Confidence and decision*

Brian Hill
GREGHEC, HEC Paris & CNRS†

Abstract

Many real-life decisions have to be taken on the basis of probability judgements of which the decision maker is not entirely sure. This paper develops a decision rule for taking such decisions, which incorporates the decision maker's confidence in his probability judgements according to the following maxim: the larger the stakes involved in a decision, the more confidence is required in a probability judgement for it to play a role in the decision. A formal representation of the decision maker's confidence is proposed and used to formulate a family of decision models conforming to this maxim. A natural member of this family is studied in detail. It is structurally simpler than other recent models of decision under uncertainty, which may make it easier to apply to practical decisions, whilst being axiomatically sound, permitting the separation of beliefs and tastes, and allowing comparative statics analysis of attitudes to choosing in the absence of confidence.

Keywords: Confidence; multiple priors; confidence ranking; cautiousness coefficient; ambiguity.
JEL classification: D81, D83.

*I would like to thank Mohammed Abdellaoui, Itzhak Gilboa, Edi Karni, Massimo Marinacci, the editors and reviewers who have commented on the paper, and the participants of the FUR XIV, RUD 2010 and the HEC Decision Sciences Seminar, for stimulating discussion and helpful feedback.

†1 rue de la Libération, 78351 Jouy-en-Josas, France. Tel: + 33 1 39 67 72 65. Fax: + 33 1 39 67 71 09. E-mail: hill@hec.fr.

1 Introduction

1.1 General motivation

A regional governor is faced with the decision whether to permit the construction of a factory in his area. To evaluate a major uncertainty in the choice – the possibility of damage to the district farming area (which specialises in maize) – he commissions meteorologists to estimate the probability that the gases emitted by the proposed factory, which are known to damage crops, reach the farming area. Although there is a fair amount of disagreement among the experts, the governor considers that a value of at most 10^{-5} for the probability is quite representative of the opinions. On the basis of this probability estimate and his utilities for the relevant consequences, which he determines using standard decision-analysis techniques, he decides to grant permission for the construction. However, the project falls through, and a different project comes before him for the same land, namely to use it for GM maize crops (which are wind pollinated). Evaluating the possible consequences as before, it becomes clear that, while many aspects are analogous to those for the factory, the orders of magnitude are far greater: both the advantages of allowing the development and its main potential undesirable consequence – cross-pollination of GM crops with non-GM crops – are larger by a factor of a thousand, in the governor's opinion. As concerns the probability that pollen from the GM crops arrives at the non-GM farming area, the experts guarantee that this is the same as the probability of fumes from the factory arriving at the maize fields, since both depend on the same meteorological factors. Were the governor to decide as before, using his utilities and the 10^{-5} estimate adopted previously, he would approve the GM project. However, he is uncomfortable with this choice. Although he was happy to rely on the probability estimate of at most 10^{-5} for the decision involving the factory, he is not sure enough in this estimate when it comes to deciding about the GM crops: after all, given the potentially grave consequences of GM infection, the stakes are higher in the latter decision.

At first glance, it may seem that the governor is violating one of the tenets of decision theory, namely the separation of beliefs and tastes. The standard normative theory of decision prohibits altering the beliefs, or probability judgements, used depending on the decision faced. However, this theory ignores the fact that decision makers may be more or less *confident* in the beliefs which inform their decisions, and that confidence in beliefs is potentially relevant for decision making. Rather than using different beliefs in the two decisions, the governor could be understood as demanding different levels of confidence in the beliefs he uses. When the stakes are mediumly high – at worst the agricultural yield suffers for a (relatively) short period due to pollution – he allows himself to rely on a judgement – that the probability is at most 10^{-5} – which he is not confident enough in to invoke when the stakes are higher – when the worst that can happen is that the GM genes infect the non-GM population. In the light of these considerations, the governor ceases to appear irrational. Indeed, he can be understood as invoking the following, perfectly reasonable maxim: the more important the decision, the more confidence is required in a belief for it to play a role in that decision.

The main aim of this paper is to develop a decision rule based on this maxim. This rule should be normatively plausible, as well as simple and tractable, and hence applicable to decisions such as the one considered above. Such a decision rule would be able to incorporate the intuitions just mooted, without violating general principles of rationality, such as the separation of beliefs and tastes.

A decision rule of this sort may be useful for prescriptive purposes in several domains. In decisions about policies to adopt in the face of climate change, for example, the stakes are high and the probabilities are uncertain; the same goes for other domains where the Precautionary Principle has been an object of debate.[1] A decision rule of the sort considered here naturally applies in such cases, and recommends only relying on probability judgements in which one has enough confidence to match the gravity of the potential consequences.

Moreover, the rule may have implications for the use made of conclusions of statistical studies (in classical statistics). Often, confidence intervals at a standard fixed level (for example, the 95% level) are used to inform decisions. The proposed decision rule would suggest that one may vary the confidence intervals one uses in decisions: whereas if relatively little is at stake one may use confidence intervals at, say, the 90% level, if there is a lot at stake one might insist on relying only on intervals that are at, say, the 99% level.

Finally, the theory developed below may be descriptively relevant. Consider the case of an investor who invests in a start-up drug firm whose major product is a new vaccine which has just come out on the market amid some controversy, though he refuses to let his daughter take the vaccine. This is naturally taken account of under the proposed rule: the investor is confident enough in his judgement about the prob-

[1]One specificity of such decisions, which plays a role in many analyses, is the irreversability of the consequences of actions. This of course implies that the stakes are higher than they otherwise would be: one cannot undo tomorrow what has been done today.

ability that the drug will be a success to take the investment decision on the basis of that judgement, but not confident enough to rely on it in the more important personal decision.

In this paper, we propose a formal model of a decision maker's confidence in his beliefs and use it to develop a decision rule that adheres to the maxim that the higher the stakes involved in a decision, the more confidence is required in a probability judgement for it to play a role in that decision. In reality, a family of such decision rules can be defined on the basis of the model of confidence proposed, and the main purpose of the paper is to promote this family. To this end, and for the sake of concreteness, we perform an in-depth axiomatic study of a single, natural member of the family. Perhaps surprisingly given the simplicity of the model and the maxim on which it is based, it corresponds to relatively attractive properties of preferences, which are essentially stakes-corrected versions of standard axioms. Moreover, it provides a full separation of utilities, belief attitudes (including confidence in beliefs) and attitude to choosing in the absence of confidence. Relations to the standard criteria for comparative ambiguity aversion are also examined. Prior to the formal analysis, we further motivate the need for a new model to capture the role of confidence in beliefs in decision by relating it to a well-known behavioural axiom, and present informally the main aspects of our proposal, as well as connections with existing literature.

1.2 Confidence and behaviour: an example

The relevance of confidence in beliefs for decision making in accordance with the maxim mentioned above is particularly evident behaviourally in certain violations of Gilboa and Schmeidler's (1989) C-independence axiom. Under the assumption of linear utility for money, this axiom basically states that preferences between a pair of acts (that is, functions from states to consequences) are unaffected by multiplying the values of the consequences delivered by the acts by a positive number and by adding a constant to them. In order to see the relevance of this property for confidence, let us consider a precise example that can be thought as a stylized version of the governor's decision described at the beginning of the paper.[2]

Suppose that the governor has linear utility,[3] and consider an urn in which he is told

[2]Formal definitions and technical details in a general framework, as well as related discussion, are given in Section 2.

[3]Of course, this assumption is not essential to the example, but is made for ease of presentation. To dispense with it, it suffices to interpret the values of consequences presented below as given in utiles, or consider that the ball drawn from the urn determines the composition of a second urn in such a way that the

that there are one million balls, each of which is coloured either red or blue. He knows for sure that at least 990 000 balls in the urn are blue, and that at least one ball in the urn is red. Moreover, his advisers (who are experts on urns, but have not been able to count the balls in this urn) estimate that at most ten of the balls in the urn are red. Given this situation, consider how he prices the bet f given in Figure 1 – that is, consider the value for which he is indifferent between receiving that value for sure and playing the bet. Suppose for example that he prices the bet at $-\$0.1M$ (this is given in Figure 1 in the line p_f). It follows from C-independence that the governor should be indifferent between $\alpha f + \eta$ and $\alpha p_f + \eta$ in Figure 1 for any positive α and any η.[4] For example, to satisfy this axiom, he must price the bet g, which yields $\$10000$ if the ball is blue and $-\$1M$ if the ball is red, at $-\$100$ (take $\alpha = 10^{-3}$ and $\eta = 0$). More generally, C-independence implies that whenever the governor gives f a negative price (and so prefers p_0 in Figure 1, which yields $0 for sure, to f) then he must do the same for the bet g. This is at odds with the behavioural pattern described in the previous section. The urn could be thought of as a stylized representation of the meteorological conditions and the governor's information concerning them, the bet g could be considered as a stylized version of the option of allowing the construction of the factory, and the bet f could be considered as analogous to the option of allowing planting of GM crops. As noted previously, it does not seem unreasonable for the governor to reject the GM project (ie. to prefer p_0 to f), whilst accepting the construction of the factory (ie. preferring g to p_0), and this is incompatible with C-independence.

Since the governor's pattern of choice in this example violates C-independence, it cannot be captured by decision models satisfying this axiom, such as the standard expected utility model and the maxmin expected utility model of Gilboa and Schmeidler (1989).[5] Indeed, for the purposes of gaining intuition, it may be useful to consider the choices in the light of the latter model, which evaluates an act by the lowest expected utility calculated over a fixed set of probability measures. In the example considered, the lowest expected utility is attained with the highest probability of the 'bad' event, namely the drawing of a red ball. The indifference between f and p_f is consistent (under the assumption that utility is linear) with the use of 0.01 as the highest possible

utility values of the second lottery coincide with those given.

[4] Under the assumption of linear utility, these acts correspond to mixture of f (resp. p) with $\frac{\eta}{1-\alpha}$, in the sense that will be defined in Section 2.1.

[5] A large class of models satisfying this axiom – those of invariant biseparable preferences – is studied in Ghirardato et al. (2004). As they note, C-independence essentially corresponds to a property of functional representations called constant linearity.

Figure 1: Bets (values in dollars, 'M' stands for 'million')

	Colour of ball drawn from urn	
	Blue	Red
f	10 M	-1 000 M
p_f	-0.1M	-0.1M
$\alpha f + \eta$	10 M $\times \alpha + \eta$	-1 000 M $\times \alpha + \eta$
$\alpha p + \eta$	- 0.1 M $\times \alpha + \eta$	- 0.1 M $\times \alpha + \eta$
g	10 000	-1 M
p_0	0	0

Preferences incompatible with the maxmin EU model: $f \prec p_0$ and $g \succ p_0$.
Preferences incompatible with the smooth model: above, $f \sim p_f$ and
$\alpha f + \eta \sim \alpha p_f + \eta$ whenever $-1000M \times \alpha + \eta < -10M$

probability of getting a red ball. So, in pricing the bet this way, the governor takes the worst probability out of those that, for all he knows, may be relevant; in doing so, he ignores his advisers' estimates. (Recall that he knows for sure that there are at most 10 000 red balls in the urn.) If, on the other hand, he gives g a positive price, this would correspond in the maxmin expected utility model to him taking as worst-case probability of drawing a red ball a value which is lower than 0.01 (and more precisely, lower than $\frac{10}{1010}$). This does not seem unreasonable: indeed, if he relied on his advisers' estimates in that decision, the worst-case probability would be 10^{-5}, and he would price the bet at \$9989.90. Hence, although the maxmin expected utility representation does not allow the use of different worst-case probabilities for different decisions, one can understand why one might want to use them: in the high-stakes decision (concerning f), one is not confident enough in the advisers' estimates of the worst-case probability to rely on them, whereas in the lower-stakes decision (concerning g), one is confident enough in the estimates to use them, and hence to work with a lower worst-case probability for the 'bad' event.

Although some more recent decision models do not satisfy C-independence, it is not clear that they can capture the sorts of violations that appear to be related to confidence as just described. To illustrate this, consider the popular smooth ambiguity model proposed by Klibanoff et al. (2005) and others (see Section 1.4), and the following extension of the example given above. Suppose that, beyond the preferences detailed above, the governor is indifferent between $\alpha f + \eta$ and $\alpha p_f + \eta$ whenever the worst consequence of the bet $\alpha f + \eta$ is below $-\$10M$. As noted, this can be understood as him ignoring his advisers' estimates and evaluating the bet using the worst probability value out of those that may be relevant given what he knows for sure about the urn. Moreover, such preferences appear to be coherent with his other behaviour: decisions where he could lose more than $\$10M$ are considered to involve very high stakes, and he is not confident enough in his advisers' estimates to rely on them in such decisions. However, as shown in Appendix B, whenever the decision maker is ambiguity averse, the set of preferences just described (see Figure 1) is inconsistent with the smooth ambiguity model. To the extent that these preferences do not seem *prima facie* unreasonable, and may even match some basic intuitions about confidence, this suggests that the smooth ambiguity model, although it allows for some violations of C-independence, cannot fully capture the role of confidence in decision making.

As we shall see in Section 2.2, the weakening of the C-independence axiom, in a sense compatible with the considerations made above, turns out to be the central axiomatic difference between the model proposed here and comparable models. It is to this proposal that we now turn.

1.3 The Proposal

The first part of the proposal consists of a representation of a decision maker's confidence in probability judgements, that is, in statements concerning probabilities, such as 'the probability of failure of a central element in a nuclear installation is greater than 0.0002'. For this, we use a nested family of sets of probability measures, which we call a *confidence ranking*. The sets in the nested family can be thought of as corresponding to levels of confidence – the larger sets correspond to higher levels of confidence. To each set in the confidence ranking, there corresponds a set of probability judgements, namely those which hold for all probability measures in the set. This can be thought of as the set of probability judgements which the decision maker holds with the corresponding level of confidence. So for each level of confidence, the confidence ranking delivers a set of probability judgements in which the decision maker has this much confidence. Note that the larger the confidence level in question, the larger the corresponding set of probabilities measures and thus the fewer probability judgements hold for all probability measures in this set. Hence, under this representation, for higher confidence levels, fewer probability judgements are held with those levels of confidence, as one would expect. A probability judgement that holds for some set in the confidence ranking is endorsed by the decision maker, although he may have only very little confidence in it. Conversely, for a given probability judgement, the more sets in the family which are such that the probability judgement holds for all probability measures in the set, the more confident the agent is in that judgement. Just like the probability measure in the standard Bayesian theory of decision under uncertainty, a natural reading of a confidence ranking is as a subjective element capturing the decision maker's beliefs and his confidence in them.

To illustrate, suppose that a decision maker has data to estimate the probability that a certain element in a nuclear reactor will fail under extreme conditions. Upon calculation, he finds that the confidence interval is $[0.0002, 0.0005]$ at the 95% level, and that it is $[0.00005, 0.00065]$ at the 99% level. If he forms his judgements solely on the basis of these data, one might expect him to be more confident that the probability of failure ($p(F)$) is less than 0.00065 than that the probability is greater than 0.0002. This can be represented by a confidence ranking where, for every set in the ranking such that $p(F)$ is greater than 0.0002 for all probability measures in the set, $p(F)$ is less than 0.00065 for all probability measures in the set. Such a confidence ranking is illustrated in Figure 2: the points are probability measures, the concentric circles represent the sets of probability measures in the confidence ranking and the shaded areas are the sets of probability measures where $p(F)$ is greater than 0.0002 and less than 0.00065 respectively.[6] The fact that $p(F) \leq 0.00065$ holds for more of the sets in the confidence ranking that $p(F) \geq 0.0002$ captures the higher confidence in the former probability judgement than in the latter one.

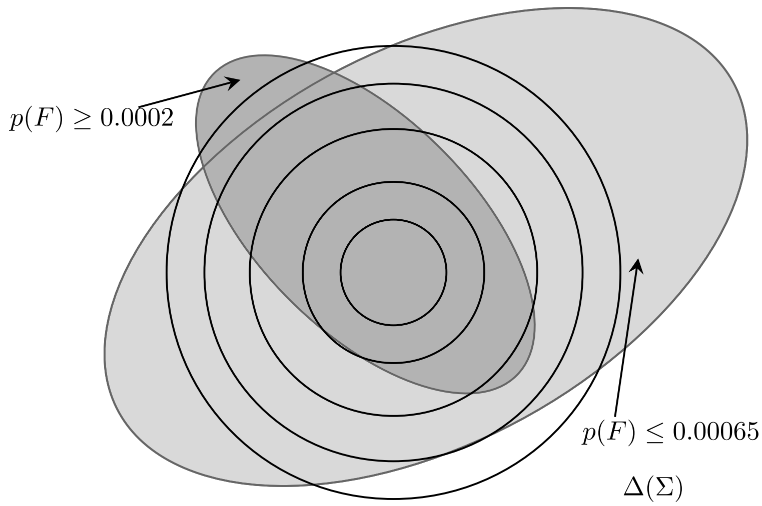
A noteworthy property of this representation of confidence in beliefs is that it involves only an ordinal structure on the space of probability measures.[7] In this sense, the model is simpler than other recent models of decision under uncertainty, which, as discussed in Section 1.4, require a structure on the set of probability measures with a certain degree of cardinality. Consequently, if one wished to use the model to aid decision making, one need only elicit an ordinal structure from a decision maker to capture his confidence, rather than deal with cardinal confidence comparisons.

The second part of the current proposal is a decision rule which uses this representation of confidence and which translates the maxim that the higher the stakes, the more confident one must be in a probability judgement for it to play a role in the deci-

[6]It would have been more accurate to represent these latter sets of probability measures by half-planes; the current diagram is preferred because it makes the comparison between them clearer and more suggestive.

[7]As shown in Proposition 2 in Appendix A, a confidence ranking corresponds to an order on the space of probability measures. See also Section 2.1.

Figure 2: Confidence ranking and confidence in probability judgements



sion. Consider a function that associates to each possible value of the stakes involved in a decision a set in the confidence ranking. We call such a function a *cautiousness coefficient*. Since a set in the confidence ranking corresponds to a level of confidence, this function can be thought of as representing the level of confidence required for a probability judgement to play a role in a decision with given stakes. (As will be shown in Section 3, this function captures the decision maker's attitude to taking decisions in the absence of confidence.) In particular, for any probability judgement that holds throughout the set of probability measures picked out by the cautiousness coefficient, the decision maker is confident enough in that judgement to use it in a decision with the given stakes. Hence a decision rule that uses such a function to pick out a set of probability measures and uses this set of probability measures in the decision captures the maxim mooted above.

What has been said so far constitutes the nub of the current proposal. It does not yield a single decision model but rather a recipe for generating models. In particular, two elements remain to be specified: the decision rule that the decision maker uses once he identifies the appropriate set of probability measures, and the notion of stakes he uses to assign to a decision the stakes involved in it. Different choices of decision rules and notions of stakes yield different decision models conforming to the maxim proposed above. As stated previously, the main concern here is to promote this family; however, for concreteness, we shall focus the analysis on one particular member. Namely, we concentrate on the simplest and most common decision rule using sets of probability measures, the maxmin expected utility rule (Gilboa and Schmeidler, 1989). We take as the stakes involved in the choice of an act the worst consequence that the act could yield. Both of these choices are reasonable: on the one hand, the maxmin expected utility rule has been widely vaunted as characterising careful decision making, whereas on the other hand it seems reasonable that the worse the worst consequence of an act, the more the decision maker has to lose on choosing it, and so the higher the stakes. Section 4 contains a brief discussion of other possible decision rules and notions of stakes.

Concretely, for a cautiousness coefficient D, and a function σ assigning to each act a level of stakes that varies inversely with the worst potential consequence of choosing the act, we consider a decision rule such that the agent (weakly) prefers act g to act f if and only if:

$$(1) \qquad \min_{p \in D(\sigma(f))} \sum_{s \in S} u(f(s)).p(s) \leq \min_{p \in D(\sigma(g))} \sum_{s \in S} u(g(s)).p(s)$$

To illustrate this rule, consider the decision about the sort of technology to use in a nuclear installation. To decide according to the rule, a decision maker has to first ascertain the stakes involved in the choice, which, in the specific model considered here, correspond to the worst possible consequence. For example, the stakes could correspond to a major nuclear accident. Then he needs to determine how much confidence to require in a probability judgement to use it in a decision of such gravity. If his confidence over probability measures is calculated solely on the basis of data using confidence intervals (as in the example above), this corresponds to assigning a confidence level which he will use. For example, he may decide that the appropriate set of probability measures will be those corresponding to the confidence interval at the 99.9% level. Then he would evaluate a technology by the minimum expected utility taken over this set, which, given the very low utility of a nuclear accident, boils down to using the worst-case probability for an accident in the interval.

Note that, were the decision maker considering whether to use a similar technology in another installation with extreme conditions – a smelting plant, for example – this procedure allows him to use the same data differently. For example, if the consequences of failure in the smelting plant are deemed less serious than the consequences of failure in a nuclear installation, then he might allow himself to use probability judgements in which he is less confident. Concretely, this would translate into using, say, the confidence interval at a 95% level, which may lead to different choices of technology.

It is clear that this sort of rule can capture the sort of behaviour mentioned in the example given at the beginning of the paper and developed in Section 1.2. Since the worst consequence of allowing the planting of GM crops (or of bet f in Figure 1) is worse than that of allowing the construction of the factory (respectively, of g), the stakes are higher, so more confidence is required, which implies the use of a larger set of probability measures. As noted in Section 1.2, the use of a more pessimistic probability estimate, which is an immediate consequence of a larger set of probability measures in the context of the maxmin expected utility rule, is entirely consistent with the behaviour described in these examples.

1.4 Related literature

The proposed model can be thought of as a refinement of the maxmin expected utility model of Gilboa and Schmeidler (1989), according to which the decision maker maximises the minimum expected utility over a single, fixed set of probability measures. This model is too simple to capture the role of confidence in decision: were it to be thought of as a model of confidence in probability judgements, confidence would be an all-or-nothing affair – for any probability judgement, either the decision maker is fully confident of it or entirely unsure about it. Something more than a single set of probability measures is required to capture different levels of confidence; the model proposed here goes 'one level up' and takes a family of sets. One consequence of using a single set of probability measures is that the maxmin model cannot capture the relation between the confidence in the probability judgements used and the stakes involved in the decision, which is central here; another consequence is that it does not admit a clean separation of beliefs and attitude to ambiguity, whereas, as shall be shown in Section 3, the model proposed here does. Finally, as noted in Section 1.2, the maxmin expected utility representation implies that preferences satisfy C-independence, which appears incompatible with certain patterns of behaviour that are naturally related to confidence; by contrast, the proposed model only satisfies a weakened version of C-independence (see Section 2.2), and can comfortably account for such patterns.[8]

By contrast with the maxmin model, other, more recent models in the literature

[8] The maxmin expected utility model is discussed here because it is the closest to representation (1). Many of the points hold for similar models, such as the α-maxmin model, which can be used to develop other members of the family of decision models described above. See also Section 4.

do involve a notion of 'degree' on probability measures, but they are all structurally richer than the one proposed here. For example, Klibanoff et al. (2005); Nau (2006); Seo (2009); Ergin and Gul (2009) propose models that use a probability measure over the set of probability measures rather than a nested family of sets. In this sense, their models are cardinal at the second-order level, whereas ours is ordinal. As anticipated above, the structural parsimony of the current model may be seen as an advantage, in particular for prescriptive applications. For someone who wishes to decide using the rules just cited, he needs to fix upon a probability measure over the set of probability measures and a transformation function (above and beyond the utility function) that assigns to each expected utility value a real number. By contrast, all that is required under the proposed model is to pick out, given the stakes involved in the option, the set of probability measures that corresponds to the confidence level that the agent deems appropriate for those stakes.

Similar points hold for the variational preferences model of Maccheroni et al. (2006) as well as the confidence preferences model of Chateauneuf and Faro (2009). Like the second-order probability models mentioned above, these models require some cardinal structure on the space of probability measures, whereas only an ordinal structure is needed here. Hence, for application in decision analysis, the model proposed here promises to be more tractable. Moreover, these models fail to separate beliefs from tastes for ambiguity, whereas the model proposed here does.

None of the models discussed above explicitly propose a dependence between the probability judgements used in a decision and the stakes involved in the choice; the role of various probability measures is always determined in some more complicated way. It is thus not surprising that the current model is not a special case of the aforementioned ones, as can be seen from the axiomatisation and the discussion in Section 1.2. The same holds for the models proposed in the literature on robustness and in particular for the multiplier and constraint preferences models proposed by Hansen and Sargent (2001) (which can be thought of as special cases of the variational and maxmin models discussed above). Gajdos et al. (2008) propose an extension of the Gilboa and Schmeidler (1989) model that allows different sets of probability measures to be used in the evaluation of acts, but these sets are derived from an 'objective' set of probability measures (modelling information) which is given exogeneously, rather than being generated endogeneously from the stakes involved in the act, as is the case under the current proposal.

2 Representation

2.1 Preliminaries

We use the standard Anscombe-and-Aumann framework, as adapted by Fishburn (1970). Let S be a non-empty finite set of states, with Σ the algebra of all subsets of S, which are called *events*. $\Delta(\Sigma)$ is the set of probability measures on (S, Σ). Where necessary, we use the Euclidean topology on $\Delta(\Sigma)$. X is a nonempty set of outcomes; a *consequence* is a probability measure on X with finite support. $\Delta(X)$ is the set of consequences. Acts are functions from states to consequences; $\mathcal{A}$ is the set of acts. So, for an act f, and a state s, $f(s)$ is a lottery over X with finite support; for a utility function u over X, we will denote the expected utility of this lottery by $u(f(s)) = \sum_{x \in supp(f(s))} f(s)(x)u(x)$. $\mathcal{A}$ is a mixture set with the mixture relation defined pointwise: for f, h in $\mathcal{A}$ and $\alpha \in \Re$, $0 \leq \alpha \leq 1$, the mixture $\alpha f + (1-\alpha)h$ is defined by $(\alpha f + (1-\alpha)h)(s, x) = \alpha f(s, x) + (1-\alpha)h(s, x)$ (Fishburn, 1970, Ch 13). We will often write $f_\alpha h$ as short for $\alpha f + (1-\alpha)h$. With slight abuse of notation, a constant act taking consequence c for every state will be denoted c and the set of constant acts will be denoted $\Delta(X)$.

We assume a preference relation $\preceq$ on $\mathcal{A}$; $\sim$ and $\prec$ are defined to be the symmetric and asymmetric components of $\preceq$. Null events and null states are defined in the usual way.[9] We use the following notation.

Definition 1. For $f \in \mathcal{A}$, $\hat{f} \in \Delta(X)$ is any $f(s)$ that is $\preceq$-minimal over the non-null $s \in S$.

Intuitively, $\hat{f}$ is a constant act that is $\preceq$-equivalent to the worst consequence that f takes on non-null states.[10] Recall that this corresponds to one way of fleshing out the notion of the stakes involved in the choice of an act f: the worse the worst possible consequence of the act, the more the decision maker has to lose in choosing it and the higher the stakes involved in the choice. The particular decision rule considered below will take the stakes involved in the choice of an act to be the lowest utility that could result from the act after resolution of the uncertainty about the state of the world

[9] An event A is null if, for any acts f and f' differing only on A, $f \sim f'$. A state s is null if the singleton event $\{s\}$ is null.

[10] Although $\hat{f}$ is not uniquely defined, it is defined up to $\sim$, which is all that is required for the axioms to have unambiguous meaning. This allows us to speak of $\hat{f}$ in the singular, which is useful for expositional purposes.

(that is, the utility of $\hat{f}$).[11] As noted in the Introduction, we consider this notion of stakes solely for the purposes of concreteness, and with no intention to suggest that it be considered the best or only reasonable notion of stakes; see Section 4 for discussion of other possible notions. In the exposition below, we shall say that f has higher, lower or the same stakes as g if $\hat{f} \preceq \hat{g}$, $\hat{f} \succeq \hat{g}$ or $\hat{f} \sim \hat{g}$ respectively.

The following notion is central.

Definition 2. A *confidence ranking* Ξ is a nested family of closed, convex subsets of $\Delta(\Sigma)$. A confidence ranking Ξ is *continuous* if, for every $C \in \Xi$, $C = \overline{\bigcup_{C' \subsetneq C} C'} = \bigcap_{C' \supsetneq C} C'$.[12] It is *centered* if it contains a singleton set.

As mentioned in the Introduction, confidence rankings represent decision makers' beliefs, and in particular their confidence in probability judgements. The sets in the confidence ranking can be thought of as corresponding to levels of confidence. The higher the level of confidence in question, the larger the set: this translates the fact that one is confident of fewer probability judgements to that level of confidence. Confidence rankings are ordinal structures; indeed, Proposition 2 in Appendix A shows that, to any confidence ranking there corresponds a weak order on the space of probability measures with appropriate properties and to any such order there corresponds a confidence ranking.

The convexity and closedness of the sets of probability measures in the confidence ranking are standard assumptions in decision rules involving sets of probabilities. The continuity property guarantees a continuity in one's confidence in probability judgements: it ensures, for example, that one cannot be confident at a certain level that probability of an event A is in $[0.3, 0.7]$ and then only confident that the probability is in $[0.1, 0.9]$ at the 'next' confidence level.[13]

The centredness property of the confidence ranking implies that, if the decision maker were forced to give his best estimate for the probability of any event, he could come up with a single value (and these values satisfy the laws of probability), although he may not be very confident in this value. In some situations this assumption may seem reasonable, whereas in others it may be less acceptable; in the representation

[11] The non-nullness clause in Definition 1 ensures that acts that differ only on null states are considered as having the same stakes, as seems reasonable.

[12] For a set X, $\overline{X}$ is the closure of X. Note that the union of a nested family of convex sets is convex.

[13] The continuity property used in the definition has been recognised as the appropriate notion of continuity for nested families of closed sets in Kuratowski (1938, pp 19-20). We thank Massimo Marinacci for this reference.

result (Theorem 1), there is an axiom that is necessary and sufficient for this property, so the property receives a behavioural characterisation.

Finally, we define a *cautiousness coefficient* for a confidence ranking Ξ to be a surjective function $D : \Re \to \Xi$ that preserves order, that is, such that for all $r, s \in \Re$, if $r \leq s$, then $D(r) \subseteq D(s)$. D assigns to a given level of stakes involved in the choice of an act a level of confidence that is required in probability judgements in order to use them in the evaluation of the act. This level of confidence corresponds to the appropriate set of probability measures in the confidence ranking. The fact that D is order-preserving corresponds to the intuition that the higher the stakes, the higher the confidence level required of probability judgements for them to play a role in the choice (and so the larger the relevant set of probability measures). Surjectivity of D basically attests to the behavioural nature of the confidence ranking: it implies that for each set of probability measures in the ranking, there will be a level of stakes, and hence an act, for which it is the relevant set for evaluating that act.

We consider a decision rule according to which the decision maker ranks acts f by the following criterion:

$$\min_{p \in D(\sigma(f))} \sum_{s \in S} u(f(s)).p(s)$$

where $\sigma(f) = -u(\hat{f})$ is the stakes involved in the choice of f. As noted in the Introduction, this translates the idea that for each level of stakes ($\sigma(f)$), the decision maker selects those probability judgements which he believes with confidence corresponding to those stakes (the judgements which hold for all probability measures in $D(\sigma(f))$) and only uses those in evaluating the act (he takes the maxmin over $D(\sigma(f))$).

2.2 *The Representation*

Consider the following axioms.

Axiom A1 (Weak order)**.** For all $f, g, h \in \mathcal{A}$: if $f \preceq g$ and $g \preceq h$, then $f \preceq h$; and $f \preceq g$ or $g \preceq f$.

Axiom A2 (S-Independence)**.** For all $f \in \mathcal{A}$, $c, d \in \Delta(X)$ and $\alpha \in (0, 1)$,

(i) if $d \succeq \hat{f}$, then $f \succeq c$ implies $f_\alpha d \succeq c_\alpha d$

(ii) if $d \preceq \hat{f}$, then $f \preceq c$ implies $f_\alpha d \preceq c_\alpha d$

Axiom A3 (Continuity)**.** For all $f, g, h \in \mathcal{A}$, the sets $\{\alpha \in [0,1] | \; f_\alpha h \preceq g\}$ and $\{\alpha \in [0,1] | \; f_\alpha h \succeq g\}$ are closed in $[0,1]$.

Axiom A4 (S-Monotonicity)**.** For all $f, g \in \mathcal{A}$, $c, d \in \Delta(X)$ and $\alpha \in (0,1]$ with $\hat{f} \sim \widehat{g_\alpha d}$ and $g_\alpha d \sim c_\alpha d$, if $f(s) \preceq g(s)$ for all $s \in S$, then $f \preceq c$, and if $f(s) \succeq g(s)$ for all $s \in S$, then $f \succeq c$.

Axiom A5 (S-Uncertainty Aversion)**.** For all $f, g \in \mathcal{A}$, $c, d \in \Delta(X)$ and $\alpha, \beta \in (0,1)$ with $\hat{f} \sim \hat{g} \sim \widehat{(f_\alpha g)_\beta d}$, if $f \sim g \sim c$ then $(f_\alpha g)_\beta d \succeq c_\beta d$.

Axiom A6 (Non-degeneracy)**.** There exist $f, g \in \mathcal{A}$ such that $f \succ g$.

Axiom A7 (Centering)**.** There exists $c, d \in \Delta(X)$ with $d \succ c$ such that, for all $f, g, h \in \mathcal{A}$ with $\hat{f}, \hat{g}, \hat{h} \succeq c$ and all $\alpha \in (0,1)$, $f \preceq g$ iff $f_\alpha h \preceq g_\alpha h$.

The axioms A1, A3 and A6 are standard. To motivate and explain the axioms A2, A4 and A5, it is helpful to compare them to the corresponding axioms for the maxmin expected utility model (Gilboa and Schmeidler, 1989), which are as follows:

Axiom A2' (C-Independence)**.** For all $f, g \in \mathcal{A}$, $d \in \Delta(X)$ and $\alpha \in (0,1)$, $f \preceq g$ iff $f_\alpha d \preceq g_\alpha d$.

Axiom A4' (Monotonicity)**.** For all $f, g \in \mathcal{A}$, if $f(s) \preceq g(s)$ for all $s \in S$, then $f \preceq g$.

Axiom A5' (Uncertainty Aversion)**.** For all $f, g \in \mathcal{A}$, $\alpha \in (0,1)$, if $f \sim g$ then $f_\alpha g \succeq f$.

Gilboa and Schmeidler (1989) established that A1, A3, A6, A2', A4' and A5' characterise maxmin expected utility preferences (which differ from representation (1) only in that the set of priors over which the decision maker is minimising is fixed rather than variable according to the stakes). As suggested in Section 1.2, the central axiomatic difference between the standard maxmin model and the model considered here lies in the independence axiom.[14] Note firstly that mixing an act f with a constant act d is a way of changing the stakes involved: in particular, if d is below the worst consequence of f ($\hat{f}$), then mixing with d raises the stakes (the worst consequence of $f_\alpha d$ is worse than the worst consequence of f); conversely, if d is preferred to $\hat{f}$, then mixing with d lowers the stakes. In the light of this, C-independence (A2') basically

[14]Since C-independence is weaker than many of the independence axioms in the literature on decision under uncertainty, such as the standard independence axiom (Anscombe and Aumann, 1963) and comonotonic independence (Schmeidler, 1989), the discussion here also applies to the comparison with those models.

states that the evaluation of an act is independent of the stakes involved. In particular, it implies that if $f \sim c$ for some constant act c, then $f_\alpha d \sim c_\alpha d$ for all α and d, which is tantamount to saying that if f is considered as indifferent to c, then they will be considered as indifferent even if evaluated after shifting the stakes up or down by any amount (by taking appropriate mixtures). However, the maxim which guides the proposal made here consists precisely in the idea that the evaluation of acts, and thus preferences between acts, *may change* upon changes in the stakes involved. From the point of view of this maxim, the C-independence axiom establishes an unwarranted independence of preferences from stakes, and as such must be weakened.

The S-independence (for Stakes-corrected independence) axiom A2 captures exactly the sort of weakening one would expect. The first clause basically says that if one *lowers* the stakes, by mixing an act with a constant act that is preferred to its worst consequence, then the act *cannot* be evaluated any *worse* than it was before (if it is preferred to a constant act, its mixture is preferred to the appropriate mixture of the constant act). The second clause deals with the other case: it says that if one *raises* the stakes, by mixing an act with a constant act that is less preferred than its worst consequence, then the act *cannot* be evaluated any *better* than it was before (if it is less preferred than a constant act, its mixture is less preferred than the appropriate mixture of the constant act). If one endorses the maxim proposed above and the caution inherent in the maxmin expected utility rule, this axiom is eminently reasonable: as stakes increase, acts may be evaluated as worse because one may no longer rely on probability judgements in which one does not have the required level of confidence, and likewise as stakes fall, acts may be considered as better, because one allows oneself to invoke probability judgements in which one did not previously have sufficient confidence.

To illustrate the difference between S-independence (A2) and C-independence (A2'), note that both imply that, for a risk-neutral decision maker,[15] if he weakly prefers a bet yielding a \$1000 loss if event A occurs and nothing if not to a sure \$250 loss (acts f and c in Figure 3), then he would weakly prefer the bet yielding a \$50 loss if event A occurs and nothing if not to a sure \$12.5 loss (because the latter two acts are obtained from the former two by taking a 0.05-mix with the constant act yielding \$0). This is equivalent to saying that if he strictly prefers the sure loss to the bet where he could lose \$50, then he strictly prefers the sure loss when the amounts at stake are twenty

[15] Alternatively, the example can be reformulated in terms of utiles instead of dollars, or using an urn whose composition is determined by whether the event occurs or not, so that it holds irrespective of risk attitudes.

Figure 3: S-independence

	A	A^c
f	-1000	0
c	-250	-250
g	-50	0
d	-12.50	-12.50

C-independence: $d \succ g \Leftrightarrow c \succ f$.

S-independence: $d \succ g \Rightarrow c \succ f$.

times larger. However, C-independence also imposes the converse relationship: if the decision maker strictly prefers the sure loss when he could lose $1000 by betting, then he strictly prefers the sure loss if he could only lose $50 by betting. This is not necessarily reasonable, for the fact that he avoids the bet when the stakes are high (when he could lose $1000) does not imply that he will not be willing to bet when the stakes are lower (when he has less to lose). S-independence only implies a relationship between the preferences in one direction, but not in this other, more controversial, direction.

As anticipated above, this axiom is the main difference between the proposed decision rule and many of the standard rules in the literature on decision under uncertainty. In particular, if S-independence is replaced by C-independence, the representation (1) reduces to the standard maxmin expected utility representation.

The S-monotonicity (for Stakes-corrected monotonicity) axiom A4 can be understood as a distilled version of the standard monotonicity axiom A4', extracting the aspects that are independent of stakes. The standard axiom establishes an implication from a (state-wise) dominance relation between a pair of acts to the appropriate preference over them; however, in the light of the intuitions mooted above, there are two considerations supporting it. One involves a possible difference in stakes: since g dominates f, the stakes involved in g can only be lower than those involved in f, and this might contribute to it being preferred. The other is independent of this difference: even if g had the same stakes as f, it would be preferred. S-monotonicity (A4) separates out the latter reason – which corresponds more closely to the standard intuition behind the monotonicity axiom – from the former. Indeed, it states, in essence, that if g dominates f then g is preferred to f when evaluated as if it had the same stakes as f, and likewise for the case where g is dominated by f. To see this, note that, since mixing acts with constant acts may shift the stakes, one can use such mixtures to glean information about preferences at different levels of stakes. For example, for an act g and a constant act c, the preferences over the mixtures $g_\alpha d$ and $c_\alpha d$ indicate the preferences between g and c at the stakes level given by the worst consequence of $g_\alpha d$. We can thus speak of preferences over g and c at a given stakes level, understanding this as shorthand for speaking of preferences over appropriate mixtures of g and c. Read in this way, the assumption in the S-monotonicity axiom states that g is indifferent to c at the stakes level corresponding to f; in other words, c can be thought of as the 'certainty equivalent' of g when evaluated at the stakes level corresponding to f. And so the first clause of the axiom states that if g dominates f, then it is preferred to f when it is evaluated at the stakes level of f, whereas the second clause states that if g is dominated by f, then f is preferred to it when it is evaluated at the stakes level of f. Hence S-monotonicity can indeed be thought of as encapsulating the intuition behind the standard monotonicity axiom, whilst refining it to account for the issue of stakes. In the presence of the other axioms (notably A2), it is easily seen to be a strengthening of the standard axiom.

Similar remarks apply to the S-uncertainty aversion (for Stakes-corrected uncertainty aversion) axiom A5: it can be read as stating essentially the same thing as the standard uncertainty aversion axiom A5', but for a fixed stakes level. In this case, the restriction on the stakes level enters twice. On the one hand, the standard axiom applies to acts that are indifferent, whereas the acts may have different stakes (and would not necessarily be indifferent if evaluated as if they had the same stakes). S-uncertainty aversion only applies to acts f and g that have the same stakes. On the other hand, the standard axiom demands that the mixture $f_\alpha g$ be preferred to f; but, as in the case of the monotonicity axiom, such preferences could be defended by appeal to the stakes involved in the acts (the stakes involved in the mixture can only be lower than those involved in the initial acts), or by appeal to the preferences after having been 'corrected for' stakes. S-uncertainty aversion isolates the second aspect, only demanding that $f_\alpha g$ be preferred to f at the stakes level corresponding to f (where the notion of preference at a given stakes level is cashed out in terms of mixtures as explained above). Summing up these two points, the axiom basically states that whenever two acts having the same stakes are indifferent, then any mixture is weakly preferred to the two acts when evaluated as if it had their stakes level; as claimed, this is essentially the content of the standard uncertainty aversion axiom, restricted to a single stakes level. As such, A5 is faithful to the principal intuition behind the standard axiom, which concerns preference for hedging. Indeed, the main cases where it differs from the standard axiom involve

Figure 4: Uncertainty aversion

	A	A^c
f	400	-100
g	50	-25
h	312.50	-81.25

Standard uncertainty aversion: $f \sim g \Rightarrow h \succeq f$.

no hedging. An example is given in Figure 4: whilst the standard axiom implies that if a risk-neutral decision maker[16] is indifferent between acts f and g in Figure 4, then he prefers h (which is a $\frac{3}{4}$-$\frac{1}{4}$ mixture of f and g) to both, S-uncertainty aversion is more liberal – it allows this pattern of preferences, without requiring it. As such, the axiom proposed here seems more reasonable: whilst the preference pattern seems perfectly *permissible*, it is hard to see why it should be an *obligatory* constraint on preferences, as the standard axiom demands. Certainly, given the absence of hedging considerations in this example, the normal intuition behind uncertainty aversion cannot explain why a decision maker who shows indifference between the first pair of acts *must* prefer the mixture in this case.

Like their counterparts elsewhere in the literature on decision under uncertainty, A2, A4 and A5 are fully behavioural, and in principle testable.

The final axiom, Centering (A7), states that the standard independence axiom is satisfied on a subset of acts where one cannot do worse that a certain consequence. (The condition of existence of a $d \succ c$ guarantees that this subset is non-trivial.) That is, it states that, for low enough stakes, the independence axiom is satisfied. As will be clear from the theorem, the axiom is not required for the representation, but if added guarantees that the confidence ranking representing the preferences is centered.

These axioms are necessary and sufficient for a representation of preferences by a decision rule featuring confidence, as represented by an confidence ranking.

Theorem 1. *The following are equivalent:*

(i) $\preceq$ *satisfies A1–A6,*

(ii) there exists a nonconstant utility function $u : X \to \Re$*, a continuous confidence*

[16] As above, the example could be reformulated in terms of utiles instead of dollars to avoid assumptions about risk attitudes.

ranking Ξ and a cautiousness coefficient $D : \Re \to \Xi$ such that, for all $f, g \in \mathcal{A}$, $f \preceq g$ iff

$$\text{(1)} \qquad \min_{p \in D(\sigma(f))} \sum_{s \in S} u(f(s)).p(s) \leq \min_{p \in D(\sigma(g))} \sum_{s \in S} u(g(s)).p(s)$$

where $\sigma(f) = -u(\hat{f})$ for all $f \in \mathcal{A}$.

Furthermore, for any other nonconstant utility function u', continuous confidence ranking Ξ' and cautiousness coefficient D' representing $\preceq$ as in (ii), there exists a positive real number a and a real number b such that $u' = au + b$, $\Xi' = \Xi$ and $D'(x) = D(\frac{x}{a} - \frac{b}{a})$ for all $x \in \Re$.

Finally, under the conditions above, $\preceq$ satisfies A7 iff Ξ is centred.

Proofs of all results are in Appendix A.

3 Confidence, cautiousness and ambiguity

The proposed model involves a strict separation of beliefs and tastes. Of the three elements in the model – the utility function, the confidence ranking and the cautiousness coefficient – the first, as is standard, captures the decision maker's tastes over outcomes. As discussed above, the confidence ranking represents not only the probability judgements endorsed by the decision maker (his beliefs) but also the confidence with which he endorses them (his confidence in his beliefs). The third element, the cautiousness coefficient, captures the decision maker's attitude to choosing in the absence of confidence. The aim of this section is to demonstrate that these interpretations are indeed valid.

First, let us consider whether the confidence ranking has separate behavioural meaning from the cautiousness coefficient, by asking when two preference relations satisfying the axioms in Theorem 1 are represented by the same utility functions and confidence rankings (though potentially different cautiousness coefficients). The following definition and proposition provide an answer.

Definition 3. Let $\preceq^1$ and $\preceq^2$ be preference relations satisfying the axioms A1–A6. $\preceq^1$ and $\preceq^2$ are *confidence equivalent* if (i) for all $c, d \in \Delta(X)$, $c \preceq^1 d$ iff $c \preceq^2 d$ and (ii) for each $d \in \Delta(X)$, there exists $d' \in \Delta(X)$ and $\alpha \in (0, 1)$ such that, for all $f \in \mathcal{A}$ and $c \in \Delta(X)$ with $\hat{f} \sim d$, $f \succeq^1 c$ iff $f_\alpha d' \succeq^2 c_\alpha d'$.

Proposition 1. *Let $\preceq^1$ and $\preceq^2$ be preference relations satisfying the axioms A1–A6 and represented by utility functions, confidence rankings and cautiousness coefficients (u_1, Ξ_1, D_1) and (u_2, Ξ_2, D_2) respectively. $\preceq^1$ and $\preceq^2$ are confidence equivalent if and only if u_1 is a positive affine transformation of u_2 and $\Xi_1 = \Xi_2$.*

Confidence equivalence can be understood as the behavioural translation of the fact for two decision makers to have the same confidence *in preferences*. If f is preferred to c at a given stakes level but g is not preferred to d at that level, whilst it is preferred to d at a lower stakes level, then this can be interpreted as indicating that the decision maker is more confident in his preference for f over c than in his preference for g over d. In light of the interpretation of preferences at a given stakes level in terms of preferences between mixtures introduced in Section 2.2, confidence equivalence thus states that the two decision makers share the same relative judgements of confidence in their preferences: one is more confident in the preference for f over c than in the preference for g over d if and only if the other one is. To the extent that confidence in preferences is determined by confidence in beliefs and confidence in utilities (which is trivial in this model, due to the use of a single utility function), one would expect that if two decision makers have the same confidence in preferences, then they have the same confidence in beliefs and utilities. Proposition 1 is the formal confirmation of this intuition.[17]

Confidence equivalence guarantees that the decision makers' confidence rankings are the same, but it does not guarantee that they are used in the same way. In particular, it says nothing about which sets in the confidence ranking are used to evaluate acts with given stakes. This is exactly where their attitudes to choosing in the absence of confidence comes in. To introduce this notion, consider the standard approach to risk attitudes.

In one standard treatment of risk attitude (Pratt, 1964), risk attitude is explored

[17] The notion of confidence equivalence is sufficient for the purposes of this section, where the focus is on separating attitudes towards choosing in the absence of confidence from confidence itself. Nevertheless, it is possible to behaviourally separate utility from confidence in beliefs in much the same way as, for example, Savage (1954) does. Under his theory, decision makers 1 and 2 have the same beliefs if, for any consequences c, c', d, d' such that $c \prec_1 d$ and $c' \prec_2 d'$, 1's preferences over acts whose outcomes are in $\{c, d\}$ are the same as 2's preferences over 'corresponding' acts whose outcomes are in $\{c', d'\}$. (More precisely, if f, g take values in $\{c, d\}$ and f', g' are defined by: for all $s \in S$, $f'(s) = c'$ if $f(s) = c$ and $f'(s) = d'$ if $f(s) = d$, and likewise for g', then $f' \preceq_2 g'$ iff $f \preceq_1 g$.) A similar condition, extended to consequences in $\{\alpha c + (1-\alpha)d | \alpha \in [0, 1]\}$ and $\{\alpha c' + (1-\alpha)d' | \alpha \in [0, 1]\}$ respectively, and relativised to stakes as in Definition 3, ensures that decision makers have the same confidence rankings, although potentially different utilities.

by considering the decision maker's attitudes to the adding of small risks to a sure amount. This sort of approach is not usually followed in the case of decision under uncertainty because of the difficulty of getting a clean notion of what it would be to add a particular 'uncertainty', especially given the dependence of the perceived uncertainty on the decision maker's beliefs. However, in the context of the current theory, where confidence enters into play via the stakes involved in the act, there is a simple analogue to adding risks: increasing the stakes.

Consider two decision makers, 1 and 2, with the same utility function and confidence ranking, and consider an act f and a constant act c. Since 1 and 2 have the same confidence rankings, at sufficiently small stakes they will both agree on the preferences between f and c; suppose that f is preferred to c at small stakes. As the stakes increase, the decision makers may switch and cease to prefer f to c; the difference between them lies in the stakes at which they cease to exhibit this preference. For example, if there is a stakes level where 1 no longer prefers f to c but 2 still does, this means that at this stakes level, 1 does not have enough confidence in his probability judgements to sustain the preference, whereas 2 does. But since they have the same confidence ranking over probability judgements, this must be because 1 is demanding more confidence in probability judgements for them to play a role at these stakes than 2. In other words, he is less comfortable than 2 in choosing with limited confidence at this stakes level; he is exhibiting more aversion to choosing in the absence of confidence than 2. If this is the case for all acts f and constant acts c, we say that 1 is more averse to choosing in the absence of confidence than 2. This leads to the following comparative notion of aversion to choosing in the absence of confidence.

Definition 4. Suppose that $\preceq^1$ and $\preceq^2$ satisfy axioms A1–A6 and that they are confidence equivalent. Then $\preceq^1$ is *more averse to choosing in the absence of confidence* than $\preceq^2$ if, for all $f \in \mathcal{A}$, $c, d, e \in \Delta(X)$ and $\alpha \in (0, 1]$, if $f_\alpha d \succeq^1 c_\alpha d$ whenever $\widehat{f_\alpha d} \succeq^1 e$, then $f_\alpha d \succeq^2 c_\alpha d$ whenever $\widehat{f_\alpha d} \succeq^2 e$.

This definition translates formally the motivation above: it states that 1 is more averse to the absence of confidence if whenever he prefers f to c for a given level of stakes and lower, then so does 2.

As noted in Section 1.4, the model proposed here can be thought of as a contribution to the literature on ambiguity. In this literature, discussions of comparative ambiguity aversion are generally motivated by considerations of the sets of acts that are preferred to particular constant acts (as in Yaari (1969)). This often leads to notions of comparative ambiguity aversion where, under assumptions guaranteeing that the agents have the same beliefs, a decision maker is more ambiguity averse than another if, every time the former prefers an act to a constant act, then the latter does too. The following is a definition of comparative ambiguity aversion which is true to the spirit of those notions.[18]

Definition 5. Let $\preceq^1$ and $\preceq^2$ be preference relations satisfying the axioms A1–A6 and suppose that they are confidence equivalent. Then $\preceq^1$ is *more ambiguity averse* than $\preceq^2$ if, for any $f \in \mathcal{A}$ and $c \in \Delta(X)$, if $f \succeq^1 c$ then $f \succeq^2 c$.

The two comparative notions of aversion – to choosing in the absence of confidence and to ambiguity – are equivalent. Moreover, they are characterised entirely in terms of the cautiousness coefficient.

Theorem 2. *Suppose that $\preceq^1$ and $\preceq^2$ satisfy axioms A1–A6 and are confidence equivalent. Suppose that they represented by (u, Ξ, D_1) and (u, Ξ, D_2) respectively. The following are equivalent:*

(i) $\preceq^1$ *is more averse to choosing in the absence of confidence than* $\preceq^2$

(ii) $\preceq^1$ *is more ambiguity averse than* $\preceq^2$

(iii) $D_1(r) \supseteq D_2(r)$ *for all* $r \in \Re$.

We conclude that the cautiousness coefficient captures the attitude to confidence and ambiguity, and that analysis of such attitudes can be carried out in the proposed model.

4 Variants and related models

Reduced to the essentials, the current proposal consists of two features: firstly, a representation of confidence in beliefs by an ordinal structure on the space of probability measures, and secondly, the idea that in the evaluation of options the decision maker uses the set of probability measures in this structure which corresponds to the stakes involved. As noted in the Introduction, this delineates not a single decision model but

[18] This literature is too rich to be discussed in detail here. Schmeidler (1989) and Epstein (1999) are early treatments of the question; Ghirardato and Marinacci (2002) is closest to that used here. The assumption about beliefs is used by those whose models admit separation of beliefs and tastes (for example, Klibanoff et al. (2005)), but not by those whose models do not (for example, Maccheroni et al. (2006) and Chateauneuf and Faro (2009)).

a family of models, which differ according to the decision rule and the notion of stakes used. The principal aim of this paper is to promote this family, rather than any particular member. In the previous sections, we have considered one member with an eye to illustrating the intuitiveness of models in the family, and the interesting and attractive properties that they may have. Let us now briefly consider some other members of the family.

On the one hand, there are decision models that use a rule other than the maxmin expected utility rule employed in representation (1). As is well-known, this rule yields an ambiguity-averse model of decision, and the representation (1) is ambiguity averse for the same reason. One can obtain an ambiguity-seeking model of the sort proposed here by replacing the maxmin expected utility rule with the maxmax expected utility rule (where the minimum in representation (1) is replaced by a maximum). An analysis similar to that carried out in Section 3 would apply to such a model, yielding a relation of 'more absence-of-confidence seeking'. Other decision models are obtained by replacing the maxmin expected utility rule with the α-maxmin or Hurwicz rule, which evaluates acts by a mix of maximum and minimum, or by a generalisation of that rule with non-constant α, such as that studied by Ghirardato et al. (2004).

Such modifications in the decision rule change relatively little as regards the main points made in the previous sections. Conceptually, the same representation of confidence in beliefs (confidence ranking) is used, and the set of probabilities is determined on the basis of the stakes via the cautiousness coefficient. Accordingly, there continues to be the same strict separation of beliefs (represented by the confidence ranking) and tastes (represented, at least, by the utility function and the cautiousness coefficient[19]). Practically, in the application of such rules to real-life decisions, one still needs to determine a set of probabilities on the basis of the stakes involved in the act; the difference is in how this set is used.

It is also worth noting other potentially interesting members of the family that can be seen as special cases of the model considered in the previous sections. For example, adding the appropriate analogue of co-monotonic independence would force the sets of probabilities in the confidence ranking to correspond to convex capacities (Schmeidler, 1986). In this case, the representation (1) would be equivalent to a stakes-sensitive Choquet expected utility representation, where the decision maker chooses according to the Choquet expected utility rule, but with a capacity that varies with the stakes. This

[19]For certain decision rules, there may be additional factors on the taste side; this is the case for example for the a function in the rule proposed by Ghirardato et al. (2004).

model is particularly interesting from an operational point of view, since standard techniques that have been developed for eliciting capacities could perhaps be applied here to elicit the confidence ranking. Another potentially interesting specification involves confidence rankings where the sets are balls of different sizes around a fixed probability measure (ie. the sets are of the form $\{q \in \Delta(\Sigma) | d(p, q) \leq \epsilon\}$ for different ϵ, where d is the metric on the space of probability measures). These easily-parametrised models can be thought of as confidence versions of ϵ-contamination models studied elsewhere in the literature (Gajdos et al., 2008; Kopylov, 2009).

On the other hand, other members of the family of decision models may differ in the notion of stakes used. The motivation for the notion of stakes employed here is obvious, if not beyond controversy, and there are a myriad of other notions of stakes which could be considered: for example, the best consequence of the act, the difference between the best and worst consequences of the act, the minimum (or maximum) probability that the act takes values below a certain threshold (calculated with the respect to smallest or alternatively the largest set of probability measures in the confidence ranking) or the minimum expected utility of values below the threshold (calculated with the respect to smallest or the largest set of probability measures in the confidence ranking). This is not to mention notions of stakes that are not functions of individual acts but of the menu of available acts; for example, one could take as stakes the worst consequence out of all the acts on the menu. A full discussion of all the options is beyond the scope of this paper. However, many of the points made above continue to hold for other notions of stakes. Changing the notion of stakes used makes little difference conceptually: confidence in beliefs, represented by confidence rankings, continues to play a role in choice via the cautiousness coefficient, with only the allocation of stakes to acts (the function σ in (1)) changing. Accordingly, the strict separation of beliefs and tastes is retained. Finally, the application of the model in practice proceeds as described above, the only difference being the properties of acts that are relevant in determining the set of probabilities used.

Let us close with two remarks. First of all, notice that, in principle, the decision rule on sets of probabilities and the notion of stakes can be varied independently. For example, any combination of maxmin expected utility and maxmax expected utility rule with stakes as worst consequence and stakes as best consequence is possible. Whilst this yields a very large family of decision models, only some will be of interest. For example, one might wonder whether there is any interest for someone who takes as stakes the best consequence to be so pessimistic as to use the maxmin expected utility rule. Secondly, notice that both of these issues are behaviourally meaningful. Both the specific decision rule used and the particular notion of stakes employed will have consequences for properties of preferences and so may, in principle, be gleaned from choices. It remains a largely open question how to axiomatise different combinations of decision rule and notion of stakes or whether there is a general axiomatic framework in which all models in the family can be treated.

5 Conclusion

This paper proposes a model of a decision maker's confidence in his probability judgements in terms of a confidence ranking – a nested family of sets of probability measures. This model can be used to formulate a family of decision rules which allow the probability judgements that play a role in a decision to depend on what is at stake. Under each of the decision models in this family, the decision maker evaluates acts using a set of probability measures that is determined by the stakes involved and his cautiousness coefficient. Models in this family differ according, firstly, to the way the set of probability measures is used in the evaluation and, secondly, to the notion of stakes employed. However, all share the advantages of the general framework: structural simplicity of the representation of confidence, relative tractability in application to aiding decision making, and separation of beliefs and tastes. A particular model in the family is studied in detail and found to have a reasonable axiomatisation. Finally, a notion of comparative aversion to choosing in the absence of confidence is defined, characterised in terms of the cautiousness coefficient and related to standard notions of ambiguity aversion.

A Proofs of results

Throughout the Appendix, B will denote the space of all real-valued functions on S, and $ba(S)$ will denote the set of real-valued set functions on S, both under the Euclidean topology. Recall that, under this topology, $ba(S)$ is locally convex (Aliprantis and Border, 2007, §5.12). B is equipped with the standard order: $a \leq b$ iff $a(s) \leq b(s)$ for all $s \in S$. For $a \in B$, we define $\sigma(a) = -\min_{non-nulls \in S} a(s)$, and for $x \in \Re$, we define x^* to be the constant function taking value x. Finally, for any $a \in B$ and $r \in \Re$, let $e_a^r \in B$ be such that $e_a^r(s) = a(s) + \sigma(a) - r$ for all $s \in S$ and let $x_a^r \in \Re$ be $r - \sigma(a)$. By definition, $\sigma(e_a^r) = r$ and $a = e_a^r + x_a^{r*}$.

Proof of Theorem 1

The main part of the result is the sufficiency of the axioms for the representation (the direction (i) to (ii)), the proof of which proceeds as follows. As is standard, the axioms imply the existence of a utility function u on $\Delta(X)$, with image K, and a functional I on $B(K)$ such that, for any $f, g \in \mathcal{A}$, $f \preceq g$ iff $I(u \circ f) \leq I(u \circ g)$. The main lemma, Lemma 3, establishes that for any stakes level there exists a closed and convex set of probability measures such that each act with these stakes is evaluated by the maxmin expected utility over this set. By Lemma 4, these sets form a nested family, with larger stakes yielding larger sets, and hence is a confidence ranking. By Lemma 5, this confidence ranking is continuous.

We now proceed with the proof of the Theorem. We assume first (i); we will show (ii). Consider firstly the following lemmas, all of which are under the assumption of axioms A1–A6.

Lemma 1. *There exists a non-constant utility function u representing the restriction of $\preceq$ to the constant acts. Moreover, for $K = u(\Delta(X))$ and $B(K)$ the set of functions in B taking values in K, there exists $I : B(K) \to \Re$ such that, for any $f, g \in \mathcal{A}$, $f \preceq g$ iff $I(u \circ f) \leq I(u \circ g)$.*

Proof. First note that, for all $c, d \in \Delta(X)$, if $c \sim d$ then $\frac{1}{2}c+\frac{1}{2}e \sim \frac{1}{2}d+\frac{1}{2}e$ for all $e \in \Delta(X)$. If $e \succeq c \sim d$, then, by the first clause of A2, it follows that $\frac{1}{2}c+\frac{1}{2}e \succeq \frac{1}{2}d+\frac{1}{2}e$; moreover, by the symmetric argument, $\frac{1}{2}d+\frac{1}{2}e \succeq \frac{1}{2}c+\frac{1}{2}e$, whence $\frac{1}{2}c+\frac{1}{2}e \sim \frac{1}{2}d+\frac{1}{2}e$, as required. If $e \preceq c \sim d$, then the second clause of A2 can be applied in a similar way to obtain $\frac{1}{2}c + \frac{1}{2}e \sim \frac{1}{2}d + \frac{1}{2}e$, as required. Given A1, A3 and A6, it follows that the axioms of the Herstein and Milnor (1953) mixture theorem are satisfied, hence the existence of the required utility function.

Moreover, by A1, A3 and A4, for any act f there exist constant acts c, c' such that with $c \preceq f \preceq c'$. (If $f(s) \sim \hat{f}$ for all non-null $s \in S$, then take $c = c' = \hat{f}$. Now suppose that this is not the case. Take any $c, c' \in \Delta(X)$ such that $c \preceq f(s) \preceq c'$ for all $s \in S$. By A3, there exists $\alpha \in (0, 1]$ such that $\hat{f} \sim c_\alpha c'$; A4 implies that $f \succeq c$. Moreover, if $c \prec \hat{f}$, then $\alpha \in (0, 1)$, so A4 implies that $f \preceq c'$. Finally, if $c \sim \hat{f}$, then applying the previous argument on $f_\beta c'$ yields that $f_\beta c' \preceq c'$ for all $\beta \in [0, 1)$; by A3, $f \preceq c'$.) By A3 and A1, using a standard argument, u can be extended to a function $J : \mathcal{A} \to \Re$ such that, for any $f, g \in \mathcal{A}$, $f \preceq g$ iff $J(f) \leq J(g)$. It also follows from A4 that $J(f) = J(f')$ whenever $f(s) \sim f(s')$ for all $s \in S$, so there exists a function $I : B(K) \to \Re$ such that $I(u \circ f) = J(f)$ for all $f \in \mathcal{A}$, as required. $\square$

Without loss of generality, it will be assumed that $0 \in K$ but not on its boundary. So K is an interval on the real line containing 0 and such that 0 is not on its boundary. Let $-K = \{r \in \Re | -r \in K\}^o$.[20] Note finally that, by construction, $I(x^*) = x$.

The following lemma will prove useful.

Lemma 2. *For all $a, a', b \in B$ and $\alpha > 0$, such that $a' - \sigma(b)^*, a - \sigma(b)^* \in B(K)$, $\sigma(a) = \sigma(a') = 0$ and $a = \alpha a'$, $I(a - \sigma(b)^*) + \sigma(b) = \alpha(I(a' - \sigma(b)^*) + \sigma(b))$.*

Proof. Consider the case of $\alpha \in (0, 1]$ (the case of $\alpha > 1$ follows as a consequence). Take $g \in \mathcal{A}$ such that $a' - \sigma(b)^* = u \circ g$, $x, y \in \Delta(X)$ such that $u(x) = -\sigma(b)$ and $g \sim y$, and $f \in \mathcal{A}$ such that $f = \alpha g + (1 - \alpha)x$. So $u \circ f = \alpha u \circ g + (1 - \alpha)u \circ x = \alpha(a' - \sigma(b)^*) - (1 - \alpha)\sigma(b)^* = a - \sigma(b)^*$. Then A2 applies, yielding that $f = \alpha g + (1 - \alpha)x \sim \alpha y + (1 - \alpha)x$, whence $I(a - \sigma(b)^*) = I(u \circ f) = \alpha I(u \circ y) + (1 - \alpha)I(u \circ x) = \alpha I(a' - \sigma(b)^*) + (1 - \alpha)I(-\sigma(b)^*)$, and so $I(a - \sigma(b)^*) + \sigma(b) = \alpha(I(a' - \sigma(b)^*) + \sigma(b))$ as required. □

The following is the central lemma of the proof.

Lemma 3. *For every $r \in -K$, there exists a closed convex set $\mathcal{C}_r \subseteq \Delta(\Sigma)$ such that, for every $a \in B(K)$ such that $\sigma(a) = r$, $I(a) = \min_{p \in \mathcal{C}_r} \sum_{s \in S} a(s)p(s)$.*

Proof. For each $b \in B(K)$ with $\sigma(b) \in -K$, we construct a probability measure p_b such that $I(b) = \sum_{s \in S} b(s)p_b(s)$ and $I(a) \geq \sum_{s \in S} a(s)p_b(s)$ for all a with $\sigma(a) = \sigma(b)$. To this end, take any $b \in B(K)$ with $\sigma(b) \in -K$ and define $I_b : B \to \Re$ as follows:

(i) $I_b(x^*) = x$ for all $x \in \Re$;

(ii) $I_b(a) = I(a - \sigma(b)^*) + \sigma(b)$ for all $a \in B$ such that $a - \sigma(b)^* \in B(K)$ and $\sigma(a) = 0$;

(iii) $I_b(a) = \alpha I_b(a')$ for all $a \in B$ such that $\sigma(a) = 0$ and $a = \alpha a'$ with $a' - \sigma(b)^* \in B(K)$.

(iv) $I_b(a) = I_b(e^0_a) + I_b(x^{0*}_a)$ for all $a \in B$ such that $\sigma(a) \neq 0$.

Lemma 2 guarantees that I_b is well-defined (and in particular, that there is no contradiction between the clauses (ii) and (iii)), and that it is homogeneous of degree one

[20]For a set X, X^o is its interior.

on $\{a \in B|\ \sigma(a) = 0\}$. Moreover, it is superadditive: that is, for all $a, \bar{a} \in B$, $I_b(\frac{1}{2}a + \frac{1}{2}\bar{a}) \geq \frac{1}{2}I_b(a) + \frac{1}{2}I_b(\bar{a})$. This can be seen as follows.

Consider firstly the case where $\sigma(a) = \sigma(\bar{a}) = 0$. If $I_b(\bar{a}) = 0$, then the desired inequality holds by A2, A4 and the definition of I_b, and similarly for $I_b(a) = 0$; so we may henceforth suppose that $I_b(a), I_b(\bar{a}) > 0$. Without loss of generality, it can assumed that $a - \sigma(b)^*, \bar{a} - \sigma(b)^*, \sigma(\beta a + (1-\beta)\bar{a})^* - \sigma(b)^* \in B(K)$; if not, multiply the non-$\sigma(b)$ terms below through by a small positive factor and use the fact that I_b is homogenous of degree one on $\{a \in B|\ \sigma(a) = 0\}$. If $I_b(a) = I_b(\bar{a})$, then, for any $\beta \in (0,1)$, $I_b(\beta a + (1-\beta)\bar{a}) = 2I_b(\frac{1}{2}(\beta a + (1-\beta)\bar{a}) + \frac{1}{2}(\sigma(\beta a + (1-\beta)\bar{a}))^*) + I_b((-\sigma(\beta a + (1-\beta)\bar{a}))^*) = 2I\left(\frac{1}{2}(\beta(a - \sigma(b)^*) + (1-\beta)(\bar{a} - \sigma(b)^*)) + \frac{1}{2}(\sigma(\beta a + (1-\beta)\bar{a})^* - \sigma(b)^*)\right) + \sigma(b) - \sigma(\beta a + (1-\beta)\bar{a}) \geq 2\left(\frac{1}{2}I(a - \sigma(b)^*) + \frac{1}{2}I(\sigma(\beta a + (1-\beta)\bar{a})^* - \sigma(b)^*)\right) + \sigma(b) - \sigma(\beta a + (1-\beta)\bar{a}) = I_b(a)$, where the first, second and last equalities follow from the definition of I_b and its homogeneity of degree one on $\{a \in B|\ \sigma(a) = 0\}$, and the middle inequality results from an application of A5 (with acts corresponding to $a - \sigma(b)^*$, $\bar{a} - \sigma(b)^*$, $\sigma(\beta a + (1-\beta)\bar{a})^* - \sigma(b)^*$, and $I(a - \sigma(b)^*)^*$ in the place of f, g, d and c respectively). If $I_b(a) > I_b(\bar{a})$, let $\alpha = \frac{I_b(\bar{a})}{I_b(a)}$, so that $\sigma(\alpha a) = \sigma(\bar{a}) = 0$ and $I_b(\alpha a) = I_b(\bar{a})$. It follows, by the reasoning for the case where $I_b(a) = I_b(\bar{a})$, that $I_b(\beta \cdot \alpha a + (1-\beta)\bar{a}) \geq I_b(\alpha a) = \beta I_b(\alpha a) + (1-\beta)I_b(\bar{a}) = \beta \cdot \alpha I_b(a) + (1-\beta)I_b(\bar{a})$ for every $\beta \in (0,1)$. Substituting $\frac{1}{1+\alpha}$ for β and multiplying both sides by $\frac{1+\alpha}{2\alpha}$ yields the desired inequality. The case where $I_b(a) < I_b(\bar{a})$ is treated similarly.

Consider finally any $a, \bar{a} \in B$. Note that, by the definition of σ, $\sigma(\frac{1}{2}a + \frac{1}{2}\bar{a}) = \frac{1}{2}\sigma(a) + \frac{1}{2}\sigma(\bar{a}) + \sigma\left(\frac{1}{2}(a + \sigma(a)) + \frac{1}{2}(\bar{a} + \sigma(\bar{a}))\right)$. So $e^0_{\frac{1}{2}a+\frac{1}{2}\bar{a}} = \frac{1}{2}a + \frac{1}{2}\bar{a} + \sigma(\frac{1}{2}a + \frac{1}{2}\bar{a}) = \frac{1}{2}a + \frac{1}{2}\bar{a} + \frac{1}{2}\sigma(a) + \frac{1}{2}\sigma(\bar{a}) + \sigma\left(\frac{1}{2}(a + \sigma(a)) + \frac{1}{2}(\bar{a} + \sigma(\bar{a}))\right) = \frac{1}{2}e^0_a + \frac{1}{2}e^0_{\bar{a}} + \sigma(\frac{1}{2}e^0_a + \frac{1}{2}e^0_{\bar{a}}) = e^0_{\frac{1}{2}e^0_a+\frac{1}{2}e^0_{\bar{a}}}$. By the definition of I_b and x^0_a, we thus have that $I_b(\frac{1}{2}a + \frac{1}{2}\bar{a}) = I_b(e^0_{\frac{1}{2}a+\frac{1}{2}\bar{a}}) - \sigma(\frac{1}{2}a + \frac{1}{2}\bar{a}) = I_b(e^0_{\frac{1}{2}e^0_a+\frac{1}{2}e^0_{\bar{a}}}) - \sigma(\frac{1}{2}e^0_a + \frac{1}{2}e^0_{\bar{a}}) - \frac{1}{2}\sigma(a) - \frac{1}{2}\sigma(\bar{a}) = I_b(\frac{1}{2}e^0_a + \frac{1}{2}e^0_{\bar{a}}) - \frac{1}{2}\sigma(a) - \frac{1}{2}\sigma(\bar{a})$. By the result established above, $I_b(\frac{1}{2}e^0_a + \frac{1}{2}e^0_{\bar{a}}) - \frac{1}{2}\sigma(a) - \frac{1}{2}\sigma(\bar{a}) \geq \frac{1}{2}I_b(e^0_a) + \frac{1}{2}I_b(e^0_{\bar{a}}) - \frac{1}{2}\sigma(a) - \frac{1}{2}\sigma(\bar{a})$. So $I_b(\frac{1}{2}a + \frac{1}{2}\bar{a}) \geq \frac{1}{2}I_b(a) + \frac{1}{2}I_b(\bar{a})$, as required.

To construct the required probability measure, there are two cases to consider.

Case i: $I(b) > I((-\sigma(b))^*)$. For $\bar{b} = b + \sigma(b)^*$, let $D^b_1 = \{a \in B \mid I_b(a) > I_b(\bar{b})\}$ and $D^b_2 = co(\{a \in B \mid a \leq I_b(\bar{b})^*\} \cup \{\bar{b}\})$.[21] Evidently, both sets are convex and nonempty (in the case of D^b_1, convexity is ensured by the superadditivity of I_b).

[21] For a set X, $co(X)$ is the convex hull of X.

We now show that D_1^b and D_2^b are disjoint; that is, for any $a \in B$ and $\beta \in [0,1]$, if $a \leq I_b(\bar{b})^*$, then $\beta a + (1-\beta)\bar{b} \notin D_1^b$, ie. $I_b(\beta a + (1-\beta)\bar{b}) \leq I_b(\bar{b})$. The result is immediate for $\beta = 0, 1$, so consider $\beta \in (0,1)$. Let $a' \in B$, $x \in \Re$ be such that $\beta a + (1-\beta)\bar{b} = a' + x^*$, with $\sigma(a') = 0$. Without loss of generality, it can assumed that $a' - \sigma(b)^*, (1-\beta)(\bar{b} - \sigma(b)^*) + \beta(I_b(\bar{b}) - \sigma(b) - \frac{x}{\beta})^*, (-\beta I_b(\bar{b}) - \sigma(b) + x)^* \in B(K)$; if not, multiply the non-$\sigma(b)$ terms below through by a small positive factor and use the fact that I_b is homogenous of degree one on $\{a \in B | \, \sigma(a) = 0\}$. Since $a \leq I_b(\bar{b})^*$, $a' + x^* \leq \beta I_b(\bar{b})^* + (1-\beta)\bar{b}$. Hence $a' - \sigma(b)^* \leq (1-\beta)(\bar{b} - \sigma(b)^*) + \beta(I_b(\bar{b}) - \sigma(b) - \frac{x}{\beta})^*$. Moreover, $\frac{1}{2}\left((1-\beta)(\bar{b} - \sigma(b)^*) + \beta(I_b(\bar{b}) - \sigma(b) - \frac{x}{\beta})^*\right) + \frac{1}{2}(-\beta I_b(\bar{b}) - \sigma(b) + x)^*) = \frac{1-\beta}{2}\bar{b} - \sigma(b)^*$ and $\sigma(\frac{1-\beta}{2}\bar{b} - \sigma(b)^*) = \sigma(a' - \sigma(b)^*)$. By Lemma 2, $I(\frac{1-\beta}{2}\bar{b} - \sigma(b)^*) = \frac{1-\beta}{2}I(\bar{b} - \sigma(b)^*) - \frac{1+\beta}{2}\sigma(b)$. Given the preceding facts, and the fact that $\frac{1-\beta}{2}I(\bar{b} - \sigma(b)^*) - \frac{1+\beta}{2}\sigma(b) = \frac{1}{2}\left((1-\beta)I(\bar{b} - \sigma(b)^*) + \beta(I_b(\bar{b}) - \sigma(b) - \frac{x}{\beta})^*\right) + \frac{1}{2}(-\beta I_b(\bar{b}) - \sigma(b) + x)^*)$, A4 can be applied (with the acts corresponding to $a' - \sigma(b)^*$, $(1-\beta)(\bar{b} - \sigma(b)^*) + \beta(I_b(\bar{b}) - \sigma(b) - \frac{x}{\beta})^*$, $(1-\beta)I(\bar{b} - \sigma(b)^*) + \beta(I_b(\bar{b}) - \sigma(b) - \frac{x}{\beta})^*$ and $(-\beta I_b(\bar{b}) - \sigma(b) + x)^*$ in the place of f, g, c and d respectively) to yield the conclusion that $I(a' - \sigma(b)^*) \leq (1-\beta)I(\bar{b} - \sigma(b)^*) + \beta(I_b(\bar{b}) - \sigma(b) - \frac{x}{\beta})$. Since $(1-\beta)I(\bar{b} - \sigma(b)^*) + \beta(I_b(\bar{b}) - \sigma(b) - \frac{x}{\beta}) = I_b(\bar{b}) - \sigma(b) - x$, it follows that $I_b(a') + x \leq I_b(\bar{b})$; so $I_b(a' + x^*) \leq I_b(\bar{b})$, as required.

Given the aforementioned properties, a separation theorem (Aliprantis and Border, 2007, Theorem 7.30) implies that there is a nonzero linear functional on B, P_b, and a real number α such that $P_b(c) \leq \alpha \leq P_b(a)$ for all $a \in D_1^b$ and $c \in D_2^b$.

$I_b(\bar{b})$ is strictly positive, by the choice of b and the definition of I_b; it follows by the definition of D_1^b and D_2^b that α is strictly positive. Without loss of generality, it can be assumed that $\alpha = I_b(\bar{b})$. Since $I_b(\bar{b})^* \in D_2^b$, $P_b((I_b(\bar{b}))^*) \leq I_b(\bar{b})$; but since $(I_b(\bar{b}))^*$ is a limit point of D_1^b, $P_b((I_b(\bar{b}))^*) \geq I_b(\bar{b})$ and so $P_b((I_b(\bar{b}))^*) = I_b(\bar{b})$. Hence, by linearity of P_b, $P_b(1^*) = 1$. Moreover, for all $E \in \Sigma$, $(I_b(\bar{b}))_{E^c}0 \leq (I_b(\bar{b}))^*$ (where $x_E y$ is the function taking value x on E and y elsewhere) since $I_b(\bar{b})$ is positive; so $(I_b(\bar{b}))_{E^c}0 \in D_2^b$ and hence $P_b((I_b(\bar{b}))_{E^c}0) \leq I_b(b)$. Since $P_b((I_b(\bar{b}))_E 0) + P_b((I_b(\bar{b}))_{E^c}0) = P_b((I_b(\bar{b}))^*) = I_b(\bar{b})$, it follows that $P_b((I_b(b))_E 0) \geq 0$. Hence, by the linearity of P_b, that $P_b(1_E 0) \geq 0$. By a classic duality result, there is a probability measure p_b such that $P_b(a) = \sum_{s \in S} a(s)p_b(s)$ for all $a \in B$.

We now show that, for all $a \in B(K)$ such that $\sigma(a) = \sigma(b)$, $P_b(a) \geq I(a)$, with equality for $a = b$. For any such a, and for all $x > I_b(\bar{b}) - I_b(a + \sigma(b)^*)$, $I_b(a + \sigma(b)^*) + I_b(x^*) > I_b(\bar{b})$, and so $I_b((a + \sigma(b)^*) + x^*) > I_b(\bar{b})$ by the definition of I_b; hence $(a + \sigma(b)^*) + x^* \in D_1^b$. By construction of P_b, we thus have $P_b((a +$

$\sigma(b)^*) + x^*) \geq I_b(\bar{b})$; by linearity of P_b, it follows that $P_b(a + \sigma(b)^*) + P_b(x^*) \geq I_b(\bar{b})$. By the continuity of P_b (letting $x \to I_b(\bar{b}) - I_b(a + \sigma(b)^*)$), it follows that $P_b(a + \sigma(b)^*) \geq I_b(a + \sigma(b)^*)$. By the linearity of P_b and definition of I_b, this implies that $P_b(a) \geq I(a)$. Since $\bar{b} \in D_2^b$, $P_b(\bar{b}) \leq I_b(\bar{b})$, so $P_b(\bar{b}) = I_b(\bar{b})$ and hence $P_b(b) = I(b)$ as required.

Case ii: $I(b) \leq I((-\sigma(b))^*)$. Let $\bar{b} = b + \sigma(b)$, and take $D_1^b = \{a \in B \mid I_b(a) > 0\}$ and $D_2^b = \{\beta\bar{b} | \, \beta \in [0,1]\}$. Evidently, both sets are convex and nonempty (in the case of D_1^b, convexity is ensured by the superadditivity of I_b). Note furthermore that, by A4, $I(b) = I((-\sigma(b))^*)$, so $I_b(\bar{b}) = 0$; moreover, it follows from Lemma 2 that $I_b(c) = 0$ for all $c \in D_2^b$, so D_1^b and D_2^b are disjoint.

By a separation theorem (Aliprantis and Border, 2007, Theorem 7.30), there is a nonzero linear functional on B, P_b and a real number α such that $P_b(\bar{b}) \leq \alpha \leq P_b(a)$ for all $a \in \overline{D_1^b}$. Since $0^* \in D_2^b$ and it is a limit point of D_1^b, α must be 0.

For all $E \in \Sigma$, $0^* \leq 1_E 0$ (where $x_E y$ is the function taking value x on E and y elsewhere), so, by A4 and A2, $0 \leq I_b(1_E 0)$, hence $P_b(1_E 0) \geq 0$. By a classic duality result, there is a probability measure p_b such that $P_b(a) = \sum_{s\in S} a(s)p_b(s)$ for all $a \in B$.

Using a similar argument to that used in case (i), it can be shown that, for all $a \in B(K)$ such that $\sigma(a) = \sigma(b)$, $P_b(a) \geq I(a)$, with equality for $a = b$.

For every $r \in -K$, define $\mathcal{C}_r$ to be the closure of the convex hull of $\{p_b \mid \sigma(b) = r\}$. By construction, $I(a) = \min_{p\in\mathcal{C}_r} \sum_{s\in S} a(s)p(s)$ for any $a \in B(K)$ such that $\sigma(a) = r$, as required.

□

The following two lemmas guarantee that the sets of probability measures mentioned in Lemma 3 have the properties needed to constitute a continuous confidence ranking.

Lemma 4. *For all $r, s \in -K$, $\mathcal{C}_r \subseteq \mathcal{C}_s$ if $r \leq s$.*

Proof. For any $a \in B(K)$ and for any $r, s \in -K$ such that $r \leq s$, there exists $x \in K$ and $\alpha \in (0,1]$ such that $\sigma(\alpha(e_a^s) + (1-\alpha)x^*) = r$. Since $x \geq -s$, A2 implies that, for any $y \in K$, if $I(e_a^s) \geq y$, then $I(\alpha(e_a^s) + (1-\alpha)x^*) \geq \alpha y + (1-\alpha)x$. By the linearity of the maxmin functional, it follows that $\min_{p\in\mathcal{C}_r} \sum_{s\in S} a(s)p(s) = \frac{1}{\alpha} \min_{p\in\mathcal{C}_r} \sum_{s\in S} \left(\alpha(e_a^s) + (1-\alpha)x^*\right)(s)p(s) + (x_a^s - \frac{1-\alpha}{\alpha}x) \geq \frac{1}{\alpha}\left(\alpha \min_{p\in\mathcal{C}_s} \sum_{s\in S}(e_a^s)(s)p(s) + (1-\alpha)x\right) + (x_a^s - \frac{1-\alpha}{\alpha}x) =$

$\min_{p\in\mathcal{C}_s}\sum_{s\in S}a(s)p(s)$. Similarly, by considering a multiple of the reflexion of a in 0^* ($a'\in B(K)$ such that, for some α, $a(s)=-\alpha a'(s)$ for every $s\in S$), we have that $\max_{p\in\mathcal{C}_r}\sum_{s\in S}a(s)p(s)\leq\max_{p\in\mathcal{C}_s}\sum_{s\in S}a(s)p(s)$. So $[\min_{p\in\mathcal{C}_r}\sum_{s\in S}a(s)p(s),\max_{p\in\mathcal{C}_r}\sum_{s\in S}a(s)p(s)]\subseteq[\min_{p\in\mathcal{C}_s}\sum_{s\in S}a(s)p(s),\max_{p\in\mathcal{C}_s}\sum_{s\in S}a(s)p(s)]$ for all $a\in B(K)$. It follows, by Ghirardato et al. (2004, Proposition A.1), that $\mathcal{C}_r\subseteq\mathcal{C}_s$, as required. □

Lemma 5. *For any $r\in -K$, $\mathcal{C}_r=\bigcap_{r'>r}\mathcal{C}_{r'}=\overline{\bigcup_{r'<r}\mathcal{C}_{r'}}$.*

Proof. We show that $\mathcal{C}_r=\bigcap_{r'>r}\mathcal{C}_{r'}$; the proof that $\mathcal{C}_r=\overline{\bigcup_{r'<r}\mathcal{C}_{r'}}$ is similar. By Lemma 4, $\mathcal{C}_r\subseteq\mathcal{C}_{r'}$ for all $r'>r$. Suppose, for reductio, that $\mathcal{C}_r\subsetneq\bigcap_{r'>r}\mathcal{C}_{r'}$ for some $r\in -K$, so that there exists a point (probability measure) $p\notin\mathcal{C}_r$, but $p\in\bigcap_{r'>r}\mathcal{C}_{r'}$. By a separation theorem (Aliprantis and Border, 2007, 5.80), there is a nonzero (continuous) linear functional ϕ on $ba(S)$ and $\alpha\in\Re$ such that $\phi(p)\leq\alpha<\phi(q)$ for all $q\in\mathcal{C}_r$. Since S is finite (so B is finite-dimensional), B is reflexive, and, by the standard isomorphism between $ba(S)$ and B^*, it follows that $ba(S)^*$ is isometrically isomorphic to B (Dunford and Schwartz, 1958, IV.3); hence there is a real-valued function $a\in B$ such that $\phi(q)=\sum_{s\in S}a(s)q(s)$ for any $q\in ba(S)$. Without loss of generality ϕ, a and α can be chosen so that $\alpha\in K$, $a\in B(K)$ and $\sigma(a)=r$. Taking $f\in\mathcal{A}$ such that $u\circ f=a$ and $c\in\Delta(X)$ such that $u(c)=\alpha$, we have, by construction, that $c\prec f$ but for any $d\in\Delta(X)$ with $d\prec c$ and $\sigma(d)>r$ and any $\alpha\in(0,1)$, $f_\alpha d\preceq c_\alpha d\prec c$, contradicting A3.

□

Conclusion of the proof of Theorem 1. To conclude the direction (i) to (ii), note that, by Lemma 1, there exists a utility function u representing preferences over constant acts, and a functional I representing $\preceq$ as specified. Consider the case where K is bounded above and below; the other cases are treated similarly. Lemma 3 guarantees that there exists sets $\mathcal{C}_r$ for all $r\in -K$, such that $I(a)=\min_{p\in\mathcal{C}_r}\sum_{s\in S}a(s)p(s)$ for all $a\in B(K)$ such that $\sigma(a)=r$. Define $\Xi=\{\mathcal{C}_r|\ r\in -K\}\cup\{\bigcap_{r\in -K}\mathcal{C}_r\}\cup\{\overline{\bigcup_{r\in -K}\mathcal{C}_r}\}$. Since $\mathcal{C}_r$ is closed and convex for all $r\in -K$, and since, by Lemma 4, the sets $\mathcal{C}_r$ are nested, Ξ is a confidence ranking. By Lemma 5, Ξ is continuous. D is defined as follows: for $r\in -K$, $D(r)=\mathcal{C}_r$, for $r<s$ for all $s\in -K$, $D(r)=\bigcap_{s\in -K}\mathcal{C}_s$ and for $r>s$ for all $s\in -K$, $D(r)=\overline{\bigcup_{s\in -K}\mathcal{C}_s}$. Order preservation and surjectivity of D are immediate from the definition and Lemma 4.

Finally, for any $c\in\Delta(X)$ having the property described in the centering axiom A7, the restriction of $\preceq$ to $\{f\in\mathcal{A}\mid\hat{f}\succeq c\}$ is non-trivial and satisfies independence. By standard results, in the presence of independence, this implies that $\mathcal{C}_{-u(c)}$ is a singleton. So A7 implies that Ξ contains a singleton, as required.

The direction from (ii) to (i) is generally straightforward. The only interesting cases are S-monotonicity (A4) and continuity (A3). As regards S-monotonicity (A4), let $f, g \in \mathcal{A}$, $c, d \in \Delta(X)$ and $\alpha \in (0,1]$ satisfy the conditions of the axiom. Since $\hat{f} \sim \widehat{g_\alpha d}$, $D(\sigma(f)) = D(\sigma(g_\alpha d))$; let us call this set $\mathcal{C}$. Since $g_\alpha d \sim c_\alpha d$, it follows from the representation (1) that $\min_{p\in\mathcal{C}} \sum_{s\in S} u(g_\alpha d(s))p(s) = u(c_\alpha d)$ and hence that $\min_{p\in\mathcal{C}} \sum_{s\in S} u(g(s))p(s) = u(c)$. If $f(s) \preceq g(s)$ for all $s \in S$, then $\min_{p\in\mathcal{C}} \sum_{s\in S} u(f(s))p(s) \leq \min_{p\in\mathcal{C}} \sum_{s\in S} u(g(s))p(s)$; so $\min_{p\in\mathcal{C}} \sum_{s\in S} u(f(s))p(s) \leq u(c)$ and $f \preceq c$ as required. A similar argument establishes the case where $f(s) \succeq g(s)$ for all $s \in S$. We now consider continuity (A3). Take any $f, g, h \in \mathcal{A}$ and consider the set $\{\alpha \in [0,1] |\ f_\alpha h \preceq g\}$; the other case is shown similarly. Suppose that α^* is a limit point of this set, and consider a sequence (α_i) of elements in the set with $\alpha_i \to \alpha^*$. If there is a subsequence of (α_i) tending to α^* such that $\widehat{f_{\alpha_{i_n}} h} \succeq \widehat{f_{\alpha^*} h}$ for all α_{i_n} in this subsequence, then the result follows from the fact that D is order-preserving and the continuity of the maxmin EU representation (with a fixed set of probability measures). Suppose now there is no such sequence. In this case, there exists $c \in \Delta(X)$, $c \prec \widehat{f_{\alpha^*} h}$; moreover, for any such c, there exists an interval I containing α^* with $\widehat{f_\beta h} \succeq c$ for any $\beta \in I$. Hence, for each such c, there exists a subsequence $(\alpha^c_{j_n})$ of (α_i), tending to α^*, with $\widehat{f_{\alpha^c_{j_n}} h} \succeq c$ for all $n \in \mathbb{N}$. By the fact that D is order-preserving and the continuity of the maxmin EU representation, it follows that $\min_{p\in D(-u(c))} \sum_{s\in S} u(f_{\alpha^*} h(s)).p(s) \leq \min_{p\in D(-u(\hat{g}))} \sum_{s\in S} u(g(s)).p(s)$. Since this holds for every $c \prec \widehat{f_{\alpha^*} h}$, and since, by the continuity of the confidence ranking and the surjectivity of D, $D(-u(\widehat{f_{\alpha^*} h})) = \bigcap_{c \prec \widehat{f_{\alpha^*} h}} D(-u(c))$, it follows that $f_{\alpha^*} h \preceq g$ as required.

Finally consider the uniqueness clause. Uniqueness of u follows from the Herstein-Milnor theorem. As regards uniqueness of Ξ, proceed by reductio; suppose that u, Ξ_1, D_1 and u, Ξ_2, D_2 both represent $\preceq$ according (1), with $\Xi_1 \neq \Xi_2$. By surjectivity of the cautiousness coefficient, for some r, $D_1(r) \neq D_2(r)$. Suppose, without loss of generality, that $p \in D_1(r) \setminus D_2(r)$. By a separation theorem (Aliprantis and Border, 2007, 5.80), there is a nonzero linear functional ϕ on $ba(S)$ and $\alpha \in \Re$ such that $\phi(p) \leq \alpha < \phi(q)$ for all $q \in D_2(r)$. As in the proof of Lemma 5, there is a real-valued function $a \in B$ such that $\phi(q) = \sum_{s\in S} a(s)q(s)$ for any $q \in ba(S)$. Without loss of generality ϕ, a and α can be chosen so that $\alpha \in K$, $a \in B(K)$ and $\sigma(a) = r$. Taking $f \in \mathcal{A}$ such that $u \circ f = a$ and $c \in \Delta(X)$ such that $u(c) = \alpha$, we have that $\min_{p\in D_1(\sigma(f))}\sum_{s\in S}u(f(s))p(s) \leq \min_{p\in D_1(\sigma(f))}\sum_{s\in S}u(c)p(s)$, whereas $\min_{p\in D_2(\sigma(f))}\sum_{s\in S}u(f(s))p(s) > \min_{p\in D_2(\sigma(f))}\sum_{s\in S}u(c)p(s)$, contradicting the assumption that both u, Ξ_1, D_1 and u, Ξ_2, D_2 represent $\preceq$. A similar argument establishes the uniqueness of D on $-K$; outside of $-K$, the value of D is determined (to be the smallest or largest set in Ξ), since D is order-preserving.

□

Proofs of other results

Proposition 2. *For Ξ a (continuous) confidence ranking, $\leq_\Xi = \{(p,q) \in \Delta(\Sigma)^2 |$ for all $C \in \Xi, q \in C$ implies $p \in C\}$ is a quasi-convex, lower semi-continuous (and locally non-satiated except at the extremes) weak order on $\Delta(\Sigma)$.*[22]

Likewise, for $\leq$ a quasi-convex, lower semi-continuous (and locally non-satiated except at the extremes) weak order on $\Delta(\Sigma)$, $\Xi_\leq = \{\{q|\ q \leq p\}|p \in \Delta(\Sigma)\}$ is a (continuous) confidence ranking.

Proof. Straightforward.

□

Proof of Proposition 1. The right to left direction is evident. For the left to right direction, the clause concerning utilities follows immediately from clause (i) of Definition 3. The clause concerning confidence rankings follows from clause (ii) of the definition, in the light of the construction and uniqueness result in the proof of Theorem 1. □

Proof of Theorem 2. Let the assumptions of the theorem be satisfied.

(i) implies (iii). Suppose that $D_1(r) \not\supseteq D_2(r)$ for some $r \in (-u(\Delta(X)))^o$, and that $p \in D_2(r)\setminus D_1(r)$. Since $D_1(r)$ is a closed convex set, by a separating hyperplane theorem (Aliprantis and Border, 2007, 5.80) and as in the proof of Lemma 5, there is an $f \in \mathcal{A}$ and a $c \in \Delta(X)$ such that $-u(\hat{f}) = r$, $\min_{p\in D_2(r)}\sum_{s\in S}u(f(s)).p(s) < u(c)$ and $\min_{p\in D_1(r)}\sum_{s\in S}u(f(s)).p(s) \geq u(c)$. So for all $f_\alpha d$ with $\widehat{f_\alpha d} \succeq \hat{f}$ $f_\alpha d \succeq^1 c_\alpha d$ whereas this is not the case for $\preceq^2$ (taking $\alpha = 1$), contradicting (i); hence (iii) is established for $r \in (-u(\Delta(X)))^o$. It follows from the order-preserving and surjectivity properties of D_1 and D_2 that (iii) holds on $r \notin (-u(\Delta(X)))^o$.

(ii) implies (i). Straightforward.

[22] An order $\leq$ is quasi-convex if the lower contour sets ($\{q \in \Delta(\Sigma)|\ q \leq p\}$) are convex for all $p \in \Delta(\Sigma)$; it is lower semi-continuous if the lower contour sets $\{q \in \Delta(\Sigma)|\ q \leq p\}$ are closed for all $p \in \Delta(\Sigma)$; and it is locally non-satiated except at the extremes if, for any $p \in \Delta(\Sigma)$ such that there exists $p' > p$ and for every neighbourhood around p, there exists q in the neighbourhood such that $q > p$.

(iii) implies (ii). Consider $f \in \mathcal{A}$, with $u(\hat{f}) = r$. Since $D_1(r) \supseteq D_2(r)$, $\min_{p\in D_1(r)} \sum_{s\in S} u(f(s)).p(s) \leq \min_{p\in D_2(r)} \sum_{s\in S} u(f(s)).p(s)$, from which it follows that if $f \succeq_1 c$, then $f \succeq_2 c$, as required.

□

B The example in Section 1.2 and the smooth ambiguity model

The smooth ambiguity model (Klibanoff et al., 2005) evaluates acts by the following functional:

$$V(f) = \int_{\Delta(\Sigma)} \phi\left(\int_S u(g(s))d\pi(s)\right) d\mu \tag{2}$$

where u is a von Neumann-Morgenstern utility function, μ is a countably additive probability measure on $\Delta(\Sigma)$, and ϕ is a continuous, strictly increasing real-valued function on the image of u, which is sometimes called the transformation function. As Klibanoff et al. (2005) show, under appropriate assumptions (and with an appropriate notion of ambiguity aversion), a smooth ambiguity decision maker is ambiguity averse if and only if his ϕ is concave. Note that, if ϕ is linear, then representation (2) reduces to standard expected utility, and hence satisfies the C-independence axiom.

To establish the claim in Section 1.2 that the preferences in Figure 1 are inconsistent with (2) if the decision maker is ambiguity averse, we use Proposition 3 below. This proposition states that, under certain conditions, an ambiguity averse smooth ambiguity decision maker's transformation function ϕ must be linear on a particular interval. It follows in particular that he satisfies C-independence on that interval.

Proposition 3. *Let $\preceq$ be represented according to (2) with a continuous utility function u, a continuous, strictly increasing and concave transformation function ϕ and countably additive probability measure μ. Suppose that, for some non-constant $f \in \mathcal{A}$ and $b, \underline{c}, \overline{c} \in \Delta(X)$, we have that $f \sim b$ iff $f_\alpha c \sim b_\alpha c$ for all $\alpha \in (0,1]$ and $\underline{c} \preceq c \preceq \overline{c}$. Let $\underline{x} = \inf\{x \mid \exists \Pi \subseteq \Delta(\Sigma)$ with $\mu(\Pi) > 0$ and such that $\forall \pi \in \Pi,\ \int_S u(f(s))d\pi(s) \leq x\}$ and $\overline{x} = \sup\{x \mid \exists \Pi \subseteq \Delta(\Sigma)$ with $\mu(\Pi) > 0$ and such that $\forall \pi \in \Pi,\ \int_S u(f(s))d\pi(s) \geq x\}$. Then ϕ is linear on $(\min\{\underline{x}, u(\underline{c})\}, \max\{\overline{x}, u(\overline{c})\})$.*

Proof. For any act g, define the real-valued measurable function on $\Delta(\Sigma)$, $\tilde{g}$, by $\tilde{g}(\pi) = \int_S u(g(s))d\pi(s)$ (this is close to the definition of second-order act given in

Klibanoff et al. (2005)); with slight abuse of notation, for a constant act b, we also use $\tilde{b}$ to denote $u(b(s))$. Noting that $\widetilde{\alpha f + (1-\alpha)c} = \alpha\tilde{f} + (1-\alpha)\tilde{c}$, it follows from representation (2) and the assumptions in the proposition that $\int_{\Delta(\Sigma)} \phi(\alpha\tilde{f}(\pi) + (1-\alpha)\tilde{c})d\mu(\pi) = \phi(\alpha\tilde{b}+(1-\alpha)\tilde{c})$ for all $c \in \Delta(X)$ with $\underline{c} \preceq c \preceq \overline{c}$ and all $\alpha \in (0,1]$. We first show that $\tilde{b} = \int_{\Delta(\Sigma)} \tilde{f}(\pi)d\mu$. As noted in Hardy et al. (1934, §3.18), the derivative of a continuous convex function exists except at most enumerably many points. There thus exists $\alpha \in (0,1]$ and $\underline{c} \preceq c \preceq \overline{c}$ such that $\underline{c} \preceq b_\alpha c \preceq \overline{c}$ and such that the derivative of ϕ exists at $\alpha\tilde{b} + (1-\alpha)\tilde{c}$. By the assumption in the proposition, for such α and c we have that

$$\int_{\Delta(\Sigma)} \phi(\beta(\alpha\tilde{f}(\pi) + (1-\alpha)\tilde{c}) + (1-\beta)(\alpha\tilde{b} + (1-\alpha)\tilde{c}))d\mu(\pi) = \phi(\alpha\tilde{b} + (1-\alpha)\tilde{c})$$

for every $\beta \in [0,1]$. It follows that

$$\int_{\Delta(\Sigma)} \frac{\phi\left(\beta(\alpha\tilde{f}(\pi) + (1-\alpha)\tilde{c}) + (1-\beta)(\alpha\tilde{b} + (1-\alpha)\tilde{c})\right) - \phi\left(\alpha\tilde{b} + (1-\alpha)\tilde{c}\right)}{\beta\alpha(\tilde{f}(\pi) - \tilde{b})}.(\tilde{f}(\pi) - \tilde{b})d\mu = 0$$

and so, taking the limit as $\beta \to 0$,

$$\int_{\Delta(\Sigma)} \frac{d\phi}{dx}(\alpha\tilde{b} + (1-\alpha)\tilde{c}).(\tilde{f}(\pi) - \tilde{b})d\mu = 0$$

Where the derivative exists in the light of the remarks above. Since ϕ is strictly increasing, it follows that $\tilde{b} = \int_{\Delta(\Sigma)} \tilde{f}(\pi)d\mu$, as required. Therefore $\gamma\tilde{b} + (1-\gamma)\tilde{d} = \int_{\Delta(\Sigma)} \gamma\tilde{f}(\pi) + (1-\gamma)\tilde{d}\, d\mu$ for all $\gamma \in [0,1]$ and all $d \in \Delta(X)$.

We thus have that $\int_{\Delta(\Sigma)} \phi(\tilde{f}(\pi))d\mu = \phi(\tilde{b}) = \phi\left(\int_{\Delta(\Sigma)} \alpha\tilde{f}(\pi)d\mu\right)$. Hence McShane (1937, Theorem 5) can be applied to $\tilde{f}$, the set of second-order acts (in the sense of Klibanoff et al. (2005)), the integral with respect to μ and the function ϕ to conclude that, for all $\pi \in \Sigma(\Delta)$ except at most those belonging to a set $\mathcal{S}$ such that $\mu(\mathcal{S}) = 0$, $\tilde{f}(\pi)$ belongs to an interval I on which ϕ is linear. It follows from the definition of $\underline{x}$ and $\overline{x}$ that ϕ is linear on $(\underline{x}, \overline{x})$. Moreover, by the same argument applied to $f_\alpha c$ and $b_\alpha c$ for $\alpha \in (0,1)$ and $\underline{c} \preceq c \preceq \overline{c}$, we have that ϕ is linear on $(\alpha\underline{x} + (1-\alpha)u(c), \alpha\overline{x} + (1-\alpha)u(c))$. Since these intervals overlap, for appropriate choices of α and c, we have that ϕ is linear on $(\min\{\underline{x}, u(\underline{c})\}, \max\{\overline{x}, u(\overline{c})\})$, as required.

□

We use this proposition to establish the claim in Section 1.2 as follows. (Recall that, for ease, we are assuming that u is linear.) As a point of notation, let s_1 be the state where the ball drawn is blue and s_2 the state where it is red (other notation is taken from Section 1.2; see in particular Figure 1). First of all, since $g \sim d$ for some constant act d yielding $\$d$ with $d > 0$,[23] it follows from representation (2) that μ must give non-zero weight to the set of probability measures π for which $\int_S u(g(s))d\pi(s) \geq u(d)$. Since $d \succ \$0$, there exists $\alpha > 0$ and η such that $(\alpha g + \eta)(s_2) \prec -\$10M$ and $\alpha d + \eta \succ \$15000$. (It suffices to take $\eta = 0$ and α sufficiently large.) It follows from the properties established above that $\mu(\{\pi \in \Delta(\Sigma) \mid \int_S u((\alpha g + \eta)(s))d\pi(s) > u(\$15000)\}) > 0$. We shall now apply Proposition 3 on $\alpha g + \eta$; note that the fact just established implies that $\overline{x} \geq u(\$15000)$ (where $\overline{x}$ is as defined in the statement of the proposition, for the act $\alpha g + \eta$). Since $\alpha g + \eta = (10^{-3} \times \alpha)f + \eta$, the preferences specified in Section 1.2 (see Figure 1), and notably the fact that $\alpha' f + \eta' \sim \alpha' p_f + \eta'$ whenever $(\alpha' f + \eta')(s_2) \prec -\$10M$, imply that $(\alpha g + \eta)_\beta c \sim (\alpha(-\$100) + \eta)_\beta c$ whenever c is a constant act with $c \prec -\$10M$. Hence the assumptions of Proposition 3 are satisfied for $\alpha g + \eta$ with any $\underline{c} \prec \overline{c} \prec -\$10M$. It follows from the Proposition that ϕ is linear on $[u(-\$1000M), u(\$10000)]$. So representation (2) reduces to expected utility – and in particular satisfies constant linearity – for acts taking values in that interval. The act $10^{-3} \times f + (-\$10M)$ takes values in the interval just described; moreover, since $(10^{-3} \times f + (-\$10M))(s_2) \prec -\$10M$, it follows from the preferences specified in Section 1.2 that $10^{-3} \times f + (-\$10M) \sim -\$10000100$. Since g also only takes values in the interval described, and $g = 10^{-3} \times f$, it follows from constant linearity of the representation in this interval that $g \sim -\$100$, contradicting $g \succ \$0$. So the preferences specified are indeed incompatible with representation (2), under the assumption that the decision maker is ambiguity averse, as claimed.

References

Aliprantis, C. D. and Border, K. C. (2007). *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer, Berlin, 3rd edition.

Anscombe, F. J. and Aumann, R. J. (1963). A Definition of Subjective Probability. *The Annals of Mathematical Statistics*, 34:199–205.

[23] Recall that we are using the same symbol for the constant act and the consequence it yields. Giving that utility is linear, we further abuse of notation we use d to also denote the monetary value yielded by the constant act d.

Chateauneuf, A. and Faro, J. H. (2009). Ambiguity through confidence functions. *J. Math. Econ.*, 45:535–558.

Dunford, N. and Schwartz, J. T. (1958). *Linear Operators. Part I: General Theory*. Interscience Publishers, New York.

Epstein, L. G. (1999). A Definition of Uncertainty Aversion. *Rev. Econ. Stud.*, 66(3):579–608.

Ergin, H. and Gul, F. (2009). A theory of subjective compound lotteries. *J. Econ. Theory*, 144(3):899–929.

Fishburn, P. C. (1970). *Utility Theory for Decision Making*. Wiley, New York.

Gajdos, T., Hayashi, T., Tallon, J.-M., and Vergnaud, J.-C. (2008). Attitude toward imprecise information. *J. Econ. Theory*, 140(1):27–65.

Ghirardato, P., Maccheroni, F., and Marinacci, M. (2004). Differentiating ambiguity and ambiguity attitude. *J. Econ. Theory*, 118(2):133–173.

Ghirardato, P. and Marinacci, M. (2002). Ambiguity Made Precise: A Comparative Foundation. *J. Econ. Theory*, 102(2):251–289.

Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *J. Math. Econ.*, 18(2):141–153.

Hansen, L. P. and Sargent, T. J. (2001). Robust Control and Model Uncertainty. *The American Economic Review*, 91(2):60–66.

Hardy, G., Littlewood, J. E., and Polya, G. (1934). *Inequalities*. Cambridge University Press, Cambridge.

Herstein, I. N. and Milnor, J. (1953). An Axiomatic Approach to Measurable Utility. *Econometrica*, 21(2):291–297.

Klibanoff, P., Marinacci, M., and Mukerji, S. (2005). A Smooth Model of Decision Making under Ambiguity. *Econometrica*, 73(6):1849–1892.

Kopylov, I. (2009). Choice deferral and ambiguity aversion. *Theoretical Economics*, 4(2):199–225.

Kuratowski, C. (1938). Sur les familles monotones d'ensembles fermés et leurs applications à la théorie des espaces connexes. *Fundamenta Mathematicae*, 30:17–33.

Maccheroni, F., Marinacci, M., and Rustichini, A. (2006). Ambiguity Aversion, Robustness, and the Variational Representation of Preferences. *Econometrica*, 74(6):1447–1498.

McShane, E. J. (1937). Jensen's Inequality. *Bulletion of the American Mathematical Society*, 43:521–527.

Nau, R. F. (2006). Uncertainty Aversion with Second-Order Utilities and Probabilities. *Management Science*, 52:136–145.

Pratt, J. W. (1964). Risk Aversion in the Small and in the Large. *Econometrica*, 32:122–136.

Savage, L. J. (1954). *The Foundations of Statistics*. Dover, New York.

Schmeidler, D. (1986). Integral Representation Without Additivity. *Proceedings of the American Mathematical Society*, 97(2):255–261.

Schmeidler, D. (1989). Subjective Probability and Expected Utility without Additivity. *Econometrica*, 57(3):571–587.

Seo, K. (2009). Ambiguity and Second-Order Belief. *Econometrica*, 77:1575–1605.

Yaari, M. E. (1969). Some remarks on measures of risk aversion and on their uses. *Journal of Economic Theory*, 1(3):315–329.