BETTER IMPRECISE DECISIONS 159

Seamus Bradley

August 22, 2011 168

`This is a draft. Please do not cite it.`

1 Imprecise decisions

Various arguments have been put forward that *sets* of probabilities do a better job of representing our uncertainty than do single probability measures.[1] One major stumbling block to the widespread adoption of the "IMPRECISE CREDENCE" model is the difficulty of decision making in this approach. I hope to offer some suggestions as to what good imprecise decision rules would look like.

Example 1 Let's say I offer You[2] the choice of two tickets c and d (on the understanding that You could choose neither n).

c Gain £10 if X, lose £5 otherwise

d Lose £5 if X, gain £10 otherwise

n Gain £0 whatever happens

	X	$\neg X$
c	10	–5
d	–5	10
n	0	0

Table 1: The decision of Example 1

Which of these tickets should You pick? Well, You might say, it depends on how likely You think X is. Let's say You know *absolutely nothing* about X: no idea what

[1] For philosophical arguments see Joyce (2005, 2011); Kaplan (2010); Levi (1974); Sturgeon (2008). Introductions to formal models of imprecise belief can be found in Walley (1991) and Cozman (n.d.)

[2] "You" are an arbitrary intentional system invented by I.J. Good.

sort of proposition it might be. In this instance You are in a situation of "DECISION UNDER IGNORANCE". Decision theorists have been exploring this sort of problem for some time and the theory of decision under complete ignorance is pretty well covered (see Peterson 2009, chapter 3). So there are some possible events (in the example above, we have X and $\neg X$) and which obtains determines what Your payout will be. You know nothing about which event is the case, so how ought You choose?

Probably the simplest decision rule is "Maximise minimum possible gain". I call 228
this rule "WALD".[3] So You evaluate each act by its lowest possible outcome, and You act to maximise that. This rule recommends going for n in Example 1. Is this a reasonable decision rule? For example, consider Table 2. WALD sanctions taking

238

	X	$\neg X$
a	100	1
b	2	2

Table 2: An unwarranted probabilistic assumption?

b in this case. The following reasoning might support choosing a over b: "Unless X is very unlikely, a is the better bet." The point is that in decision under ignorance, You can't rule out the possibility that X is very unlikely. So choosing a on this basis would be to make an unwarranted probabilistic assumption about how likely X could be. Criticism of particular ignorance choice rules often implicitly makes unwarranted assumptions about the probabilities of the various states.

If WALD seems a rather pessimistic way of looking at things, perhaps the HUR- 262
WICZ rule would suit better. It says "Maximise a weighted sum of maximum and minimum possible gains". That is, if the worst outcome of act a is $\underline{a}$ and the best outcome is $\overline{a}$, then HURWICZ says "act to maximise $\rho\underline{a} + (1-\rho)\overline{a}$." If $\rho = 1$ then this is just the same as WALD. For all but the biggest values of ρ, HURWICZ opts for b in the decision problem in Table 2. In Example 1, for $\rho \leq \frac{2}{3}$, c and d are both preferred to n, otherwise the preference is reversed.

A different way to look at Table 2 is as follows: "even when b is better than a, 275
the difference is small. But when a is the better act, it is *much* better. This points to another possible ignorance rule: SAVAGE. Also known as "minimax regret", this rule says You should evaluate acts in terms of their "regret" and aim to minimise Your maximum possible regret. Regret is defined as the difference between what You got from Your act and what You *could have* got in that state by choosing a different act. So for Table 2, the maximum regret of a is 1, because if X turns out false, You could have got 2 by choosing b rather than the 1 You actually got. b's

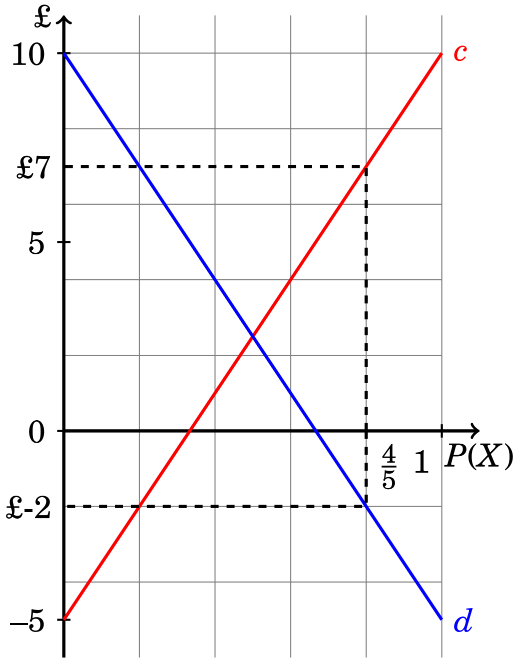
[3] I borrow this naming scheme from Milnor (1951) 231

maximum regret is 98 because if X is true, You could have got 100 from a rather than the 2 You actually did from b. Savage opts for n in Example 1.

Another way You might try and evaluate acts is not in terms of their regret, 292
or in terms of the maximal and minimal outcomes, but in terms of the average outcome. This is the essence of Laplace. How this rule is normally formulated is to say “treat each state as equally likely, and maximise expectation”. This comes out the same as saying “maximise average outcome”. So assign probability of X and $\neg X$ to be a half each and then maximise expectation.

There is some debate over which ignorance rule is the best. I find the debate 302
somewhat strange: if You know literally nothing about the situation You find Yourself in, why assume there should be *some* rationally determined course of action? Might it not be the case that You simply don't have enough information to make an informed choice? I will return to this point later.

Ignorance is one extreme of decision problems. To return to Example 1, the 311
other extreme is where you know exactly the probability of X.[4] Perhaps You know that X is the proposition “The next flip of this biased coin will be heads”. In this case, again, decision theory has you covered: “decision under risk” typically says You should maximise expected value.[5] Let's say You have examined this biased coin and found the probability of heads for it to be $\frac{4}{5}$. Under these circumstances



336

Figure 1: c and d as functions of degree of belief

[4]Well, the other *extreme* is where you know whether or not X is true, but that is uninteresting for decision theory 316

[5]There are dissenting voices, I shall ignore them 322

c will have expected value £7 and d will have £-2 (see Figure 1). So c is a better bet in this case. I don't have much to say about this case, except that it serves as a kind of "ideal limit" for imprecise decisions.

Now we come to ambiguity or imprecision: let's say You aren't allowed to examine the coin. You know it's biased, but You don't know which way it's biased, or by how much. Any particular probability function fails to represent the ambiguity of your evidential state. It has been argued that there is a set of probability measures compatible with your "ambiguous" evidence, and that you should use this set as representing your uncertainty. I don't wish to argue this point here. So let's take it for granted that there is some set of probability measures that represents Your credence in X; call this Your REPRESENTOR. How ought You make decisions on the relative merits of c and d from Example 1 in this situation?

You could always resort to "decisions under ignorance" discussed above. But this seems unsatisfactory: if you know *something* about the probabilities, this information should feed into your decision making. This seems to be the guiding norm of decision theory: your decisions should be guided by your current state of knowledge. Ignoring the probability information You do have in Your representor seems to violate this norm. This "use Your evidence" norm will be important later.

You can't use the same maximising rule that "strict probabilists" can, because there's no guarantee that there will be one act that maximises expectation with respect to all probabilities in the representor. But when there is one act that maximises with respect to all probabilities, then You should certainly choose it.

To return to the example I discussed earlier, which of the bets c or d ought You take? Let's say Your belief in X ranges over the interval $[0.2, 0.8]$. Note, this range of values that probabilities in the representor assign X is a "summary statistic": the representation of belief is done by the representor — the set of probability functions — not by the collection of values it assigns to the event.[6] For each probability in your representor there is some expected value assigned to each event. For example, in Figure 1, the probability that assigned $\frac{4}{5}$ to X has expected value £7 for c and £–2 for d. We can consider the sets of expected values. For each act, there is a range of expected values in the representor. Act c's expectation covers the range $[-2, 7]$ as does act d. Act n is such that all probabilities in the representor give it expected value £0 (see Figure 2).

There is no clear best act in this scenario. So how best to make decisions? The point is that in the strict Bayesian model there is some act that *maximises* expected value. Or some collection of acts each of which maximises expectation. Figure 2 for the strict Bayesian would just be a series of dots, and it is obvious that the highest dot is the best. In the imprecise case, intervals can overlap, so what is "best" is not as clear. Acts whose expectations overlap are "incommensurable". It is not clear how to compare them to each other, thus evaluating them with respect to each other is difficult.

[6] For examples of why this point is important see Joyce (2011)

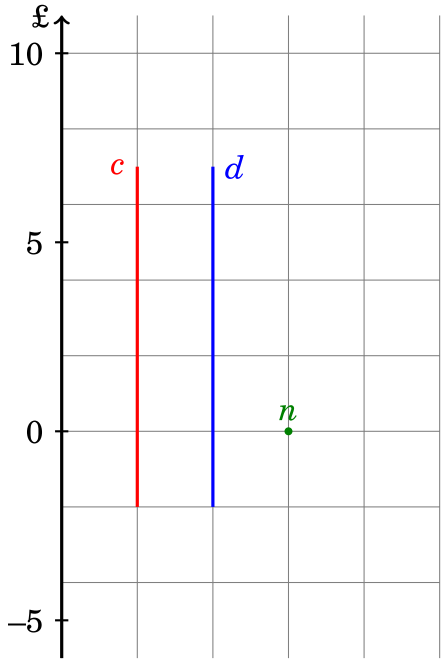


Figure 2: The range of expected values

This imprecise belief model is, in a sense, "between" these two better studied modes of decision making. If there is just a single probability in Your representor You'd better act like a maximiser. If Your representor contains all possible probability functions — if all possible probability functions are compatible with Your evidence — then You are effectively in a state of complete ignorance and You'd better act accordingly.[7]

Imprecise probabilities aren't a radically new theory. I see them as merely a slight modification of existing decision theory for situations of ambiguity. Often Your credences will be precise enough, and Your available actions will be such that You act more or less *as if* you were a strict Bayesian. I see imprecise probabilities as the "Theory of Relativity" to the strict Bayesian "Newtonian Mechanics": all but indistinguishable in all but the most extreme situations.[8]

There are a variety of issues that I will not be talking about, I briefly mention them here. The decisions I consider take place *at a time*. So there are no *sequences* of decisions. Nor will I be discussing updating or learning in the imprecise frame-

[7] This is deliberately vague: whichever decision-under-ignorance rule You endorse, You'd better act *that* way when all probability measures are in your representor

[8] This analogy goes deeper: in both cases, the theories are "empirically indistinguishable" in normal circumstances, but they both differ radically in some counceptual respects. The role of absolute space in Newtonian mechanics/GR; how to model ignorance in the strict/imprecise probabilist case.

work.[9] I also keep Your utility function determinate: I will not allow imprecise utilities for the time being.[10] I will also avoid the issue of whether sets of probability ought to be convex, I don't think they need be, although in this paper I normally assume they are.[11] I also make some standard decision theory assumptions: there is no risk aversion and Your utility is linear with money.

2 Imprecise decision rules 438

In this section I highlight some possible imprecise decision rules. Once I've out- 441
lined a number of rules, I will turn to assessing which might be the best.

Before continuing, we will need some terminology. The basic insight of the im- 446
precise probabilities approach is to represent belief by a *set* of probabilities instead of a single probability. So instead of having some Pr represent Your belief, You have $\mathcal{P}$, a set of such functions. Your degree of belief in a proposition, X, is given by $\mathcal{P}(X) = \{\Pr(X) : \Pr \in \mathcal{P}\}$.

From this we can derive an "imprecise expectation". If $E_{\Pr}(f)$ is the expected 455
value of act f with respect to probability Pr, then $\mathcal{E}_{\mathcal{P}}(f)$ is:

$$\{E_{\Pr}(f) : \Pr \in \mathcal{P}\}$$

From these concepts we can define some "summary statistics" that are often 461
used in discussions of imprecise probabilities. Your "lower envelope" of X is: $\underline{\mathcal{P}}(X) = \inf \mathcal{P}(X)$. Likewise, Your "upper envelope" is $\overline{\mathcal{P}}(X) = \sup \mathcal{P}(X)$. In the same way we can talk about the bounds of the spread of Your expectation. Your "lower prevision" is defined as $\underline{\mathcal{E}}(f) = \inf \mathcal{E}(f)$ and Your upper prevision ($\overline{\mathcal{E}}$) defined in the obvious way.

One might want to call "lower prevision" "lower expectation", but we must be 467
careful: Walley (1991) and Cozman (n.d.) use "lower expectation" to mean something slightly different. The difference won't be relevant here, so I stick with the more intuitive terminology of "expectation".

Remember also that these summary statistics are not properly representative 474
of Your belief. Information is missing from the picture. Recall Example 1: all the probabilities in Your representor that give a high expectation to c give a low expectation to d and vice versa. This complementarity is lost in Figure 2: the intervals for c and d simply look the same.

---

[9] But see Seidenfeld and Wasserman (1993); White (2010) for the extra problems with updating 430
imprecise credences and see Joyce (2011); Wheeler (forthcoming) for responses

[10] But see Bradley (2009) 431

[11] But see Kyburg and Pittarelli (1992, §4) for some discussion of problems with convexity. See also 434
Bradley (2009) for comments on convexity when both probabilities and utilities are imprecise.

2.1 Ignorance analogues 481

So our first decision rule is probably the simplest possible rule: "maximise $\underline{\mathcal{E}}$". We might term this $\mathcal{E}$-WALD after the way it mirrors our "maximise minimum gain" rule for decisions under ignorance. Like WALD in the ignorance case, this rule maybe seems overly pessimistic. 485

Maybe some rule that takes account of both the top and bottom of the interval would do the trick? Take an "ambiguous analogue" of the Hurwicz criterion: "Maximise $\rho\,\underline{\mathcal{E}} + (1-\rho)\,\overline{\mathcal{E}}$". Call this $\mathcal{E}$-HURWICZ. This is actually a whole class of different decision rules depending on choice of ρ. If $\rho = 1$ then we recover maximise lower prevision ($\mathcal{E}$-WALD). But again, we can run into problems. Indeed, any decision rule that is sensitve to only the set of expected values looks like it will lead to problematic results. Consider the following situation. 491

Example 2 There is a coin of unknown bias. You are offered the choice between these two bets: 501

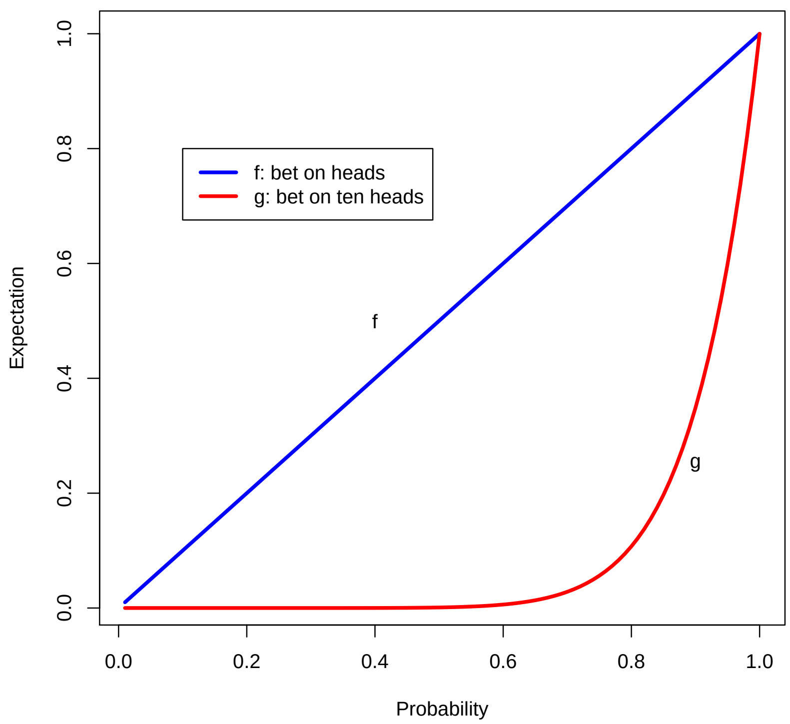
- f Win £1 if the next toss lands heads 505
- g Win £1 if the next ten tosses all land heads 506

Intuitively, the first of these is clearly better. However, the set of expected values for each is $[0, 1]$ so a choice rule that only pays attention to the set of expectations will not be able to discriminate between these bets. The important aspect that allows discrimination between these bets is that every $\Pr \in \mathcal{P}$ is such that expectation with respect to Pr is at least as high for the first as for the second (and almost always higher). The first act "dominates" the second.[12] As well as knowing how acts fare as Pr varies, we need to pay attention to how the acts *compare* for a fixed Pr. 510

What about imprecise analogues of our other two ignorance rules? LAPLACE is tricky, since there isn't an obvious way to derive the "average expectation". What we need to do in the ignorance case is construct a probability function that makes each state equally likely and maximise expectation with that. But imagine if there were uncountably many states. How do we make them all equally likely? We end up with Bertrand's paradox: two descriptions of the same set of states lead to two different incompatible probability functions (see Gillies 2000, pp.37–49). 529

We have the same problem in the imprecise case. You can't make all probabilities in Your representor "equally likely" because of problems like Example 2, and partly because the whole point of the imprecise probability set up is that any particular "higher order" probability would be an unwarranted extra assumption. The problem is that there is no principled way to put a measure on the set of probabilities. Or rather, there are many different measures we could put on the set of probabilties none of which is in any sense "more consistent with the evidence". 542

[12] But perhaps this isn't enough: what if we replaced the prize in the second bet by £1 + ε? Then the second act wouldn't dominate, but it would still be intuitively dispreferred. 519



524

Figure 3: Graphs of Example 2

One could argue that the "objective Bayesian" tradition offers a principled method 554
of choosing "the right probability" from those in Your representor to make decisions with (Williamson 2010). The "Calibration" norm sanctions a set of probabilities consistent with the evidence (p.28). It is only the extra norm of "Equivocation" that forces the objective Bayesian onto a single function. There is also the "Maximum Entropy" approach championed by E.T. Jaynes (Jaynes 1973, 2003). Williamson also discusses this.

Another approach with some of these characteristics is the "Transferable Belief 565
Model". Smets and Kennes (1994) argue that You should maintain a set of probabilities in a "credal set" but that when called on to make a decision, You should collapse this into a "pignistic probability" and maximise expectation with respect to that. Seidenfeld (2004) ends up taking something like this line, but for rather different reasons.

It seems if You were to get one probability out of Your representor, then You 574
wouldn't really be in an imprecise probability framework. If there's one probability function that always "trumps" the others in decisions, what work are the other probabilities in Your representor doing? What's the point of having a set of probabilities representing Your degree of belief if they are completely otiose when in comes to decisions? And in contexts where we have sequences of decisions, the problem becomes worse: sequential consistency forces the single probability to be consistent with Bayesian conditioning. So the redundancy of the representor is even more apparent. Elga (2010) makes a similar point.

One could argue that there is more to representation of belief than its use in 586
decision making. I have some sympathy for this point, but it seems that divorcing belief and decision to the extent required by LAPLACE-style rules is going too far. Recall the "guiding norm" of decision theory I discussed above: your best available evidence should be informing your decisions. This sort of LAPLACE approach seems to be going beyond the evidence by collapsing the representor into some single probability function. I claim that this is just as much a violation of this guiding norm as ignoring the representor and reverting to decision under ignorance.

Maybe we could reinterpret what it means to "use all the information in Your 595
representor". Perhaps what we should require is that all the information of the representor is used in coming up with the single probability which is then used for decision. The idea is that the choice rule should make use of all of the representor in coming to a decision. This "procedural" character might allow TBM and even objective Bayesianism to avoid charges of not properly taking account of all of the representor, depending on exactly how they arrive at their single probability functions. I have my doubts as to whether the Objective Bayesian norm of "Equivocation" is suitably procedurally sound in this respect. But one shouldn't perhaps rule out all single probability solutions on these grounds.

There is a lot of work on judgment aggregation that could be of interest here. 611
For example, Lehrer and Wagner (1981) offer a thorough treatment of aggregating numerical opinions. Unfortunately, a lot of it depends on there being only finitely many "opinions" to average. Whether any results or methods can be translated

into the current potentially uncountable situation is an open question.

There's also a worry about exactly which probabilities should be averaged. Con- 620
sider Example 2 again: should we average over the chance of heads or over the chance of ten heads? This choice has an impact on which act will be considered as optimal. Williamson (2010) discusses this issue in chapter 9.

In what follows I have less to say about this sort of approach. It seems there are 627
epistemological problems with deciding how to "average" Your probability judgements, but once that is achieved, the decision theory is wholly orthodox.

Finally, what about a $\mathcal{E}$-Savage rule? For example, we could define an imprecise 632
analogue of "regret" as follows:

$$\mathbf{Reg}(b) = \max_{\Pr\in\mathcal{P}} \left\{ \max_{a} \{\mathrm{E}_{\Pr}(a)\} - \mathrm{E}_{\Pr}(b) \right\}$$

For an act b, we look at how far b is from the act that performs best for that probability function. We do this for all probabilities, and take the maximum of these "probabilistic regrets". We then try and minimise this. That is, for each probability in Your representor, subtract the expectation of act b from the "top act" for that probability. The maximum of those values over all probabilities in Your representor is Your maximum regret for act b. You should act to minimise that.

2.2 Ruling out acts 650

Let's now take a different tack to constructing decision rules. Let's look at restrict- 653
ing the space of reasonable acts. Instead of considering which acts are the "best" in some sense, I want to focus on which acts You can justifiably rule out. Or alternatively, I want to restrict the space of acts that I am considering, by ruling out acts it would be unreasonable to perform. There are a couple of restrictions that have been discussed in the literature.

Recall that in Example 2 we criticised rules that focus only on the set of expecta- 663
tions for not taking into account that one act is always better than the other. This suggests our first restriction: "Choose only among the undominated acts". That is, choose only among acts that have no act that is always better. Call an act f "dominated" if there is another act g such that, for every member of the representor, g has higher expectation than f. It seems reasonable that in this circumstance we should prefer g to f and in fact that we can safely ignore f when making our final decision. Call the set of undominated acts $\mathcal{U}$.

Another restriction of the act set — "E-admissibility" — is due to Isaac Levi (Levi 678
1974, 1986). An act is E-admissible if there is some probability in Your representor such that that act maximises expectation with respect to that probability function. In picture terms an act is E-admissible if it is "on top" for some probability value in the representor. Levi argues that You should only choose among E-admissible acts.

As it stands, the definition of E-admissible isn't quite good enough. Recall Example 2 where we had the choice between a bet on heads and a bet on 10 heads in a row. The latter maximises expectation for $\Pr(H) = 0$ and $\Pr(H) = 1$ so it is E-admissible. This act is, however, *weakly* dominated. I expect that Levi would want to rule out choosing weakly dominated acts, even if they maximise for some Pr. So let's call the set of acts that are undominated and which maximise for some probability in the representor $\mathcal{L}$. 686

There are undominated acts that are not E-admissible, for example n in Example 1. We also know that $\mathcal{L} \subseteq \mathcal{U}$. 697

These two rules are *very* permissive. They don't establish an ordering among the acts as did our imprecise analogues of ignorance rules did. 701

2.3 Penalising spread 705

Another property one might intuitively like to pay attention to is the *spread* of expectations. $\mathcal{E}$-HURWICZ sort of does this, in that it takes into account the top and bottom ends of the set of expectations. However, acts can be evaluated the same by $\mathcal{E}$-HURWICZ and have different spreads. For $\rho = \frac{2}{3}$ in Example 1, $\mathcal{E}$-HURWICZ evaluates c and n the same. But n has a much lower spread of expectations. 708

Now obviously "minimise $\overline{\mathcal{E}} - \underline{\mathcal{E}}$" would be a crazy decision rule, but can we construct some method of evaluating acts that penalises acts for offering a large spread of values? 718

Let's consider what properties we'd like this "spread penalty" to have. Consider a function **SP** that penalises acts with large spreads of expectation. 723

- $\mathbf{SP}(f)$ should be 0 when $\overline{\mathcal{E}}(f) = \underline{\mathcal{E}}(f)$ 727
- $\mathbf{SP}(f)$ should be 0 when $\overline{\mathcal{E}}(f) - \underline{\mathcal{E}}(f) = \overline{\mathcal{E}}(g) - \underline{\mathcal{E}}(g)$ for all g. 728
- $\mathbf{SP}(f + B) = \mathbf{SP}(f)$ for constant $B \in \mathbb{R}$ 730
- Multiplying all the outcomes by $A \in \mathbb{R}$ multiplies all the spread penalties by A. 731

The following function seems to satisfy these criteria: 735

$$\mathbf{SP}(f) = \overline{\mathcal{E}}(f) - \underline{\mathcal{E}}(f) - \min_g\{\overline{\mathcal{E}}(g) - \underline{\mathcal{E}}(g)\}$$

So we could use a rule that says something like "Maximise $\mathcal{E}$-HURWICZ[13] minus some spread penalty". You should act to maximise: 740

$$\rho\,\underline{\mathcal{E}}(f) + (1-\rho)\,\overline{\mathcal{E}}(f) - \sigma\,\mathbf{SP}(f)$$

[13] $\mathcal{E}$-WALD is a special case of $\mathcal{E}$-HURWICZ so let's adopt the more general rule as a basis. 743

Where σ is some number which measures how important we think the spread is. We could also write this as:

$$\overline{\mathcal{E}} + (\rho + \sigma)(\underline{\mathcal{E}} - \overline{\mathcal{E}}) + \sigma \min\{\overline{\mathcal{E}} - \underline{\mathcal{E}}\}$$

Note that this rule is still only sensitive to the set of expectations, so it has all the problems associated with Example 2.

To summarise, there are a great many rules we could pick. You need to decide 759 whether to choose among all of the acts, just the undominated ones ($\mathcal{U}$) or the E-admissible ones ($\mathcal{L}$). Then You need to decide *how* to choose among that collection of acts. We have seen the $\mathcal{E}$-Wald, $\mathcal{E}$-Hurwicz and $\mathcal{E}$-Savage rules as analogues of decision rules under ignorance. We could call the rule "maximise Hurwicz expectation among undominated acts" "$\mathcal{U}$-Hurwicz". And likewise for other combinations of "act space" and tiebreaker. We would also need to specify a ρ value to pick out a particular rule.

How do we choose among this plethora of decision rules? This is the topic of the 771 remainder of the paper.

3 Criteria for good choice 774

There are a great many ways we could approach the question of how best to settle 777 on an imprecise decision rule. I survey some ways here.

One major source of inspiration is the literature on decision under ignorance. 781 Just as this was a source of inspiration for decision rules, so it will serve again for criteria to choose among them. The most important sources here will be Milnor (1951) and Luce and Raiffa (1989).

Another important inspiration will be work on social choice theory: how to ag- 788 gregate many individuals' preferences into a single group choice. If we think of each probability in Your representor as a member of a "credal committee" that has to vote on what You should do, then the parallel between imprecise decision and social choice becomes clear. Here I will draw on Arrow's theorem (Gaertner 2009) and the work of Amartya Sen (Sen 1970, 1977).

There are two ways one might frame the discussion: in terms of the choice rule 798 determining an ordering over the acts (Arrow, Milnor), or in terms of the choice rule determining a set of optimal acts (Luce and Raiffa, Sen). I favour the second, but the two views are connected: the maximal elements of the ordering determine the optimal set, f is preferred to g just in case f is optimal in the choice between the pair $\{f, g\}$.

The general optimal set of acts is called $\mathcal{A}$. The set of "undominated" acts will 808 be denoted $\mathcal{U}$ and E-admissible acts will be denoted $\mathcal{L}$ as above.

3.1 The basics 812

What are the most absolutely basic properties that *any* decision rule should satisfy? 815
Milnor (1951), Luce and Raiffa (1989) and Arrow all have a condition that amounts to "You should never pick a dominated act". In social choice, this criterion is called "Pareto" and I adopt that name. Choosing a dominated act seems a strange thing to do: if f is dominated by g, every credal committee member agrees that g is better than f. So picking f despite this unanimity cannot be rational. This is the intuition behind restricting attention to $\mathcal{U}$. Perhaps we could go further and also demand that optimal acts should always be E-admissible.

Another thing that our decision rule should not do is fail to make any sort of 828
discrimination. We don't want a decision rule that can deliver us an empty optimal set. So Luce and Raiffa's first axiom is that the optimal set is always non-empty.

Arrow and Milnor also agree that the choice rule should determine a complete 834
and transitive ordering of the acts. I agree that to fail to have transitive preferences is an indicator of irrationality, but I am a little bit sceptical of completeness. I am sceptical of the claim that there is always some best *rational* act to perform, no matter what your informational state. What I find implausible about this is that it seems perfectly possible that You might have two acts and not have any preference between them. That is, You might not know enough about the situation to decide which would be the better act. That said, in many circumstances when You need to make a decision, a partial order is not enough. How to choose between those acts that are "optimal" in this loose sense? They needn't all be commensurable, so You are stuck with the intractable question of how to choose among incommensurable options. This isn't a problem specific to imprecise decsions: incommensurable options, whenever they appear, are tricky beasts.

So perhaps it's best to keep two projects separate. First we want to know what 849
rationality demands of imprecise decision; second we want to know how we should act in cases of imprecision. The answer to the former question might not fully determine an answer to the latter. But in those cases, we still need an answer to the second question. We will have to accept that the answer might not be fully rational. That's not to say that it will be *irrational* — violating rationality —- but just that it will be *arational*: without rationality. Rationality can only get us so far.

Another property that Luce and Raiffa share with Milnor is "linearity" (Luce 861
and Raiffa Axiom 2). The optimal set (or the preference if you like) is not affected by choice of zero or unit for the utility scale. This is a pretty standard assumption in decision theory. Binmore (2008) takes issue with this property in Chapter 9, but I don't follow up on this here.

Another indispensable property is Luce and Raiffa's Axiom 4 which says that if 869
$a_1 \in \mathcal{A}$ and $a_2 \succeq a_1$ then $a_2 \in \mathcal{A}$. What this says is that it can't be the case that an optimal action is dispreferred to a non-optimal one.

One might go further and say that You should be indifferent between optimal 874
actions: "if $a_1, a_2 \in \mathcal{A}$ then $a_1 \sim a_2$". This would solve the problem of choosing

among incommensurable acts: if we had a rule that determined an optimal set and You were indifferent among the optimal acts, then there would be no problem of choice between them. This seems to be going too far, however. At least for $\mathcal{U}$ or $\mathcal{L}$, the whole problem is that we can't expect to be indifferent between the acts in that set.

3.2 Expansion and contraction consistency 885

What happens when a new act is added to a decision problem or an old act removed 888
from it? How ought a new option affect the optimality of an act? There are many similar properties in this area.

Let's first distinguish three dimensions. First there are properties that discuss 893
whether addition of an act (expansion) can affect optimality, second there are properties that discuss whether removal of an act (contraction) can affect optimality. The second dimension is seen in the difference between turning an old optimal act into a non-optimal one, and turning an old non-optimal act into an optimal one. Finally, there is the question of what properties the added/removed act has. Is it dominated? Would it be optimal? And so on.

Arrow had "Independence of Irrelevant Alternatives" which, among other things, 907
entails a form of expansion and contraction consistency. Sen has his α and β properties, which we will see in due course. Luce and Raiffa have their axioms 6 and 7 and its strengthenings. Milnor had "row adjunction" and "special row adjunction". All of these will be fitted into this framework.

So, what properties might we want our decision rules to have with respect to 915
expansion and contraction of the acts? If the added act is dominated, then it seems like it shouldn't make a difference. Adding a dominated act shouldn't turn a non-optimal act into an optimal one or vice versa. Nor should removing a dominated act turn an optimal act non-optimal or vice versa. These two properties I term "Special Expansion Consistency" and "Special Contraction Consistency" to echo Milnor's "Special Row Adjunction" which says that adding a dominated act shouldn't change anything. Since You basically "ignore" dominated acts, they shouldn't change anything about the decision problem.

For completeness, consider another property, "$\mathcal{L}$-expansion consistency" which 930
says that if an expansion makes a optimal act into a non-optimal one, then the new act is E-admissible. Likewise, "$\mathcal{L}$-contraction consistency" says that a contraction can make a non-optimal act optimal only if the removed act was E-admissible.

In fact, it seems that no expansion should be able to make a non-optimal act 937
optimal. Nor should any contraction be able to make an optimal act non-optimal. These I term "expansion consistency" and "contraction consistency". Note these terms are what Gaertner uses for Sen's α and β. While my contraction consistency is the same as Sen's α, his β is *much* stronger than my expansion consistency. My expansion consistency corresponds to Luce and Raiffa's Axiom 7.

Let's focus on expansion for a moment. We have seen properties that rule out 948 any sort of expansion turning non-optimal acts into optimal ones. So let's look at what sorts of expansions we might want to forbid from turning optimal acts into non-optimal ones. We have already seen that special expansion consistency (SpEC) rules out dominated acts turning optimal acts non-optimal.

Let's consider an example of an expansion where turning an optimal act non- 956
optimal seems reasonable. A new act can make some old optimal act no longer optimal. That is, if I add h to a decision problem where f used to be the best, and $h \succ f$, then f is no longer optimal. So perhaps we should demand that the new act's being optimal is a necessary condition for the old act changing from optimal to non-optimal. Let's call this "strong expansion consistency". This is Luce and Raiffa's first strengthening of Axiom 7. They call it Axiom $7'$.

We could go even further, to Luce and Raiffa's Axiom $7''$, also known as Sen's 968
β property. I shall call it "all-or-nothing expansion consistency". This says that if an old optimal act is made non-optimal, then *all* old optimal acts are made non-optimal. As Luce and Raiffa show, this "all-or-nothing" character of $7''$ makes sense only when You are evaluating the acts on a single scale.

Imagine trying to choose what restaurant to go to. You've already disregarded 976
all the restaurants that don't serve good food. Some options are cheap, and some do really excellent food. The choice is between a cheap restaurant with good food (f), or a more expensive restaurant with really excellent food (g). Now one of the group remembers another possibility: a restaurant that does *even better food* than g for the same (high) price (h). h dominates g: it has better food and is the same price. But h doesn't affect f: f is still the cheapest. So the choice is now between f (cheap) or h (really excellent food). That the addition of h hasn't also stopped f being "optimal" illustrates that more than one criterion is in play when deciding where to eat: cost and quality are both important.

In the case of imprecise decisions, we evaluate acts in terms of their expectation 994
across a whole range of probability values. So this all-or-nothing version of axiom 7 is not reasonable. That some very reasonable rules violate All-or-Nothing EC could be taken as evidence that we are evaluating imprecise decisions on a number of dimensions. The question then is, what are those dimensions?

Now we can return to contractions and define similar properties for them. "Strong 1002
contraction consistency" says that removing an act can make a non-optimal act optimal, only if the removed act was optimal. Note that "all-or-nothing contraction consistency" doesn't really make any sense.

Table 3 summarises what the various properties make illegal. As we will see, 1009
not all of our rules satisfy all of them.

3.3 Miscellaneous properties 1036

A couple of other relevant properties of Milnor (1951) and of Luce and Raiffa (1989) 1039
are the following: Convextiy, Continuity and the Sure Thing Principle.

Convexity says that "if $a_1, a_2 \in \mathcal{A}$ then $xa_1 + (1-x)a_2 \in \mathcal{A}$ for $x \in [0,1]$". This is 1043

Expansion	$a \in \mathcal{A} \to a \notin \mathcal{A}$	$a \notin \mathcal{A} \to a \in \mathcal{A}$
b dominated	SpEC	SpEC
$b \in \mathcal{U} \setminus \mathcal{L}$	$\mathcal{L}$-EC	EC
$b \in \mathcal{U} \cap \overline{\mathcal{A}}$	StEC	EC
$b \in \mathcal{U}$		EC
Contraction	$a \in \mathcal{A} \to a \notin \mathcal{A}$	$a \notin \mathcal{A} \to a \in \mathcal{A}$
b dominated	SpCC	SpCC
$b \in \mathcal{U} \setminus \mathcal{L}$	CC	$\mathcal{L}$-CC
$b \in \mathcal{U} \cap \overline{\mathcal{A}}$	CC	StCC
$b \in \mathcal{U}$	CC	

1014

Table 3: Summary of expansion and contraction consistency rules

Luce and Raiffa's Axiom 9. Milnor has a version of this that says that mixtures of acts are weakly preferred. That is, it says "if $a_1 \sim a_2$ then $xa_1 + (1-x)a_2 \succeq a_1, a_2$ for $x \in [0,1]$." Binmore (2008) takes issue with this property. He says that "anticonvexity" is just as reasonable a property: "a mixture of two acts is optimal only if the two actions are optimal". He goes on to suggest that a reformualtion of this concept of convexity makes it contradictory: the axiom "if $a_1 \sim a_2$ then $xa_1 + (1-x)a_2 \sim a_1, a_2$ for $x \in [0,1]$" is incompatible with Binmore's list of indispensible axioms. Binmore's indispensible axioms, however, contain axioms that I don't consider reasonable for imprecise decision, so we needn't worry about this inconsistency.

Continuity is about sequences of decision problems. A sequence of decision problems is defined by a collection of sequences of outcomes. I borrow Binmore's gloss on Milnor's "continuity" axiom here: 1063

> Consider a sequence of decision problems with the same acts and states, in all of which $a_j \succ a_i$. If the sequence of matrices of outcomes converges, then its limiting value defines a new decision problem in which $a_j \succeq a_i$. (Binmore 2008, p. 137) 1068

Continuity is going to be an interesting property. It says that if we keep changing the payoffs slightly and the preference remains stable, then in the limit of our fiddling, the preference will still be stable. The change from "$\succ$" to "$\succeq$" is to allow this to be consistent with a case where act a_j tends to a_i. Or rather, each consequence of a_j tends to the corresponding consequence of a_i. In the limit, they are equal, so there cannot be strict preference between them. Or in terms of optimal acts, if we keep changing the payoffs slightly and the set of optimal acts remains stable, then in the limit, the set of optimal acts is stable. 1076

The sure thing principle is often said to be violated by people's actual choices in games like those of the Allais paradox and the Ellsberg paradox. The property 1090 does, however, have some intuitive appeal, so it is worth discussing with a view to perhaps *limiting* if not eradicating violations of it. Luce and Raiffa's Axiom 8 corresponds to the "sure thing principle". It says "if $f \geq g$ then $xf + (1-x)h \geq xg + (1-x)h$ for $x \in [0,1]$." Or, as Luce and Raiffa put it: " Consider a probability mixture of two decision problems with the same actions and states. If the payofs in the second problem do not depend on the act chosen, then the optimal set in the mixed problem is the same as in the first problem."

Perhaps the best way to understand STP is with an example. 1106

Example 3 I am going to flip a coin. If it lands heads the game is over and You win nothing. If it lands tails, You have the choice to bet on some proposition X. 1107

This is an instance of Sure Thing Principle: the game is a mixture of the game "do nothing" and the game "bet with X". Say f and g are bet on or bet against X. And h is the "do nothing" act. Now, imagine for a moment that in both games the options available are "bet on X" and "bet against X". In the "do nothing" game, You don't actually get a chance to bet, so Your choice doesn't affect Your payout. So which act You choose in the full game depends only on which act You'd choose in the "bet with X" branch of the game. So if You prefer f to g, You prefer f mixed with h to g mixed with h. 1112

Seidenfeld (2004) argues that STP gives us reason to prefer $\mathcal{L}$ to $\mathcal{E}$-WALD since the former satisfies STP while the latter doesn't. 1127

3.4 Equivalent acts 1130

Instead of considering ways we might try to distinguish between acts, let's think of when we *wouldn't* want to distinguish two acts: when You should be indifferent between those acts. Recall Example 1. It seems that we would be wary of a rule that distinguishes between c and d. The two acts seem "mirror images" of each other and if Your probability function is suitably "symmetric" then it would be suspicious if c were evaluated differently from d. That is, for every credal committee member who thinks c has a particular expectation, we can find a committee member who thinks that about d. 1133

Cashing out exactly what this intuition amounts to is a little tricky. In Example 2 we had two acts that we *don't* want to evaluate the same (heads versus 10 heads), but we can find a one to one correspondence between credal committee members who think X about the first act and ones that think X about the second act.[14] So we have to be careful how we articulate this mirror image intuition. This has something of the flavour of Bertrand's paradox I discussed earlier. 1145

But consider deciding between a bet on X and a bet on Y. If You know nothing about either event, or know the same thing about each event, then surely a decision rule shouldn't favour one act over the other. At the very least if two acts differ 1157

[14]This problem only occurs for infinite representors... 1151

only on whether the prizes are determined by X or by Y, and every $\Pr \in \mathcal{P}$ has $\Pr(X) = \Pr(Y)$ then the acts should be evaluated the same way.

4 Assessing the rules 1165

Let's look at how the various rules we have discussed fare with respect to the criteria we have outlined. 1167

4.1 Permissive rules 1170

First let's consider two permissive rules: $\mathcal{U}$ is optimal; $\mathcal{L}$ is optimal. 1173

Which properties do these rules satisfy? Let's consider $\mathcal{U}$. Obviously $\mathcal{U}$ satisfies "Pareto": that is what it was designed to do! It is certainly the case that $\mathcal{U}$ is always non-empty. 1177

It also satisfies Axiom 4, which says that optimal acts can't be dispreferred to non-optimal ones: no dominated act is preferred to any undominated one. I don't think it is wise to go further and say that You should be indifferent between all undominated acts. 1183

What about expansion and contraction consistency? $\mathcal{U}$ certainly fails all-or-nothing expansion consistency. 1189

Example 4 I offer You a choice between: 1191

f Gain £10 if X, nothing otherwise 1194

g Gain nothing if X, £10 otherwise 1195

Now I add a third act: 1197

h Gain £11 if X, £1 otherwise 1199

h dominates f, so in the expanded decision problem, f is not optimal. However, g is still undominated, so this violates all-or-nothing. The same example works for $\mathcal{L}$. It should be obvious that "$\mathcal{U}$ is optimal" and "$\mathcal{L}$ is optimal" both satisfy SpEC and SpCC. In both cases, dominated acts can't do anything to change what is optimal. 1202

A dominated (not E-admissible) act cannot be made undominated (E-admissible) by the addition of new acts. So $\mathcal{U}$ and $\mathcal{L}$ both satisfy EC. Likewise, both satisfy CC. 1212

The only way an act can stop being optimal (Undominated, E-admissible) is if a new act is better than it (everywhere, where it maximises). Thus both $\mathcal{U}$ and $\mathcal{L}$ satisfy StEC. Likewise, both satisfy StCC. 1217

Turning now to the miscellaneous properties, convexity first. A mixture of undominated acts can be dominated (See Table 4). Now each of a_1 and a_2 are undominated. But the mixture is dominated by a_3. 1222

So convexity is not true for $\mathcal{U}$. Both of a_1 and a_2 in Table 4 are E-admissible, but their mixture (a_4) is not. So convexity can be violated by $\mathcal{L}$ too. 1242

	s_1	s_2
a_1	2	–2
a_2	–2	2
a_3	1	1
$0.5a_1 + 0.5a_2 = a_4$	0	0

1229

Table 4: A mixture of undominated acts can be dominated

	s_1	s_2
a_1	0	1
a_2	x	x

1263

Table 5: A sequence of undominated acts with a dominated limit

Would it be possible to construct a rule that never picks an act whose mixtures 1247
with other acts can be dominated? Would it be desirable? Any decision rule that allows $a_1 \succeq a_3$ and satisfies Pareto is going to violate convexity. Is this a reason to prefer rules that decide in favour of a_3 as against a_1 and a_2?

Continuity next. A sequence of undominated acts can have a dominated limit. 1255
For example, consider the decision problem in Table 5. For $x > 0$, a_2 is undominated, but as you let x tend to 0, it becomes dominated in the limit. And I take it for granted that performing dominated acts is irrational. Now, would this be a potential case of a failure of continuity? For that to be the case, we'd have to stipulate for all $x > 0$, we have $a_2 \succ a_1$. Because then, in the limit we have $a_1 \succ a_2$ thus violating continuity. Is it reasonable that $a_2 \in \mathcal{A}$ for all $x > 0$? $\mathcal{U}$ and $\mathcal{L}$ both agree with that. But maybe that just shows that they aren't reasonable decision rules on their own. But then, making that claim seems to rely on some sort of unwarranted intuition about "higher order probabilities" in the same way that criticisms of maximin rely on unwarranted intuitions about the chances of the events like we saw for ignorance in Table 2. It also takes us beyond what our simple permissive rules actually say. Intuitively, for small enough ε, $x = \varepsilon$ would make a_2 dispreferred to a_1. But does this intuition rely on some unwarranted assumption about some minimum probablity that s_1 could have?

Even if this intuition can be bolstered, is it the case that *all* the "potential fal- 1294
sifiers of continuity" are such that the preference switches at some point before we reach the limit? If we want to satisfy continuity, we have to impose a restriction of sequences of acts that tend to a dominated act. This would say something like "if the limit of the sequence of acts is dominated then at some point before reaching the limit, the act becomes dispreferred to the act that dominates in the limit". Some of the decision rule I look at will be such that they build in this sort of restriction.

How do $\mathcal{U}$ and $\mathcal{L}$ deal with the Sure Thing Principle? Mixing f and g with the 1306 same thing can't change the relation between the expectations of f and g for each Pr. So $\mathcal{U}$ satisfies STP. And since whatever maximises will remain maximal if each act is mixed with the same thing, $\mathcal{L}$ does too.

4.2 Simple extrema rules 1313

By "extrema rules" I mean those that pay attention only to the extrema of the set of expectations. That is, $\mathcal{E}$-Wald, $\mathcal{E}$-Hurwicz, $\mathcal{U}$-Wald, $\mathcal{U}$-Hurwicz,$\mathcal{L}$-Wald and $\mathcal{L}$-Hurwicz. The simple rules are those that don't do any funny business restricting to a smaller space of acts: $\mathcal{E}$-Wald and $\mathcal{E}$-Hurwicz. The former is a special case of the latter, but given that "maximin" is suggested independently of Hurwicz criterion style rules, I deal with it separately. As we will see, treating with only the minimum value will lead to pathological behaviour. 1316

First, $\mathcal{E}$-Wald: this isn't a very good rule. $\mathcal{E}$-Wald can't discriminate between dominated and undominated acts. 1327

Example 5 You have a choice between: 1329

f Get nothing 1333

g Get £1 if X, nothing otherwise 1334

Both of these acts have $\underline{\mathcal{E}} = 0$ so they are evaluated the same by $\mathcal{E}$-Wald. Despite this, every $\Pr \in \mathcal{P}$ gives g at least as high expectation. For any $\rho < 1$, $\mathcal{E}$-Hurwicz satisfies Pareto at least for "straight" acts. If we allow acts like those in Example 2,[15] then $\mathcal{E}$-Hurwicz also violates Pareto. That is, in the choice between betting on Heads or betting on 10 Heads in a row, both acts have the same $\overline{\mathcal{E}}$ and $\underline{\mathcal{E}}$ if the representor covers the whole $[0, 1]$ interval. So they are evaulated the same by $\mathcal{E}$-Hurwicz for all ρ. 1337

These rules deliver non-empty optimal sets. $\mathcal{E}$-Wald and $\mathcal{E}$-Hurwicz both determine a complete and transitive ordering on the acts. They also satisfy all of the expansion and contraction consistency properties. 1351

$\mathcal{E}$-Wald satisfies convexity and continuity, but violates STP. 1357

Example 6 I offer You the choice between: 1358

f Gain £10 if X, nothing otherwise 1361

g Gain £1 if X, £11 otherwise 1362

Now imagine mixing 50–50 each of f and g with g. 1364

$\mathcal{E}$-Wald supports doing g over f. But $.5f + .5g \succ g$ by the lights of $\mathcal{E}$-Wald. So STP is violated. For $\rho > 0.5$ the same example works to show $\mathcal{E}$-Hurwicz violates the Sure Thing Principle. For $\rho < 0.5$ consider mixing f and g with f. $\rho = 0.5$ is not immune to this problem either. 1381

[15] That is, acts that aren't "linear in probability" 1342

	X	$\neg X$
f	10	0
g	1	11
$.5f + .5g$	5.5	5.5

1368

Table 6: Example 6

$\mathcal{E}$-Hurwicz violates all the properties that $\mathcal{E}$-Wald does, and in addition, for small ρ, it violates convexity.[16] In Table 4, both of a_1 and a_2 are $\mathcal{E}$-Hurwicz-optimal for suitably small ρ. Mixtures of them will not be optimal. 1389

4.3 Restricted extrema rules 1397

Let's now turn to the remaining extrema rules: $\mathcal{U}$-Wald, $\mathcal{U}$-Hurwicz, $\mathcal{L}$-Wald, $\mathcal{L}$-Hurwicz. 1400

As we've already seen, all of these rules satisfy Pareto and deliver a non-empty optimal set. These rules also all determine a complete and transitive order on the acts and satisfy linearity. They all satisfy Axiom 4 as well. 1403

$\mathcal{U}$-Wald and $\mathcal{U}$-Hurwicz can violate the "optimal acts are E-admissible" property, but $\mathcal{L}$-Wald and $\mathcal{L}$-Hurwicz can't. 1409

Because both $\mathcal{U}$ and $\mathcal{L}$ violate all-or-nothing expansion consistency, all of the above rules do too. But as we've discussed, this is no real flaw. 1412

All the rules satisfy SpEC and SpCC, since they restrict themselves to undominated acts anyway. Adding an act can't make a non-optimal act $\mathcal{U}$-Hurwicz-better. So $\mathcal{U}$-Wald and $\mathcal{U}$-Hurwicz satisfy EC. Nor can removing an act make an act $\mathcal{U}$-Hurwicz-worse, so $\mathcal{U}$-Wald and $\mathcal{U}$-Hurwicz satisfy CC as well. 1416

Because non E-admissible acts can be optimal for $\mathcal{U}$-Hurwicz, this rule violates $\mathcal{L}$-EC and $\mathcal{L}$-CC. It does, however, satisfy StEC and StCC. 1423

The outlook is not so good for the $\mathcal{L}$-type rules. Isaac Levi champions $\mathcal{L}$-Wald as the best imprecise decision rule. He argues that only those acts that have some chance of being the best are worthy of consideration. He argues for the "maximise minimum expectation" part by analogy to what happens in precise cases of equal expectation. 1427

He considers a bet like the following: I offer You a bet where You win £1 if heads comes up, but You lose £1 if tails lands up. Should You accept this bet, or refuse it? Both acts (accept, refuse) have the same expectation — £0 — so how do You choose between them? Levi suggests that in this situation you should maximin over the acts that maximise expectation. He says that the reason to refuse the bet is: 1434

> not that refusal is better in the sense that it has higher expected utility than accepting the gamble. The options come out equal on this kind of 1441

[16]Smaller ρ means more weight given to $\overline{\mathcal{E}}$. 1391

> appraisal. Refusing the gamble is "better" however, with respect to the security against loss it furnishes. (Levi 1974, p.411)

He suggests the same reasoning works in the imprecise case. We should use "security" as a tiebreaker.

$\mathcal{L}$-Wald violates EC.

Example 7 You are offered the following choice:

f	£10 if X, nothing otherwise
g	£3 if X, £3 otherwise

I now add another act

h	£–1 if X, £8 otherwise

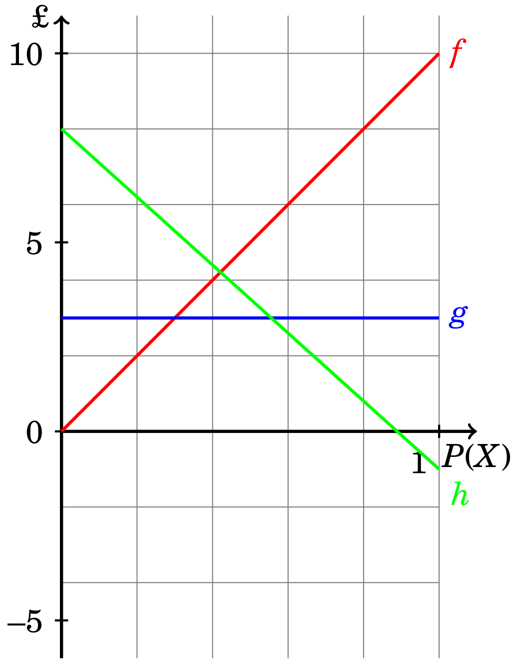


Figure 4: Graph of Example 7

In a choice between f and g, it is g that does best by $\mathcal{L}$-Wald. However, adding h means that g is no longer E-admissible and of f and h, f does better. So the addition of h turns f from a non-optimal act into an optimal one. It also turns g from optimal into non-optimal without the added act being optimal. So $\mathcal{L}$-Wald violates EC and StEC. Removing h from the above example shows that $\mathcal{L}$-Wald violates CC and StCC as well. It does, however, satisfy $\mathcal{L}$-EC and $\mathcal{L}$-CC. For large values of ρ, $\mathcal{L}$-Hurwicz inherits all the same problems. The smaller the ρ, the harder it is to make $\mathcal{L}$-Hurwicz violate EC and CC.

The trick that causes $\mathcal{L}$-Hurwicz to violate all these properties is as follows: some act g is E-admissible in virtue of maximising for some value of $\Pr(X)$. And it also maximises minimum expectation at some *other* value of $\Pr(X)$ for which g *does not maximise expectation*. So You can make it fail E-admissibility while still having it maximise minimum expectation. Since you have this independence of the two parts of this "lexicographic" rule, you get all the strange behaviour associated with lexicographic orderings.

Now, for ρ small enough — take the case of $\rho = 0$: maximax— you can no longer do this: the places where it wins the Hurwicz criterion test are also the places where it is E-admissible. So the trick that works for $\rho = 1$ is harder to pull off, and in the limit, impossible. 1500

What we saw in Example 7 is that g maximises minimum expectation but an act that performs worse on this metric can still make it not E-admissible and therefore not optimal. This means that some other previously non-optimal act must become optimal. 1509

$\mathcal{U}$-Wald violates continuity, since $\mathcal{U}$ does. $\mathcal{U}$-Hurwicz does too, for large ρ. For small ρ, $\mathcal{U}$-Hurwicz is much less likely to violate continuity. In the $\rho = 0$ limit it is impossible: if a sequence of decision problems has a stable maximax act, then in the limit, that act is still optimal. This is because the "dominated limit" trickery of Table 5 won't work. The act is E-admissible *at the point where it also maximises*. This satisfaction of continuity comes at a cost, however: recall that for small ρ $\mathcal{E}$-Hurwicz and thus $\mathcal{U}$-Hurwicz violate convexity. It is the same story for $\mathcal{L}$-Wald and $\mathcal{L}$-Hurwicz. The same example that shows $\mathcal{E}$-Hurwicz violates STP works for $\mathcal{L}$-Hurwicz and thus for $\mathcal{U}$-Hurwicz. 1515

The problem with $\mathcal{L}$-Wald is that an act can maximise minimum expectation, but not be E-admissible. This is basically the root of its problems with expansion and contraction consistency. An act that *maximises* maximum expectation has to be E-admissible. So, for a given decision problem; the smaller the ρ, the less likely it is that you see this awkward behaviour. For large ρ, You have problems with convexity; for small ρ, You have problems with continuity and expansion consistency. 1531

4.4 Savage 1543

I turn now to the rules in the mould of Savage's minimax regret rule: $\mathcal{E}$-Savage and $\mathcal{L}$-Savage. There is no "$\mathcal{U}$-Savage" since the **Reg** function is never maximised by a dominated act, so $\mathcal{E}$-Savage is already restricted to $\mathcal{U}$. So $\mathcal{E}$-Savage satisfies Pareto. It always defines a complete, total ordering and a non-empty optimal set. It also satisfies linearity and Axiom 4. 1546

$\mathcal{E}$-Savage does not do very well with expansion and contraction consistency. Given that the assessment of each act depends on maxima over the set of acts, it shouldn't be surprising that adding or removing acts can affect how the acts are evaluated. 1555

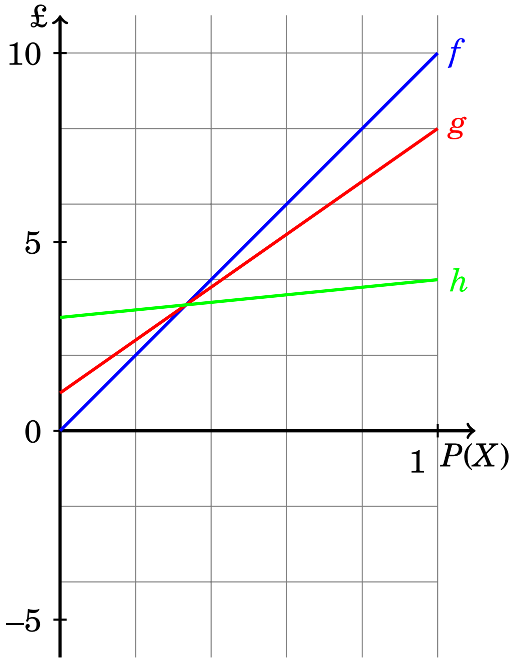
Example 8 Consider the following choice. 1562

f	Gain 10 if X, nothing otherwise	1565
g	Gain 8 if X, 1 otherwise	1566

Now add this new act. 1568

h	Gain 4 if X, 3 otherwise	1570

graph-menufail.tex



End graph-menufail.tex 10

Figure 5: Example 8 as a graph of expectation against probability

Now, in a choice between f and g, f wins, since it is at most 1 from the best act (when $\Pr(X) = 0$) while g drops to 2 less than the top act (when $\Pr(X) = 1$). All this changes when h becomes an option. Now, when $\Pr(X) = 0$, f is 3 from the top act while g's maximum distance from the top stays at 2. So the addition of h has switched the evaluation of the two acts round! Indeed, it is even odder than that: h makes g no longer E-admissible, but it also makes g preferred (by the lights of $\mathcal{E}$-Savage) to both the other acts! So $\mathcal{E}$-Savage doesn't always pick E-admissible acts to be optimal. In fact, you can even construct an example where h dominates g but makes g preferred to f. However, in this case, h is "$\mathcal{E}$-Savage-optimal", so that is perhaps less weird. In short, $\mathcal{E}$-Savage violates all the expansion and contraction consistency conditions except the special ones. 1579

$\mathcal{E}$-Savage does satisfy continuity and convexity. It fails STP though, for the same reason as the other rules. In Example 6 g is preferred to f. But a 50–50 mixture of f and g is preferred to a 50–50 mixture of g with g. 1598

$\mathcal{E}$-Savage isn't a great rule, but it will perhaps make a good "regret penalty" for modifying the other rules like $\mathcal{U}$-Hurwicz. The above discussion is summaraised in Table 7. 1604

	Undom	Levi	EWald	EHurwicz	ESavage	UWald	UHurwicz	LWald	LHurwicz
Pareto			X	X					
Opt in Levi	X		X	X	X	X	X		
Opt neq emptyset									
Ordering	X	X							
Linearity									
SpEC									
EC					X			X	X[1]
Levi-EC			X	X	X	X	X		
StEC			8		X			X	X
All-or-Nothing	X	X			X	X	X	X	X
SpCC									
CC					X			X	X
Levi-CC			X	X	X	X	X		
StCC					X			X	X
Convexity	X	X		X[2]		X	X[2]	X	X[2]
Continuity	X	X				X	X[1]	X	X[1]
STP			X	X	X	X	X	X	X

Table 7: Summary of rules violations

[a] For large ρ 32
[b] For small ρ 33

5 More complex rules 1614

Let's look at some more complicated decision rules. Here I will look at rules with 1617
a spread penalty, rules with multiple values of ρ, and rules that use $\mathcal{E}$-Savage as a penalty.

The idea behind using these rules is two-fold: first it seems a flaw that the spread 1621
of expectation for an act does not count against it; second, "perturbing" the simple rules above allows them to avoid violating continuity.

Consider $\mathcal{E}$-Wald and $\mathcal{U}$-Wald. Both of these rules had problems. $\mathcal{E}$-Wald vio- 1627
lated Pareto, $\mathcal{U}$-Wald violated continuity. In both cases, the problem disappears if the acts are changed very slightly.

Take, for example, $\mathcal{U}$-Wald. This rule violates continuity because a sequence of 1633
undominated acts can have a dominated limit, and the maximin part does nothing to help. And $\mathcal{U}$-Hurwicz can be afflicted with the same problem for any ρ with some suitably tricky acts. But it seems unreasonable that the acts in the sequence should be preferred to the act that dominates in the limit, *no matter how close You get to the limit*. And the $\mathcal{E}$-Savage rule has the nice property that as You tend towards a dominated limit, the act does less well, and stops being optimal before You reach the limit. So adding a "regret penalty" helps to mitigate the effect of dominated limits. This does however open You up to the more serious violations of expansion and contraction consistency that afflict $\mathcal{E}$-Savage. However, if the penalty is suitably small, then these problems will only occur in very extreme cases. The question becomes: "How important is it to satisfy continuity? Is it worth the potential violations of EC?"

Take Example 2, with the choice between Heads and 10 Heads in a row. But 1653
modify it so that the latter bet wins £$1 + x$. Now as x tends to 0, this latter act tends to a dominated act. However, for every value of ρ it does better on a $\mathcal{U}$-Hurwicz evaluation than the bet on Heads does. If we add a "regret penalty" to this evaluation, the 10 Heads act stops being optimal before it becomes dominated, since it has a higher regret than the 1 Heads bet. In this way, the regret penalty limits the violations of continuity.

Another approach is to consider $\mathcal{U}$-Hurwicz rules, but to assess acts on the basis 1664
of a number of values for ρ. For example, when $\mathcal{U}$-Hurwicz rules violate continuity, this violation is very sensitive to the value of ρ: tiny changes of ρ will cause the violation continuity to disappear in that particular case. It seems that You don't want to have Your decision rule be so sensitive to choice of ρ, so looking at some small collection of values of ρ might suggest a way to avoid violating continuity.

6 Conclusion 1675

The list of properties I have discussed here is by no means complete. Nor will 1678
all of the properties I have discussed strike everyone as reasonable. However, I

hope to have moved us towards a better imprecise decision rule. One conclusion I think I can draw from this is that Levi's rule of "maximise minimum expectation among E-admissible acts" ($\mathcal{L}$-Wald) is not a good rule. Neither the restriction to $\mathcal{L}$, nor the exclusive focus on the minimum expectation, is reasonable. I think the restriction to $\mathcal{U}$ is unequivocally a good thing. That leaves us with Hurwicz or Savage based rules. $\mathcal{E}$-Savage's violation of pretty much all the contraction and expansion consistency properties rules it out as a potential good rule. It might however serve as a penalty modifier. So $\mathcal{U}$-Hurwicz rules seem to come out best.

My own preference is towards some kind of $\mathcal{U}$-Hurwicz based rule, with a "regret 1695
penalty" based on $\mathcal{E}$-Savage. I think with a suitable tuning of the ρ parameter and the σ parameter for the regret penalty, this rule can be made to fit with intuitions in many examples like Table 2, and will only suffer from violations of EC, convexity and the like in extreme circumstances.

$$\rho\,\underline{\mathcal{E}} + (1-\rho)\,\overline{\mathcal{E}} + \sigma\,\mathbf{Reg}$$

That's not to say that there may be some alternative rule that doesn't fit into 1706
the categories I have discussed which does better than all the rules I have looked at so far. And of course, extending this work to sequential decisions or decisions with learning may suggest a different conclusion. But for the simple cases I have explored, I think this rule does best.

The regret penalty brings with it dangers of violating EC and friends, but as 1713
long as the value of σ is suitably small, these violations will only happen in extreme cases.

I have explored a variety of decision rules for imprecise decision. I have ar- 1717
gued for restricting Yourself to undominated acts and against the restriction to E-admissible acts. I have argued for some sort of Hurwicz based rule, and against the extreme pessimism of Wald. I have argued that despite the danger of violations of expansion consistency, a "regret penalty" based on $\mathcal{E}$-Savage is a useful way of limiting violations of continuity.

References 1726

All figures drawn by the author in TikZ. 1726

Binmore, K. (2008). *Rational Decisions*. Princeton University Press. 1726
Bradley, R. (2009). "Revising Incomplete Attitudes". *Synthese* 171, pp. 235–256. 1726
Cozman, F. (n.d.). *A Brief Introduction to the Theory of Sets of Probability Measures*. 1726
http://www.poli.usp.br/p/fabio.cozman/Research/CredalSetsTutorial/quasi-bayesian.html.
Elga, A. (2010). "Subjective Probabilities should be Sharp". *Philosophers' Imprint* 10. 1726
Gaertner, W. (2009). *A Primer in Social Choice Theory*. Oxford University Press. 1726
Gillies, D. (2000). *Philosophical theories of probability*. Routledge. 1726
Jaynes, E. T. (1973). "The well-posed problem". *Foundations of Physics* 3, pp. 477–493. 1726

Jaynes, E. T. (2003). *Probability theory: The logic of science*. Early draft version available: http://omega.math.albany.edu:8008/JaynesBook.html. Cambridge University Press. 1726

Joyce, J. M. (2005). "How Probabilities Reflect Evidence". *Philosophical Perspectives* 19, pp. 153–178. 1726

– (2011). "A Defense of Imprecise Credence". *Oxford Studies in Epistemology* 4. forthcoming. 1726

Kaplan, M. (2010). "In defense of modest probabilism". *Synthese* 176, pp. 41–55. 1726

Kyburg, H. and M. Pittarelli (1992). *Set-based Bayesianism*. Tech. rep. UR CSD;TR407. http://hdl.handle.net/1802/765. University of Rochester, Computer Science Department. 1726

Lehrer, K. and C. Wagner (1981). *Rational consensus in science and society*. Reidel. 1726

Levi, I. (1974). "On Indeterminate probabilities". *Journal of Philosophy* 71, pp. 391–418. 1726

– (1986). *Hard choices: decision making under unresolved conflict*. Cambridge University Press. 1726

Luce, R. D. and H. Raiffa (1989). *Games and Decisions*. Dover. 1726

Milnor, J. (1951). *Games against Nature*. Tech. rep. RAND corporation. 1726

Peterson, M. (2009). *An Introduction to Decision Theory*. Cambridge University Press. 1726

Seidenfeld, T. (2004). "A contrast between two decision rules for use with (convex) sets of probabilities: Γ-maximin versus E-admissibility". *Synthese* 140, pp. 69–88. 1726

Seidenfeld, T. and L. Wasserman (1993). "Dilation for sets of probabilities". *Annals of Statistics* 21, pp. 1139–1154. 1726

Sen, A. (1970). *Collective choice and social welfare*. Holden-Day. 1726

– (1977). "Social choice theory: A re-examination". *Econometrica* 45, pp. 53–89. 1726

Smets, P. and R. Kennes (1994). "The Transferable Belief Model". *Artificial Intelligence* 66, pp. 191–234. 1726

Sturgeon, S. (2008). "Reason and the grain of belief". *Noûs* 42, pp. 139–165. 1726

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Vol. 42. Monographs on Statistics and Applied Probability. Chapman and Hall. 1726

Wheeler, G. (forthcoming). "Dilation Demystified". In: *The Continuum Companion to Epistemology*. Ed. by A. Cullison. Continuum. 1726

White, R. (2010). "Evidential Symmetry and Mushy Credence". *Oxford Studies in Epistemology*. 1726

Williamson, J. (2010). *In Defense of Objective Bayesianism*. Oxford University Press. 1726