

Indeterminate survival.

Most views of personal identity allow that sometimes, facts of personal identity can be *borderline* or *indeterminate*. Bernard Williams argued that regarding questions of one's own survival as borderline "had no comprehensible representation" in one's emotions and expectations. Whether this is the case, I will argue, depends crucially on what account of indeterminacy is presupposed.

I outline and motivate a general class of *rejectionist* accounts of indeterminacy. I argue that Williams' concerns can be given precise focus if this account of indeterminacy is presupposed. The ultimate upshot of his arguments, I suggests, is to place constraints on theorizing about the kind of indeterminacy characteristic of personal identity – arguably undermining the application of standard accounts such as supervenience to this case.

Most views of personal identity allow that sometimes, facts of personal identity can be *borderline* or *indeterminate*. Bernard Williams argued that regarding questions of one's own survival as borderline "had no comprehensible representation" in one's emotions and expectations. Whether this is the case, I will argue, depends crucially on what account of indeterminacy is presupposed.

I outline and motivate a general class of *rejectionist* accounts of indeterminacy. I argue that Williams' concerns can be given precise focus if this account of indeterminacy is presupposed. In particular, we can argue that if one knows it is an indeterminate matter whether X is the same person as me (in the "forensic" sense), then one should not take X's interests into account in determining what actions are prudentially rational. Nor should one feel anxious on one's own account about future harms accruing to X, nor feel proud of X's triumphs. X will be a *de se* emotional blank.

Williams suggests that indeterminate personal identity is incoherent. While the required attitudes to X are strange and alien, I do not support his verdict. I examine three responses. First is simply to follow the argument where it leads---strange and alien though this is. The second is to follow Parfit in denying that *any* two-place

“forensic” relation is what matters in survival. The third is to undermine the puzzle by denying the rejectionist conception of indeterminacy on which it rests.

All three responses are possible, but the first seems ad hoc, and the second radically revisionary. I conclude by suggesting that the case be turned around: we should view examination of the case as providing constraints on acceptable accounts of the nature of indeterminacy involved in personal identity. To illustrate the case, I sketch the consequences of adopting an *epistemicist* conception of personal identity, which avoids the problems. It would, however, alter what is standardly assumed to the dialectical effect of declaring problematic cases of personal identity indeterminate – on an epistemicist construal, an acknowledgement of the indeterminacy of survival is no in principle block to pursuing an inquiry into what the truth value of the indeterminate claims are.

1. Indeterminate personal identity

Might it *indeterminate* or *borderline* whether I survive some episode? Can we allow that certain episodes of psychological or physical disruption, where there is no fact of the matter whether *I* am the person emerging out of the episode? Many would say *yes*.

First reason: indeterminacy in criteria of identity. The kind of criteria often associated with personal identity over time – a sufficient degree of psychological continuity and connectedness perhaps, or sufficient continuity of matter and brain function – are the sort of things for which we can envisage borderline cases. The borderline cases may arise in a number of ways. First source: the notion of “sufficient degree” seems paradigmatically vague – what degree of continuity suffices for having a *sufficient* degree of continuity? Second source: what it takes to constitute having a given degree of continuity may be unsettled. Psychological theorists of personal identity, for example, often say that connectedness between A and B is a matter of A and B having *appropriately related* psychological characteristics where the psychological characteristics of B are *the right kind of causal product* of the

corresponding characteristics of A.¹ Liberals might say that the qualifier “right kind” here is redundant – so long as B’s characteristics are the causal upshot of related characteristics of A, a degree of connectedness is established. Conservatives might say that the “right kind” of causal connectedness requires a special “immanent” kind of causal relationship. Moderates might say that the “right kind” of causal relationship involves something in between---something that rules out “deviant causal chains” but which does not commit one to making sense of immanent causation. Indeterminacy in: (i) what is a cause of what; (ii) what is an immanent cause of what (ii) what is the “right kind” of cause of what, plausibly lead to cases where it is indeterminate whether the conditions for personal identity are met for the respective theorists.

An important case for us will be where it is *determinate* that persons exist at times t1 and t2, but indeterminate whether they stand in the relations of continuity and connectedness that are necessary and sufficient for them being *the same person*.

Second reason: the emptiness of hard cases. The personal identity literature throws up many “hard cases” – often in an effort to diagnose the different predictions of different criteria of personal identity. Such hard cases include teleportation scenarios: the material from which I am made is destroyed on Earth, causing by some reliable mechanism a copy to be created a few seconds later on Mars. Here we have no continuity of matter, but paradigmatic continuity of psychology. Do I survive such an episode? In the same spirit are “psychological transfer” cases, where psychological states originally hosted by body A are transmitted to body B and vice versa. Psychological theories typically see this as a case of persons switching bodies. Bodily theories typically see this as a case of persons switching psychological states.

Some elaborate these cases in the hope of prompting intuitions favouring either the psychological or materialistic criteria of personal identity. But – especially where the scenarios described are “science fiction” rather than something that we might realistically confront in real life – many suspect that debate here is fruitless. One *principled* rationale for such suspicion is the claim that our concept of “person” is not

¹ What is the “appropriate relation”? One thought would simply that it consists in similarity of psychological states – which would make connectedness an extremely temporally limited relation. A frequent rival suggestion is that “direct” connectedness involves an experience and an apparent (q-) memory of that experience; or between a (q-)intention and an act that fulfils that intention.

sufficiently determinate to decide between the opposing views. Such sceptics ask: why think that a concept of *same person as* that has primary application to *normal* cases of survival, should settle how it should be applied to pathological cases such as these? The idea that arguments between psychological and bodily theorists of personal identity over hard cases is “empty”, is well captured by the claim that in such cases it that in such scenarios there is *no fact of the matter* about who will survive as who.²

Third reason: fission cases. Consider a version of the *teleportation* case as described above, but where two copies of me are created on Mars.³ On a psychological account of personal identity, at least, both are good candidates for being me. But – the thought runs – certainly they can’t *both* be me, for then they’d have to be the same person as one another!

The case just described is tailored to the psychological continuity view of personal identity. But fission scenarios confront other positions on personal identity. We are familiar with the idea that amoeba undergo a *physical* fission. What if the biology of some person were relevantly amoeba-like? Post-fission, we would have to viable organisms, each with a substantial degree of material continuity with the original pre-fission case.⁴ Or again, let’s suppose that a person could survive even though one

² See for example Sider “Criteria of personal identity and the limits of conceptual analysis” *Philosophical Perspectives* 15 (2001) 189-209. Many other authors also suggest this kind of indeterminacy. An relatively early example is Measor, “Persons, Indeterminacy and Responsibility”, *Philosophical Review* 87(3) 1978 pp.414-422.

There are alternative reasons for suspicion of such cases – for example, even if one takes there to be facts of the matter, one might think the methodology of canvassing intuitions about hard cases is not a good way to find out what the facts are.

One might think that it is indeterminate *whether one would survive teleportation* without admitting that there are possible cases of indeterminate survival. This would be consistent if one could locate all the indeterminacy in vagueness in the implicit counterfactual. If the “hard case” is underdescribed, so that different elaborations give rise to divergent scenarios, which are each either determinate cases of survival or determinate cases of non-survival, then perhaps one could say that indeterminacy in what would happen in the underdescribed hard case is a matter of indeterminacy in which scenario is picked out by the initial, underdescribed hard case. (Cf. Cartwright 1993: “On two arguments for the indeterminacy of personal identity” *Synthese* 95 pp.231-273). But – on behalf of the sceptic at issue here – I will suppose that the “empty question” reaction would persist *even when the cases are fully described*.

⁴ One obvious response is to argue that the *degree* of material continuity in such cases will never be “sufficient” for personal identity in each of the cases. We cannot investigate this further without a definite proposal for what material continuity consists in. But the obvious proposal – that material continuity requires continuity of a sufficient *volume* of one’s parts – threatens to return counterintuitive verdicts in cases where a large volume of my parts are instantaneously vaporized,

hemisphere of their brain is destroyed---indeed, could survive even when the surviving hemisphere is implanted into a new host body. Consider then a case where the two hemispheres are separated, and implanted into distinct host bodies.

One might class such outlandish scenarios with the “hard cases”, above, and hold there is no determinate answer to questions of who the original person survives as. That of course, is to concede the point at issue here. If one gives up this *general* response to “hard cases”, a range of responses are on offer. One might try to defend the view, that *both* post-fission persons are me.⁵ Or one might try to defend the view that I would be *destroyed* by fission, and neither post-fission case is me. Or one might argue that there is a *brute, determinate fact of the matter* about which one is me – even though the two stand in the same intrinsic relations to my pre-fission self.⁶ None of the options is easy. An attractive answer – even for those who are prepared to argue for a determinate answer to questions of survival in other “hard cases” – is that there is no fact of the matter about who I survive as.⁷

Indeterminacy and the forensic concept of identity

Notice that very little in the above discussion traded on the thesis that personal identity was genuinely a matter of *identity*. Some theorists wish to distinguish between two readings of “x being the same person as y”. On one reading, this requires

without impairing my biological functioning. (Consider, though, Hudson’s proposal, wherein I am a minimal psychology-supporting part of my body).

⁵ Notice that this isn’t the same as defending the view that both the fission products are persons who existed on Earth before the fissioning event (cf. Lewis 1976 “Survival and Identity” reprinted with appendices in his *Philosophical Papers*, CUP 1983 pp.55-78). For even if two persons existed before the fission, one would have to argue in addition that *both* are me. And if “me” has a unique referent at a single time, this couldn’t happen. Thinking of “me” as indeterminate in reference between the two creatures that will later diverge is a natural option on the Lewisian view.

The “stage theory” of persistence defended by Ted Sider (e.g. *Four-dimensionalism* OUP 2002) and Katherine Hawley (*How Things Persist* OUP 2001),, in a version that allows for a single stage to have multiple temporal counterparts, does better at defending the multiple survival of a single pre-fission thing. Temporary identity theorists can also defend this line.

⁶ For a section of the philosophical community, the canonical response to the fission puzzles is Lewis’s *worm-theoretic* view. Lewis argues that both the fission products are stages of persons who pre-exist the fission--but that these two persons *share a common temporal part* until the fission event occurs. So we have, in one sense, “double survival” through the fission event. However, that does not yet address the question: which of the surviving persons is me? Semantically ascending: which of the multiple persons does “I” pick out? Answering “both” would require that singular terms be allowed multiple referents. Answering “neither” would leave it mysterious how “I am a person” could be made true. If both are rejected, then we are left with a choice between there being a brute fact over which of the two candidate persons “I” picks out – which is hardly in the Lewisian spirit – or it being indeterminate, a matter of semantic indecision, which it refers to. I take it that most Lewisians would go for this last option.

x and y to be *one and the same thing*, and for each to be persons. On a second reading, calling two things “the same person” just says they stand in a particularly intimate sort of relationship, but doesn’t require that x and y be identical. For example, Eric Olson⁸ argues at length that we are all human animals, with animalistic persistence conditions. Thus, the transplant case described earlier can be thought of as a case where psychology is switched between two continuant animals. Olson is unequivocally committed to the facts of *personal identity* matching these facts of *animalistic persistence*. But there is a role in Olson’s account for the relation among persons-at-times that the *psychological* theorist would think of as personal identity: psychological continuity and connectedness could be, for Olson, the proper locus of moral responsibility and similar relations of concern (compare Parfit on whether “identity is what matters”⁹). We might call the former the “metaphysical” concept of *same person as*, and the latter the “forensic” concept of *same person as*.

Notice that many of the cases for indeterminate personal identity apply equally to the forensic as to the metaphysical conception. If there is vagueness in what degree of continuity is “sufficient” for the forensic relation to hold, or if the concepts break down when faced with sci-fi cases, or if fission poses problems, we have indeterminate survival *in the forensic sense*.¹⁰

Indeterminate survival questioned.

Bernard Williams thought the idea of indeterminate survival problematic. He considered a putative situation in which one is sure that hurt is to be inflicted on a particular individual, who one knows to be a borderline case of being *oneself*:

To be told that a future situation is a borderline one for its being myself that is hurt, that it is conceptually undecidable whether it is me or not, is something which, it seems, I can do nothing with; because, in particular, it has no

⁸ *The Human Animal*, OUP, 1997.

⁹ “Personal Identity”, *Philosophical Review*, 80(1) 1971, pp.3-27; and *Reasons and Persons* OUP 1984 sec III *passim*.

¹⁰ The one case that might make a real difference is in the fission scenario (which, indeed, is Parfit’s primary case for distinguishing between the metaphysical and forensic concepts). Some of the pressure here comes from the thought that the *logic of identity* forbids my being identical to each of y and z, given they are distinct from each other. But there’s no logical tension to holding this pattern of relationships with respect to “same person as” in the forensic sense.

comprehensible representation in my expectations and the emotions that go with them. (Williams, B. 1970, p.58) ¹¹

Such worries are the central theme of this paper. I will argue that *given certain assumptions about how to think about indeterminate cases in general* Williams' concerns can be given a sharp formulation. The implication will be that *rationality forces us* to take a strange and alien attitude to people such that it is *indeterminate* whether they are identical to us.

The case centres in the first instance on particular sorts of *de se* care we have for ourselves. But the problem is not exclusive to the first-personal case. Often, concern *for another* has a principle role in our mental lives. How should we react to the prospect of their survival being indeterminate in one of the ways described above? Again, the argument will be that granted a certain conception of indeterminacy, the rationally required attitudes are strange and alienating.

I will be arguing for the untenability of indeterminate personal identity *under certain assumptions*. One may, therefore, regard the argument to follow as a reductio of those assumptions about indeterminacy, rather than a reductio of indeterminate personal identity. ¹²

2. Conceptions of indeterminacy

To this point, we have outlined a range of cases in which *personal identity* may be indeterminate. But what do we mean by this? There is a rich literature of different logical and semantic proposals for how vagueness and/or indeterminacy could be analyzed. I want to abstract from this logico-semantic detail (valuable though it is) and outline two broad *conceptions* of indeterminacy. The conceptions will be

¹¹Bernard Williams "The Self and the Future" *The Philosophical Review*, Vol. 79, No. 2 (Apr., 1970), pp. 161-180.

¹² I will not survey here Evans' notorious arguments against indeterminate identity in general. For one think, it contentious whether *personal identity* really has as its topic *identity*, as opposed to what Lewis calls the "I-relation" relating distinct objects (such moves are often favoured those who adopt an eternalist metaphysics of time). Even on a view that does allow for strict identity across time, it is not clear that the Evans argument is applicable. For one might describe it instead as cases where post-Episode, it is not determinate what individual *exists*—whether it the individual who pre-existed fission continues to exist, or whether a wholly new individual has come into being. But then an attempt to name the post-episode individual will at best be indeterminate in reference---and notoriously such contexts block moves in the Evans argument. For discussion, see (ANON).

individuated by the *cognitive role* they assign to indeterminacy judgements: *rejectionism* and *the ignorance view*. (I do not suppose these to exhaust the options – indeed, in an appendix I sketch other possible views. But these will serve to (a) delimit broad range of accounts that substantiate Williams’ worries; and (b) provide a contrast case which does not substantiate the worries).

Rejectionism. Suppose one asks whether some colour patch, out of sight to you, is *red*. Your informant might say that it is red, she might say that it is not red. Or she might say that there is *no fact of the matter* about whether or not it is red---it is simply a borderline case.

Here is an attractive description of how you should react. The response that has been given to you is *complete*. There’s no point pursuing the inquiry as to the redness or otherwise of the patch. Nor is there any point to wondering whether it is red or hoping that it is red *despite there being no fact of the matter as to its redness*. Nothing more needs to or can be said---the end of inquiry, as to whether or not the patch is red, has been reached.¹³

This kind of sketch would be explained by *rejectionism*:

Rejectionism (full-belief version)

It is (pro tanto) irrational to both believe that it is indeterminate whether p, and to believe that p.

¹³ See in particular Field (2001) and Field (2008) for an articulation and defence of this rejectionist conception. Many theorists hold that vague expressions are context sensitive---so that in one context we might rightly call patch x red, and in a different context we might rightly call the same patch not red. This is compatible with the view being put forward here. See Soames *Understanding Truth* ch 7 (OUP 1999). for an explicit combination of rejectionism and contextualism.

Context-sensitivity of “same person as”, in the context of personal identity, is connected to the idea that one might shift the facts about who is identical to who by “conventional stipulation” (cf. B Williams op cit). But matters are delicate. Take a context-sensitive expression, such as “that is flat”. Suppose we are in a context in which it’s true to say “the lawn is flat”. In a more demanding context, it might have been true to say “the lawn is not flat”. But, in the original undemanding context, it needn’t be ok to say “If we were in a more demanding context, the lawn would not have been flat”. The world-shifting counterfactual standardly isn’t thought to shift the index on which the extension of “flat” depends. On the other hand, one might adopt the view on which sensitivity to context behaved in the opposite fashion. “same person as” might just mean “same person as by the stipulated standards”. And now, counterfactuals concerning different stipulations might well shift the extension of this complex relation. [Mention Wright against verdict exclusion]

Rejectionism (partial-belief version)

It is (pro tanto) irrational to both have credence 1 that it is indeterminate whether p, and have non-zero credence in p.¹⁴

Rejectionism would explain and unifies the phenomena pointed to above. If you know p is indeterminate, you know exactly what level of credence to assign to p and to $\sim p$ -- zero in each case. Further inquiry as to whether p is thus pointless (unless it undermines one's confidence that p is indeterminate) --- one already knows that one's credences in p cannot be lifted, on pain of irrationality. Rejectionism also explains why it would be inappropriate to hope, fear, or feel anxious that the patch is red, given the information that this is an indeterminate matter -- or so I will be arguing shortly.

Rejectionism specifies a cognitive role for indeterminacy judgements. There are some in the literature who explicitly endorse it (Hartry Field for one). But many theories of indeterminacy do not explicitly address the question of what the cognitive role of indeterminacy is. It is a natural construal of the cognitive role that goes along with truth-value gap theories (where p is indeterminate iff "p" is neither true nor false) and excluded-middle rejecting theories (where acceptance that p is indeterminate requires rejection of "p or $\sim p$ "). In the first case, the relevant bridge principle would be that one should reject things known not to be true; in the latter case, the relevant bridge principle would be that rejecting a disjunction commits one to rejecting each disjunct.

That such widespread theories are naturally thought of as rejectionist is enough to make investigating rejectionism interesting. A stronger thesis would be to claim that the logico-semantic frameworks involved in such accounts not only *suggest* rejectionism, they *require* it. This would mean that if rejectionism were untenable, truth-value gap or excluded-middle rejecting theories would be untenable. I think a decent case for this stronger thesis can be made, by elaborating the kind of bridge principles mentioned above.

¹⁴ We will ultimately use a more general formulation of partial-belief rejectionism, but this will serve our purposes for the time being. Notice that modelling credences via classical probability theory is incompatible with rejectionism in this sense --- for according to classical probability theory, the probability of p and of $\sim p$ must sum to 1. There are non-classical "probability theories" that can be used to flesh out various forms of rejectionism.

To get a flavour of how this might go, consider *Truth Value Gap* theories. On such theories, to say p is indeterminate is to say that p is neither true nor false. Plausibly, truth norms belief: beliefs that are not true fail to achieve their aim. So if one knows p is indeterminate, then one knows that neither p nor $\sim p$ is true, and so one should not believe either. The truth norm for full belief has a gradational analogue: the instruction to minimize the difference between the truth value of p ($t(p)$, which equals 1 if p is true, and 0 otherwise) and one's credence in p . On the truth value gap conception, $t(p)=0$ and $t(\sim p)=0$, and so in each case one minimizes inaccuracy by having credence 0.¹⁵ Modulo the gradational norm, then, truth-value gap theories require rejectionism.

Another way of arguing for similar results is to look into connections between the logical revisionism characteristic of both truth value gap theories and LEM-rejecting theories, and argue on this basis for rejectionism. This I think presents the strongest case for the strong link between logico-semantic theories and rejectionism, and I pursue it in more detail in an appendix to this paper. There, I present in more detail a number of theories I think are plausibly rejectionist (including supervaluationism, non-classical truth-value gap accounts, and logically revisionary accounts where we reject instances of excluded middle, such as (certain) intuitionist views and the strong Kleene based systems recently advocated by Hartry Field). I also present arguments that they *are* in each case committed to rejectionism – based on the respective logics that they induce, rather than the truth-norm argument just sketched – and explore ways in which such theorists might argue against this conclusion.

To emphasize, though: these issues concern only how widespread the *impact* of the discussion below will be on logico-semantic frameworks for theorizing about indeterminacy. The discussion in the main text can proceed in terms of rejectionism *per se*, without taking a stance in this question.

The Ignorance View. Rejectionism gives one picture of the cognitive role of indeterminacy, and will be our primary focus in what follows. But it is useful to

¹⁵ For the gradational analogue, see Joyce “A non-pragmatic vindication of probabilism”, *Philosophy of Science*, vol. 65, 1998. See in particular Joyce's axiom of dominance, which effectively incorporates the above norm within his broader accuracy norm for credence functions in general.

provide a foil: an illustration of an *alternative* take on the cognitive role of indeterminacy judgements. Again, suppose that one asks of an unseen colour patch: Is it red? And one receives the answer: It's indeterminate whether the patch is red. On the uncertainty view, this conveys the responders *ignorance* of whether or not the patch is red.

In this case, it helps to get specific early on. The leading version of the ignorance view is *epistemicism* --- I will work with the classic defence of this view given by Williamson (1994). The epistemicist insists that there are facts of the matter in indeterminate cases--- a fact of the matter as to whether the colour patch is red, for example. The indeterminacy is rather a reflection of our (in-principle) ignorance of what those facts are. Epistemicism is, of course, deeply controversial – but we set aside its evaluation for another day.¹⁶ Thus the ignorance conveyed by saying “it's indeterminate whether that is red” is not mere happenstance --- it is (for Williamson) something we cannot hope to overcome.

Epistemicism nicely illustrates the non-rejectionist treatment of indeterminacy. The epistemicist thinks of indeterminacy as a kind of ignorance. And *uncertainty* as to whether p seems exactly the sort of attitude that is called for when one knows one is ignorant.

On the rejectionist conception, receiving the information that p is indeterminate “closes off” various questions relating to p. There is no rational space to wonder whether p or hope that p, nor to pursue further inquiry as to p. On the ignorance/epistemicist conception, I will argue that all these are possible.

Wondering and hoping seem clearly open. There is, after all, a fact of the matter as to the colour patch is red. Even if we can never know it, there's simply no reason to think we should not wonder about it, or take various attitudes to it. That *further inquiry* as to whether p is in principle intelligible might seem less clear. After all, on the ignorance view, we can't ever hope to achieve knowledge as to whether the colour

¹⁶ The classic defence is in Williamson *Vagueness* (Routledge, 1994), esp. chs. 7 and 8.

patch is red or not. If inquiry aimed at knowledge alone, then inquiry would be inappropriate.

But inquiry needn't aim at knowledge. If the question of whether p matters to you, then an inquiry whose result was to to *increase the evidential probability of p* , or allow you to increase your level of justified confidence in p , might be very much worth pursuing. And there's nothing in the epistemic view that rules out inquiry of this form into indeterminate cases.

We can illustrate this with the following example. For the epistemicist, there will be a fact of the matter concerning which number is the remainder, on division by 100, of the minimal wavelength of red light in nm. Let k name that number. Despite there being a fact of the matter as to what number it picks out, the red/non-red division is a paradigmatically vague one --- as good epistemicists, we can suppose that for all anyone can know, this number could pick out any one of the numbers from 1 to 100.

We will use this name to construct a sentence which is both indeterminate, and for which we have strong evidence of truth. A fair 100-ticket lottery is held. For any particular ticket n , the evidential probability that *ticket n is a loser* is 99/100---and this is also the appropriate level of confidence to invest in such propositions. But "ticket k " just is a name of one particular ticket, so no matter which one it names, our evidential probability for "ticket k is a loser" is 99/100 and the appropriate level of confidence in it is 99/100. But "ticket k is a loser" is indeterminate – for it is indeterminate which number " k " picks out, and so indeterminate whether it picks out the winning ticket (29, say).

This is a proof-of-principle that we can have quite definite evidence pertaining to questions even where these are known to be indeterminate, on the epistemic conception. Further inquiry into epistemically indeterminate matters makes perfect sense.

3. Indeterminate survival and rational action

Let us turn to the consequences of rejectionism for cases of indeterminate survival. What counts as a case of indeterminate survival will vary from theory to theory. Psychological theorists will think of cases of massive psychological change following some trauma; bodily identity theorists strange transplant cases. Readers are invited to tailor a case to their preferred theory with the following characteristics:

- (1) BEFORE (B) is a person, existing at t1.
- (2) AFTER (A) is a person, existing at t2, later than t1.
- (3) In B's future, and in A's past, is an Episode (E).
- (4) In virtue of the events in E, it is indeterminate whether B survives as A.
- (5) If B survives as anyone at t2, she survives as A.
- (6) B knows and is certain of (1-5).

Now consider B's attitude to A. Since she knows that it is indeterminate whether she survives as A, by *Rejectionism*, she is committed to assigning zero credence to:

I survive E

I am the same person as A

Likewise, suppose that A comes by a great fortune at t2. Before is committed to assigning zero credence to:

I will come by a great fortune at t2.

Of course, B also has credence zero in each of:

I fail to survive E

I am not the same person as A.

I will not come by a great fortune at t2

We now consider what actions are rational for B to take *from an entirely self-interested* point of view. That is, we are interested in the "prudential rationality" of various actions open to B.

Prudential action blanks

Suppose that B is offered the chance to buy investments, for a small sum now, which will pay out vast amounts anyone emerging out of Episode. Now, if B were altruistic, she might be happy to see this as a good thing to do even setting aside the issue of whether it benefits *she herself*. But because we're interested in *prudentially rational* action, this is irrelevant. We need to ask: considering the action only insofar as it benefits the agent herself, should she invest the money?

The answer to this for the *rejectionist* should I think be clear. What can B expect to gain from the transaction? She is sure of giving up the small amount (ten pounds, say) she now invests. Is this sure loss outweighed by expected returns? Not at all. For, since she's certain that it's indeterminate whether the beneficiary is the same person as her, she should have zero credence in benefiting from the money. And though a vast sum is involved, the expected returns from B's point of view are that vast sum (one million pounds, say) *multiplied* by B's credence that she will benefit---zero. So the expected returns are zero.¹⁷ The *net* return from taking the investment opportunity

¹⁷ This is the answer we get if we calculate by the formula given in Halpern, *Reasoning Under Uncertainty*, MIT Press, 2005 as defining the "lower expected utility" in the setting of Dempster-Shafer belief function theory, which as already noted is a generalization of classical probability theory well-suited to model "supervaluation-style" rejectionism. So the Dempster Shafer theory "lower expected utility" bears out the informal sketch given in the text.

There is another quantity discussed in the literature: the upper expected utility, a dual notion that is defined in terms of "upper beliefs", where one's degree of upper belief in p is defined as one minus the quantity we have been calling your credence in p. Dempster Shafer theory is usually deployed as a measure of *uncertainty*, with "upper" and "lower" beliefs being a measure of the extent of one's agnosticism over p. And so both upper and lower expected utilities may be argued to define a range of values *which for all one knows could be the expected value of one's action*. But in the present setting, there is no obvious understanding available of what the defined quantity is: one's upper belief in certain inconsistencies will be 1, for example. I doubt whether on the *indeterminacy* interpretation of the formalism of Dempster-Shafer theory, we can really make sense of a positive graded doxastic attitude to p measured by one minus one's credence in p. Unlike the agnosticism interpretation of the same formalism, I do not think we should take the quantities corresponding to "upper expected utilities" into account when figuring out which actions are rationalized by one's doxastic states.

[Note added 20/2/09. Following discussions with Ofra Magidor and Cian Dorr (among others) I'm no longer sure that this is the right way to think of things. The lower expected utility corresponds to the following pessimistic methodology: start with the a minimal value of utility, and then increase by increments, according to how likely you believe it to be that you'll obtain that incremental increase. The upper expected utility corresponds to the following optimistic methodology: start with a maximal value of utility, and the decrease by increments, according to how likely you think it is that you won't achieve that increment. These methods are equivalent when degrees of belief behave like classical probability functions, but they are inequivalent in the Dempster-Shafer setting, as are infinitely other mixed methods, starting with intermediate values and increases/decreasing appropriately. I now think that none of these methods is the right way to go, and that we should rethink the kind of decision theoretic machinery involved here. In particular, unlike standard Dempster-Shafer setting, I think we need to have a "mass utility function", which'll directly assign utilities e.g. to situations where something is indeterminate. The argument above will go through so long as the "self-interested" mass

is therefore a sure loss of 10 pounds. Contrast this with the net returns on not taking the investment opportunity – she loses nothing and gains nothing. The investment returns are zero.

The moral is clear: from a purely self-interested point of view, B should decline the investment opportunity. Given the case as we've set it up, this should hold no matter how small the cost to B, nor how large the prospective return. We can summarize the situation as follows. A is, for B, a “prudential action blank”. When calculating the expected returns in the above situation, from a self-interested point of view, A can for all practical purposes be treated like someone determinately distinct from B.

4. Indeterminate survival and rational emotions.

We've just argued that, for B, A is a “prudential action blank”. I shall now argue that, on the same assumptions, A should be a “de se emotion blank” for B.

We focus on emotions with intentional content---in particular, those that can have *de se* intentional content. A sample includes:

X is happy that she is sailing.

X is sad that she is sick.

X is angry that she slept in.

X is disgusted that she forgot about her appointment.

X fears that she be in pain.

X hopes that she will get some money.

X worries that she won't be able to pay the rent.

X wishes that she could avoid harm

The content of each of these emotional states is specified in *de se* terms---the content concerns *the very individual* who is having the emotion, as such.

utility function attached to a situation where it is indeterminate whether I gain any benefit, is zero. The grounds for this would be that I have zero credence that I benefit in those situations. But the issues here are rather complex! I'd be very glad to talk about this further.]

Some preliminary observations. First, certain of these intentional emotional attributions display a distinctive feature. For it to be true that Ted is happy that p, it must be that p itself is true---that is, “happiness that” appears to be a *factive* verb (compare: knowing that p). If Ted is in a state indistinguishable for him from his being happy that p, but p is false, we typically resort to some circumlocution: for example, we say that Ted is happy *because* he believes/thinks (falsely) that p. Second, as well as the use of emotive verbs in intentional attributions as above, there is a second, *intransitive* use of the verbs. Sometimes we simply say that a person is happy, fearful, angry, worried, without specifying something that the person is happy/fearful/angry/worried *about*. Relatedly, we say that a person *feels* angry, fearful etc. An overall account of these states should presumably deal with both the intentional ascriptions and intransitive cases. But it is the former that will primarily occupy us here, since it is these that are most naturally connected to central propositional attitudes such as belief and desire.

What sort of connections between belief and desire and intentional emotions might there be? Here is a first pass at characterizing common elements to those we have listed: each seems to involve some kind of valence, or preference between alternatives (being happy, hoping, or wishing that p involves preferring p to relevant alternatives; being sad, worrying or regretting that p involves preferring alternatives over p). Some may be complex, requiring the presence of related beliefs (so, for example, disappointment that p pretty clearly involves both a prior preference for some alternative over p, but also a realization that p). And of course, each has associated with it a distinctive phenomenological and physiological profile (which we might seek to capture, in central cases, by notions such as: *feeling happy, feeling sad, feeling hope, etc*).

Let us concentrate on the relations between beliefs and intentional emotional ascriptions. Set aside for a moment problem cases where one might suggest that it is proper to ascribe an intentional emotional state without relevant accompanying beliefs (we turn to this issue shortly). Even focusing on “garden variety” cases, we can distinguish certain classes. As representatives, let us contrast being happy that p, hoping that p and wishing that p (to keep things simple, we shall only be considering cases where p concerns past events).

Suppose one believes that the train was on time. This is something one can be happy about. But – unless something strange is going on – one cannot rationally believe that the train was on time and simultaneously hope that it was on time. On the other hand, suppose one believes that the train was late. Then one could wish that it had been on time; but one cannot be happy that it was on time (!) or even hope that it was on time. Finally, suppose that one thinks that the train might have been on time, or it might not. Then – without strangeness – one cannot be happy that the train was on time, nor wish that it had been; but one can hope. Very roughly, then, “happiness that p” requires belief that p; “wishing that p” requires belief that $\sim p$. And “hoping that p” requires belief that each of p and $\sim p$ is possible.¹⁸ Let’s use “belief-requiring”, “opposite belief-requiring” and “possibility-requiring” as classifications for emotion-types that---at least in garden-variety cases---display the respective relationships to belief.¹⁹

But do beliefs need to be in play in the first place? One source of worry about this claim, however, comes from our emotional responses to fiction.²⁰ Suppose I watch a horror film. I might feel afraid that Jason will kill Ted, or happy that Ted escapes, and so forth. It seems that imagining that Jason might kill Ted, or that Ted has escaped, can underpin the ascription of emotional states in the way that (in “garden variety” cases) belief in these contents do.

What seems unarguable is that many of the affective responses characteristic of fear are present in these cases. I certainly wouldn’t argue that we can “feel fear” in this scenario. But arguably, the full functional role of fearing that someone might be killed

¹⁸A few pointers to complications. First, the emotive verb “happy”, as well as being belief-entailing, seems to be factive---someone’s being happy that p entails that p. Second, one might try to subsume factivity and requirement for belief under a single heading---as the requirement for knowledge (cf. Gordon, *The Structure of Emotions*, CUP, 1987). Third, there are various options in formulating the condition on “hope”. Obviously the current formulation requires that the reader charitably interpret the appeal to possibilities – requiring that the subject believe that p be logically possible would be far too weak! The relevant sense is one of epistemic or doxastic possibility. It might be better, then, to reformulate this directly in terms of degrees of belief: hoping that p requiring non-zero degree of belief in each of p and of $\sim p$. That seems ok so far as it goes, though it may need strengthening to capture the full force. Finally, I’ve been talking about entailment among attitudes “so long as nothing surprising is going on”. One might want to strengthen this to claim that the emotional attitudes entail the various doxastic attitudes mentioned; or rather formulate the constraints in terms of rationality.

¹⁹ Cf. Gordon (op cit) on “Factive” vs. “Epistemic” emotions.

²⁰ For relevant discussion, see Meskin and Weinberg ‘Emotions, Fiction, and Cognitive Architecture,’ *British Journal of Aesthetics*, 43 (2003), 18–34 and the references therein.

is not---I am not spurred to action in any way, for example. Concentrating on the emotional ascriptions, is someone who feels afraid watching a horror film truly *fearing* that Jason will kill Ted? The form of words is natural to use. But this kind of evidence is inconclusive: in the same context it's also natural to say that the viewer *believes* that Jason might kill Ted (or even that he *knows* that this is a possibility). I don't think this should lead us to question standard accounts of belief or knowledge – in part precisely because the imaginative engagement doesn't have the connection to action characteristic of true belief-states. What seems to be happening in the latter case is that there are analogues, within imaginative engagement with fiction, of ordinary belief and knowledge – and that somehow we use the words for the paradigmatic state for the fictional analogue. One might well extend this model to reports on emotional engagement with fiction, as again, typical connections to action associated with paradigm cases of fearing that p are missing. We don't believe that Jason might *really* kill Ted, and so we don't fear that Jason will *really* kill Ted. I shall assume this is the right response to this case in the formulations to follow.

Another challenge to the concentration on belief states comes from cases where affective responses are prompted by mental states *more primitive than* belief. Thus, perhaps *perceiving* that a snake is (or might be) nearby might prompt a fear-reaction, without the perception having first to generate a perceptual belief with that content. One response is to parallel the suggested move in the case of fiction: the intentional state associated with the fear-affect here isn't really a fear that a snake is nearby. But this defensive reaction seems less persuasive than in the fiction case, since here the connections to action, as well as the fear-affect, do seem to be present.²¹

I am inclined to think that the moral of these cases is that the sense in which an intentional emotion state “requires” belief isn't as a matter of necessity or a *cause*. Rather, the requirement is one of rationality, and the requirement should be that *so long as the agent has an opinion at all on the matter*, the agent will be irrational if she is happy that p without believing p, fears that p without believing it might be that p, and so on. Whatever the etiology, it does seem plausible that it one who continued to

²¹ Challenges from these kind of cases are discussed in Griffiths *What Emotions Really Are* (University of Chicago Press, 1997), ch. 2.

fear that a snake was present *and* took the view that there was no possibility that a snake was present, would be in an irrational state.

Let me be clear what I am not assuming. I am not appealing to connections between, say, fearing that a dog will bite me and the belief that dogs are dangerous – rather, I am appealing to a connection between fearing that that dog will bite me and a belief that the dog might bite me (in combination with relevant preferences against the dog biting me). I am not assuming that individual emotions can be *defined* in terms of belief and desires, independently of affective, physiological and phenomenological states. I have no ambition to “reduce” the phenomenological aspect of feeling emotions to belief and desire. Indeed, I’m not even saying that beliefs and desires feature in a specification of the causal-functional role of emotions like fear – the case of perceptually-based fears may provide counterexamples to this.²²

What I am assuming is that, for the reflective agent, there are normative constraints between what one fears, is happy about, regrets and so on, and doxastic states with matching (or closely connected) content.

Emotions given known indeterminate survival

Presupposing this kind of normative connection between emotions and beliefs, we turn again to the case of indeterminate survival. It certainly follows from rejectionism that B shouldn’t have any *belief-requiring* de se emotional attitudes on the basis of A’s predicament. All the following are ruled out:

B being happy that she will get money.

B being sad that she will feel pain.

B being angry that she’ll be so different after Episode.

²² For critical discussion of many “cognitive” theories of emotions that do require these theses, see Griffiths (op cit, ch.2). See particularly the list of six types of objections to cognitive accounts on pp.28-30. I don’t believe that the theses defended here are directly susceptible to the types of objection Griffiths surveys. In one sense, this is unsurprising, since I don’t pretend to be offering a *theory* of emotions here – and so the constraint that I rely on is compatible with all sorts of different approaches to emotions themselves.

All these belief-requiring emotions are inappropriate, given that A positively disbelieves the propositional content associated with each one (and so, a fortiori, neither believes nor knows it to be the case).

But this is no real news. We could say much the same thing if Episode were replaced by Episode*, which rather than producing (with probability 1) the indeterminate-survivor B, produces someone who is *determinately* A (with probability ½) or determinately not A (with probability ½). Given the *uncertainty* in this case, the belief-requiring *de se* emotions will again be unavailable to the rational agent.

But the case of possibility-requiring emotions is distinctive. Consider the following, which seem to fall in this class:

B fearing that she will feel pain

B being anxious about whether she will feel pain

B hoping that she will get a large payout

These kind of emotions *would* be available in the Episode* variant on our story, where B is genuinely *uncertain* whether she will survive. Given rejectionism, are such *de se* emotions available to B?

I say not. To take p to be a possibility in the relevant sense is incompatible with utterly rejecting p. More specifically, I endorse the following:²³

The basic constraint on possibility-requiring emotions

It is (pro tanto) irrational to both fear that p, and to utterly disbelieve p --- i.e. have credence zero in p.

Combining the basic constraint with rejectionism, we have a consequence prefigured in our initial discussion. Saying that there was “no fact of the matter” about p was to “close off” certain issues: including hoping whether p. It is only once we have the

²³ Those worried about the fiction cases discussed earlier might prefer to formulate this in terms of the fear that *really*, p, or “real-fear that p”, rather than the simple formulation. As mentioned above, I am supposing that such qualifications are unnecessary.

basic constraint on possibility-requiring emotions in place that we have reached the stage when we can account for why it seems silly to hope that a certain patch is red, given you're sure there's no fact of the matter as to whether or not it's red. The basic constraint seems a natural part of the rejectionist picture.

In the particular case at hand, and granted the basic constraint, indeterminacy will make fearing post-Episode pain *irrational* for B --- and likewise for every other possibility-requiring emotion with similar de se content. So A is not only a prudential action blank for B, but also a rational de se emotion blank, for the whole range of belief and possibility-requiring intentional emotions – happiness, sadness, relief, despair, fear, hope, anxiety, panic, etc.²⁴

Emotional blocks

It is tempting to think that as far as B is concerned, A might just as well be someone else, determinately distinct from B. After all, B feels no prudential concern for A, cannot rationally feel afraid upon learning that A will undergo severe pain (though of course, B might feel sympathy for A as for a friend or relative in a similar predicament).

If this were the right response, then it would be natural to say that for the rejectionist *indeterminacy is as good as death*, as far as what matters in survival goes.

However, it isn't quite true that A is "just like everyone else" for B, as far as de se emotions are concerned. It's true that B should be (de se) emotionally and prudentially unconcerned for A. But the indeterminacy over whether B survives as A does show up in B's emotional profile in a distinctive way.

Suppose B believed she would not survive Episode. A natural response would be for B to wish that she could survive; or to feel despair that she won't survive. She might rationally feel relieved that her suffering will soon be over; or angry that she won't be

around to see her projects through. A rich range of emotional responses focusing on her *non-survival* would be open to her.

In the case where B knows she (determinately) won't survive two things become available: *opposite belief requiring emotions* concerning her survival (e.g. wishing that she would survive) and *belief requiring emotions* concerning her lack of survival (e.g. despair, relief, anger). But in the case with which we are concerned, such emotion reactions are rationally *blocked*, since B should have *zero* credence that she won't survive.

A is an emotional blank for B in virtue of B's commitment not to self-identify with A. But A is also an emotional *block* for B in virtue of B's commitment not to self-*disidentify* with A. All sorts of de se emotions that B might naturally feel about her predicament if A wasn't around or replaced by someone determinately distinct from B are unavailable.²⁵

5. Inevitable irrationality?

B's attitude to A's future harm is a strange and alien one. On the one hand, there is no shade of the concern one would expect to find associated with identity. On the other hand, one cannot conceptualize the situation on the model of determinate non-survival---for de se emotions characteristic of *non-survival* also fail to be present. Williams' comment that indeterminate survival "has no comprehensible representation in my expectations and the emotions that go with them" seems to capture the case nicely.

²⁵ The case just made focuses on self-interestedly rational action, and de se emotions. These make the case particularly vivid. But our concern for the survival of persons goes beyond our own case. Suppose that C is someone who cares for B – who is, perhaps, prepared to give B's interests great weight when deliberating on what to do. C would fear for B if he suspected that B would suffer some harm; be relieved if B were to avoid harm; be happy if B were to receive some great reward. What should C think in the scenario earlier, where (C knows that) it is indeterminate whether B survives Episode? It would be irrational for C to be anxious that B will undergo great harm, since C's credence in this proposition is zero. But C cannot feel relief that B will not undergo great harm – for C also totally disbelieves this. And, if all C were concerned about were B's benefit, it would still not be worth investing tiny amounts for the great future benefit of A---since the expected value of this *for B* is zero. (C maximizes value for B better by spending the small amount buying her penny sweets). So self-concern is not the only way in which indeterminate survival gives rise to the alienating pattern of attitudes: de re concern for another gives rise to an exactly similar puzzle.

Williams writes (op cit):

In the present case, fear seems neither appropriate, nor inappropriate, nor appropriately equivocal.

There is something appealing about this statement. Fearing future harm *should be a problem for B*. But given a rejectionist treatment of indeterminacy, we have argued that it should, definitively, be *inappropriate*.

Can we go further, and argue that the attitudes indeterminate survival prompts are not only alien, but also, irrational? Williams appears to think so:

Central to the expectation of S is the thought of what it will be like when it happens – thought which may be indeterminate, range over alternatives, and so forth. When S involves me, there can be the possibility of a special form of thought: the imaginative projection of myself as a participant in S.

.... the subject has an incurable difficulty in how he may think about S. If he engages in projective imaginative thinking (about how it will be for he himself) then he implicitly answers the necessarily unanswerable question. *If he thinks that he cannot engage in such thinking, it looks very much like he also answers it, though in the opposite direction*. Perhaps he must just refrain from thinking; but is he just refraining from it, if it is incurably undecidable whether he can or cannot engage in it?

(Williams, op cit. p.61; italics added)

This is directly relevant to the scenario we have found ourselves in. For we have argued that fear is *inappropriate* – and the reflective and sufficiently informed agent will know that this is so. So “merely” refraining is not an option: B can know that he cannot (or anyway, should not) engage in the kind of distinctive imaginative thinking under discussion. Williams says that in this case “it looks very much like he answers (the necessarily unanswerable question) though in the opposite direction” i.e., that this implicitly commits B to thinking that he will *not* feature in situation S – the future harming of A.

But why should believing that *de se imaginative thinking is inappropriate* commit one to this? One might argue that if all the characteristic *de se* attitudes are absent, then *for all practical purposes* it is as if one does not survive. But we have already seen that the rejectionist conception does not lead to this conclusion. For the indeterminate identity *blocks* various emotional attitudes that would be characteristic of non-identity. So the claim that the bar on *de se* imaginative thinking is the same “for all practical purposes” as non-identity simply doesn’t hold true, at least in that setting.

Are there other ways of arguing the situation is not only strange, but incoherent? A psychological theorist of personal identity might have particular problems here. For if it would be irrational for B to believe that she will survive as A, it would be irrational for her to intend to perform various actions as A. And likewise, since it would be irrational for A to regard herself as having a prior life as B, she should not rationally endorse apparent memories of actions performed by B. But the connection between intentions and future actions, and memories and past experience, is one of the mainstays of psychological treatments of personal identity. The suggestion might then be that any psychological connections that do tie A to B will be infected with irrationality, and so *should not* be present. This is not to say that they *are not* present: but for a certain kind of theorist it might undermine the assumption that A and B can be fully rational *and* stand in the relations required for there to be a case of indeterminate identity. Such thoughts are speculative, however, and tied to a particular, contentious characterisation of persistence conditions, so I do not attach too much weight to them.

For completeness, I cover one final thought that would turn the case from mere strangeness into unsustainability. The key idea here will be that – once one is appropriately informed – *failing* to have certain emotional responses can be just as criticisable as having inappropriate emotional responses. Suppose that one knows that one has an exam tomorrow, on which a great deal hangs. Suppose one is unsure about whether one will pass. It would seem (pro tanto) wrong *not* to be anxious about it.

Call the follow the *contentious constraint on possibility-requiring emotions* (here formulated for the special case of anxiety).

Suppose one is actively considering the possibility that great harm will befall someone. Then *failing to be anxious* that one will suffer great harm is (pro tanto) irrational unless one has sufficient confidence that one will not suffer great harm.

Granted this kind of constraint on lack of emotions, B is in a bind, when she actively contemplates A's future harm. B is rationally committed to having zero credence that she will not suffer great harm, and so, by the above principle she will be (pro tanto) irrational if she fails to be anxious. However, by the earlier "basic constraint" on anxiety itself, we would be able to conclude that if B feels anxiety about the future pain, she will be (pro tanto) irrational.

Supposing both the basic and the contentious constraints, it is tempting to conclude that B faces inevitable irrationality here. For either she is anxious about the future pain or she isn't. If she is anxious about this, she is (pro tanto) irrational by the basic constraint. If she fails to be not anxious about this, she is (pro tanto) irrational by the contentious constraint. So (arguing by cases) she will be (pro tanto) irrational whatever happens.²⁶

If we endorsed the contentious constraint, we'd have a case for inevitable irrationality. But I don't think this will be particularly worrying for the rejectionist, since I don't see independent grounds for thinking that they should *endorse* the contentious constraint.

6. Evaluation

²⁶ To argue in this way presupposes that excluded middle holds for fearing---that no matter what state A is in, she is either anxious about the harm, or she isn't. But that can't be assumed without argument in this context. For perhaps A could put herself in a state which is indeterminate between anxiety and non-anxiety, and *under certain conceptions of indeterminacy* that would block the appeal to an instance of excluded middle ("either she fears or she doesn't") which is required to argue for inevitable irrationality.

But not all rejectionist theorists of indeterminacy allow for failures of excluded middle. And even for those who can allow for failures of excluded middle, and argue that indeterminate emotional states are the called-for response, there is still the problem of prudential action blanks. Indeed, the above might even make such puzzles worse. For consider: we are now supposing that the right stance for A to adopt to B's future harm involves *not determinately not fearing* that she herself will be harmed. But can A coherently adopt that stance, while still not being willing (on a prudential basis) to invest small amounts for the benefit of the one who suffers that harm?

If one denies the contentious constraint just discussed, then no case has been made for *inevitable* irrationality in the indeterminate survival scenario – even granted rejectionism. Yet still, the state into which we are forced is an alien one. What sort of state is it, in which we are neither anxious about ourselves suffering a known future pain, nor able (rationally) to reassure ourselves and feel relief that we won't be in pain? The obvious model is a circumstance in which we *ignore the question* of whether we will suffer the pain we know is in the offing, so that the question of anxiety or relief never arises. But the present case is exactly supposed *not* to be like this: B has full knowledge of all the relevant facts, and *still* is allowed no positive de se emotional attitude towards the event.

First reaction: follow the argument where it leads.

The emotional and prudential blank/block reaction is not, I suspect, how we do or *would* react to cases of indeterminate survival. However, the rejectionist may embrace the above results – perhaps as *normative* rather than *descriptive* of our attitudes to cases of vague survival. Harold Noonan writes, of the predicament of one such as B:

Why should such an [appropriate emotional] response be possible? Surely the correct reply to Williams here is to question the assumption that one must be able to produce an appropriate emotional response to any genuinely possible situation which one can foresee. For if the situation is of a type one has never before encountered there is no reason why this should be so. If what I wish for is that *P*, but what is going to happen will make it neither the case that *P*, nor the case that not-*P*, then relative to my desire that *P* neither elation nor disappointment is an appropriate response to what is going to happen... what to think about the situation will simply be an irresolvable problem for me. But that cannot show that situations of this borderline type are impossible.²⁷

Noonan suggests that we refrain from imaginatively projecting in cases of the relevant kind; and that this is what we are beholden to do (“neither elation nor disappointment is an appropriate response...”). He challenges others to say what goes wrong with the description of the case, which, as we have seen, rejectionism enforces.

²⁷ In his *Personal Identity* Routledge, 2nd edition.

Unless further principles are defended, such as the “contentious constraint” mentioned earlier, it seems to me that Noonan’s suggestion is consistent line to adopt. But consistent responses to a puzzle need not be satisfying. In particular it may not be dialectically stable. For it leaves unexplained the puzzling feature of the cases. Methodologically, we should look for broader accounts that remove, explain or make intelligible the puzzling feature of these cases – for *ceteris paribus* these will provide more satisfying overall theories. The bullet-biting response is available, but should be no more than a fall-back option. The next two responses we consider are more ambitious.

Second reaction: Revisionism.

The second reaction to the puzzling cases is to see them as symptomatic of a more general kind malaise in self-concern. Rather than adopting a revisionism of self-concern in the *particular* case of indeterminate self-identity, this response will advocate a *general* revisionism about self-concern. By putting the particular case in the more general setting, the hope will be to avoid charges of *ad hocery* that can be laid against the earlier view.

Note that our results target in the first instance *prudentially* rational action (in the sense of strictly self-interested action), and *de se* emotions. As we have seen, parallel cases can focus on the notion of action targeted at maximizing value for another individual, and the corresponding *de re* emotional states. It is beyond doubt, I think, that something very much like these evaluations and attitudes have a hugely significant psychological role. For example, the *phenomenology* of *de se* emotional states (e.g. fear for pain in one’s own future) is – for the great majority of us at least – distinctively different from the phenomenology of *de re* fear for pain in others’ future. And again, the fear *de re* of future harm to a loved one is different from *de dicto* fear that *someone* with thus-and-such characteristics might suffer harm. In connection with action, conflicts between altruism and self-interest are an everyday occurrence; and acting to the benefit of another – to the detriment to one’s own or the general good – marks an important category of action.

There are more or less radical lines the rejectionist could take to argue that states of *de se* and *de re* care should not be accorded the role they appear to have in our mental lives:

- (1) [Less radical] What plays (or should play) that privileged psychological role is not *self*-interest, or *de se* emotions, but something non-self-involving, which are not compromised by vague cases of personal identity.
- (2) [More radical] There are no “natural” distinctions to be drawn between pursuing self-interest and pursuing the interests of others; or between *de se* emotions and *de dicto* emotions about what happens to others. Revising our mental economy to get rid of “self-interest” and “*de se* emotions” would not leave any *natural* distinctions unrecognized.

Proposal (1) leads to skirmishing that does not advance the rejectionists’ cause at all. In effect, it switches attention from the *metaphysical* to a purported *forensic* conception of personal identity (call it relation R). Rather than focus on “strict” *de se* emotions and prudentially rational action – what benefits *myself* in the future, or attitudes directed towards *my* being such-and-such – it offers a replacement: what benefits *someone R-related to me* or anxiety/happiness that someone R-related to me will be such-and-such. Following Olson, relation R might be a matter of psychological continuity (or perhaps, a sufficient degree of connectedness), rather than the animalistic continuity that allegedly underlies real identity over time.

The idea would be that caring about R plays the role, emotionally and prudentially, that one would initially think is played by care for oneself. But as we noted at the beginning of this essay, a forensic two-place relation R, analyzed in terms familiar from the general personal identity literature, will be just as susceptible to cases of indeterminacy as “real” identity over time. So option [1] above simply invites us to replay the story again, now with B and A being indeterminately R-related. Let us therefore set this aside.

Proposal (2) is more promising. It builds on Parfit’s insistence that *what we care about in survival* comes in *degrees*. (For Parfit, the relation R will be one of psychological continuity *and degree of connectedness* – where the measure reflects

the number and strengths of causal connections between the relata.) On this picture, we should not be talking about whether or not two individuals are R-related, but rather *to what degree* they are R-related. And the natural description of the cases we've been focussing on from this perspective is that they are cases where R obtains between A and B *to a lesser degree than normal*. Thus, the best description of the case would be that B's investing money for A's future benefit is "R1 rational"--- where R1 is the relation of standing R to at least some low degree – and "R2 irrational" – where R2 is the relation of standing in R to at least some fairly high degree. Paralleling these evaluations of action, one might distinguish different kinds of contents of emotion: anxiety that someone R1-related to oneself will suffer harm; anxiety that someone R2-related to oneself will suffer harm, etc.

This picture gives us a fine-grained set of distinctions. But we haven't located within this set of distinctions something that correlates with those central features of our mental life---*self interest* and *de se emotion* (and indeed, *interest for another* and *de re emotional concern*).

Of course, we *could* rehabilitate this notion. We could say that "R" in an all-or-nothing sense holds between x and y when they are *sufficiently strongly R-related*.²⁸ And R would then be a candidate for explicating a "forensic" notion of "same person as" of the kind discussed earlier. We would then retain the letter of our original puzzle---B shouldn't have *de se* (or R) emotions directed on the events of A's life, nor should he act, for self- (or R-) interest, in ways that benefit A greatly, for marginal initial cost to B.

The puzzle can be stated. But is it still puzzling? Imagine that we simply dropped the all-or-nothing "same person as" or "R-related" concepts from both thought and talk. Parfit's idea, I take it, is that *we would lose nothing of importance* ---there are no natural joints to be carved by such notions. So by Parfittian lights there's no reason to pay attention to *prudentially* rational action or have particular phenomenological accompaniments to *anxiety for one's own harm*. Such contents or evaluations just fix

²⁸ Perhaps a "no-branching" condition should be added---though I'm personally convinced by Lewis's contention that once we distinguish between identity as such and the "I-relation" that interrelates stages of a single person, then apparent problems of non-Euclideanness dissolve.

on a rather arbitrary joint. If one accords them a special role in determining, all-in, what one should do or feel, then one would be left in the original puzzling situation. But the Parfit suggestion, I take it, is that such notions should have no *special* importance. The puzzles introduced above are just one more illustration of why all-or-nothing concepts like *personal identity* are not the proper basis for systematic thinking about *what one should do*.

In short, the Parfittian idea I am canvassing here is stronger than the contention that *personal identity* (strictly construed) is not what matters---it is that there is no replacement “forensic” relation in the vicinity that distinctively matters. And the thought is that once we describe the situation in the fundamental, underlying terms – of degrees of R-relatedness – there’s no puzzle left to address.

The Parfittian account is an intriguing proposal for “dissolving” our puzzles. It provides a principled explanation of the source of our troubles: far more satisfying than simply following the argument blindly where it leads. But – as Parfit himself is concerned to emphasize – it is far from a low-cost option. The visceral phenomenology of fear of pain in *one’s own* future and the pull of *selfish interests* – or the corresponding fear of harm for *that person one loves* or actions that benefit him or her – on the current proposal, such attitudes are generating an illusion that there is an important distinction, not a matter of degree, between “oneself” or “that person” and “others”.

Much of our standard self-conception is at stake here. Some might simply reject the revisionary picture out of hand. Even those of us with no particular axe to grind should be cautious of advocating too revisionary a picture. Endorsing that picture is *risky* – who knows what elements of our standard thinking rely in unobvious ways on the usual platitudes. That will not discourage the bold from advocating Parfittianism; but those of us more methodologically conservative are still in search of a less revisionary response.

Third response: Rejecting rejectionism

The final way of responding to the puzzle cases is to reject the presupposition on which the case was based. This was the idea that personal identity could be indeterminate *under a rejectionist conception of indeterminacy*.

To see the crucial role this plays, suppose, contra the rejectionist, that the appropriate attitude to borderline cases is one of *uncertainty* – exactly the same kind of attitude that we adopt, for example, towards the proposition that a fair coin has landed heads, when it lands out of sight.²⁹ The case for the prudential irrationality of B’s investing money for A’s future benefit would collapse – for if B is *uncertain* whether she will be A, and so gain the benefit, she can still give *some* weight to the possibility when calculating her expected returns from the investment. Likewise, if B’s attitude is one of uncertainty, E-type situations will be relevantly like E*-situations: uncertainty on whether she will suffer harm will be exactly a situation in which fear, anxiety, hope and the rest will be appropriate.

So the case for *prima facie* tension here does rest squarely on something like the rejectionist conception of indeterminacy. The question becomes: are alternatives available? And what would they be like?

We have already canvassed one alternative: that of the *epistemicist*, who thinks of indeterminacy as a kind of ignorance. Shifting from a rejectionist to an ignorance-based conception of indeterminacy can have important practical consequences for debates on personal identity. Consider those who argue that “science fiction” cases of personal identity are ones where the question of survival is *indeterminate*. Such a view is often offered as a diagnosis of why a literature purporting to support one side or another is *misguided* --- the thought is that further debate on the question, once it is acknowledged that the questions are indeterminate, would be silly. This is exactly the right thing to say about such cases if rejectionism were right. But if the ignorance view is correct, there is no such inquiry-ending moral to be had. Inquiry could sensibly continue, seeking to uncover evidence raising the likelihood that the

²⁹ It is not uncontroversial that uncertainty should be compared to the fair-coin case, though it is a common model. Another model would be: our attitude to the proposition that the number of magpies in region R is k, where k is some plausible-sounding number and k some precisely delimited region. In the latter case, we don’t have any obvious guidance as to the objective probability of the proposition being true.

admittedly indeterminate proposition is true. Shifting conceptions of indeterminacy can thus change our conception of how debates about personal identity should be conducted.

The ultimate suggestion of this paper, therefore, is that rather than see Williams' puzzles about indeterminate identity as providing challenges to theories of personal identity that allow for indeterminate cases, we should see them as providing data that constrain one's wider account of what indeterminacy is. Viewed in this light, the moral of Bernard Williams' puzzles is that rejectionism leads to difficulties. Indeterminate cases of personal identity, and our thinking about them, provide potentially crucial leverage on accounts of indeterminacy in general.³⁰

³⁰ Among the important non-rejectionist, non-epistemicist conceptions of indeterminacy is the *degree theoretic* conception.

I am assuming that the degree theorist will say that rational credences should conform to known degree-of-truth (or determinacy) of a given proposition. Once one knows the degree of truth, therefore, one knows exactly how to apportion credences, and further inquiry would be futile. Thus saying that "B survives as A" is half-true would be to settle matters. On the other hand, there's no obstacle from the basic constraints on emotions to hoping, being anxious etc about some knowably indeterminate p, since one assigns an intermediate degree of belief to p.

(Here, though, the distinction between the initial gloss on possibility-requiring attitudes and the "basic constraint" becomes important. The basic constraint assumes that it suffices to assign "I will be harmed" a significant degree of belief. But given the degree theory, one can do this while *giving zero weight to every possibility where it is more than 0.5-true that I will be harmed*. Given the latter, it is not clear one *does* believe that there is a possibility that one will be harmed; and so not obvious that the possibility-requiring emotions are appropriate.)

A different sort of account of indeterminacy is defended in (ANON)

Appendix: Rejectionist theories of indeterminacy.

One of the most distinctive reactions people have to borderline cases is to reject principles that seem to presuppose that the case must fall into one of two polar options. So for example, presented with a borderline red/orange patch, and asked whether the patch is *red* or whether it is rather *orange*, many feel that they are being presented with a false dichotomy. It's tempting to respond "it's neither---its borderline red/orange".

What is this "neither" doing? At it's most literal, we can interpret it is saying that the patch is neither red nor orange. It follows from this relatively uncontroversially from this that:

The patch is not red, and the patch is not orange.

Now consider the same reaction where the issue is not the red/orange borderline but the red/non-red borderline. The same reaction is tempting. In this case we get:

The patch is not red, and it's not the case that the patch is non-red.

But this apparently has the form $\sim P \& \sim \sim P$ ---an explicit contradiction.³¹

It's hard to believe that people mean to commit themselves to contradictions in the presence of borderline cases of vague predicates. Given the distinction we drew earlier between *accepting the negation of p* and *rejecting p*, a charitable reconstruction suggests itself. What is being conveyed by the response above is the *rejection* of the following:

Either the patch is red, or it is non-red.

³¹ One potential point of resistance is the assumption that the "predicate negation" in "non-red" should be formalized by the sentential negation "it is not the case that" of standard first order regimentations.

If we endorsed the negation of this sentence, then we would have the problems just described (in most logics). But if we're simply *rejecting* it, then the problematic reasoning cannot get started.

But what is being rejected here is (more or less) an instance of a classical tautology. To reject it is ipso facto to reject the classical setting which would enforce it. Thus, those who take the rejection of the above seriously are challenged to provide a *revised* account of logic.

Many different accounts are available---of various strengths. Two of the most famous are *Intuitionism* and the *Lukasiewicz* logics. They differ in what principles they admit. For example, we have noted from the denial of the above $\sim(R\vee\sim R)$, we get an explicit contradiction. Thus *reductio* would tell us we can negate this assumption---to get $\sim\sim(R\vee\sim R)$. The intuitionist accepts this reasoning, but holds that one cannot eliminate the double negation---hence accepting $\sim\sim(R\vee\sim R)$ is compatible with rejecting $R\vee\sim R$. Lukasiewicz settings allow double negation elimination, but disallow *reductio*—despite the fact that a contradiction follows from $\sim(R\vee\sim R)$, we reject $\sim\sim(R\vee\sim R)$ just as we reject $R\vee\sim R$.

The generic idea here is that *utter rejection* of an instance of the law of excluded middle is a distinctive mark of indeterminateness. Thus we might think that we should be able to characterize some operator “It is indeterminate whether” (I), such that if one accepts I_p , one should reject $p\vee\sim p$. There's no obvious way to do this in the intuitionistic setting—but there is such an operator in the Lukasiewicz-omega setting, and in others more general logics such as Field's LCC.

Notice that we have focused aspects of non-classical *logics* – e.g. the acceptance of rejection of excluded middle, of rules of inference such as *reductio*, and so forth. In the context of vagueness, many of these logics are introduced *semantically*. For example, the Lukasiewicz-omega logic is often presented in terms of a *continuum valued semantics* interpreted in terms of *degrees of truth*. It is important to distinguish this *semantically loaded* interpretation of the logics from what is being proposed here. For example, in such settings, the law of excluded middle in borderline cases will be “half true” – but the suggestion here is that we *utterly reject* (assign credence zero) to

such instances. These two thoughts seem in tension – but the believer in nonclassical logic might reasonably query *why* he should be asked to endorse the degree-theoretic gloss on what he is up to – or indeed, whether a semantically loaded interpretation is owing at all.³²

Suppose that rejecting instances of LEM for p is distinctive of regarding p as indeterminate. Then to argue from this to rejectionism proper, all we need is one minimal assumption about the logic, and a way of reading off (pro tanto) constraints on rational belief from validities.

The constraint is that *rational credences cannot decrease across a valid argument*, or more precisely:

If $A \models B$, (and one is aware of this?) then it is pro tanto irrational to assign higher credence in A than one assigns to B .

The particular piece of logic we assume is simply that $p \models pvq$.

Suppose we're certain that p is a borderline case. Then by the current account, we should utterly reject (assign credence 0) to $p \vee \sim p$.³³ This means we cannot assign non-zero credence to anything that logically entails this (in obvious ways). But p entails $p \vee \sim p$. So we cannot assign non-zero credence to p . But this is just rejectionism: if we're certain that p is borderline, we should utterly reject p .

Truth value gap theorists.

³² Field discusses this point at length in connection to the interpretation of non-classical logics in treating semantic paradoxes. See H. Field *Saving Truth from Paradox* OUP 2008; e.g. ch.5.

³³ How exactly to regiment this step is unclear if we haven't got an "indeterminacy" operator in our language. It is perhaps best in settings lacking this operator (e.g. intuitionism) to read "certainty that p is indeterminate" as shorthand for "utter rejection of LEM for p "---trivializing our first step. In languages like L-inf and Field's LCC, which have an indeterminacy operator, typically the following will be valid: $A \vee \sim A \models DA$. Given this, if we are certain that A is indeterminate, and hence utterly reject DA , we can appeal another instance of the logical constraint on rational credences to argue that $A \vee \sim A$ should also be rejected.

We earlier noted that the various logical systems above need not be accompanied by a “loaded” semantic interpretation. But famously, many of the above systems can be characterized in a many-valued model theory, and various accounts of indeterminacy have been built up by taking such semantics seriously.

Let us focus on the simplest case. We start by generalizing the classical setting. There, we often think of the logical operators as characterized by means of tables that describe how the truth values of sentences compounded from logical operators are determined by the truth values of their parts. Suppose one “adds a truth value”---so we have true, false and neuter. The possible combinations of truth statuses for the ingredients of a two-place operator increases from four to nine. A “3-valued” truth table for e.g. “and” tells us which combinations lead to “A and B” being respectively true, false, and neuter.

Let us note two characteristics of this semantics-first approach. To begin with, there are *more* two place operators --- 27, as opposed to the classical 16. Some of these truth-functions are named after logical systems they help define, and there is likely to be controversy over which truth-functions correspond to which operators. Thus we have strong and weak Kleene tables for conjunction and disjunction, rival accounts of conditionals including the L3 conditional and the K3 conditional; and so forth. Furthermore, there are *several* properties of argument types that one might identify and take a theoretical interest in – guaranteed preservation of truth, guaranteed preservation of non-falsity, etc.

The characteristic mark of indeterminate sentences, I shall suppose, is that they take the neuter truth status. But this doesn’t yet tell us what attitude it is appropriate to take to sentences that have this status. In the previous setting, we had two bits of traction on the analogous question. One was that we were supposed to *reject* instances of LEM. The other was that we were working with a theory of *validity* which we could assume to constrain credences.

But in the current setting, we haven’t said enough to settle such questions. To be sure, there are ways of defining validity in this setting (K3 tables and validity as truth-

preservation) that extensionally coincide with the K3 logic discussed earlier. But should this property of truth-preservation constrain credences? And even if this is granted, are the tables the right one to choose, and should the LEM be rejected?

There are, I think, rival ways to answer the questions. The first option---which is non-rejectionist---I shall call *degree theoretic*. This glosses the “neuter” truth status as a truth value *intermediate* between truth and falsity: “half truth”. Such a notion may seem obscure, but if we combine this with the thesis that one’s degree of belief in p should match p ’s *expected truth-value*, then we simultaneously get a distinctive thesis about the appropriate attitude to certain borderline cases – that they should be according credence 0.5 – and also a sense of the significance of calling the neutral truth value “half truth”.

In complete contrast to this, we have a *rejectionist* interpretation of the neuter truth value. Think of the neuter truth value as simply the absence of either truth or falsity – so that having the neuter value is a way of being *untrue*. This fits neatly with the view that when p has this status -- when it is neither true nor false – one shouldn’t invest any confidence in it. In contrast to the degree theoretic reading of the formalism, this is a *truth value gap* interpretation of indeterminacy. So, alongside the LEM-rejecting theories, we have the truth-value gap conception just given.³⁴

(Some familiar criticisms of many-valued theories miss their mark when directed at a gap-theorist. For example, when A is neuter valued, on the most common systems $A \& \sim A$ will be neuter valued. But how can a contradiction be anything other than completely false? When this complaint is directed against the degree theory, it hits its mark – for representing $A \& \sim A$ as neuter-valued is, according to the degree theory, to say that it is half-true, and that we should invest credence 0.5 in it (the same credence as we invest in a fair coin turning up heads when flipped). This equivalence is hard to

³⁴ In fact, the truth value gap setting will also end up rejecting instances of LEM for borderline p , as untrue. Moreover, once we have these interpretations, we have leverage on how to choose between the variety of possible definitions of validity. If our interest is on validity that corresponds to constraints on credences, than a natural choice for the gap theories will be truth preservation (for then, tautologies will correspond to things that we should believe no matter what, and inconsistencies to things that we should reject no matter what). If our interest is on the degree setting, then the situation is rather different. An appropriate logic for constraints on belief, a “no drop in degree of truth value” definition (the logic Field labels S3, if the Kleene connectives are chosen) is most suitable.

accept. But in the context of the gap theory, it becomes simply the observation that such explicit contradictions can be untrue rather than false – but since we reject all untruths utterly, there is no problematic implication of “half-accepting” contradictions to get one’s head around.

It is not always clear when philosophers put forward a three-valued theory of indeterminacy which of these two conceptions is being propounded (or whether they have a third in mind). But for the reasons given in this essay, and also in the context of the familiar criticisms just sketched, such interpretative issues are of the first importance.)

Strong Supervaluationism.

The last of the rejectionist theories I here sketch I will call “strong Supervaluationism”. It is one way of interpreting one of the most popular accounts of indeterminacy.

“Supervaluationism” is in fact multiply ambiguous. One popular view associates it with a “semantic indecision” view of indeterminacy. On this view, there are multiple precise interpretations of our language not ruled out by meaning-fixing facts. These are the “precisifications” of the language. Supervaluationism refuses to choose between them – what is true (supertrue) is what is true on every precisification. What is false (superfalse) is what is false on every precisifications. Indeterminacy is then a truth-value gap (or at least a gap between supertruth and superfalsity) – it arises when a sentence is true on one precisification and false on another.³⁵ Within the object-language, we have “Determinately p” true (at s) iff “p” is true at all precisifications.

What attitude should we take to borderline cases? I suggest that the situation here is similar to that facing a proponent of three-valued truth tables – the formalism of precisifications, is compatible with several interpretations. A manifestation of this is that just as in the previous case, we have several competing definitions of *validity* to contend with: supertruth-preservation (global validity); truth-on-a-precisification-

³⁵ More generally...

preservation (local validity); non-superfalsity preservation, etc. The supervaluational formalism needs to be supplemented by a story that guides us in choosing amongst these options.

There is a rejectionist reading of the above. We can start with a certain *logical* system that enforces rejectionism – and then interpret Supervaluationism to fit with this. This leads to the reading I call *strong Supervaluationism*.

The logical system (call it SSV) we begin with is very close to the classical system. Indeed, every classical tautology is a SSV tautology. So unlike the LEM-rejecting setting, each instance of the law of excluded middle is accepted. But it contains a sentential operator D (determinacy) that leads to departures from strict orthodoxy. A key result is that the following is valid:

$$p \models Dp$$

Application of classical metarules would allow us to conclude that $\sim Dp \models \sim p$, and indeed would ultimately allow us to argue that p and Dp were intersubstitutable. However, the logic gives up these metarules: proof by cases, reductio, conditional proof, etc.³⁶

The pattern above allows us to argue for rejectionism. For if we are certain that p is indeterminate, we are thereby certain that $\sim Dp$, and so utterly reject (assign credence 0) to Dp. Since Dp is a consequence of p, by our earlier principle that credence can never decrease across a valid argument, our credence in p must be 0. So we must utterly reject p.

Now SVV is the supervaluationist’s “global consequence”—the logic we get if we look at those arguments that preserve truth in every model. So if we find an interpretation of the supervaluationist’s formalism that fits with according a central role to this notion, we have a rejectionist supervaluational package. A natural thought

³⁶ In fact, one often considers a weaker logic, SVV-, which lacks this rule. That logic still contains the closely related principle that p and $\sim Dp$ are inconsistent--- i.e. $p, \sim Dp \models \perp$. By a generalization of the theses governing logical constraints on rational credences---covering the multi-premise case – we can get the same results.

is to emphasize the *identification* of truth with supertruth---since then SVV will be the logic corresponding to *guaranteed truth preservation*.

An alternative, non-rejectionist reading of supervaluationism (let's call it weak Supervaluationism) emphasizes instead the role of *precisifications*. Relative to each precisification, a sentence expresses a perfectly determinate proposition (and we might have a determinate credence in each such proposition). So, compatibly with S being indeterminate, one might seek for further information about S --- what the pattern of probabilities of the *set of its precisifications* is. One might think of the appropriate attitude to an indeterminate S as given by a kind of “superposition” of the credences one has in S's precisifications---with salient qualities of such superposed states including, e.g. the upper and lower limit of credences assigned to some one of the propositions it expresses on some precisification.

The weak Supervaluationist's attitudes are unbound by rejectionist strictures---and certainly global validity has no real relevance to her attitudes. One of the other definitions of validity---local validity especially---will fit better with her approach. For similar reasons, there's no real motivation for giving supertruth the honorific *truth* on this approach.

We have here the beginnings of two rival interpretations of the supervaluational formalism. If each is internally consistent, there is no real point in arguing about which is “really” supervaluationism – we should simply ask self-described supervaluationists what they have in mind! What is important here is that on one of these disambiguations – the one built around SVV logic – Supervaluationism is a form of rejectionism.

We have seen two forms of “truth value gap” accounts of indeterminacy – the nonclassical and supervaluationist. It is perhaps worth finishing by presenting a little argument that shows that *every* such theory will have to give up some plausible thesis if it is to avoid rejectionism (I use T for a truth operator).

Premises:

1. It is indeterminate that p iff $\sim Tp \& \sim T\sim p$.
2. If necessarily and a priori, if p is true then q is true, then q follows from p
3. p and q are inconsistent if an explicit contradiction follows from them.
4. If q and p are inconsistent, then (on pain of pro tanto irrationality) the sum of one's credences in each cannot exceed 1.

Given these assumptions, p and “it is indeterminate whether p ” are inconsistent. For by premise 1, from the truth of the latter we have $T(\sim Tp \& \sim T\sim p)$. If this holds, then $T(\sim Tp)$. And from this, by the factivity of truth, we have $\sim Tp$. But this explicitly contradicts the truth of p . By (2), this means that an explicit contradiction follows from “ p ” and “it is indeterminate whether p ” together. But by (3) this means that they are inconsistent.³⁷

Granted this, by (4), then (on pain of pro tanto irrationality) the sum of one's credences in “ p ” and “it is indeterminate whether p ” cannot exceed 1. So if our credence in the latter is already 1, ceteris paribus we must have credence 0 in p —which is to say that rejectionism must hold. (Indeed, we have a variety of partial results—that one's confidence in p cannot exceed 1 minus one's confidence that p is indeterminate.)

Now, this little argument need not sway a truth-value gap theorist who wishes to reject rejectionism – since they can always deny one of our premises (or hold, for example, that the argument equivocates on the notion of “following from”). But since this can only be done by giving *something* natural up, it does disclose, I think a tendency for truth-value gap theorists across the board to be rejectionists.

³⁷ Strictly, we have not yet shown that an explicit contradiction *follows from* our pair, since for this we would have to derive the truth of an explicit contradiction. But explosion---the principle that from an explicit contradiction, everything follows---we get from $\sim Tp \& \sim T\sim p$ to anything we like, including in particular $T(q \& \sim q)$ for any q we choose.

