Decision making under indeterminacy

J. Robert G. Williams [*]

May 8, 2013

Contents

1 Introduction. **3**
1.1 The Cabinet . . . 4
1.2 The broker's offer . . . 5
1.3 Patterns of action . . . 6

2 Uncertainty and Inconstancy **9**
2.1 Supervaluational indeterminacy . . . 9
2.2 Indeterminacy-induced uncertainty . . . 10
2.3 Caprice and Randomize . . . 12
2.4 Application to speech acts and the sorites argument . . . 15
2.5 Diachronic puzzles and hyperplanning . . . 17
2.6 Incrementalizing Caprice and Randomize . . . 19
2.7 Application to the forced march sorites . . . 21

3 Elga's puzzle **22**
3.1 Planning . . . 25
3.2 Incremental hyperplanning . . . 27
3.3 Conditionalizing on future choices . . . 28
3.4 Mind-making . . . 29
3.5 Choosing to randomize and randomly choosing . . . 32

Appendix A Indeterminacy-induced mushy beliefs **35**

Appendix B Indeterminacy-induced mushy desires **35**

Appendix C Sharpenings recommending action **36**

[*]This paper grew out of work on indeterminate survival I worked on while on an AHRC research leave award 'Intrinsic survival, Multiple survival, Indeterminate survival'. Material relevant to this was presented in several places: in Oxford, Cambridge, Leeds, MIT, USC, Brown, Freiburg and Glasgow. I'd like to particularly thank Ofra Magidor and Cian Dorr for helping me see that the decision-theoretic side of that early version needed further study. Thanks to participants in a CMM seminar for feedback on part of this paper; and in particular to Elizabeth Barnes, Seamus Bradley, Ross Cameron and Agustin Rayo. I'd also like to thank Adam Elga for his comments on the paper at FEW (in USC) and Laura Buchak, Michael Caie, Sarah Moss, Katie Steele for very helpful and constructive discussion. This work was supported by a British Academy Research Development Award (BARDA: 53286) and was supported by the workshops funded by Spanish Government grant FFI2008-06153 (MICINN).

Abstract

This papers studies a decision puzzle with indeterminacy at its centre. It is indeterminate whether Alpha survives as Omega; but Alpha has the chance to invest at small cost to greatly benefit Omega. Furthermore, Alpha is entirely self-interested. Should he invest? What patterns emerge if we repeat the experiment?

Addressing such questions is a central challenge in explicating the cognitive role of indeterminacy. But there is little consensus in the literature about even such mundane questions as: what attitude to *p* is appropriate, when one knows that *p* is indeterminate? This paper develops an answer on the model of imprecise credence treatments of uncertainty, generating successful predictions about the sorites. A diachronic puzzle for the framework due to Adam Elga is set out and explored.

1 Introduction.

Sometimes, the facts run out. Faced with a reddish-orangish colour patch, it seems to many that it is *indeterminate*, or *indefinite*, or *borderline*, or *unsettled*, whether the patch is red. The notion of indeterminacy is deployed all over philosophy—in discussions of vague predicates, theory change, future contingents, incompleteness in mathematics, conditionals, presupposition failure, the paradoxes of self-reference, and more. But what is this notion? Is there even a single unified concept that covers all the cases just mentioned? Much sophisticated work has been done outlining semantico-logical frameworks for indeterminacy. But they do not directly address one of the most pressing questions about describing something as indeterminate. Suppose we accept that p is indeterminate; how should this belief affect our other doxastic states and wider mental life—in particular, what attitude can we rationally adopt to p, while holding the belief that p is indeterminate? What is the cognitive role of indeterminacy-judgements?[1]

This is one instance of a more general topic. We believe, desire, hope, fear, and act under vague guises. Our best accounts of the interrelation between such attitudes (for example, the interrelation between belief and desire set out in decision theories) often presuppose a classical backdrop that is under fire in the literature on vagueness. The methodology below is to approach the direct cognitive role question indirectly, through looking at the interplay of beliefs, desires and action in the presence of indeterminacy. (Nothing I say here will depend on whether the *source* of the indeterminacy in question is 'in the world' or 'due to language', or the like. I will, however, appeal throughout to indeterminacy as an operator, rather than a predicate of linguistic items. Our cognitive and conative attitudes to *it being indeterminate whether Harry is bald* is one thing; the linguistic expression 'Harry is bald' having some particular semantic status another. A basic desire to bring about the first is not the same as a basic desire to bring about the second; altering the meanings of words is one way to achieve the latter; the former would require manipulations of Harry's head.)

The paper is divided into three parts. The first introduces our stalking horse: a decision puzzle with indeterminacy at its heart. We describe the setup (1.1), the decision puzzle (1.2) and the possible patterns of reaction (1.3). A dispositionally inconstant pattern of response is identified as target.

In the second section, a model of mind is developed that supports such inconstant responses. (2.1, 2.2) introduce the basic machinery of sharpenings and sharpening-induced mental states. (2.3) links mushy mental states to action, via the decision rules Caprice and Randomize. These sections are relatively informal, and appendices A-C regiment and generalize the proposals. (2.5, 2.6) present and resolve a theoretical worry for the account to do with diachronic coordination. (2.4, 2.7) apply the framework to talismanic puzzles of indeterminacy: the sorites and forced-march paradoxes.

Section 3 explores an objection due to Adam Elga. The net effect of this is to bring into sharp relief two competing interpretations of the framework. One (discussed in 3.2 and 3.3) involves steadfast indeterminacy-induced uncertainty, supplemented by pragmatic bookkeeping to ensure one's actions across time do not conflict. The other (3.4), the mind-making proposal, involves agents genuinely making a judgement—albeit on an arbitrary basis—-when forced to act under indeterminacy.

The conclusion highlights the most characteristic features of the account of decision making under indeterminacy developed here: its anti-compromise stance.

[1]For work with an emphasis on this question (albeit developed in ways incompatible with the approach developed here) see, inter alia, (Smith, 2010; Dorr, 2003; Barnett, 2009; Wright, 2001; Schiffer, 2003; Field, 2003).

1.1 The Cabinet

The survival of a person across time can be an indeterminate matter, as van Inwagen (1990) emphasizes:

> One's life may be disrupted in various ways. If a pin is stuck into one's finger, one's life goes on. If one is blown to bits by a bomb, then—even if God immediately puts the bits back together again—one's life has ended. . . . If, at the extremes of a spectrum along the length of which are arranged more and more radical disruptions of lives, we can find definite cases of the end of a life and definite cases of the continuation of a life, then it seems reasonable to suppose that somewhere between the extremes will be found disruptions of which it is not definitely true or definitely false that they constitute the end of a life. And if this is so, then there are possible adventures of which it is not definitely true or definitely false that one would survive them. Let us call such episodes 'indeterminate adventures'. Not everyone—perhaps hardly anyone—will agree with my contention that one survives an adventure if and only if one's life persists through that adventure. But anyone who thinks that people are complex material organisms will be hard put to it to deny that possibility of indeterminate adventures. (van Inwagen, 1990, p.243)

Many cases of this kind involve episodes after which it is indeterminate whether there is any person at all around. But another class are those where it is definite that there are persons before and after the episodes, but it's indeterminate whether it is the same person throughout. To discuss such cases without getting into the nitty-gritty of how to set things up for this or that theory of persistence, we'll follow van Inwagen in appealing to 'the cabinet':

> Suppose that a person, Alpha, enters a certain infernal philosophical engine called the Cabinet. Suppose that a person later emerges from the Cabinet and we immediately name him 'Omega'. *Is* Alpha Omega? . . . Let us suppose the dials on the Cabinet have been set to provide its inmates with indeterminate adventures. (We need not agree on what would constitute an indeterminate adventure to suppose this. Let each philosopher fill in for himself the part of the story that tells how the dials are set). Alpha has entered and Omega has left. It is, therefore, not definitely true or definitely false that Alpha is Omega. (van Inwagen, 1990, p.243-4)

If one wants to discuss the impact of indeterminacy on belief, desire and action, vague personal identity is a good place to start. The reason is that our own survival *matters* to us. One desires, inter alia, that good things happen to oneself in the future; and fears the bad things that may happen. And these fears and desires are intrinsic, rather than instrumental—it is not, in the most usual cases, that one fears one's future pain because being in pain is a signal that something else fearful is happening. It is the pain itself one desires to avoid. Of course, one might also desire that good things happen to loved ones, or that that the welfare of humanity in general is improved. But there's a particular kind of self-interested concern that's extremely psychologically salient.

We can imagine that Alpha takes this to extremes—all that he cares about is the good or bad things that are going to happen to him. Our question is: how should Alpha then think about known goods and evils that happen to Omega? To test this, let some broker offer Alpha an investment opportunity. He's certain, we'll assume, that he will soon be subject to the Cabinet. For a small investment now by Alpha (100 dollars perhaps), Omega will benefit to the tune of thousands.[2] Alpha is entirely self-interested, but not exclusively present-self-interested: he's prepared to give up goods now if doing so brings him great gains in the future. So if he was certain that Omega was him, this would be a no-brainer: he'd take the investment like a shot. But, caring only about his own benefit, if he was certain that he was not Omega, he'd keep his cash and spend it on a few final nights of partying. What should he do? What would *you* do, in his shoes?

1.2 The broker's offer

How should Alpha think about the broker's offer? It's illuminating to compare the situation to a parallel one that does not involve indeterminacy. Perhaps the makers of the Cabinet also make a machine that gives people chancy, rather than indeterminate, adventures. With the dials suitably set, anyone who enters this second machine has a 50/50 chance of passing through unscathed. If the dice roll against an inmate, then he undergoes destruction and reconstruction that (according to one's favoured view of personal identity) produces a person determinately distinct from the one who enters. Faced with the certainty he's about to enter the chancy cabinet, and knowing the relevant chances, the investment decision facing Alpha would be comparatively familiar. At first pass, he would calculate the expected utility of investing as opposed to not-investing, and choose whichever maximizes utility.

A first pass at the decision table (not attempting to assign numerical values at first) would be the following:

Chancy cabinet	$\alpha = \omega$	$\alpha \neq \omega$
Invest	Long life, No party, Riches	Short life, No party. No riches.
Reject	Long life, Party, No riches	Short life, Party, No riches.

Expected utility theory requires we associate numbers with the outcomes described in the cells above. The determinants of utility of the outcome described in each cell includes: whether an individual has a long or short life; whether or not they get a party; and whether or they get riches (we assume that in all other respects the prospects are equally good). Let's suppose that the life up to the point of encabination on its own contributes -100 utils; the extra years come with a boost of $+100$ utils on top of this. We'll assume that having riches comes with a boost of $+100$ utils, and a party with a boost of $+10$. Lacking riches, and lacking a party, don't add or subtract any utils to the life. We're then in a position to figure out the net utility for ech cell in the table above. The top left cell includes long life, i.e. a life up to encabination plus extra years $(-100+100)$ and riches $(+100)$, which will give a net utility of $+100$ The bottom right cell includes a short life (-100), a party $(+10)$, and thus has a utility of -90. Filling in the other cells likewise, we get the following:

Chancy cabinet	$\alpha = \omega$	$\alpha \neq \omega$
Invest	$+100$	$-100.$
Reject	$+10$	$-90.$

[2]The broker's offer adapts a case that Bernard Williams discusses in a related context (Williams, 1970, p.48).

Alpha doesn't know the exact outcome of either of the actions open to him. But he does know the relevant chances—we can list the probability he attaches to each at the head of each column—and using this he can aggregate the utilities of the two possible outcomes of each action, to arrive at an overall ranking. The recommended choice is the action that maximizes expected utility:

Chancy cabinet	P($\alpha = \omega$)=0.5	P($\alpha \neq \omega$)=0.5	Expected utility
Invest	$+100$	-100	0
Reject	$+10$	-90	-40
Verdict:			**Invest**

Let us now turn back to the original cabinet, which delivered indeterminate rather than chancy adventures. Once more, we can assess the four outcomes individually. What Alpha values remains the same—having more years, partying and riches all add utility. And so, I will be assuming, the same utilities should be entered into the four possibilities represented on the decision table.

(In making this assumption, I'm glossing over some differences between the cases that you might think should make a difference to the utilities assigned. In the case of the chancy cabinet, supposing Alpha survives, he will survive unchanged psychologically and biologically. With the indeterminate cabinet, even on the hypothesis one survives, one knows that one undergo radical changes; perhaps some (but not all) memory links to early childhood are raised and some (but not all) of ones present plans for the future are erased. In the light of that, you might think that possibilities featuring extra years should deliver less of a utility boost for Alpha in the indeterminate cabinet scenario, compared to the chancy cabinet scenario. The exact numbers won't matter for the points to be made below, and so we could indeed factor this in. But for ease of comparison, I'll be assuming that we add detail to the indeterminate cabinet scenario so a utility boost of $+100$ for the extra years remains appropriate. For example, perhaps the psychological changes will include erasing traumatic memories as well as welcome ones; perhaps the erasure of some of Alpha's plans will be compensated by raising the chance of success in others.)

The table we arrive at takes a familiar form:

Indeterminate cabinet	$\alpha = \omega$	$\alpha \neq \omega$
Invest	+100	-100.
Reject	+10	-90.

The crucial difference is that this time we have no recipe for aggregating the rankings of the two columns. What Alpha knows is that it is indeterminate which of the columns describes reality aright—but this gives little steer about how to proceed.[3] Our question is how to make decisions in the peculiar kind of uncertainty generated by indeterminacy.

[3]The situation is reminscient of the classic puzzle of decision making under *uncertainty* rather than *risk*—with expected utility theory being the inheritor of the latter tradition. A classic text on these matters is Luce & Raiffa (1989).

1.3 Patterns of action

One way to think about what sort of advice we might give Alpha, is to imagine the dispositions to act that agents *could in principle have* faced with this decision situation, and variations of it—and then try to pick one among them as the appropriate one. I will describe this space of options, and identify one in particular for which it will be our task in what follows to provide rational underpinnings.

In any particular instance of the broker's situation, Alpha must either invest or reject the offer (doing something indeterminate between investing and rejecting is not an option at all). So it looks like there are only two options he must decide between. But we should also remember the possibility of *randomly* choosing what to do among the available options. To detect these random dispositions to act, suppose that we repeat the experiment many times over (perhaps with duplicates of Alpha; or perhaps the Cabinet-makers present him with the investment decision, then take him aside, wipe his memory, and rerun the experiment). Advice to invest, advice to reject, or advice to pick one or the other with certain chances predict different patterns under this repeat examination. Alpha might invariably invest, invariably not invest, or behave in an inconstant manner, sometimes investing and sometimes not investing (with the long-run frequency matching up to the chances in the mixed strategy). To gain even more information, we could include among the repetitions variations in the parameters of the setup. We could, for example, raise or lower the prize given to Omega if Alpha invests.

There are four salient possibilities here for what the overall pattern will look like:

Universal acceptance Alpha displays constant behaviour, and always accepts the broker's offer no matter how much money is offered to Omega.

Universal rejection Alpha displays constant behaviour, and always rejects the broker's offer no matter how much money is offered to Omega.

Threshold This pattern of behaviour features constant behaviour for a fixed level of prize. But when we varying the prizes awarded to Omega, there is a a tipping point. Prizes below this threshold invariably lead to not-investing, prizes above it lead to investing.

Inconstant Even for fixed prizes, Alpha acts inconstantly, sometimes investing, sometimes rejecting duplicate contracts. Subvarieties of this pattern involve a fixed long-run frequency of investing vs. rejecting; or a varying one (depending on the exact contract offered).

Subsequent sections explore the Inconstancy view, and this is the main focus of the paper. That focus requires some motivation, and in the remainder of this section I provide an argument for favouring Inconstancy over its rivals. Not every advocate of Inconstancy as the rational response to indeterminate decision-making will accept what follows as the reason for favouring their approach, or even accept the premises. But I hope it helps the reader see how one line of thought can lead us to that position.

Consider the following pair of theses:

(A) Personal identity is what matters in survival.

(B) There is no need to revise standard (classical) logic and semantics, even when propositions are (non-epistemically) indeterminate.

Each of these is widely discussed in the respective literatures on personal identity and indeterminacy.[4] Using these as premises, we can argue (by elimination of alternatives) that Inconstancy describes the appropriate dispositions for Alpha to have faced with the broker's offer. I sketch the arguments below.

First, it is hard to combine Universal Acceptance or Rejection with (A). (A) tells us that personal identity matters in survival; one's self-interested actions should be sensitive to the welfare possessed by y in outcomes iff y is oneself. But suppose one is a Universal Rejector. Then it looks like one's self-interested actions simply discount the interests of y, whenever it's indeterminate whether y is oneself. Modus tollens applies: one is not y. Thus, on the assumption that we're Universal Rejectors, indeterminate survival collapses into non-survival. There's a dual worry for Universal Acceptors—if what happens to indeterminate y always matters to one's self-interested action, then modus ponens on the same biconditional gives us that y is oneself. So indeterminate survival collapses to survival. Further, if each premise of the above argument is determinately true, the conclusions hold determinately—and that amounts to a reductio.

(A) and (B) together also eliminate the threshold view. The argument I have in mind goes as follows:

1. Truth is bivalent, even in cases of (non-epistemic) indeterminacy.
2. If the threshold view is correct and the case of Alpha and Omega involves non-epistemic indeterminacy, then the right semantic treatment will be non-bivalent.
3. $\therefore$ The threshold view is incorrect.

The argument is valid, and (B) delivers the first premise directly. So the tenability of the threshold view (under our assumptions) turns on premise (2). In (Author), I defend this in detail. The key claim is that the best rationale for threshold views (consistent with assumption (A), that personal identity matters) involves picture of indeterminacy involving infinitely many *degrees of truth*, rather than two exclusive and exhaustive truth values. That is, in effect, the upshot of David Lewis's treatment of these cases of personal identity, which aims exactly to show that the thesis (A) can be reconciled with Parfit's model of rational self-interested action.[5] I think that if the threshold view is correct, Parfit's model is the way to account for it; and if you want to simultaneously endorse (A), then you better understand it in Lewis's way.[6] (Author) examines all this in detail, and in particular argues that there is a real incompatibility between the 'degrees of truth' and 'bivalence-friendly' models of indeterminacy, and so a genuine conflict between Lewis's model and (B). So (2) is secured, and the threshold view is eliminated by (A) and (B) together.

Assume these arguments are successful. Then exploring the Inconstancy view should be interesting even to one who harbours doubts over one of (A) or (B). For what the arguments

[4]The *locus classicus* of discussion of (A) is of course (Parfit, 1971); see also (Lewis, 1976; Parfit, 1984). Thesis (B), non-epistemicist classicism, was until relatively recently a fringe view, but in the last ten years or so has gained many advocates. See (Fine, 1975) (in his discussion of the conceptual priority of penumbral truth); McGee & McLaughlin (1994) (though they regard it only as one disambiguation of 'truth' talk); Dorr (2003), Greenough (2008), Barnett (2009), Barnes (2006, 2010); Barnes & Cameron (2009), (Author), and Eklund (2010).

[5]Parfit's model, recall, involves thinking that Omega's welfare matters to Alpha at a *discounted rate*, where the discount factor is fixed by the strength and extent of the psychological connections between Alpha and Omega. That produces threshold behaviour, since the key question will simply be whether the future benefits to Omega (appropriately discounted) exceed the present sacrifice—and by raising or lowering the benefits to Omega you can flip the answer from 'yes' to 'no' in exactly the way that the threshold view requires.

[6]Note that I think that the Parfittian model is the best way for accounting for indeterminacy under the assumption that the indeterminacy in question is non-epistemic. If epistemicism were correct, I would expect threshold behaviour just as in the case of the chancy cabinet; and that model would be compatible with bivalence.

show us, framed neutrally, is the joint incompatibility of (A), (B) and the denial of Inconstancy. One who accepts (A) and (B) must then defend the view that Inconstant behaviour is rationally appropriate. But if Inconstant behaviour has no rational underpinnings, we can then conclude that one of (A) or (B) must be given up, giving us a new way of arguing against one or other of these widely-discussed philosophical theses.

I've outlined motivations for exploring the Inconstancy view that trade on theoretical connections between decision making under indeterminacy and theses elsewhere in philosophy. I haven't argued that Inconstancy *requires* (A) and (B), and one could in principle endorse the framework to be developed in section 2 for quite different reasons. Indeed, there's an obvious route to favouring inconstancy that needs no high-level theoretical backing. One might simply compare the predictions of the Inconstancy account against what strikes us pre-theoretically as appropriate attitudes or behaviour, and find the fit appealing. I applaud that methodology too, but of course, we can't implement it until we know what the Inconstancy view *does* predict when we apply it to a range of cases. This sort of theory-neutral support for Inconstancy will emerge piecemeal throughout this paper—I will draw out the consequences of the Inconstancy account for linguistic behaviour in two famously and intensively studied cases—the sorites argument and a forced march sorites.[7])

2 Uncertainty and Inconstancy

Our goal is to find rational underpinnings for inconstant behaviour in the presence of indeterminacy.[8] First I want to make explicit the semantic skeleton on which the account will be built.

(A housekeeping note. In what follows I will be attributing semantic properties to truth-bearers. I'll also be talking about the consequences these semantic properties have for attitudes such as *believing that Alpha is Omega* where the embedded claim is indeterminate. One might be worried about a mismatch here, since usually truth-conditional theorizing focuses on the semantic properties of sentences; whereas beliefs have propositions as their objects. But there are plenty of ways in which the mismatch could be avoided. Perhaps the vehicles of truth and falsity in semantic theory are sentences; and beliefs consist in a relation to a (mentalese) sentence.[9] Or perhaps beliefs and desires are in the first instance relations to Fregean thoughts; and then the truth-conditional theory should be construed as setting out a theory of the truth-conditions of such entities. But it doesn't matter that much how it is achieved; what I need is that there is such a match.[10]

[7]For more on this lack of consensus and it's source, see (Author). For the attractions of something like an inconstancy view, see Crispin Wright and Stewart Shapiro.

[8]As already mentioned, my ideas in this area are shaped by Crispin Wright's work (see especially (Wright, 1976, 2001, 2003); though the way I develop the idea is very different. My thinking below—especially the association of a degree theory with partial dispositions to act in conflicting ways, which grew into the weightings of Randomize—grew directly out of engagement with the characterization of vagueness-related degrees of partial belief in Schiffer (2003)—especially the brief characterizations of their characteristic *ambivalence*. Schiffer assures me, however, that his account should be intepreted as excluding the kind of inconstant behaviour I appeal to. Closest to the model I develop is the contextualist account of Shapiro (2006), which faces a quasi-supervaluational notion of conversational score.

[9]See the belief-star relation of (Field, 1978)

[10]The natural way to reject the matching thesis is to hold that propositions (the objects of belief) have precise truth-conditions, so that they are never indeterminate in truth-value; but it is indeterminate which proposition a vague sentence expresses. One such account of propositions would be the Lewis-Stalnaker model on which beliefs are relations to sets of possible worlds. Another might be a Russellian view on which when *N* is indeterminate in reference, it is indeterminate which Russellian singular proposition '*N* is *F*' picks out. Of course, in light of puzzles

2.1 Supervaluational indeterminacy

I will work with a broadly *supervaluational* treatment of indeterminacy. Indeterminacy, on this conception, involves a kind of *unsettledness* between candidate truth-value assignments to propositions. An epistemicist thinks there's One Correct way of scattering truth and falsity over sentences or propositions, given the way the world is (including, inter alia, over the claim that Alpha is Omega—this comes out either as true, or as false). The supervaluationist agrees that these are decent candidates to be truth-value assignments, but holds that the relevant facts do not select one in particular as uniquely correct. Some candidates can be thrown away—those that give clearly wrong truth values to e.g. 'Alpha is human', or say that some borderline colour patches are red, while colour patches even redder than them are not red. But even once all this information is in, there remain many candidate classical truth value assignments, $S = \{s_1, s_2, \ldots, s_n\}$.[11] We call the members of S *sharpenings*. The classical case is simply the limiting case where S is a singleton; but indeterminacy characteristically is manifested by the presence of multiple sharpenings. A sentence or proposition is indeterminate iff some of these sharpenings declare it true, and others classify it as false. It is determinately true if all agree it is true; and determinately false if all agree it is false.[12]

There are some neat features of this picture. To start with, all classical tautologies (like the law of excluded middle) are true on each classical interpretation and so a fortiori, on all the s_i in our set of sharpenings. So classical tautologies will be determinately true. Supervaluationism is closer to classical logic than many other semantic models of indeterminacy. Furthermore, we can make room within the model for talk of degrees of truth (or degrees of determinacy, if the former vocabulary makes one queasy). A sentence or proposition p is true to degree k iff $k = \frac{|\{s \in S : p \text{ is true on } s\}|}{|S|}$.[13] Such degrees of truth will be structured like a classical probability—something we will exploit below.[14]

This much structure I will suppose. But I am neutral on many other matters that supervaluationists need to take stances on, particularly on matters of truth and logic. I do not say what formal construct within the semantics is to be identified with truth; nor how to define logic in this setting. My own view is supervaluationism is not a single view but a whole family, and the right way to individuate family members is to describe their interactions with belief, desires and decision making as we are about to. Questions of what counts as truth, and what logic, should be settled after one has got clear on the roles that the basic features of the model play in our wider life. (My own view is that the package to follow is naturally paired with a classical, bivalent treatment of truth and logic. That's certainly desirable if the model is to fit with thesis (B) from the previous section. But nothing in the technical development to follow will depend on such issues.[15])

of logical omniscience and Frege puzzles, both these views of the objects of belief are very controversial; and often advocates complicate the story by throwing in additional elements to the full story about belief ('guises', for example). It may be even in those apparently hostile settings, a version of the discussion here can be reconstructed *when we see the full framework*.)

[11] I'll standardly work with finite sets; though in actual applications, we expect there to be infinitely many sharpenings. Everything I will say can be easily generalized to that case.

[12] Some locus classici: (van Fraassen, 1966), (Fine, 1975), (Keefe, 2000).

[13] in the infinite case, we need to appeal to a (normalized) *measure* over the set of sharpenings, and then let the degree of truth of p be the measure of the set of sharpenings that makes p true.

[14] These supervaluational 'degrees of truth' (or something like them) are explored in (Lewis, 1970; Edgington, 1997; Kamp, 1975; Cook, 2002) and (referenced ommitted). For criticism, see (Smith, 2008). I should note that the use I put the machinery too below is clearly not what was intended by most of these authors.

[15] Pairing bivalent truth with a supervaluational model is a familiar theme—see Fine (1975) for an early exploration. (Williamson, 1994, ch.5) argues against this being a version of supervaluationism.

2.2 Indeterminacy-induced uncertainty

In the classical, bivalent setting, one familiar model of 'ideal psychological states' associated with propositions are point-like credences: full confidence in p if its true; no confidence if it is false. Cases of *uncertainty* are similarly pointlike, represented by real numbers strictly between 1 and 0. Corresponding, the desirability or utility of a proposition can be represented by real numbers (like those entered into our representation of the chancy cabinet earlier). Overall, a classical state of mind will be a pair of a probability and utility function, $\langle p, u \rangle$.

There is an alternative model of uncertainty: the 'imprecise' or 'mushy' credence model favoured by, inter alia, Isaac Levi (1974).[16]. The idea here is to represent a person's doxastic state not by a single, precise credence, but by a whole *set* of them, represent the agents' 'open mindedness' or 'uncertainty' between the various more particular views. Likewise for desirability/utility. So we expect to associate an agent not with a single classical state of mind, but a whole set of them: $\{\langle p_1, u_1 \rangle, \ldots, \langle p_n, u_n \rangle\}$. You are in some sense undecided between the views expressed by the classical states of mind within this set.

On this 'mushy credence' model, there are two dimensions of uncertainty available. One can have a particular level of confidence in a question, which doesn't reach the poles of complete certainty or utter rejection. Or one might be in the dark even about particular levels of confidence. Faced with a fair coin, midway through being flipped, one can justifiably be half-confident that it will land heads. But (say the advocates of mushy credences) the question of the state of the economy a few years down the line, where statistics have proved a poor guide in the past, demands a more radical kind of uncertainty—one aptly captured by representing the range of levels of confidence one is open to.

There's clearly an analogy between the way that supervaluationism generalizes classical semantics and the way that Levi-like mushy credences generalize classical states of mind. This hasn't gone unnoticed—indeed, Levi describes his view as one involving 'indeterminate beliefs'. But this kind of generalization of classical states of mind needn't (and usually is not) tied to a response to indeterminacy. However, the view to be explored below makes exactly this connection.[17]

The rough idea, then, is that if p is known to be indeterminate, then one's attitude to p is represented not by a single level of confidence, but a whole mushy set of them (or more generally, one's doxastic state is not represented by a single probability function, but a set of them). But to tighten this up, we need to ask: what probabilities, in particular, make it into the set? It is here that the supervaluational model of indeterminacy comes in.

Let's go back to our worked example. Alpha, we may assume, knows exactly which particle-configurations will arise depending on how he acts. The difficulty is that 'Alpha survives as Omega' is true at one of these possibilities relative to one sharpening, and false at the same possibility relative to other sharpenings. So even though Alpha has perfectly definite credences over the possible worlds, the proper degrees of belief he should have over *sentences* or (fine-grained) *propositions* cannot be read off.[18] But here is the rub: holding fixed a sharpening, one can read off a definite truth value assignment for sentences or propositions at those worlds. And

[16]There are too many varieties to survey here, but see Joyce (2011) for the state of art within philosophy, and Weatherson (Manuscript) for references to some of the wider literature and the variety of approaches being pursued. (Jeffrey, 1992) is perhaps the version closest in spirit to the approach I will be developing here.

[17]I should note that a connection of this kind—sometimes accompanied by the additional gloss that it is indeterminate which of the classical beliefs described by elements in the mushy state the agent subscribes to—is very common in conversation. I do not know of places in the literature where it is examined in detail however.

[18]Notice: I'm hear thinking of worlds as describing directly particle configurations—think of them as Lewisian space-times, or ersatz sets of sentences in a world-making language that is expressively restricted to talk of fundamental features of the world.

so, relative to each sharpening, there's a natural proposal for reading off appropriate degrees of belief in each vague sentence/proposition: it should be equal to the credence he invests in those worlds at which the proposition is true according to that sharpening.

For example, relative to an $\alpha = \omega$ sharpening, and with Alpha having underlying credence 1 that he will enter the cabinet and Omega will emerge, the appropriate degree of belief for him to have in surviving encabination is 1. Relative to the $\alpha \neq \omega$ sharpening, the appropriate degree of belief in the same proposition is 0. If Alpha were only 0.5 confident that he will enter the cabinet tomorrow morning (the alternative being that it was all a bad dream, and he'll definitely survive) the appropriate degree of belief in him being around tomorrow evening would be 1 on the $\alpha = \omega$ sharpening, and 0.5 on the $\alpha \neq \omega$ sharpening. In short, relative to an underlying credence distribution over worlds, the sharpenings induce a *set* of degrees of belief. Details are given in appendix A.

This kind of projection works equally well for utility. Alpha, we said, cares only about himself getting benefits. Relative to a sharpening, we know exactly who Alpha is across times and worlds (and whether he gets particular, vaguely-specified, benefits there). So relative to that sharpening we can assign specific utilities to outcomes. But of course, rival sharpenings may lead to very different utility assignments. We see that in Alpha's decision puzzle. One sharpening assigns Omega having money high utility; the other assigns situations where Omega gets money no extra utility.

More abstractly: if we desire P intrinsically, the rule might be that any world such that P gets $+1$ utility. If it's indeterminate whether P is true at w, we don't have a straightforward way to pick out a single utility function over worlds. Relative to a sharpening, however, we can construct a utility function in this way. We can combine this with the indeterminacy-induced belief state described above to get a mushy mental-state: a set of belief-desire pairs. Details are given in appendix B. Thus the supervaluational machinery, together with the vaguely specified utility-determination recipe, generates a set of utility functions. So our model predicts the desirability-analogue of uncertainty. When I talk of an agent's (mushy) mental state, I will be talking of this set of belief-desire pairs.

On this model indeterminacy will be characterized by a certain distinctive kind of indeterminacy-induced uncertainty. Now, *typically*, such uncertainty will show up in the very proposition one knows to be indeterminate. But this is not always the case: sometimes we can know that something is indeterminate and yet have a perfectly settled level of confidence in it, consistently with the model just described.[19] The cognitive role of indeterminacy in this setting manifests in global constraints on mental states, but it does not requires a *specific* kind of uncertainty-related attitude to the proposition that is known to be indeterminate.[20]

[19] In appendix A, I give one example: one can know that it is indeterminate whether p, since p is indeterminate at both the worlds one regards as open possibilities; and yet meet the coherence conditions above by having mushy credences over the worlds, while retaining degree of belief 1 in the known-indeterminate proposition. One might be sceptical about whether particular combination this is a permissible belief state, however (I won't need to suppose it is until the model discussed in the closing subsection—it will not arise, as far as I can see, if we require agents to have a definite credence distribution over worlds). But even so, we can find examples where the level of confidence in known-indeterminate propositions is the same throughout the mushy mental state. Suppose 'hails' is indeterminate in reference between 'heads' and 'tails'. I am 0.5 confident that the fair coin will lands heads, and 0.5 confident that it will land tails. On either sharpening, therefore, the degree of belief appropriate to 'the coin will land hails' is 0.5. So here there is no mushy uncertainty, but a pointy degree of belief, in the proposition in question.

[20] A last note about the indeterminacy-induced belief states described above. Unlike Levi, and much of the literature on mushy belief, the belief states above are not closed under convex combinations. To insist on such closure would be utterly ad hoc, in our setting—however natural it is for uncertainty generated by other sources. This does have distinctive consequences, and is implicated in the problems we face in the final section. Some believers in mushy mental states for reasons other than indeterminacy, such as Richard Jeffrey, already reject

2.3 Caprice and Randomize

With this account of Alpha's mental state in hand, look back to the Broker's offer and the choice that Alpha faces. Earlier we represented the information available as follows:

Indeterminate cabinet	$\alpha = \omega$	$\alpha \neq \omega$
Invest	+100	-100.
Reject	+10	-90.

We are now in a position to describe what this represents. One can think of the columns in such a decision table as representing a partition of Alpha's mushy mental state—ones that agree on enough of the facts to assign the same utility to each action. The numbers in each cell of that column then just the agreed desirability (expected utility) of the relevant action. Equivalently, you can think of the column as a set of sharpenings—the sharpenings that induce the respective elements of the mushy mental state.[21] Given what Alpha knows about the Broker situation, the mental states induced by a given sharpening will agree on their evaluations of the utility of an act just in case they agree on whether Alpha is Omega. Hence, what we learn from the decision table above is that Alpha's mushy mental state is mushy exactly on the crucial issue: whether investing is preferable to rejecting the broker's offer, or vice versa.

What this gives us is an interpretation of what's going on in the table. It regiments the problem, but we can't yet read off any implications for how Alpha may permissibly act. Quite generally, it's one thing to have formal tools for representing belief (or mental) states, quite another to say what the relevance of these states are for one's wider psychology or behaviour. And that of course is precisely what we need to reach a recommendation for whether he should accept the Broker's offer.

There are many possible decision rules that may be associated with the mushy mental state representation. One common basis for decision rules given a mushy mental state is the following 'dominance' principle: that if an action is impermissible[22] on every belief-desire pair in the mushy mental state, then it is impermissible simpliciter. Contrapositively, each permissible action must be optimal by some belief-desire pair one's uncertainty leaves one 'open' to.

The decision-rule that Weatherson (Manuscript) labels 'Caprice' says that the dominance principle is both necessary and sufficient for permissibility. Decision-making under uncertainty, on this view, characteristically leaves the agent options, since more than one act may be permissible.[23] A different option is to supplement the dominance principle with further rules. Levi's preferred approach takes this form—once we have the arena of 'admissible' options, we select among them, by choosing the one (roughly) that guarantees us best 'worst case scenario'.[24]

We wanted a model that would provide intelligible underpinnings for inconstant behaviour

convex closure as a constraint in set-like belief states.

[21] For issues about whether a sharpening will induce a unique mental state, see appendix C.

[22] does not maximize expected utility

[23] Of course, ordinary expected utility theory also leaves one options, where the expected utility of a pair of acts is tied.

[24] Notice that we can think of this second-stage rule in quite a few ways. It might be a rationally enforced choice rule, on a par with expected utility maximization in the point-like setting. It could on the other hand simply be a description of how we cope with the optionality Weatherson diagnoses. Or it might be something in between. Here's one approach I find intriguing. Think of impermissibility as a thin normative evaluation of actions, and let Caprice be the end of the story so far as that goes. Consistent with this, there may be many more thicker evaluations of actions—actions may be biased, or reckless, or overcautious, etc. So one might think of different second-level decision rules as associated with different dimensions of evaluation—Levi's rule being an articulation of what one must do to avoid the charge of recklessness, for example.

in the presence of indeterminacy. From that perspective, Caprice seems attractive. Let's look again at our tabular representation of the Broker's offer, this time listing in a final row what each *column* taken individually recommends Alpha do:

Indeterminate cabinet	$\alpha = \omega$	$\alpha \neq \omega$
Invest	+100	-100.
Reject	+10	-90.
Verdict	**Invest**	**Reject**

According to Caprice, this would make both investing and rejecting permissible—exactly what we want, if inconstant behaviour is what we're seeking to fit. So our initial conjecture can be that *Caprice* is an appropriate decision rule for indeterminacy-induced mushy mental states.

The package of Caprice and sharpening-generation of uncertainty fits nicely with Alpha's actions being inconstant when faced with the Cabinet. After all, investing looks good on one sharpening (that on which Alpha is Omega, and hence gets the returns) and seems terrible on the other. Inconstant acts are catered for, since we may interpret them as cases where Alpha capriciously opts to act on the basis of one, or the other, sharpening.

But while the package of indeterminacy-induced mushy mental states and the caprice rule is *consistent* with inconstant behaviour in decision making under indeterminacy, the package does not *predict* it. Caprice is committed only to the permissibility of a certain range of actions. That is compatible with an agent being disposed to dogmatically stick to one particular sharpening no matter what. Such an agent is not disposed to act inconstantly. Furthermore, a capricious agent might capriciously opt for the survival sharpening if the reward is over a given amount, and opt for the non-survival sharpening iff the reward is lower. So, it appears, a capricious agent is consistent with threshold pattern of behaviour under repetitions. (Such examples could be multiplied: perhaps a pessimistic agent would capriciously opt for that action whose minimum sharpening-relative utility is highest; an optimistic one for the one whose maximum sharpening-relative utility is highest: in our case these correspond to Universal rejection and acceptance respectively). If we allow such patterns, then it seems there is the potential for a whole body of different systematic dispositions to capriciously choose. In some contexts, this might be welcome flexibility. But we wanted an account that commits to inconstancy, and so shuts off these highly patterned options.

I suggest we replace/supplement the Caprice rule with a second-order rule for selecting which member of our mushy mental state to act in accordance with—one that in effect says that we must choose which way to go *arbitrarily*. This is the Randomize rule. Given a decision situation in which one must choose whether to ϕ or ψ, if k sharpenings recommend ϕing, and $1-k$ ψing, then one should choose at random—ϕing with chance k and ψing with chance $1-k$.[25] So a given piece of behaviour will be describable in two ways—as being in accordance with Caprice as originally characterized, with the operative sharpening selected truly arbitrary;

[25] Two things to note here. First, this talks of *sharpenings* issuing recommendations, but strictly speaking it is elements of our mushy mental state that evaluate options for action. In appendix C I briefly sketch why this is an acceptable slide. Second, the formulation of Randomize given in the text is applicable only when there are only finitely many sharpenings. However, in formulating the underlying supervaluational semantics earlier, I mentioned the possibility of the package including a designated measure over the sharpenings, used to define 'degree of determinacy/truth' of sentences or propositions. I propose that in general we use this to fix the odds at which the Randomize rule, so that the Randomize rule will assign a chance of ϕing equal to the measure of sharpenings that induce mental states that recommend ϕing. For simplicity, I'll continue to work with the pretence that there are finitely many sharpenings (and that any measure is a uniform distribution over these).

or alternatively as being the result of Randomize, generating a mixed act uniquely determined by the proportions of sharpenings supporting one act over another.

What Randomize tells you to do is *choose randomly* among your original set of options. It will be very important in what follows that we clearly distinguish this advice from a closely related procedure one could follow: to decide to flip a coin to determine which option to take (perhaps you delegate the actual implementation of A and B to a friend, telling her to do A if the coin lands heads, B if it lands tails). The latter involves *choosing to randomize* between A and B. The choice to randomize over A and B is *different* from choosing A, even if (as it happens) the coin lands heads and A is brought about. By contrast, choosing at random between A and B, as I'm conceiving it, either ends up with you choosing A, or with you choosing B—it's just that you don't have a settled disposition ahead of time about which to go for. On this model, randomness is a *property* of the eventual choice (choosing at random is a way of choosing, just as choosing quickly or confidently are ways of choosing)—but it isn't part of the *content* of the choice.

The distinction between random choices and choices to randomize is a real one, as we can see from other areas in which we need to appeal to it. Suppose that we are in a 'Buridan's ass' situation where we have two cupcakes, A and B, and attach equal utility to eating either (though we cannot eat both). How should we decide what to do, given we have two equally good options? It's very natural to advise someone in this situation to make an arbitrary choice—to choose at random whether to eat A or B. But it would be perverse to interpret this piece of commonsensical advice as involving (a) the expansion of the agent's conception of the available actions to include random acts such as flipping a coin to determine what to do; and (b) the advice to opt for one of the randomizations, as opposed to simply taking cupcake A. After all, flipping the coin has *exactly the same utility as taking A in the first place*—so the expected utility of the available actions are just as tied as they were in the original scenario, except with three, rather than two, options.[26] So the expansion of options doesn't help at all in giving a principled basis for action. If one is going to end up resolving the symmetry by plumping for a particular action out of tied set of options, you could just as well have recommended (say) A at the first stage. It think it's pretty clear that this is an uncharitable interpretation of the commonsensical advice, and once we recognize one can choose randomly without choosing to randomize, we have a much better account.

Confusing randomly choosing and choosing to randomize in Buridanic situations results in rather odd interpretations of commonsense advice. It's a far worse confusion in the context of indeterminacy—since (it will turn out) flipping a coin to decide between the options left open by sharpening-relative decision tables may actually be an act that is *impermissible*. For more on this, see the end of section three.

2.4 Application to speech acts and the sorites argument

Randomize has some interesting and plausible predictions about our attitudes to the sorites paradox. The most persuasive way I know to motivate the paradox is to ask people to first

[26] Of course this is a simplification, based on assuming that the expected utility of randomizing between A and B is simply a weighted average of the utilities of A and B. But the utility of flipping a coin (or using any other randomizing device) turns in part on consequences of the coin flipping itself. If a coin-hating madman is hiding in the bushes, waiting to shoot anyone who randomizes in this way, that obviously decreases the utility of flipping.

Of course, there could be madmen (equipped with appropriate brain-scanners) who are disposed to punish those who choose randomly. As a problem for Randomize, this is comparable to situations involving madmen (equipped with brain-scanners) disposed to punish those making decisions by optimizing expected utility; or those who make decisions in cases of Levi-an uncertainty by minimizing the worst outcome. How to think about situations where *following the rules of rationality* is itself punished is a very hard (and general) issue, and I won't pursue it here.

consider claims of the following form: $Fa \wedge \neg Fa'$, where a and a' are adjacent in a sorites series. F may be 'red', and a and a' different though indiscriminable shades on the continuum from red to orange, for example. The data is that such conjunctions seem flatly false—indeed, it seems they *could not* be true. So, naturally enough, we assert their negations in each case. But that is just to endorse something equivalent to the material conditional $Fa \supset Fa'$, and chaining these together along the sorites series, we just need to keep applying modus ponens to derive a contradiction—assuming we also classify some clearly scarlet patch as red, and some clearly orange patch as not red.

So what does our current account say about this? Well, it's a story about action in the first instance. But we take it that in motivating the sorites paradox, what we are asking people to do is act—to utter words. Let's model this by supposing that when S means p, then affirming S when p, has utility $+1$; affirming S when $\neg p$, has utility -1. Mutatis mutandis for the utility of denying S. Consider some particular predication of red to a borderline colour patch (Patchy) drawn from our sorites series.

Colour judgements	Patchy is red	Patchy is not red
Affirm "Patchy is red"	$+1$	-1
Deny "Patchy is not red"	-1	$+1$
Verdict:	**Affirm**	**Deny**

According to Caprice, both affirming and denying "Patchy is red" are permissible. Once we add in randomize to the mix, we get stronger predictions. The chances of affirmation will increase with the proportion of sharpenings that make Patchy red. If Patchy is pretty reddish, then many ways of placing the cutoff will make it red, and so we expect most sharpenings to declare it red, and our model predicts a high chance of affirmation. If Patchy is flat borderline, then the chances of affirmation and denial will be evenly balanced. If Patchy is close to the clear non-red cases, the chances of affirmation will be tiny, and the chances of denial high. This all seems sensible.

An even nicer prediction of randomize given this model of speech acts is its predictions about behaviour faced with the sorites itself. Let's suppose with the same utilities, we take the choice about whether to affirm or deny one of the cut-off statements.

Cutoff judgements	Cut-off between a and a'	Cut-off elsewhere
Affirm "$Fa \wedge \neg Fa'$"	+1	-1
Deny "$Fa \wedge \neg Fa'$"	-1	+1
Verdict:	**Affirm**	**Deny**

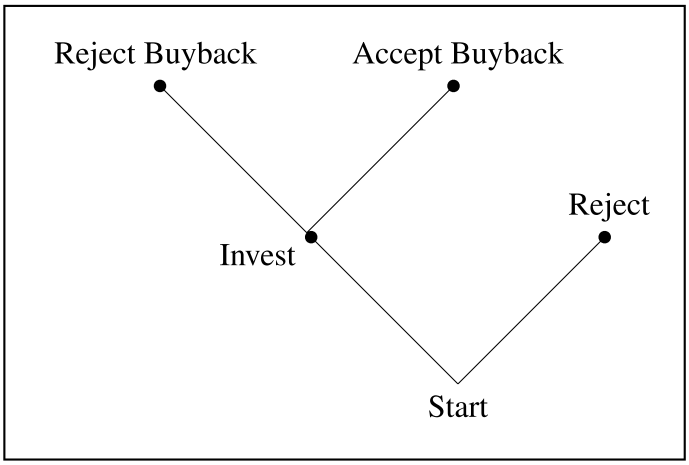
Now, the proportion of sharpenings which place the cut-off between that particular pair of indiscriminable colour patches we can reasonably take to be a tiny proportion of the available sharpenings. And so the chances of affirmation are minute, and the chances of denial almost 1. Correspondingly (I won't bother to draw out the chart this time) the chances of affirming the negation of the cutoff, or equivalently the conditional $Fa \supset Fa'$ are almost 1. So we expect agents following these roles almost always to endorse the premises of the 'longform' sorites argument. Of course, if we looked at the conjunction of all such claims, then every sharpening will think one conjunct is false, and so the whole conjunction is false. The predicted chance of judging the conjunction true is 0.

These predictions are very different from those that an epistemicist (for example) would offer. Though they should have very low confidence in the cut-off statements, they will typically demur from flat out judging it false (relative to the operative sharpenings, however, we can have confidence 1 that it is false, not merely very high credence). They also will struggle to account for the modal strengthenings of the judgements in question. The epistemicist will presumably say that for all we know, $Fa \wedge \neg Fa'$ *might* be true. There's no reason for the advocate of the current framework to agree. Indeed, I think the principled line here is to assess epistemic modal judgements in a sharpening-relative manner—under a sharpening where the cut-off for red/not-red is somewhere else, our knowledge of the colour shades of a and a' allows us to rule out $Fa \wedge \neg Fa'$ (after all, that's why on that sharpening we feel able to assert its negation). And so on such sharpenings, we should equally deny that $Fa \wedge \neg Fa'$ *might* be the case. So I think the chances of endorsing the epistemic modal statement are as low as the chances of affirming $Fa \wedge \neg Fa'$ itself.

All this supports the idea that the package developed so far is promising. But Randomize can't be quite correct as presently formulated. Well known diachronic puzzles for mushy mental states cause problems for Randomize as presently stated. However, I will argue that those same puzzles motivate a more sophisticated, diachronic version of the package.

2.5 Diachronic puzzles and hyperplanning

Here is the diachronic puzzle for Caprice.[27] Suppose the broker offers Alpha the investment opportunity at t. But if Alpha invests, the Broker is going to come back at $t+1$ to offer Alpha another deal: this time, the broker will pay Alpha half the money back (enough to throw a mediocre party before encabination), so long as Alpha agrees to alter the earlier contract so nobody gets any money. In short, Alpha faces the following decision tree:



If Alpha is at Start, then both Investing and Rejecting are permissible. Suppose Alpha invests. At $t+1$, the buyback on offer looks a good deal to Alpha exactly on the sharpening where he does not survive—so construed, it's money at no cost! On the sharpening where he does survive, Alpha is giving up thousands for a trivial gain now. So each sharpening recommends a different course of action, so again, both options are permissible by the lights of Caprice, and Randomize tells us to choose randomly amongst them.

We can verify this by writing out the decision table that faces Alpha after he has decided

[27] In fact, it doesn't affect Caprice as carefully formulated by Weatherson, but it does affect the simple form we've been working with. See (Elga, 2010) for the puzzle in the case of ordinary uncertainty, and for arguments against a range of possible reactions—including Weatherson's version of the rule (which Elga calls 'Sequence').

to Invest. Just as before, we can assign utilities to each of the various options. The utility contributed by each factor (years up to encabination; extra years; riches; the original planned party) will be exactly as before—with the addition that the mediocre party that Alpha could afford upon accepting the buyback offer produces a boost of only +5 utils; half of that attaching to the original party plans. Thus, accepting the buyback offer, on the sharpening on which Alpha is Omega, produces a net utility of +5; the resulting of summing -100 for the life to encabination; +100 for the post-encabination years; and +5 for the mediocre party. Similar calculations produce the following table:

Buyback offer	$\alpha = \omega$	$\alpha \neq \omega$
Accept Buyback	Meh party, long life, no riches: +5	Meh party, short life, no riches: -95
Reject buyback	No party, long life, riches: +100	No party, short life, no riches: -100.
Verdict	**Reject buyback**	**Accept Buyback**

As predicted, it's again a case of decision-making under indeterminacy, and Caprice allows Alpha to opt for the non-survival sharpening; hence it is permissible for Alpha to accept the later deal. Suppose Alpha does so. The net effect of Alpha's route through the deision tree is that he is down 50 dollars (or 5 utils—the difference between a great party and a mediocre one), and so is determinately operating at a loss. So it looks like we have an uncomfortable triad: permissibility of investing at t, permissibility of accepting the buyback offer at $t+1$, but the intuitive impermissibility of the conjunction of investing and accepting buyback together, which leads to a guaranteed (and determinate) loss.

One good question is what exactly the puzzle consists in. The permissibility of two actions taken individually doesn't generally entail the permissibility of their conjunction. Consider dating norms: it's permissible for A to be in a relationship with B, and permissible for A to be in a relationship with C, but typically is not permissible for A to be in relationships with both simultaneously. However, the striking thing about the current case is that the related events are *successive*, and we evaluate the later one as permissible *even bearing in mind that the earlier one has already taken place*. Given the temporal ordering, there's something weird about denying the conjunction is permissible though each conjunct is. For example: in signing the second contract, does Alpha do something permissible (qua accepting a buyback offer) or something impermissible (qua investing-and-then-accepting-buyback)?

The case is intuitively awkward, but no more, I think. Carefully described, there is nothing contradictory going on. But what we do learn is that if agents are to avoid violating norms (including norms governing conjunctions) then later actions will have to be in sync with earlier ones. Even though it's permissible in a one-off case to opt for whatever sharpening one likes, in extended chains of action, one needs to ensure that the product of the individual acts is jointly permissible. Caprice should be read as characterizing not single actions, but whole sequences, as permissible when optimal on one element within the mushy mental state. This is indeed what Weatherson's version requires.

Randomize appears to be in more trouble. If Alpha Randomizes on the first offer, he must give a 50/50 chance (say) to investing. If Alpha Randomizes on the second offer, he has a 50/50 chance of accepting the buyback offer. But since the first act has already taken place, the well-run agent will invest-and-then-buyback in a quarter of cases. But that overall course of action is uncontroversially impermissible. Randomize therefore leads to impermissible courses of action in a way that Caprice alone (due to its lack of predictive power!) does not.

The natural (and I think correct) reaction is not to abandon Randomize completely, but (like Weatherson) to switch focus from single actions to sets of actions, and to distinguish two readings of Randomize. To say, ahead of time, that the chances in the two successive acts must be 50/50 in each case *does not* mean the chances need be independent of one another. Indeed, it is compatible with the prior chance of each being 50/50 that the chance of accepting a buyback offer, given one has already invested, is zero. To get a model for this, imagine that an agent must choose the sharpening on which to act at random at the beginning of an extended series of action, and subsequently must stick with that sharpening throughout. On one reading of Randomize, the letter of its recommendation will have been satisfied.

These diagnoses and alterations to Caprice and Randomize come at considerable cost, however. No longer does our model offer advice in individual, local decision situations. Since whole courses of action are evaluated, what is recommended are whole *sequences* of action. If that has direct relevance to an agent in a choice-situation at all, it would have to be to an agent who was settling on a *plan* of action, rather than how to act in the immediate situation they face. And since we don't know which sequence of decisions situations we'll in fact be faced with, from the initial position what we'll need to choose between is complete *contingency plans*—plans of what to do in circumstances that may or may not arise. This would be of use to a fictional hyperplanner—-an agent who before taking any action whatsoever sits down to work out a complete contingency plan for any situation, choosing the optimal sequence among those that they may be faced with. In its domain, I have no wish to quarrel with its recommendations.

The trouble is that we want more from a theory of decision than advice about the normatively correct hyperplan to adopt. A theory of rational decision should return advice about the normatively correct option to take in much more local, limited decision problems. It's not in my power to now pick and commit to a hyperplan, any more than it's in my power to throw a dart twenty metres and hit a bullseye.[28] A theory of decision entailing that the thing to do in my situation was to throw the dart and hit the bullseye wouldn't be a theory with relevance to agents like me; and a theory that advised me to pick a hyperplan is similarly irrelevant.

(The point could easily be lost sight of amidst all the idealizations that are standardly made in the initial stages of a theory of decision. We work with logically omniscient agents with perfect recall, who (for example) when faced with a tree of decisions can perform backward induction reasoning of arbitrary complexity. But all these idealizations still retain their connection to real-world decision puzzles: for fixed information, goals, and available courses of action, the theory tells us what the best act is to perform. If the reasoning involved is too demanding for boundedly rational agents, that doesn't matter if we take the job to be to identify the *best* act, rather than describe how real agents should practically go about figuring out which act is best. The usual theory can be seen as an account of how my ideally rational self would advise me to act in my actual circumstances, with the options in fact open to me. To retreat to a normative theory that restricts itself to ranking hyperplans would be to give up on this project entirely.)

2.6 Incrementalizing Caprice and Randomize

What we'd like to do is incrementalize our choice procedure, allowing agents to partially commit to a course of action, leaving open what they'll do at later decision situations, while still

[28]As Michael Caie points out, one way of 'keeping track of a hyperplan' is to adopt the particular beliefs and utilities that rationalize that hyperplan. So *unless* there is already a problem with 'keeping track' of such sharp mental states, it's not clear that there is an in-principle problem. Of course, one motivating thought behind mushy credences, independently of uncertainty, is exactly the idea that it's implausible that actual agents keep track of sharp credences and utilities.

ensuring that the actions in the end are coordinated so they won't end up losing money without compensation.

Let's consider the same issue of diachronic coordination in the simpler case of Caprice. We shall assume that in the context of action there is a contextual 'score', a set that initially contains all the sharpenings.[29] When an action is carried out that is permissible on some but not all sharpenings, the score updates by eliminating those on which is it is not permissible. (As a special case, when a linguistic act is true on some but not all precisifications, those on which it is not true are eliminated). We call an action *dynamically permissible* at time t just in case it maximizes utility on some sharpening live at the score at t. The practical implementation of Caprice, is that agents should strive to make their actions dynamically permissible (as well as permissible in the absolute sense). If successful, they will have ensured that the product of their actions during the period in which the score is evolving is permissible in the absolute sense.[30]

What of Randomize? The analogous idea is that the mixed act actually implemented at t should accord with chances that evolve dynamically through the action period. Suppose that our decisions are first, between A and $\neg A$, and second, between B and $\neg B$. Suppose that AB is a sure loss outcome, and so must be avoided. The percentage of sharpenings recommending AB, $\neg AB$, $A\neg B$, $\neg A\neg B$ respectively are: $0, 0.2, 0.3, 0.5$.

The first choice is between A and $\neg A$, and in accordance with our original Randomize rule, the respective weightings are 0.3 and 0.7 (so on repetitions of the choice procedure, the relative frequency of A choices to $\neg A$ choices will be 3:7). But once the action has been carried out, just as on the Caprice model the score is updated to eliminate all sharpenings not in accord with the action taken. Suppose we end up going for the act $\neg A$ at the first choice-point. Then we face a second decision, and since all and only the sharpenings that recommended A have been eliminated, the proporitions of B-recommending sharpenings and $\neg B$-recommending sharpenings will be 2/7 and 5/7 respectively. We then locally Randomize at the second choice point at these new odds. On the other hand, suppose we chose to A at the first choice point. Then all remaining sharpenings vindicate $\neg B$, so applying the mixed strategy recipe at the second choice point we will go for $\neg B$ with chance 1.

This dynamic, local version of the Randomize rule accords perfectly with the original global Randomize rule, on the reading where it is applied once and for all before any act is taken. Dynamically, there's a 0.3 chance of getting A, and that then ensures we get $\neg B$, so the chance of ending up with $A\neg B$ is 0.3, as it should be. Likewise, there's a 0.7 chance of getting $\neg A$, and given this a 2/7 chance of B, for a 0.2 chance of $\neg AB$ overall; and a 5/7 chance of B following $\neg A$, for a 0.5 chance overall. The overall chance of getting $\neg B$ through some route or other

[29]Compare (Shapiro, 2006), the immediate inspiration for this model. Shapiro is in turn drawing on (Lewis, 1979)

[30]Compare Elga's discussion of the rule 'narrow' (of course, he is working in a rather different setting, concerned with the representation of ordinary uncertainty not the indeterminacy-induced kind). The narrowing proposal says that one's doxastic change should update upon acting, to eliminate from the representor those probability functions that would not recommend the action one in fact takes. Elga complains that this is a case where one's doxastic state updates without relevant evidence being gained. In the current setting, we have a couple of available responses. First, we can think of the doxastic state as being constant throughout the process, and a separate 'scoreboard' being updated—the sharpenings and their induced credal functions are still around in the doxastic state, but some are not 'live' in the sense that they have been eliminated from the scoreboard. It is the scoreboard, not the doxastic state, that updates and fixes what is dynamically permissible. But dynamic permissibility is only interesting because it's a way of practically ensuring that one's overall course of action is (atemporally) permissible by the the lights of the original Caprice or Randomize rule—and permissibility in this sense is fixed by the doxastic state alone. I don't see any discomfort in non-evidential updating of the scoreboard in this context. The second option is to endorse narrow, and take on Elga's complaints directly. I discuss this in the 'mind-making' interpretation of the framework in the final section, where I argue that we can identify the information that we gain that rules out the sharpenings. Thanks to NNfor pressing me on these points.

is 0.5, and likewise for B. The incremental recipe is a practical way of (locally, dynamically) implementing the mixed strategies at their original ratios.[31] This is no coincidence. Effectively, given initial chances Ch, the chance of choosing A at the first choice point is $Ch(A)$, and the chance of choosing $\neg A$ is $Ch(\neg A)$. If the first is realized, the chance of choosing B at time 2 is $Ch(B|A)$. If the second is realized, the chance of choosing B at time 2 is $Ch(B|\neg A)$. The overall chance of B is therefore $Ch(A)Ch(B|A) + Ch(\neg A)Ch(B|\neg A)$, which by the law of total probability, is simply $Ch(B)$. So the diachronic version of Randomize is a way of implementing the original atemporal Randomize strategy.

2.7 Application to the forced march sorites

I've just been arguing that for practical purposes, at least, Randomize needs to be construed as invoking chances for judging this way or that, evolving over time. This is a prediction of the account: how well does it fit data? We can test this by thinking again about the special case of linguistic action. I'll adopt the same model as before, where utilities for linguistic action are assigned based on the truth value of the claim made.

Consider the forced march sorites. This might again concern a set of colour patches, pairwise indiscriminable, taken at small intervals from the continuum from red to orange. The forced march process is to ask a subject to classify a given colour patch as red or not. One starts with clear red cases, and then marches down the sorites asking about the redness of successive patches. Presumably the subject starts by classifying scarlet patch as red. The first time she calls a patch not-red, she will have done something that seems utterly unprincipled—after judging the last patch red, she has now judged an indiscriminable one not-red. Taking the judgements together, she has effectively committed herself to the location of the red/not-red cutoff. There are many things to say about the forced march sorites, but the immediate phenomenology that everyone needs to account for is that there is enormous felt pressure to judge the *next* colour patch red, given one has judged the last one red. And this is so even when, if you were given the same patch 'cold', you'd more likely reach a different verdict. Past history seems to bias your judgement. How?

Let's build a toy model of the situation. Suppose that 100 borderline cases of redness are indexed from 1 to 99, from almost-determinately-red to almost-determinately-not-red. There are 100 ways of drawing the cut-off, and we treat these as the initial set of sharpenings. One is then asked, of patch 0, whether it is red. On 99/100 of these cutoffs, it counts as red, so almost certainly one will judge it red. The score is updated to remove the one cutoff inconsistent with this verdict. The forced marcher then moves to patch 2. On 98/99 of the remaining cutoffs, it counts as red. So chances are it'll be judged accordingly. The successive choices then give odds of finding it red 97/98, 96/97, and so on... in fact, even when one has gone almost completely through the sorites, and classified 90 of the patches as red, the chances of classifying the (distinctly orangey) 91st patch as red will still be high: 9/10. The odds don't get very low till the last few patches: the 97th patch has a 3/4 chance of counting as red, the 98th patch 2/3; and the 99th 1/2. But of course, if the forced march had been run the other way, we could get into a situation where there's a 3/4 chance that the third patch would be judged as not-red. So the account accounts for forced-march sorites in an extremely strong fashion.

It's worth emphasizing again that the chances just derived are entirely consistent with the low categorical chances assigned to judging one of the nearly-orange colour patches as red. For in each move in the forced march sorites, there is a tiny but real chance of defecting. The

[31] In the setting with infinitely many sharpenings, and consequently a measure over them, we simply need to conditionalize the measure to remove the sharpenings that have been eliminated, and then used this derived measure over the live score to set the 'local' chances for the mixed acts.

chance of judging an almost-orange colour patch red is high *conditionally on getting that far down the forced march sorites*. But the chances of getting that far in the first place is low, since the chances of defection at some previous point is, in aggregate, high. And if one has already defected, one is committed to a sharpening which classifies the nearly-orange patch as not-red. So everything works out smoothly.

This account of the forced march is strong and striking. Some caveats: it's not obvious that it makes the correct empirical predictions. For example, you might think that the conditional chances towards the end are too high. Also, a kind of retraction phenomenon has often been noted—-that is, once one has defected and called a given patch not-red, then if one is asked to reclassify the last few cases, one will start labelling them not-red.[32] It's not obvious how to fit this into the current model, which is built to ensure diachronic consistency. However, the model we've been working with is just a baseline, and could be elaborated in numerous ways—by varying the underlying measure, by accounting for pragmatic features and so forth.[33] What I found impressive on first encounter with it is the way it captured the central feature of the forced march—that felt pressure to *keep* calling patches red, despite being designed with completely different problems in mind.

Intermediate conclusion

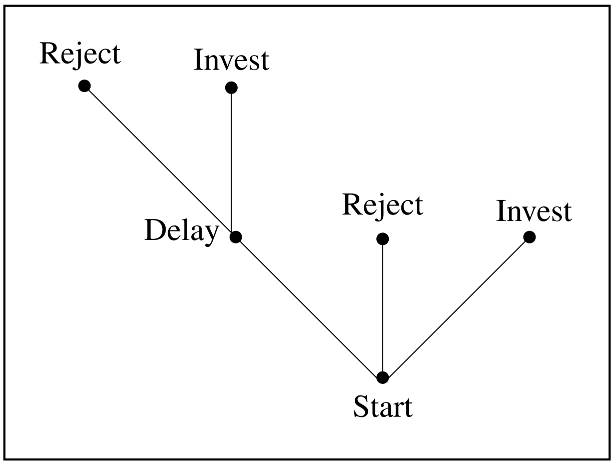
The first two sections of this section developed an account of indeterminacy-induced uncertainty. Drawing on the literature on mushy credence, this was represented as a set of mental states an agent is 'open to'. This was linked to an account of acting under (this kind of) uncertainty: Caprice (the claim that any act recommended by any sharpening is permissible to perform) and Randomize (the positive claim that one should select randomly among the acts declared permissible by the criteria just mentioned). Considerations of diachronic coherence motivated, initially, a restriction of these criteria to choices among courses of action rather than individual actions; and then an incrementalization to localized decision situations. Applications to the argument-form and forced-march sorites demonstrate the wider interest and plausibility of the framework.

3 Elga's puzzle

Everything so far looks rosey for the inconstancy view, as underpinned by Caprice and Randomize. But Adam Elga (in commentary on an earlier version of this essay) put forward a puzzle case that is well worth exploring in depth. Elga's puzzle goes like this. Suppose Broker offered Alpha a slightly different deal: he can either invest or refuse the contract, at fixed price, but he has an extra option: Alpha can take a few extra dollars and be presented with the invest/refuse decision again in five minutes (that time without the possibility of further delay). He faces a decision tree that looks like this:

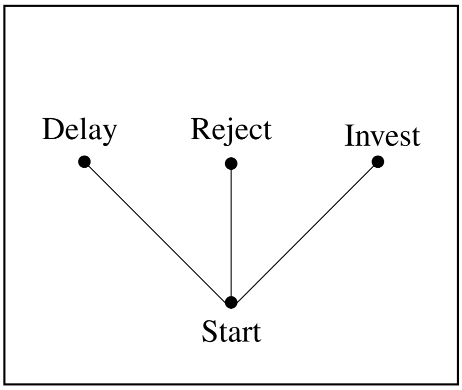
[32] See the discussion of 'backward spread' in Raffman (1994).

[33] Shapiro (2006) explores adding some features onto a model of vagueness broadly similar to the one I'm exploring here.



The payoffs at the lower Invest and Reject leaves are just as before. At the Delay node, the extra dollars are given, and a second choice presented. So the upper Reject and Invest leaves have payoffs that match those at the lower leaves, with the addition of a few dollars. Taking the delay option seems a no-brainer. After all, Alpha won't close off any of the rival options, and he'll gain some money whichever way he goes.

But prima facie our account tells Alpha not to delay at the first choice point he comes to (the node labelled 'Start', above). At that moment, he faces a choice between three options:



Alpha is self-aware, he knows that he's implementing the Randomize rule. So he knows that there's (say) a fifty-fifty chance of him investing at the later time; and a fifty-fifty-chance of him refusing at the later time. That means he's in a position to calculate the expected utility of the 'Delay' option: the expected value of his delaying is 0.5 times the expected value of investing; plus 0.5 times the expected value of refusing.

But relative to *any* element of his mushy mental state, Delaying looks like a bad deal! Relative to the sharpening on which he would judge that he is Omega, the expected value of investing (right now) far exceeds that of refusing (right now). But it also exceeds the expected value of delaying, since that runs a 50/50 risk of ending up doing the bad thing. Relative to the other sharpening, on which he would judge that he is not Omega, the expected value of refusing the contract (right now) exceeds that of investing (right now). And again, refusing the contract beats delaying, which runs the risk of ending up paying money from (what relative to this stance seems) no return.

Let's run through this more slowly. Here, again, is the original decision table, with columns representing the sharpening of the crucial indeterminacy: whether Alpha is Omega. As is now familiar, on the first column investing maximizes expected utility. On the second column, rejecting does so. And the inconstancy model we have developed says that either of these options therefore permissible to take over (contrasting with models that would try to aggregate the columns e.g. by taking a suitable weighted average across them of the utilities assigned by the column to each action).

Broker's offer	$\alpha = \omega$	$\alpha \neq \omega$
Invest	+100	-100
Reject	+10	-90
Verdict:	**Invest**	**Reject**

When we add in the "delay" option, we have to decide what utility boost the extra dollars produce. Presumably a positive one: let's just assume it adds a single utile. The table looks schematically as follows:

Delay puzzle	$\alpha = \omega$	$\alpha \neq \omega$
Invest now	$+100$	-100
Reject now	$+10$	-90
Delay	future choice+1	future choice+1
Verdict:	**?**	**?**

The utilities of the third row are determined by whatever the expected utility of you facing the *original* decision table in a minute's time—+1 for the dollars you are paid for your time. Since you know you will randomize at the future time, the chance of you going with the Invest verdict is 50/50, as is the chance of you going with the Reject verdict. That gives us all the information we need to calculate the expected utility:

Delay puzzle	$\alpha = \omega$	$\alpha \neq \omega$
Invest now	+100	-100
Reject now	+10	-90
Delay	$(0.5 \times 100 + 0.5 \times 10) + 1$ $= +56$	$(0.5 \times -100 + 0.5 \times -90) + 1$ $= -94$
Verdict:	**Invest now**	**Reject now**

So while the sharpenings disagree among themselves about what is to be done, they both condemn the 'wait and see' option. By the criteria we have been working with, therefore, delaying is impermissible.

To repeat: that seems like the wrong advice. Intuitively, delaying dominates taking the decision to invest or reject immediately: whatever you go for, you could have got a better deal by waiting. And our model of rationality also backs that up. Recall that our initial response to buy-back problems was to evaluate sequences of action spread over time for rational permissibility. From the atemporal perspective, a total course of action counted as permissible iff there was some sharpening that, for every decision situation, recommended the relevant part of the total course of action. The sequence of actions that for Alpha secures maximal utility on the $\alpha = \omega$ sharpening is to first delay, and then invest. The course of actions that secures maximal utility on the other sharpenings is to first delay, and then reject. So by our original criterion, the only permissible courses of action are the ones where we should delay (and we should randomly select which one to go with).

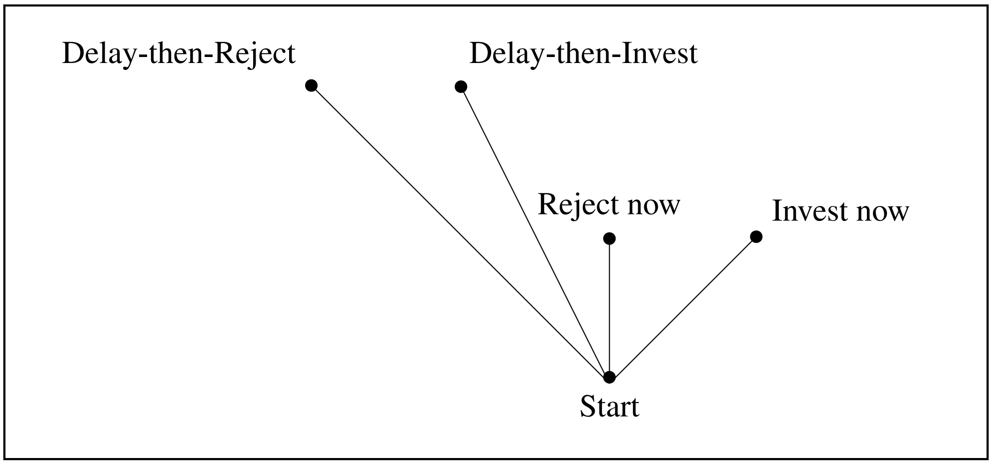
So what Elga's puzzle demonstrates is a prima facie inconsistency between the original atemporal perspective which evaluated whole courses of action (routes through a tree), and the way that we 'incrementalized' that in order to have something to say to ordinary agents about the rational choices to make in local situations (nodes in that tree). It looks like in order to achieve a rationally permissible sequence of actions, you would have to do something rationally impermissible at the first node. And that is not something I can live with in a theory of decision.[34]

How should a defender of the inconstancy model respond to this challenge? The major issue is what one thinks the *correct* verdict to be. Shall we argue that (by the lights of the inconstancy model) delaying is forbidden, and this is the correct result? Or shall we argue that the correct decision is to delay, and then find a way to get the inconstancy model to align with this, the argument to the contrary above notwithstanding? I go for the latter option. The justification for introducing an 'incrementalized' version of Randomize was to show what one would have to do at a local, choice-by-choice level, to ensure that one's course of actions collectively was a permissible one. From that perspective, Elga's challenge shows that the theory of individual, local choices introduced does not succeed in its aim, and needs to be rethought.

I will put forward four lines of response to Elga's puzzle, which make a case that suitably understood or tweaked, the incrementalized decision procedure will recommended delaying in the choice situation we have described.

3.1 Planning

When faced with the delay puzzle as originally formulated, it is tempting to think that the options for action it gives are underspecified. After all, I could make up my mind now for what I will do in a minute's time. If that's the right way to think about the choices I'm making at the node labelled 'Start', then our earlier decision tree misrepresented the situation; it should have been written like this:



Correspondingly, when evaluating the choice being made at Start, the critical third row of the decision fragments:

[34] In a sense, what we have here is an inconsistency between the theory of games (including extended single-player games in 'extensive form', as in our decision trees; and the theory of rational decision. Of course, both have been adapted here to the problematic cases of decision making under indeterminacy, so I do not claim that such tensions arise independently. But of course, this is an instance of a more general issue, that could in principle arise without indeterminacy. Here's a straightforward example: Jeffrey's evidential decision theory recommends cooperation in suitable Prisoner's dilemma's scenarios—orthodoxy in game theory recommends the opposite. That *is* a genuine tension, and one often resolved by reformulating the decision theory in causal terms.

Delay puzzle (extended)	$\alpha = \omega$	$\alpha \neq \omega$
Invest now	+100	−100
Reject now	+10	−90
Delay-then-invest	+101	−99
Delay-then-reject	+11	−89
Verdict:	**Delay-then-Invest**	**Delay-then-Reject**

In this presentation, delaying is recommended by both sharpenings, because they recommend different (and incompatible) determinants of (what is now conceived of as) that determinable action.

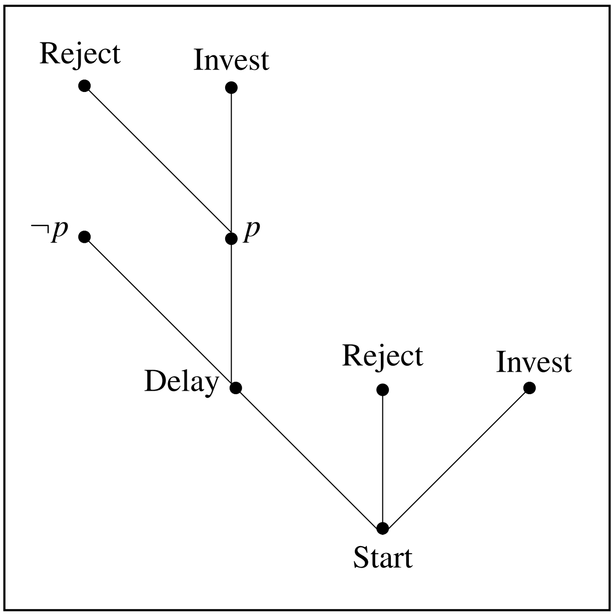
What this response highlights is an open issue in the application of decision theory—the individuation of the options we assess for expected utility, and among which we select the best—-or likewise, within the theory of games, what settles what 'extensive form' is the appropriate description of a given situation of sequential choice.[35] I'm personally quite sympathetic to the idea that these options might include descriptions of how we act at some temporal distance—we needn't think of them as something we implement immediately.[36]

So I think this might be exactly the right response to the puzzle as formulated. But we miss the force of the underlying puzzle if we rest content with it, because it won't generalize to the broader category of situations of which Elga's is a simple instance. Consider variants of the delay puzzle in which the decision to delay has different consequences. Perhaps, instead of guaranteeing that you'll be offered the same contract again (plus a few dollars), it just gives you a chance (if some scenario p comes about) of receiving a much-enhanced offer—one that gives millions to Omega, for the same small sacrifice on Alpha's part. If we choose the chances and monetary reward correctly, we can make the expected utility of this prospect equal to that of the original delay puzzle. And indeed, we can set it up to get exactly the delay puzzle structure. The natural description of the decision tree is the following, where at the delay node, 'nature' chooses whether p or $\neg p$ obtains, at fixed chances:

[35]Just to be clear here—I'm not saying that whenever game theory gives us a criterion for selecting an optimal 'strategy', then we should think of that as a game whose extensive form consists of a single node, with a branch for each strategy. In the terminology I'm using here, I see the crtierion of strategy selection in extensive form as telling us what sequence of choices are the best to take. Each node then represents a separate choice. What is represented here is a situation in which the agent literally chooses (or 'plans') how to act right at the start.

[36]Of course, overly fine-grained descriptions of actions can make nonsense of the decision tables. Should I take a gamble that I will hit the bullseye with a single dart, under pain of torture if I miss? If my option space includes 'taking the gamble and hitting the bullseye' and 'taking the gamble and missing the bullseye', then decision theory will return the verdict I should do the first, and so a fortiori take the gamble. But if I'm bad at darts and hitting the bullseye is not in my control, this seems like bad advice. Since one cannot right now guarantee that one's future self will have to do one thing rather than another, the latter two options may seem illegitimately fine-grained.

Here's one perspective on decision tables which distinguishes the bullseye option from the delay-then-act ones. Suppose that the subject matter of decision theory is not in the first instance a relation between beliefs, desires, and acts, but belief, desires, and intention-formation. Presumably the bad dartplayer cannot form an intention to hit the bullseye, if she is self-aware enough to know that this isn't under her control. But prima facie, one can form intentions, or plans, about your future acts—you control your future behaviour in a relevant sense, even if you can't guarantee you'll follow-through on your current decision. So the more fine-grained decision table above is appropriate, and the analogy to the dart case is inapt.



If our options for action at Start in this tweaked version of Elga's puzzle situation were to invest-now, reject-now or delay (as in the situation depicted above), then we have the original puzzle all over again—delaying would be ruled out. But again, this would be inconsistent with the complete course of action that each sharpening recommends from an atemporal perspective. So to generalize the solution that appeals to an enriched choice at Start, we need to redescribe the options available to the agent at Start, as involving inter alia delaying-then-accepting-if-p; or delaying-then-rejecting-if-p. Each sharpening will recommend one of these, over investing or rejecting immediately. But p could be as recherché as one likes: it could be, for example, that the United Nations in special session votes out the offer.

In ordinary situations, we are simply not choosing between plans that make explicit provision for such whacky possibilities (and this is the tip of the iceberg, since we can complicate the array of contingencies and future decisions involved arbitrarily)—only superhuman hyper-planners could face decisions situations where they really choose between such options. And to emphasize a point argued earlier: a theory of rational decision is supposed to tell us (inter alia) about the normatively correct choice *between those options open to finite, limited agents like us*, relative to our information and goals. Sure, the reasoning that determines that one among these options is the best may be beyond our capacities; but still, it is what our idealized advisor would tell us to do, and so remains tied to the question of what the *best* thing to do is in our actual situation. By contrast, advice to take options that is beyond our cognitive capacities to secure is just as irrelevant to us as advice to take options that it is beyond our physical capacities to secure.

3.2 Incremental hyperplanning

In the glosses earlier I talked about sharpenings "recommending actions" in particular decision situations. The permissible actions in a given situations are those "recommended by at least one sharpening". We need to think carefully about what this means.

Earlier, we noted that each sharpening induces a particular *degree of belief assignment* for Alpha over all propositions; and equally a *utility assignment*.[37] Put these together, and expected utility theory delivers recommendations for action in decision situations. That is—as well as credences and utilities, the sharpening recommends *choices* in decision situations.

[37] Recall: in the general setting, this was only so when we assume that all uncertainty in our mushy mental state was indeterminacy-induced. But that is perfectly fair assumption in the present dialectic.

But of course, we face the question: what are the relevant decision situations? We met in previous sections the idea of a hyperplan—something that delivers a recommendation for action in arbitrary decision situations we may meet. As we have seen, one way that sharpening-relative credences and utilities can be applied is in evaluating and choosing hyperplans. And that is all fine, in its place—but we wanted to generalize the theory so it talks about actual choice-situations we face, which involve much coarser grains option for action. The proposal was to incrementalize choices by introducing a "scoreboard", on which "live" sharpenings are recorded. The dynamically permissible courses of action when choosing whether to invest, reject or delay should be those "recommended by some live sharpening".

Elga's puzzle arises when we read this in one very natural way—-if we read off a given sharpening credence and utility functions, and then apply expected utility theory to the local decision problem (one node in a decision tree). But an alternative interpretation is the following: each sharpening recommends a hyperplan. And what a given sharpening recommends about the decision situation at node n is simply what the corresponding hyperplan recommends at n. But all hyperplans (as we saw) recommend delaying, so this is the only permissible course of action in that choice situation.

That model is still incremental, giving advice to agents who need to choose between relatively short-term local options rather than complete hyperplans. The constraint is just that any local choice that an agent makes be consistent with all the "live hyperplans". It is not subject to the criticisms of the previous section, since the options which are assessed for permissibility in a decision situation remain as localized as one wishes.

I think just this could be a stable resting place and response to Elga; but it would have significant costs. The trouble is that ordinary practical reasoning explanations of action don't appeal to fit with a broad hyperplan—they talk about what one currently believes and desires, and how this relates to the options available to you. The decision tables of ordinary expected utility theory regiment this nicely. This leads to two puzzles. First, if the response to Elga's puzzle is simply that this whole model of local applications of expected utility theory is inapplicable, and we must instead appeal to the deliverances of rationalizable hyperplans, then we lose a connection between what makes a decision rational and the (local) elements of ordinary practical reasoning. Second, we have on the table an argument that delaying was impermissible, which simply used (regimentations) of ordinary norms of practical reasoning. And while one could respond by baldly declaring that such reasoning is inappropriate in localized situations, that's terribly unsatisfying. We can do better.

3.3 Conditionalizing on future choices

Can we get the same results as the incremental hyperplanning account, but via the kind of local decision tables and local expected utility calculations that Elga uses? The trouble with the delay puzzle is that (we reasonably suppose) each induced credal state will include a 0.5 credence that the post-delay decision will be to invest (/reject) given the known 0.5 chance of making that choice in that situation. But there is a gap from c assigning probability 1 to it being chance 0.5 that P, and c assigning 0.5 probability in P itself. Bridging this requires we assume a norm such as Lewis's Principal Principle.

But the Principal Principle is not a general truth about all probability functions. And in particular, it doesn't hold under conditionalization. Even if the induced credences c obey this principle, the result of conditionalizing c on P, for example, will violate it.

So here is one way of getting the expected utility calculations to come out correctly. Rather than calculating them relative to the mental state induced by sharpening s directly, calculate them based on that mental state *updated by conditionalizing on the proposition that in future* *decision situations the agent will act as s recommends*. And so the $\alpha = \omega$ sharpening, for example, will assign probability 1 to one investing after the delay, since that it is what it would recommend in that situation. If those are the rules of the game, then we can write down local decision tables for the delay puzzle, with the (desirable) result that delaying is recommended:

Delay puzzle	$\alpha = \omega$	$\alpha \neq \omega$
Invest now	+100	-100
Reject now	+10	-90
Delay	$(1 \times 100 + 0 \times 0) + 1$ $= +101$	$(0 \times -100 + 1 \times -90) + 1$ $= -89$
Verdict:	**Delay**	**Delay**

Now, the obvious worry is that though the rule that determines the entries in the crucial cells of the third row has the form of expected utility calculations, the content is different—and objectionably ad hoc. What could justify discarding one's actual assessment of the probabilities in favour of conditionalizing on propositions you are at best agnostic about? The move has precedent: Jeffrey's evidential decision theory and various versions of causal decision theory do not weight the utility of outcomes by one's actual assessment of the probability that they will come about, but by that probability updated (in one sense or another) on the proposition that one takes the action in question.[38] But nevertheless, I do think that the current proposal remains ad hoc.

If one wished to defend this approach, I think the proper way to think about this is the following. The real story about what rationality recommends is given by the story about (incremental) hyperplanning described earlier. And what we need to do is show what impact this has on local decision tables. The principled (expected utility) story involve belief-desire pairs selecting hyperplans, sharpening by sharpening. The use of conditionalized credences to make ersatz expected utility calculations is instrumentally justified as a recipe that accords with the principled story, but is applicable to local decision situations.

In the end, therefore, I think of this proposal at it stands as not an *alternative* to the incremental hyperplanning of the previous section, but as a way of elaborating that with a kind of *error theory* of ordinary practical reasoning—one which comes equipped with instructions for tweaking usual practice so that it delivers the right results. Our final option for analyzing the situation will be far more ambitious.

3.4 Mind-making

I want to consider, finally, whether we can respond directly to the delay puzzle in its own terms. Let's suppose that we are faced with decision tree as originally formulated (pace hyperplanning); that decision theory is directly applicable to the choice faced at each node (pace incremental hyperplanning); and that they should be calculated relative to one's categorical credences (pace conditionalization on future choices). Is there any way to avoid the anti-delay conclusion?

I think there is, but it requires a far more committal interpretation of our framework than anything we've needed so far. I can't fully explore and defend it here, but I'll set it out and sketch its application.

The key thought will be this. It's quite standard in analyzing an agent's behaviour in decision

[38] See (Jeffrey, 1965) and (Joyce, 1999) respectively.

trees to assume that the agents know that they will remain rational throughout the period. After all, if I get an extra few dollars for delaying a choice, I shouldn't take it if there's a high risk that I'll go crazy in the meantime and choose the wrong option later. So in modelling our decision situations, we should assume that Alpha knows he'll be rational throughout the period in question.

On the other hand, it's not so clear what the content of this is. One option is to identify it as saying: Alpha always make a permissible choice; always randomizes as recommended. That's natural enough, and leaves us no further on than before. But here's a different suggestion. Let's suppose that the indeterminacy in whether Alpha is Omega translates to an indeterminacy in whether Alpha *should* believe that he is Omega. In other words, what the set of sharpening-induced mental states represent are various candidate mental states, such that it's indeterminate which of them rationality recommends Alpha adopt. Accordingly, it's indeterminate which actions rationality recommends when Alpha gets to a decision situation. Despite it being indeterminate what Alpha should think and do, the decision situation forces him to opt for one or another action. So he's stuck.

All Alpha has to cling to is this: any course of action licensed by some sharpening is not *determinately* irrational. Suppose Alpha performs some act A—something recommended as rationally required by some sharpenings (s') and condemned as rationally forbidden on others (s''). Then anyone who criticized Alpha as irrational would themselves be doing something that is appropriate only on *some* sharpenings (s'') and not on others (s'). That is, the critic would be manifesting an ungrounded attachment to one of the sharpenings of the indeterminacy *other* than that which Alpha chose to listen to. So acting in the way that a sharpening recommends has at least this virtue: it grants an immunity from *neutral* rational criticism.

That gives an alternative foundation for the idea of 'Permissible' actions, as used throughout this paper. Rather than viewing such acts as (determinately) rational, we regard them only as enjoying the immunity just identified—something that still makes them preferable to any act that is determinately *irrational*. In law, certain acts may receive no punishment, irrespective of whether they remain illegal. The option I am exploring here is that our 'permitted' acts have an analogous decriminalized status, which should not be confused with the question of whether they are legal/rational.

Against that background, let us follow through the consequences of assuming agents believe themselves to be rational in choice situations. Consider the post-delay situation. It's indeterminate what rationality recommends there. But the credences induced by the $\alpha = \omega$ sharpening will say that rationality requires investing; and hence (given that each credence assumes that the agent will do what is rational) those credences will assign full probability to the agent investing later on. Correspondingly, the $\alpha \neq \omega$ sharpening holds that rationality requires rejection of the offer; and will assign full probability to the agent rejecting the post-delay offer. But that's exactly the situation we need to get the right recommendations in the delay puzzle!

This is an attractive place to end up. But what costs do we incur in getting there? Well, note that the elements of your mushy mental state will feature levels of confidence in propositions out of line with their known chances of coming about. You know that the chance of your investing is 0.5. And yet you are uncertain (in the sharpening-relative sense) about whether you will invest—1 or 0 in this proposition on different sharpenings. In any case, your credences are not aligned to the known chances. But such violations are not a surprise. By making the rationality assumption, we are including assumptions about the results of future chancy processes–"inadmissible evidence" in Lewis's terminology. Exceptions to the Principal Principle such as this were already provided for in Lewis's discussion of the rule.[39]

[39](Lewis, 1980).

Even if we can coherently make the agent confident of her own rationality in future choices, can we say the same about her past choices? Suppose the agent faces the original broker situation, and decides to invest. Then the mental states appropriate to some sharpenings will declare that action irrational—but since she has the information that she did it, all the sharpenings need to give high credence to the proposition that she invested, and so (on the sharpening on which this is irrational) that she did something irrational.

Note that the dynamics of incremental choice outlined earlier ensure that only *dead* sharpenings (sharpenings no longer on the scoreboard) will condemn the agents' past choices in this way. After all, the rule which kills them off is precisely that they fail to rationalize something the agent does. But that doesn't speak to the concern, which was that the agent should believe herself rational. All sharpenings (practically relevant or not) should agree on this—but it seems they don't. One could, I suppose, restrict the rationality assumption to the deliverances of future choices. But this would both be ad hoc, and also lead to a strange epistemic state: you (determinately) believe in the future you will always act rationally, even though your inductive evidence is at best mixed.

What this forces us to do is think carefully about the significance of the scoreboard, and what we are doing when "killing off" sharpenings. I've been careful to be neutral about it so far. But it is open to at least two interpretations. The minimal interpretation has us having a constant mental state of uncertainty, when indeterminacy arises. For practical purposes, we may need to act in ways that are as-if we had formed a belief about the question at hand (whether Alpha is Omega, say)—and indeed, form a disposition to act as-if this were the case in future, to ensure diachronic consistency between our actions. But on this view, this is simply a point about how we act, and we do not *actually* change our doxastic state or attitude to the question about whether Alpha is Omega. The scoreboard is just bookkeeping.

On the other hand, we could think of the process as genuinely one of making up our mind on an indeterminate matter. To be sure, the mind-making would be groundless and arbitrary, but what results is an updated and less uncertain belief state. On this interpretation, we really do change our attitudes over time when forced by circumstance.

It is this mind-making interpretation of the framework that I think is necessary to make the rationality-response to the delay puzzle cohere. For note that if one has *made up one's mind* after acting in a certain way (investing, say), so that one's indeterminacy-related uncertainty is now captured only by the *live* sharpenings, then all the precise credences and utilities one is open to will commend your past actions as rational—and so there is an inductive fit between your attitudes to the past and the future. Indeed, the rationality assumption becomes the mechanism that can drive the incremental process of mind-making in the first place. Given the assumption that you act rationally, and your knowledge that you have acted thus-and-so, you need to update your mental state by eliminating sharpening-relative mental states incompatible with the new information.

The mind-making interpretation is far more controversial than the previous alternatives.[40] I'm somewhat attracted to it, but I wouldn't want the interest of the framework described here to rest upon it. It's good, therefore, that we have an alternative responses to the delay puzzle available—the incremental hyperplanning proposal—even if it cannot embrace expected utility reasoning as fullbloodedly as mind-making can.

[40] Though compare the treatment of imprecise/mushy credences in Jeffrey (1992), where a similar 'mind-making' is appealed to, though one free of the kind of rich rational constraints that I've been appealing to in the case of indeterminacy.

3.5 Choosing to randomize and randomly choosing

The delay puzzle has been defused. But there's a way of tweaking the example that makes it reveal one last striking feature of decision making under indeterminacy, which relates to the distinction we drew early on between choosing randomly and choosing to randomize. I finish by sketching this.

Consider the following variation on the delay puzzle: faced with the initial offer from the broker, rather than Alpha being faced with a choice between investing-now, rejecting-now and delaying, Alpha is faced with the choice between investing-now, rejecting-now, and letting the invest/reject result be settled by the flip of a fair coin (as inducement, if he chooses to flip, he'll get an extra few dollars). Call this the flip puzzle. Notice two things. First, the situation is 'consequentially equivalent' to the delay puzzle—the chances of outcomes on corresponding options in the two puzzles are the same. After all, Alpha is perfectly aware that if he delays, he will later randomly choose, at 50/50 odds, between investing and rejecting. The coin just automates this.

But faced with the flip decision situation, there's no question what our model recommends. The table looks like this:

Flip puzzle	$\alpha = \omega$	$\alpha \neq \omega$
Invest now	$+100$	-100
Reject now	$+10$	-90
Flip	$(0.5 \times 100 + 0.5 \times 10) + 1$ $= +56$	$(0.5 \times -100 + 0.5 \times -90) + 1$ $= -94$
Verdict:	**Invest now**	**Reject now**

Since there are no future choices involved, no appealing to hyperplans or rational future actions will change the result. Unambiguously, choosing to randomize is rejected.

To press home the differential treatment of random choices and choices to randomize, consider a variant where we have to choose whether to delay for a few minutes and then choose between investing and rejecting, vs. flipping a coin now to determine the result. We'll assume that one of the four analyses of the delay option outlined earlier is correct:

Delay or Flip	$\alpha = \omega$	$\alpha \neq \omega$
Delay	$+100$	-90
Flip	$(0.5 \times 10,000 + 0.5 \times 10) + 1$ $= +56$	$(0.5 \times -100 + 0.5 \times -90) + 1$ $= -94$
Verdict:	**Delay**	**Delay**

Randomly choosing is determinately preferred to flipping a coin, even though the two give rise to the same outcomes with the same chances. There's a kind of bias to authentic, first-person decision-making in the model, which seems both intriguing and puzzling.

I don't see any argument that this is the wrong result, so I regard this as a feature (and strong prediction) of the model, rather than a bug. I want to make just a couple of points about the situation.

First, thinking of randomly choosing and choosing to randomize as consequentially equivalent may be misleading. In single decision situations, as above, they do lead to the same outcomes with the same chances. But when embedded in decision trees, they are inequivalent.

To see this, consider the buyback offer mentioned earlier. One first faces the decision to invest or reject the brokers offer; but then later (if one invests) one gets an offer from the broker to cancel the obligations to Omega, for a partial refund of the purchase price. If one takes the decision oneself at the first choice point whether to invest or reject, we can be assured that one will not end up first investing and then selling it back at a loss. Our model was designed to ensure this was so: those sharpenings which would favour cancelling the contract (those that have $\alpha \neq \omega$) are killed off after the choice to invest, as not rationalizing the action just taken—so it will not be dynamically permissible at the future time to accept the buyback.

On the other hand, suppose one opted to randomize at the earlier point and ended up investing. The action you took (flipping a coin) is itself even-handed between the recommendations of the two sharpenings. You don't have grounds based on what you did, to eliminate one sharpening while keeping the other on the scoreboard. So this process doesn't get you an unambiguous recommendation to decline the buyback offer at the later time. The point is easiest to grasp under the final, mind-making interpretation of our machinery. Randomly choosing is, inter alia, to make up one's mind about a question, in a way that informs and constrains future choices. But choosing to randomize doesn't involve making up one's mind—it's best understood as an attempt to stay neutral. As we've seen, this difference matters to the outcomes you secure after whole courses of actions; so it's a good thing that our framework marks the difference.

Intermediate conclusion

Responding to Elga's delay puzzle requires us to get precise on exactly how our model of decision making under indeterminacy should apply in extended decision situations. I see two main stable positions.

The first is incremental hyperplanning. Here, decision theory operates primarily at the level at which hyperplans, rather than local decisions, are recommended. But we, finite, agents can delay committing to a particular permissible hyperplan as long as is convenient. The mushy mental states of an agent can be constant throughout a decision tree, and the scoreboard on which live sharpenings are recorded and killed off, is a notational tool and a record of changing dispositions to enact one's beliefs and desires, rather than a reflection of any underlying change of view.

The second is the mind-making picture of the evolution of mushy mental states. Here we find a far more committal picture on which in choosing how to act, we make a judgement call in the fullest sense—we form a new belief. By making up our mind in this way (through a process we know to be arbitrary and groundless—but enforced by circumstance) over time we precisify our picture of the world. I've sketched how this could be presented as a response to evidence—under the presupposition that our actions are rational, the data that we have acted this way or that implies that this or that (respectively) are rational ways to act—-data inconsistent with remaining open to sharpenings that say otherwise.

Conclusion

Alpha faced a decision puzzle. Should he invest for Omega's benefit, when it is indeterminate whether he was Omega? The upshot of the model of decision making under indeterminacy developed here is to recommend Alpha take a judgement call (in either a thin, incremental hyperplanning sense; or a thick, mind-making, sense). The judgement leads to an act, to invest or not to invest. Over time, coordination ensures that Alpha does not act inconstantly—once he has decided to act against the recommendations of a sharpening, he is disposed to disregard its advice in future. But his initial dispositions are inconstant, in a way that would be manifest under repeat trials of his duplicates, or in memory-wipe and retest scenarios.[41]

Our model provides rational underpinnings for inconstant action. It also has welcome predictions for the sorites paradox. With overwhelmingly likelihood, we declare cut-off statements flatly false—and we need not say that they are even epistemically possible. And we can account for the felt pressure to judge a patch red, conditionally on having judged its neighbour red a moment previously. To be sure, there's some chance that we permissibly draw a line, or defect during a particular stage during the forced march. But that is fine, I think: that we *sometimes* draw lines or defect is as much data as is the overwhelming rejection of such courses of action for the most part.

The model of action under indeterminacy-induced uncertainty is profoundly different from the model of acting under risk or uncertainty we are most familiar with. It refuses to hedge or aggregate desirability across the possibilities one is open to. This has practical consequences. If a patient is borderline between life and death, and one has to choose between no treatment (best in the case of already-dead) expensive-and-excellent treatment (best in case of life) and mediocre-but-cheap treatment, then *epistemic* uncertainty over whether the patient is alive or dead would aggregate the risks and rewards. The chances of helping a live patient must be weighed against the opportunity-costs of wasting money that could have helped others. It is easy to draw up models where the treatment which is mediocre-but-cheaper is recommended, as minimizing opportunity costs while still having a chance of helping the potentially-alive patient.

But if there is genuine indeterminacy over whether the patient is alive or dead, then on the model we have been developing such hedging is ruled out. One sharpening—on which the patient lives yet—will favour the expensive treatment just as much as if we were certain the patient were alive. The other sharpening, on which the patient is dead already, recommends no treatment. No sharpening favours the mediocre-but-cheap treatment. Our model has the mediocre treatment forbidden, and forces an arbitrary choice between the expensive treatment and no treatment at all. On the current model, action under indeterminacy does not tolerate compromise.

[41] The limiting relative frequency of ϕing in such trials would (on the mind-making picture) match the degree of determinacy of the claim that ϕing is rational.

Appendix A Indeterminacy-induced mushy beliefs

One model of indeterminacy-induced mushy mental states involves *projecting* credences in precise propositions in a sharpening-relative manner into degrees of belief across all propositions. The resulting set of possible degrees of belief are the mushy set that characterizes the agents mental state. To model this, suppose that c is a credence assignment over worlds $w \in W$, and let $|p|^s_w$ give the truth value (1 for truth, 0 for falsity) of (sentence or proposition) p at world w and sharpening s. Then the mushy set that c induces is:

$$\{P : \exists s \forall q [P(q) = \sum_w c(w)|q|^s_w]\}$$

The recipe just given works only if we spot ourselves the underlying c. Here is a more general formulation (to use this I will use w ambiguously, to indicate both the world w, and the proposition which, determinately, is true at w and false everywhere else). First, let the set $\mathbb{P}_s$ be the set of 'probabilities relative to s'. In our finite setting $\mathbb{P}_s$ will be the functions P from propositions to $[0,1]$ such that $1 = \sum_{w \in W} P(w)$ and for arbitrary proposition q, $P(q) = \sum_{w \in W} P(w)|q|^s_w$. Now let $\mathbb{C}$ be an arbitrary constraint on one's degrees of belief—for example, it could be that one's credences in the world-propositions w take certain specific values; it could be that one's credences in some other (possibly vague) propositions take certain specific values; or it could be specified some other way. We model it by a subset of the set of functions from propositions to $[0,1]$. Then the indeterminacy-induced belief state $\mathbb{B}^{\mathbb{C}}$ generated by constraints $\mathbb{C}$ will be given as the union of the sets $\mathbb{B}^{\mathbb{C}}_s = \mathbb{P}_s \cap \mathbb{C}$, for s a sharpening. Our assumption will be that permissible belief states in the presence of indeterminacy must be generated in this way by some $\mathbb{C}$

When $\mathbb{P}_s \cap \mathbb{C}$ is a singleton set for each s, I will say that that the belief-state involved has *only indeterminacy-induced uncertainty*. The distinction between cases where the uncertainty is entirely indeterminacy induced, and cases where some uncertainty remains even after sharpening, plays a key role in allowing us to characterize conditions under which selecting a sharpening is tantamount to selecting a recommendation for action.

Notice that in the special case where $\mathbb{C}$ is a complete (and probabilistically coherent) assignment of credences to worlds, then $\mathbb{P}_s \cap \mathbb{C}$ is a singleton set, and the union of these over s gives us exactly the mushy belief state described earlier. But we could instead have a constraint which required, inter alia, that $P(R) = 1$ and $P(w \vee u) = 1$, in the presence of two sharpenings, one of which (s') makes R true at w and false at u, the other (s'') requires the reverse. In that case, $\mathbb{P}_{s'}$ will contain only functions P such that $P(w) = 1$, and $\mathbb{P}_{s''}$ will contain only functions P such that $P(w) = 0$. So the generated belief state $\mathbb{B}^{\mathbb{C}}$ will have no settled level of confidence invested in w (nor indeed, in u). In other words, this agent has full confidence in the known indeterminate proposition R, and in virtue of that is rationally required to be uncertain, in the indeterminacy-induced sense. But the topic over which she is uncertain is not a vague proposition, but the underlying state of the world. A case in point was described earlier (where the proposition R is 'I act rationally' and the worlds in question describe the precise facts about the agent's future acts).

Appendix B Indeterminacy-induced mushy desires

Alpha's intrinsic desires are for something that can be indeterminate in particular physical situations: that he himself gets goodies. We now model the consequences of such an indeterminate desire as part of a mushy mental state—first concentrating on desirability of worlds alone, and then, in combination with degrees of belief, desirabilities across all propositions.

Suppose that $v \in V$ is a vector with one entry for each determinant of utility, and f a function from such vectors to real numbers, such that $f(v)$ gives the utility of possessing exactly the features described by v. In general, it can be vague whether a world w has the features specified by a given v. One of the determinants of utility for Alpha could be, for example, the amount of money possessed by Alpha in a week's time, where (as above) it is vague whether Alpha is even around in a week's time.

Relative to a sharpening, however, we can let v^s_w be exactly the vector that specifies the determinants of utility possessed by w on sharpening s, and this enables the utility determination function to induce a valuation of worlds: $f_s(w) := f(v^s_w)$. The mushy desire state, evaluating worlds, induced by a (potentially vague) utility-determination function f is therefore $\{u : \exists s[u = f_s]\}$.

Relative to an assignment of degrees of belief to propositions, we can then apply the rules of standard expected utility theory to determinate a set of assignments of desirability to propositions, induced by those degree of beliefs and a world-desirability drawn from the above mushy set.

For indeterminacy-induced mental states: suppose we are given, as determinants of an indeterminacy-infected mental state, a constraint $\mathbb{C}$ on degrees of belief and a utility-determination function f. The associated mushy mental state $\mathbb{M}^{\mathbb{C},f}$ will be the set of those pairs $\langle p, u\rangle$ such that for some s (i) $p \in \mathbb{B}^{\mathbb{C}}_s$; and (ii) u satisfies the axioms for expected utility theory relative to probability p and valuation of worlds $u(w) = f(v^s_w)$.

Appendix C Sharpenings recommending action

Suppose an agent is in a decision situation, with the options for action drawn from set O. The agent has a mushy mental state, each element of which assigns utility to the acts within O. Each element of the mental state recommends a particular act;[42] but these recommendations may conflict.

Caprice is then the principle that the agent may permissibly act in any of the ways reocmmended by any sharpening, given her mental state; and Randomize is the principle that she must choose the sharpening to listen to at random (more precisely: if ϕ is recommended for action by exactly measure k of the sharpening, one's chances of ϕing should be k).

In characterizing Caprice and Randomize for indeterminacy-induced mushy mental states, I switched from talking of what options for actions are recommended by particular elements of one's mushy mental state (belief-desire pairs, which assign utilities to all acts); and what sharpenings recommend. It is crucial to our model that the latter is legitimate, since in a general setting is *sharpenings* over which a measure is defined which characterize the odds at which Randomize is to operate. If a sharpening could be associated with two incompatible recommendations for action, then we wouldn't know whether to allocate the weight assigned to the sharpening to one act or the other.

When the mushy mental state is generated by an underlying credence across worlds and (potentially vague) utility-determination function, then each sharpening generates a unique member of the mushy mental state. More generally, where the uncertainty in a given mental state (in the terminology introduced in appendix A) *entirely indeterminacy-induced* then each sharpening will generate a unique member of the mushy mental state. So under these assumptions, selecting a sharpening will indirectly select a specific recommendation for action.

Where some uncertainty is not indeterminacy-induced, then we need some other story to tell. But this shouldn't be a surprise: expected utility theory, supplemented with a theory of

[42]Though note the qualification in the final paragraph of this appendix.

decision making under indeterminacy-generated uncertainty, shouldn't by itself predict how to behave under other kinds of uncertainty. So our story does exactly what it needs to.[43]

A final coda. Even granted all this, a sharpening might not recommend a particular option for action, since according to it two incompatible acts are assigned the same expected utility. I suggest the following patch: calculate the weights of sharpenings assigned to sets of optimal acts—so that if ϕ and ψ are tied on s, then the weight of s counts towards the set $\{\phi, \psi\}$ rather than $\{\phi\}$ or $\{\psi\}$ alone. If Randomize selects a non-singleton set as the operative recommendation, then the agent's options are whatever they would be in the ordinary case—perhaps either act is permissible (I do *not* want to assume that here the agent needs to randomize—perhaps she can exercise discretion here on a more voluntaristic basis).

References

Barnes, Elizabeth. 2006. *'Conceptual room for ontic vagueness'*. `https://research-repository.st-andrews.ac.uk/bitstream/10023/348/6/FinalThesis.pdf`.

Barnes, Elizabeth. 2010. 'Ontic vagueness: a guide for the perplexed'. *Noûs*.

Barnes, Elizabeth, & Cameron, Ross P. 2009. 'The open future: Bivalence, determinism and ontology'. *Philosophical studies*, **146**(2), 291–309.

Barnett, David. 2009. 'Is vagueness sui generis?'. *Australasian journal of philosophy*, **87**(1), 5–34.

Cook, Roy. 2002. 'Vagueness and mathematical precision'. *Mind*, **111**, 225–48.

Dorr, Cian. 2003. 'Vagueness and ignorance'. *Philosophical perspectives*, **17**, 83–114.

Edgington, Dorothy. 1997. 'Vagueness by degrees'. *Pages 294–316 of:* Keefe, Rosanna, & Smith, Peter (eds), *Vagueness: A reader*. Cambridge, MA: MIT Press.

Eklund, Matti. 2010. 'Vagueness and second-level indeterminacy'. *In:* Moruzzi, Sebastiano, & Dietz, Richard (eds), *Cuts and clouds*. Oxford: OUP.

Elga, A. 2010. Subjective Probabilities Should be Sharp. *Philosophers imprint*, **10**(5).

Field, Hartry H. 1978. 'Mental representation'. *Erkenntnis*, **13**, 9–61. Reprinted in Field, *Truth and the Absence of Fact* (Oxford University Press, 2001) pp. 30-67.

Field, Hartry H. 2003. 'Semantic paradoxes and the paradoxes of vagueness'. *Pages 262–311 of:* Beall, J. C. (ed), *Liars and heaps*. Oxford: Oxford University Press.

Fine, Kit. 1975. 'Vagueness, truth and logic'. *Synthese*, **30**, 265–300. Reprinted with corrections in Keefe and Smith (eds) *Vagueness: A reader* (MIT Press, Cambridge MA: 1997) pp.119-150.

[43]If one wants a sketch about how it could be integrated into wider theory of decision making under any kind of uncertainty, here's an ption: take your favoured account of acting under uncertainty when indeterminacy isn't involved—perhaps Levi's loss-minimization rule. If uncertainty remains even after sharpening, apply this story to retrieve the recommendation for action, and then treat that as the recommendation delivered by randomly selecting that sharpening.

Greenough, P. 2008. Indeterminate truth. *Midwest studies in philosophy*, **32**(1), 213–241.

Jeffrey, Richard. 1965. *The logic of decision*. 2nd edn. Chicago and London: University of Chicago Press. Second edition published 1983.

Jeffrey, Richard. 1992. 'Bayesianism with a human face'. *In: Probability and the art of judgement*. Cambridge University Press.

Joyce, James M. 1999. *The foundations of causal decision theory*. Cambridge, UK: Cambridge University Press.

Joyce, James M. 2011. A Defense of Imprecise Credences in Inference and Decision Making. *In:* Gendler, Tamar Szabo (ed), *Oxford studies in epistemology*. Oxford University Press.

Kamp, J. A. W. 1975. 'Two theories about adjectives'. *Pages 123–155 of:* Keenan, E. (ed), *Formal semantics of natural language*. Cambridge: Cambridge University Press. Reprinted in Davis and Gillon (eds) *Semantics: A reader* (Oxford University Press, Oxford, 2004) pp.541-562.

Keefe, Rosanna. 2000. *Theories of vagueness*. Cambridge: Cambridge University Press.

Levi, I. 1974. On indeterminate probabilities. *The journal of philosophy*, 391–418.

Lewis, David K. 1970. 'How to define theoretical terms'. *Journal of philosophy*, **67**, 427–446. Reprinted in Lewis, *Philosophical Papers vol. I* (Oxford University Press, 1983) 78–95.

Lewis, David K. 1976. 'Survival and identity'. *In:* Rorty, Amélie O. (ed), *The identities of persons*. University of California Press. Reprinted in Lewis, *Philosophical Papers I* (Oxford University Press, 1983) 55–72.

Lewis, David K. 1979. 'Attitudes *de dicto* and *de se*'. *The philosophical review*, **88**, 513–43. Reprinted with postscript in Lewis, *Philosophical Papers I* (Oxford University Press, 1983) 133-55.

Lewis, David K. 1980. 'A subjectivist's guide to objective chance'. *Pages 263–93 of:* Jeffrey, Richard (ed), *Studies in inductive logic and probability*, vol. II. University of California Press. Reprinted with postscript in Lewis, *Philosophical Papers II* (Oxford University Press, 1986) 83–113.

Luce, R.D., & Raiffa, H. 1989. *Games and decisions: Introduction and critical survey*. Dover Pubns.

McGee, Vann, & McLaughlin, Brian. 1994. 'Distinctions without a difference'. *Southern journal of philosophy*, **supp XXXII**, 203–251.

Parfit, Derek. 1971. Personal identity. *Philosophical review*, **80**(1).

Parfit, Derek. 1984. *Reasons and persons*. Oxford: OUP.

Raffman, Diana. 1994. 'Vagueness without paradox'. *Philosophical review*, **103**.

Schiffer, Stephen R. 2003. *The things we mean*. Oxford: Oxford University Press.

Shapiro, S. 2006. *Vagueness in context*. Oxford University Press, USA.

Smith, Nick J. J. 2008. *Vagueness and degrees of truth*. Oxford: Oxford University Press.

Smith, Nick J. J. 2010. 'Degree of belief is expected truth value'. *In:* Moruzzi, Sebastiano, & Dietz, Richard (eds), *Cuts and clouds*. Oxford: OUP.

van Fraassen, Bas. 1966. 'Singular terms, truth-value gaps, and free logic'. *The journal of philosophy*, **63**(17), 481–495.

van Inwagen, Peter. 1990. *Material beings*. Ithaca and London: Cornell University Press.

Weatherson, Brian. Manuscript. Decision Making with Imprecise Probabilities.

Williams, Bernard. 1970. 'The self and the future'. *Philosophical review*, **79**(2), 161–180. Reprinted in his *Problems of the Self*, CUP, 1973.

Williamson, Timothy. 1994. *Vagueness*. London: Routledge.

Wright, Crispin. 1976. 'Language mastery and the sorites paradox'. *Pages 223–47 of:* Evans, Gareth, & McDowell, John (eds), *Truth and meaning*. Reprinted Keefe and Smith (eds) *Vagueness: A reader* (MIT Press, Cambridge MA: 1997) pp.151-173.

Wright, Crispin. 2001. 'On being in a quandary'. *Mind*, **110**.

Wright, Crispin. 2003. 'Vagueness: a fifth column approach'. *Pages 84–105 of:* Beall, J. C. (ed), *Liars and heaps: New essays on paradox*. Oxford: Oxford University Press.