent or strengthening the consequent, and that one would weaken them by either strengthening the antecedent or weakening the consequent. Here both the antecedent and consequent are strengthened in each case, but in the case of A the antecedent is strengthened in two ways and the consequent in one, while in the case of B and C the antecedent is strengthened in one way and the consequent in two ways.

¹³ Someone might object at this point that it might conceivably be indeterminate whether one’s attitudes to a proposition are indeterminate, i.e. one’s attitudes might be higher-order-indeterminate. Perhaps, but

With C^{**} at our disposal, the reasoning runs as follows. If one's attitudes to R^* are indeterminate, then one does not determinately bear the attitude of rejection to R^* . By C^{**} , that entails that one determinately rejects that one determinately bears the attitude of rejection to R^* . Given the content of R^* , that amounts to determinately rejecting R . And that, as we have seen, gets us into trouble.

To sum up, while the indeterminist's option (v) does seem to get us out of the original pickle that the Rejecter landed us in, closely analogous reasoning about the Determined Rejecter gets us straight back into it. So the indeterminist solution, while the most promising of the ones I've considered, does not get the job done.

5 Giving up in the face of irrationality

If the indeterminist option is the most promising one, and it doesn't work, does that mean we should simply give up on solving the Rejecter paradox? Yes, I think that is exactly what it means. We should take the way of least resistance, and embrace that there's no philosophical response to the Rejecter paradox which neutralises it. In other words, we should accept that R lands the ideal cognitive agent in rationally incompatible attitudes. I do not think that this is merely the act of a desperate philosopher, though. I think it's a perfectly acceptable conclusion to draw that irrationality is sometimes unavoidable.

First, as a general consideration, we might note that it is sometimes worth sacrificing a little rationality to save something else of value, such as effort. This is a view one could take of cases like the preface paradox, for instance.¹⁴ Let's say I've written a monograph, and while I've done so with due diligence, I think that I've still probably made a mistake somewhere. How can I rationally believe the claims in my book while also believing that at least one of them is false? Well, perhaps I can't. But if the only alternative to being irrational in this way is checking and re-checking the whole book *ad nauseam*, I might be better off just living with the irrationality. And if we ought to be willing to accept that sometimes irrationality is a lesser evil, is it so much more terrible to accept that sometimes it is an unavoidable evil?¹⁵

One might have more specific worries, though. One might worry that, in some way, buying into unavoidable rationality gets one into greater trouble down the line. I can think of two such worries.

Footnote 13 continued

one would not want to endorse that option in the case of R or R^* . For if it is indeterminate whether one's attitudes are determinate, and having determinate attitudes gets one into trouble (as we have already seen) then the best result one can hope for with respect to R or R^* is that it is indeterminate whether they get us into trouble. Now perhaps one might decide to go for that result if one were convinced that nothing better could be had, but it clearly falls short of the result that the indeterminism about R was supposed to secure for us.

¹⁴ See Makinson (1965).

¹⁵ Perhaps what I describe as sacrificing rationality for something else could, in the case of the preface paradox, be described as one kind of rationality (practical rationality) trumping another (theoretical rationality). If so, take me to be speaking specifically about theoretical rationality, everywhere above. It does not really matter for the case at hand, because with R no such trumping is going on, and no other kind of rationality is involved.

Ought implies can. Is the idea of unavoidable irrationality inconsistent with the principle that ought implies can or (more immediately) its contrapositive, the principle that can't implies needn't? This depends on what view one takes of the normativity of rationality. Ought we always be rational? If so, then yes, irrationality is always a failure to do what we ought, and unavoidable irrationality would constitute a case where *can't* doesn't bring *needn't* along. So on that assumption, can't-implies-needn't (and thus ought-implies-can) is indeed violated.

If one wanted to avoid this result, one option is to claim that rationality is only defeasibly obligatory—one ought to be rational, but only where one can. In that case, the Rejecter would constitute a case in which rationality is not obligatory, because rationality is not attainable there. This option might attract those who think the notions of normativity and blameworthiness should be quite closely linked. We clearly don't always regard irrationality as blameworthy: we all know that we all probably have some incoherencies among our beliefs, and yet we're not parcelling out any blame for that. If one holds that something not being blameworthy implies that it is permissible, then one should think there's lots of permissible irrationality about.

I do however think that irrationality is always impermissible. I think there is something inherently normative about rationality. But I also think that failures to do what one ought can be excusable (i.e. blameless), and that this is the case with the Rejecter. To accept R and reject it is irrational and hence impermissible, but also excusable because unavoidable. This is a violation of ought-implies-can, sure. But I don't think we should be scared to give that principle up, once we've distinguished the impermissible from the blameworthy. The real intuitive force is with necessity-implies-blamelessness, not with necessity-implies-permissibility (i.e. can't-implies-needn't, i.e. ought-implies-can). And necessity-implies-blamelessness *is* preserved. So this first worry doesn't hold up.

Rational explosion. Accepting and rejecting the same thing seems somewhat akin to believing a contradiction. Might it therefore engender something akin to explosion? Would having such attitudes to R commit an agent to, say, accepting any proposition whatsoever, or rejecting any proposition whatsoever, or perhaps not rejecting any proposition whatsoever? It's easiest to construct a putative argument for the last of these—that accepting and rejecting the same thing commits the agent to not rejecting any proposition. Extra assumptions will then let us turn that into a putative argument for universal acceptance, but since universal non-rejection is already bad enough, I won't pursue the stronger result. Also, I should note that other arguments are possible besides the one I present; but this one seems most elegant and straightforward, and as far as I can see, others will have comparable weaknesses.

Here goes. Assume the following:

[D] If $A_1, A_2, \dots, A_n \vdash B$, then one should not accept all of A_1, A_2, \dots, A_n whilst rejecting B (for any propositions A_1, A_2, \dots, A_n and B).

[E] One ought not reject a proposition A without also accepting its negation.

D is, I think, a pretty plausible principle.¹⁶ E is considerably more contentious. As noted in Sect. 1, paracomplete and intuitionist logicians may well hold that some

¹⁶ As noted in Sect. 1, it is defended in Restall (2005).

propositions (i.e. gappy or unknowable ones) ought to be rejected along with their negations. Principle E would then commit them to accepting those propositions and their negations, not something they would be happy to do. Classical and paraconsistent logicians will not have a similar issue with E—but that is not to say that they have positive reasons for accepting it.

Let's say the agent accepts and rejects R. By E, they ought then accept $\neg R$. If they do that, they accept R and $\neg R$. By *ex contradictione quodlibet*, R and $\neg R$ imply a randomly picked proposition Q. Given assumption D, they then have two options: they ought either not accept both R and $\neg R$, or not reject Q. However, given the Rejecter paradox, they cannot but accept R, and given assumption E they ought to accept $\neg R$. So the only remaining option is to not reject Q. Since Q was randomly chosen, they ought not reject any proposition. Therefore, accepting and rejecting P commits them to not rejecting anything.

What are our options for avoiding this calamity? I can see four:

- *Give up ex contradictione quodlibet*. Many philosophers already do give up this principle, for varied reasons. I would not be averse to taking this line myself, but it is still generally regarded as radical.
- *Give up assumption D*. This assumption, to my mind, is fairly plausible and I would be loath to reject it.
- *Give up assumption E*. As noted, some logicians can be expected to take issue with this principle independently of the issue at hand. It seems to have various *prima facie* counterexamples: truth-value gaps, cases of indeterminacy, cases of in-principle ignorance. (These will not necessarily be distinct cases, of course—that will depend on one's views.) Those counterexamples might individually be controversial, but in the absence of any strong reason pushing us towards E, they should have some force. And these qualms are not exclusive to paracomplete and intuitionist logicians: classical logicians who advocate a supervaluationist treatment of indeterminacy might well hold that one ought to reject indeterminate propositions and their negations, and they would then have the same trouble with E as paracomplete and intuitionist logicians do. Paraconsistent logicians might have no issue with E, but they would be very likely to have given up *ex contradictione quodlibet*, and so are not in trouble anyway.
- *Find a loophole*. In the argument above, assumption E tells the agent to accept $\neg R$. If they do so, they land in a normative pickle: they will be obliged not to reject any proposition. But of course they might also simply disobey the commands of assumption E, and not accept $\neg R$. Sure, that would be to fail to live up to their obligations (assuming E is correct in the first place) but if they did accept $\neg E$ they would end up with a great many more obligations that they would inevitably fail to live up to (i.e. obligations to not reject anything). So why not duck out early? Of course, in the present context that means that the ideal cognitive agents we've been talking about would be involved in additional irrationality over and above violating the Principle of Self-Agreement: they would also violate assumption E.

Of these four options, the third one seems to me the most painless, most conservative, and most plausible. But whichever one prefers, it should be clear that 'rational explosion' is anything but unavoidable.

There may of course be other problems besides these two that result from embracing unavoidable irrationality. And it's quite imaginable that those will turn out to be so severe that buying into unavoidable irrationality simply isn't an option. But I am not aware of any such problems, and in their absence, I think buying into the irrationality is our best option.

6 Conclusion

In this paper I've presented a paradox and argued that it tells us something interesting about ideal—'transparent'—cognitive agents: they either violate the Principle of Self-Agreement or they fail to be truly ideal in failing to live up to assumptions A, B or C.¹⁷ The paradox does not pertain merely to ideal agents, though. Though the assumptions A, B and C are not such that ordinary cognitive agents will satisfy them in their full generality, it is not so implausible that they will satisfy them to a limited extent (i.e. with regard to some propositions, and up to some level of higher-order attitudes). They may well satisfy them to the extent of getting into trouble with the Rejecter.

But though the Rejecter gets one into trouble—in the sense of landing one in irrational attitudes and thereby stopping one from living up to one's cognitive obligations—I've argued that this is not the end of the world. Just as irrationality can sometimes be a lesser evil, it can also be necessary evil. That's life; and since no further trouble seems to result from this evil, we should be able to learn to live with it.

Apart from what it shows us, the Rejecter is also an interesting paradox in its own right. While it bears an obvious similarity to the Liar, it is independent of it—the notions of truth and falsehood never once make an appearance. And it is also interestingly more tricky than the Liar—neither the closest analogue of a glutty response (i.e. accepting and rejecting R) nor the analogue of a gappy response (i.e. not accepting or rejecting R) nor more sophisticated versions thereof (i.e. infinite-order agnosticism, indeterminism) can get a grip on it.

Acknowledgments I would like to thank Michael Bench-Capon, Jason Turner, Robbie Williams and Francesco Berto and audiences at Leeds, Manchester, Amsterdam and Aberdeen for comments on drafts of this material, as well as several anonymous referees. This paper was prepared within the 2013–15 AHRC project *The Metaphysical basis of Logic: the Law of Non-Contradiction as Basic Knowledge* (grant ref. AH/K001698/1).

¹⁷ As this is a disjunctive claim, there in principle three ways in which it could be true. One might think there could be fully transparent agents, but not ones that are simultaneously fully rational; or one might think there could be fully rational agents, but not ones that are fully transparent; or that agents are never fully transparent nor ever fully rational. I've argued for the disjunction by assuming for the sake of argument that we have a fully transparent agent, and arguing that they would end up having irrational attitudes to R. But that argument does not commit me to any one of the three options listed. One might believe, as Priest at one time did (see footnote 4) that it's metaphysically impossible to accept and reject the same proposition. If so, then the argument would commit me to believing that there are no fully transparent agents (with views on the Rejecter). But I want to merely defend the weaker, disjunctive claim.

References

- Bendall, K. (1979). Negation as a sign of negative judgment. *Notre Dame Journal of Philosophical Logic*, 20(1), 68–76.
- Caie, M. (2012). Belief and indeterminacy. *Philosophical Review*, 121(1), 1–54.
- Field, H. (2009). What is the normative role of logic. *Proceedings of the Aristotelian Society*, 83(Suppl), 251–268.
- Frege, G. (1919). Die Verneinung. (Republished in *Logische untersuchungen*, 5th edition, pp. 63–83, by G. Patzig, Ed., 2003, Göttingen: Vandenhoeck & Ruprecht.)
- Makinson, D. C. (1965). The paradox of the preface. *Analysis*, 25(6), 205–207.
- Parsons, T. (1984). Assertion, denial and the liar paradox. *Journal of Philosophical Logic*, 13(2), 137–152.
- Price, H. (1990). Why ‘Not’? *Mind*, 99(394), 221–238.
- Priest, G. (2005). *Doubt truth to be a liar*. Oxford: OUP.
- Priest, G. (2006). *In contradiction: Expanded* (2nd ed.). Oxford: OUP.
- Restall, G. (2005). Multiple conclusions. In P. Hajek, L. Valdes-Villanueva, & D. Westerstahl (Eds.), *logic, methodology and philosophy of science: Proceedings of the twelfth international congress* (pp. 189–205). London: Kings’ College Publications.
- Restall, G. (2013). Assertion, denial and non-classical theories. In K. Tanaka, F. Berto, E. Mares, & F. Paoli (Eds.), *Paraconsistency: Logic and applications* (pp. 81–99). Dordrecht: Springer.
- Rumfitt, I. (2000). ‘Yes’ and ‘No’. *Mind*, 109(436), 781–824.
- Smiley, T. (1996). Rejection. *Analysis*, 56(1), 1–9.
- Tarski, A. (1933). The concept of truth in formalized languages. (Republished in *Logic, semantics, meta-mathematics*, 2nd ed, pp. 152–278, by A. Tarski, J. H. Woodger (transl.), J. Corcoran, Ed., 1983, Indianapolis: Hackett.)
- Wright, C. J. G. (2001). On being in a quandary. *Mind*, 110(437), 45–97.