

Subjective Complexity Under Uncertainty

Quitze Valenzuela-Stookey *

September 20, 2020

Abstract

Complexity of the problem of choosing among uncertain acts is a salient feature of many of the environments in which departures from expected utility theory are observed. I propose and axiomatize a model of choice under uncertainty in which the size of the partition with respect to which an act is measurable arises endogenously as a measure of subjective complexity. I derive a representation of incomplete Simple Bounds preferences in which acts that are complex from the perspective of the decision maker are bracketed by simple acts to which they are related by statewise dominance. The key axioms are motivated by a model of learning from limited data. I then consider choice behavior characterized by a “cautious completion” of Simple Bounds preferences, and discuss the relationship between this model and models of ambiguity aversion. I develop general comparative statics results, and explore applications to portfolio choice, contracting, and insurance choice.

1 Introduction

Acts with uncertain outcomes are complicated, potentially infinite-dimensional, objects. As recognized by Von Neumann and Morgenstern (1944) and Aumann (1962), completeness of preferences over such acts is a lot to ask. In this paper I explore the connection between complexity, incomplete preferences, and choice behavior. I begin with assumptions regarding

*Economics Department, Northwestern University; email: quitze.vs@u.northwestern.edu. Many thanks to Eddie Dekel, Alessandro Pavan, and especially Marciano Siniscalchi for feedback on this project from its inception. I am also grateful to Edi Karni, Wojciech Olszewski, Francisco Poggi, Ludvig Sinander, Bruno Strulovici, Marie-Louise Vierø, participants at the Transatlantic Theory Workshop 2019, and seminar participants at Northwestern University.

which acts a decision maker finds it easy to compare, which characterizes a set of well-understood acts. From there I derive a representation of preferences in which incompleteness arises from subjective choice complexity.

First, I make assumptions on the ability of the decision maker to make comparisons between constant and non-constant acts (maps from the state space to a set of outcomes). I define well-understood acts to be those that have a certain equivalent, i.e. a constant act to which they are indifferent. I derive from these assumptions a characterization of the set of well-understood acts: there exists an integer N , determined endogenously by individual choice behavior, such that an act is well-understood if and only if it takes on no more than N distinct values. I then impose the standard EU axioms, with minor modifications, on the set of acts which possess certain equivalents, leading to the standard expected utility characterization of preferences over such acts. A final assumption implies that the value of any act can be bounded by that of “simple acts”, i.e. acts with certain equivalents, to which they are related by statewise dominance. Thus some complicated acts may also be compared. I refer to these as *Simple Bounds* preferences. Relative to standard EU preferences, Simple Bounds preferences have only one additional parameter, N , which is pinned-down by choice behavior.

The notion of complexity that arises endogenously from the assumptions on choice behavior, the number of distinct values in the range, can be interpreted as reflecting the difficulty people have in contemplating the relative likelihoods of multiple events. To formalize this intuition, I study a procedural model of decision making in which a frequentist decision maker uses a finite set of i.i.d. observations to learn about the unknown distribution of the state. I show that the axioms allowing for the characterization of complexity in terms of partition size are implied by natural assumptions on the procedural model of choice.

I consider two models of complete preferences derived from Simple Bounds. These models complement the analysis of incomplete preferences by exploring how the DM handles their lack of understanding. Preferences are *Cautious* if the DM evaluates an act f according to its simple lower bound, i.e. the best act that has a certain equivalent and is dominated statewise by f . Similarly preferences are *Reckless* if acts are evaluated according to their simple upper bound, defined analogously. I axiomatize these two preferences, relate them to the DM’s attitude towards ambiguity, and compare them to existing models of ambiguity aversion.

I provide some general comparative statics results, building on Tian (2015) and Tian (2016), which are useful in applications. Finally, I investigate three applications of the model. First, I study consumption savings problems and equilibrium asset prices. Cautious decision makers will save more, and allocate a greater portion of their savings to a safe rather than

a risky asset, compared to fully rational individuals. In equilibrium these biases lead to higher prices for safe relative to risky assets, as observed in the “equity premium puzzle” of Mehra and Prescott (1985). Second, I examine principal-agent contracting with a complexity constrained agent, identifying general features of optimal contracts. Finally, I study the choice of insurance plans. I show that the cautious model rationalizes many “behavioral” phenomena identified in the empirical literature, such as over-sensitivity to deductibles and coverage rates. It also helps explain the correlation, documented by Bhargava et al. (2017), between specific forms of dominated plan choice and both the degree of health risk and level of education. This application provides a good illustration of some of the general comparative statics results presented in Appendix C.

Section 2 introduces the setting. Section 3 presents the characterization of subjective complexity. Section 4 explores the procedural learning model. Section 5 presents the Simple Bounds characterization. Section 6 discusses complete preferences derived from Simple Bounds. Sections 7 through 9 present the applications. Related literature is discussed in Section 10. Omitted proofs, general comparative statics results, and further discussion of the representation theorems are organized in the appendices.

2 The Setting

The framework is that of Anscombe and Aumann (1963). The decision maker is characterized by a binary relation \succsim over acts, which I call the preference relation. The strict preference relation and indifference relation are defined as usual.¹ Further notation is as follows:

- Z : the set of *outcomes*.
- L : the set of $v\mathbb{N}$ - M *lotteries* (finite support distributions) over Z .
- Ω : the *state space*, endowed with an algebra Σ of *events*.
- F_c : the set of constant acts.
- F : the set of finite valued acts; Σ -measurable $f : \Omega \mapsto L$ such that $|f(\Omega)| < \infty$.

¹I take as primitive the reflexive and transitive relation, interpreted as weak preferences. An alternative approach, as in Galaabaatar and Karni (2013), would be to take as primitive a transitive and irreflexive strict partial order. I find my approach more convenient because I make use directly of the existence of certain equivalents and the transitivity of the weak preference relation.

The preference relation is assumed to satisfy the following basic conditions (see e.g. Gilboa et al. (2010)).

Basic Conditions

PREORDER: \succsim is reflexive and transitive.

MONOTONICITY: For every $f, g \in F$, $f(\omega) \succsim g(\omega) \forall \omega \in \Omega$ implies $f \succsim g$.

ARCHIMEDEAN CONTINUITY: For all $f, g, h \in F$, the sets $\{\lambda \in [0, 1] : \lambda f + (1 - \lambda)g \succsim h\}$, $\{\lambda \in [0, 1] : h \succsim \lambda f + (1 - \lambda)g\}$ are closed in $[0, 1]$.

NONTRIVIALITY: There exist $f, g \in F$ such that $f \succ g$.

C-COMPLETENESS: \succsim is complete on F_c

WEAK C-INDEPENDENCE: Let c_1, c_2 be constant acts. Then for any act f and $\alpha \in (0, 1)$, $f \succsim (\precsim)c_1$ if and only if $\alpha f + (1 - \alpha)c_2 \succsim (\precsim)\alpha c_1 + (1 - \alpha)c_2$.

The key difference between the Basic Conditions and the corresponding subset of the standard SEU axioms is of course the lack of completeness. Transitivity is preserved as a basic tenet of rational preferences. For a detailed discussion of the relationship between transitive but incomplete preferences and intransitive choice see Mandler (2005). Weak C-Independence is a weakening of the common C-Independence assumption, as it imposes that two of the acts involved be constant.

The Basic Conditions imply that there is a utility representation v on L (in fact it is an EU representation). Say that an act f is N -simple if $|v \circ f(\Omega)| \leq N$, and N -complex otherwise. Let F_N be the set of N -simple acts. I will refer to the *coarsest* partition of Ω with respect to which $v \circ f$ is measurable as f 's *partition* (so an act is N -simple iff its partition has no more than N elements). I will say that f is *N -simple on E* if $|v \circ f(E)| \leq N$. I make no behavioral assumptions regarding N -simple acts directly, but they will turn out to be closely related to perceived complexity. Intuitively, the fact that what ends up mattering is the range of $v \circ f$, rather than f , reflects that the DM does not find it difficult to evaluate individual consequences, but rather to understand the mapping from states to consequences. Moreover, distinguishing between acts with the same utility image is incompatible with Monotonicity; if $v \circ f = v \circ g$ then Monotonicity implies $f \sim g$. Monotonicity, i.e. respecting statewise dominance, I take as a basic tenet of rationality. In many applications, such as the portfolio choice and insurance applications considered here, the DM has a strict ordering on $v \circ f(\Omega)$, in which case the distinction between the partition of f and $v \circ f$ disappears.

The size of an act's partition is a common measure of complexity in both the theoretical and empirical literature (see Section 10). This notion captures the idea that the DM has trouble *i*) forming beliefs about many events and *ii*) combining a large number of potential outcomes to understand an act's value. Unlike most existing papers however, I do not assume

that partition size represents subjective complexity. Rather, I show that this measure arises endogenously as a characterization of well-understood comparisons.

3 Characterizing understanding

In order to characterize complexity as perceived by the DM, I require a choice-based measure of the types of comparisons that the DM finds it easy to make. Monotonicity gives us some information along these lines, as it says that the DM is always able to rank acts that are ordered by statewise dominance. The easiest acts to compare to an arbitrary act f , aside from those related to f by statewise dominance, are, intuitively, constant acts. The more constant acts the DM is able to rank against f the better f is understood. Motivated by this intuition, I will begin by making assumptions about comparisons to constant acts, and in particular about the set of acts for which there exists a certain equivalent, i.e. a constant act to which it is indifferent. Let F_{CE} be the set of such acts, which I call *well-understood*.²

A bet on an event E is a binary act that has a strictly better consequence on E than on the complement of E . Given any acts f, g , let f_{Eg} be the binary act which is equal to f on E and g elsewhere. Say that an event E is *non-null* if there exist $c^1, c^2, c^3 \in F_c$ such that $c_E^2 c^1 \succ c_E^3 c^1$, and *null* otherwise (the empty set is null, by reflexivity of \succ).³ The following axioms characterize the set of acts that are well-understood by the decision maker.

Simplicity Conditions. Let τ be f 's partition, with typical elements T, T' .

A0. For any event $E \in \Sigma$ there exists a bet on E that is well-understood.

A1. If f is well-understood then for any partition τ' coarser than τ , there exists a well-understood act with partition τ' .

A2. Let f be a well-understood act. Let g measurable with respect to f 's partition, and for which there exists a constant act $a \in f(\Omega)$ such that for all $\omega \in \Omega$ either $g(\omega) = f(\omega)$ or $g(\omega) = a$. Then if g is well-understood, $\alpha f + (1 - \alpha)g$ also is well-understood for all $\alpha \in (0, 1)$.

A3. If f is well-understood then for any non-null $T \neq T'$ in τ , and any 2-element partition τ' of $T \cup T'$, there exists a well-understood act with partition $\tau' \cup (\tau \setminus \{T, T'\})$.

A4. If f is well-understood and E is null then $g_E f \sim f$ for all $g \in F$.

The Simplicity Conditions will be motivated by a natural model of learning, presented in

²It is intuitive that acts which are easy to understand have certain equivalents. Conversely, I will assume further on that Independence, another intuitive property of easy to understand acts, is satisfied on F_{CE} .

³As we will see, we will be able to identify a probability P which the DM uses to evaluate well-understood acts, and null events will be zero measure events under P .

Section 4. Axioms A0-A3 are stated using the “there exists an act” qualifier, rather than “for any act” for two reasons. First, the learning model, which helps clarify the content of the Simplicity Conditions, implies the “there exists” version of the axioms, but not the “for any”. Second, to state the axioms “for any act” would essentially be to assume that complexity is a property of an act’s partition. Instead, I show exactly which assumptions characterize this type of complexity.

Intuitively, A0 says that there are no events which the DM does not understand. This allows us to identify a subjective probability measure on the state space that characterizes the DM’s preferences over binary acts. A2 is similar to the Certainty Independence assumption of Gilboa and Schmeidler (1989). Certainty Independence would imply that if f is well-understood and c is a constant act, then $\alpha f + (1 - \alpha)c$ would also be well-understood. A2 generalizes this conclusion to mixtures with non-constant acts g that are well-understood, provided g has the same partition and range as f , and is constant on the set of states on which it differs from f . The usual arguments in favor of Certainty Independence therefore apply. A3 says that if f is well-understood, then there are acts that are “close” to f that are also well understood. Here “close” has two meanings. First, the partitions of the acts can only differ on two cells. Second, since A3 only implies that there exists *some* well understood act with partition $\tau' \cup (\tau \setminus \{T, T'\})$, the values of the new act can be close to those of f . If the new act is “nearly constant” on τ' , then A3 is conceptually close to A1. This connection will be formalized in the learning model of Section 4. As we will see, the Basic Conditions, A0-A2, and S-Independence, discussed below, imply that there is an expected utility representation of preferences over well-understood acts. In light of this, A4 simply says that the decision maker is not artificially confused by changes to a well-understood act that occur with zero probability.

Under the Basic Conditions, the Simplicity Conditions characterize the set of well-understood acts. This characterization can be separated into two parts. First, A0-A2 imply that *i*) if any act with a given partition is well-understood then so are all such acts, and *ii*) with the refinement order on the space of partitions, the set of partitions for which measurable acts are well-understood is a lower-set.⁴

Proposition 1. *Under A0-A2 and the Basic Conditions, if some act with partition τ is well-understood then so are all τ -measurable acts.*

Conceptually, Proposition 1 implies that complexity is a property of an act’s partition.

⁴Recall the distinction between an act “with partition τ ” and a τ -measurable act. The former means τ is the *coarsest* partition w.r.t which the act is measurable.

One way to interpret Proposition 1 is as follows. Along any sequence of increasingly fine partitions, there is a last partition such that all acts with this partition have certain equivalents. This alone is an interesting and potentially useful characterization of understanding. If we look only at acts with partitions ordered by refinement, then Proposition 1 tells us all we need to know. In many applications, however, it will be helpful to have a bit more structure on the model of choice. This structure is delivered primarily by A3, with A4 playing more of a technical role. These assumptions imply that along any sequence of refined partitions, the last partitions with the “certain equivalent property” have the same number of non-null elements.⁵

Theorem 1. *Assume \succsim satisfies the Basic Conditions and Simplicity Conditions. Then there exists an $N \in (\mathbb{N} \setminus \{1\}) \cup \{\infty\}$ such that an act is well-understood iff its partition has at most N non-null elements.*

Another way of stating the characterization in Theorem 1 is that an act is well-understood iff it is N -simple on all but a null event. This theorem provides a choice-based foundation for the intuitive, and widely used, measure of complexity as partition size.

4 A Learning Motivation for the Simplicity Conditions

I present here a simple procedural model of a decision maker using data to inform their choice between uncertain acts. The purpose of this section is twofold. First, it demonstrates that the Simplicity Conditions are satisfied under natural assumptions on the learning model. Second, it formalizes the intuition that acts with coarser partitions are easier to understand.

The setting is that of a standard frequentist inference problem. A decision maker is endowed with a dataset χ of K i.i.d. draws from an unknown distribution P .⁶ They must compare an arbitrary simple real valued act f and constant act c . These can be thought of as the utility images of acts with outcomes in an arbitrary space. The DM uses their data to estimate the expected value of f , as well as the risk arising from sampling uncertainty. Let

⁵The extension from Proposition 1 to Theorem 1 is non-trivial. It is not the case that Theorem 1 would follow immediately if “there exists” was replaced with “for any” in axioms A1-A3.

⁶This sampling procedure can be interpreted literally as sampling from an unknown distribution. Alternatively, we can think of it as a reduced form model of contemplation in which the DM accesses a latent belief P , similar in spirit to drift-diffusion models of cognition.

the empirical distribution of the sample be \hat{P}_χ . The empirical expectation of f is

$$\hat{E}_\chi[f] = \sum_{i=1}^N f(T_i) \frac{1}{K} \sum_{x \in \chi} \mathbb{1}\{x \in T_i\}. \quad (4.1)$$

The error due to sampling uncertainty is $\varepsilon_f(\chi) := \hat{E}_\chi[f] - \mathbb{E}[f]$. Denote the true distribution of $\varepsilon_f(\chi)$ across different samples by G_f . G_f is unknown to the DM, as it depends on the unknown distribution P . Instead, the DM uses an estimate \hat{G}_f of G_f when making decisions. I assume that the DM uses the bootstrap to estimate \hat{G}_f .⁷ To complete the description of the environment, assume that the DM uses a decision rule that maps $\hat{E}_\chi(f)$ and \hat{G} to a ranking between f and c . I also allow the DM to declare that f and c are un-rankable. This is a standard problem of statistical inference. I will show that any protocol that is consistent, in a weak sense, with second order stochastic dominance rankings of the error distributions will satisfy A1. I will show further that reasonable decision rules in this environment satisfy all of the Simplicity Conditions.

First, I will discuss how A1 is related to second order stochastic dominance rankings of the error distributions. Let $\tau = \{T_i\}_{i=1}^N$ be an act f 's partition, and let $\tau' = \{T', T_3, \dots, T_N\}$ where $T' = T_1 \cup T_2$. It seems intuitive that learning about acts measurable with respect to τ' will be easier than learning about those measurable with respect to τ , since the set of events to which the DM must assign probabilities is strictly smaller (in the inclusion order). Proposition 2 formalizes the sense in which this is true. Of course, the difficulty of a comparison depends not only on the partitions of the acts involved, but also on their values. Let \tilde{f} be an act such that $\tilde{f}(x) = \mathbb{E}[f|T']$ for $x \in T'$, and $\tilde{f}(x) = f(x)$ otherwise.

For any act h , let $\varepsilon_h(\chi) = \hat{E}_\chi[h] - \mathbb{E}[h]$, and let G_h be the distribution of ε_h . Say that distribution F *strictly second-order stochastically dominates* distribution H ($F >_{SOSD} H$) if $\int u(x)dF(x) \geq \int u(x)dH(x)$ for all concave u , with strict inequality if u is strictly concave.

Proposition 2. $G_{\tilde{f}}$ *strictly second-order stochastically dominates* G_f .

⁷The bootstrap procedure for estimating G_f is as follows. Draw a sample $\bar{\chi}$ of size K from the empirical distribution (i.e. sample the data with replacement) and calculate $\bar{\varepsilon}_f(\bar{\chi}) := \hat{E}_{\bar{\chi}}[f] - \hat{E}_\chi[f]$. Do this repeatedly and use the resulting empirical distribution of $\bar{\varepsilon}_f(\bar{\chi})$ as the estimate of G_f . The idea is to treat the empirical distribution \hat{P}_χ as if it were the true distribution, for the purposes of estimating G_f . Since $n^{-1/2}\varepsilon_f$ is an asymptotically pivotal statistic, the bootstrap estimator provides an asymptotic refinement of the normal approximation, and therefore performs better in finite samples (see Horowitz (2001)). Thus it is reasonable for the DM to use the bootstrap to calculate \hat{G}_f .

The intuition for this result is straightforward. Consider the errors made in the estimation of $E[f]$ versus $E[\tilde{f}]$. Notice that for all datasets χ , $\hat{E}_\chi[f|T'] = E[\tilde{f}|T']$. The randomness of $\hat{E}_\chi[f|T']$ simply adds noise to the distribution of errors from the estimation of $E[f]$, relative to those of $E[\tilde{f}]$. A similar conclusion holds for \bar{f} , where $\bar{f} = \hat{E}_\chi[f|T']$ on T' , and $\bar{f} = f$ elsewhere.

Lemma 1. $\hat{G}_{\bar{f}} >_{SOSD} \hat{G}_f$

This follows immediately from Proposition 2, since the bootstrap estimator treats the empirical distribution as if it were the true distribution. We return now to the DM's problem of choosing between the act f and a constant act c . A natural assumption about decisions in this framework is that greater uncertainty about $E[f]$ makes it harder for the DM to compare f to constant acts (comparable in this setting meaning that the DM is willing to state a preference for one of the two acts).

CONFOUNDING SAMPLING UNCERTAINTY: For any acts f, f' with $\hat{E}_\chi[f] = \hat{E}_\chi[f']$ and $\hat{G}_{f'} >_{SOSD} \hat{G}_f$ and any constant act c , if f is comparable to c then so is f' .

Proposition 3. *Any decision rule satisfying Confounding Sampling Uncertainty will satisfy A1*

Proposition 3 is an immediate implication of Lemma 1. As a concrete example of a protocol satisfying Confounding Sampling Uncertainty, consider the following. Given data set χ the DM concludes that $f \succsim c$ if and only if $E_{\hat{G}_f} \left[\phi \left(\hat{E}[f] \right) \right] \geq \phi(c) - k$ for some increasing, strictly concave function ϕ and constant $k > 0$. Moreover $c \succsim f$ if $c \geq \hat{E}_\chi[f]$. I will refer to this as the *smooth sampling uncertainty* model. It is similar in spirit to the smooth ambiguity model of Klibanoff et al. (2005), where here the higher order uncertainty derives from sampling uncertainty, rather than subjective ambiguity. Additionally, the smooth sampling uncertainty model allows the DM to express incomplete preferences.

Clearly, f will have a certain equivalent only if \hat{G}_f is not too dispersed. In particular, f will have a certain equivalent c if and only if *i)* $\hat{E}_\chi[f] = c$, and *ii)* $E_{\hat{G}_f} \left[\phi \left(\hat{E}[f] \right) \right] \geq \phi(c) - k$. Lemma 1 implies that the smooth sampling uncertainty model satisfies Confounding Sampling Uncertainty since ϕ is concave. In fact, the model satisfies all of the Simplicity Conditions.

Proposition 4. *Smooth sampling uncertainty satisfies the Simplicity Conditions.*

Smooth sampling uncertainty is by no means the only model for which the Simplicity Conditions will be satisfied. As the proof of Proposition 4 shows, A1 and A3 are very close from a learning perspective. Given Proposition 3, A3 will be satisfied so long as the decision rule is suitably continuous in the estimated error distribution \hat{G} .

5 Simple Bounds representation

I turn now to representing preferences. My goal is to stick as close as possible in this regard to the SEU model. However some modifications to the standard SEU axioms must be made to maintain the spirit of the previous axioms. In particular, the usual independence assumption must be modified. In general, the mixture of two N -simple acts will not be N -simple. The standard independence axiom would thus expand the set of well-understood acts; it would imply $N = \infty$. The modified axiom, S-Independence, eliminates this concern. Moreover, it addresses some of the usual critiques of the independence assumption. For one, it applies only when all acts involved are well-understood. Moreover, it only applies to mixtures that do not result in too complex an act, in the sense of partition fineness.

S-INDEPENDENCE: Let f, g, h be well-understood acts. For any $\alpha \in (0, 1)$, if $\alpha f + (1 - \alpha)h$ and $\alpha g + (1 - \alpha)h$ are both well understood, then $f \succsim g$ if and only if $\alpha f + (1 - \alpha)h \succsim \alpha g + (1 - \alpha)h$.

Finally I make an assumption on how comparisons involving complex acts can be made. I write $f \geq_E g$ if $f(\omega) \succsim g(\omega) \forall \omega \in E$. If there exists a null set E such that $f \geq_{\Omega \setminus E} g$, write $f \geq^0 g$. For simplicity, I write \geq rather than \geq_{Ω} to represent statewise dominance.

UNIFORM COMPARABILITY: For any $f, g \in F$ such that $\neg(f \geq g)$, $f \succsim g$ holds iff there exist well-understood acts h, k such that $f \geq h \succsim k \geq g$.

Of course, if f and g are well understood then $f = h$ and $g = k$. In general, it seems reasonable to assume that comparisons according to statewise dominance can be made, even if the acts in question are complex. Uniform Comparability says that statewise dominance is the *only* way to compare complex acts. Thus Uniform Comparability implies the minimal extension of preferences beyond well-understood acts. The “only if” direction formalizes the idea that complexity causes incompleteness of preferences.

Uniform Comparability leads directly to a representation of preferences in which complicated acts are bracketed by well-understood acts, and $f \succsim g$ if and only if the best well-understood act dominated statewise by f is preferred to the worst well-understood act that statewise dominates g . A convenient way to describe this bracketing is through simple upper and lower bounds.

Definition. For any $f \in F$ and $N \in \mathbb{N}$, denote by $\mathbf{sim\,sup}_{N,f}$ the set of acts h satisfying: 1) $h \in F_N$, 2) $h \geq^0 f$, and 3) there is no k satisfying 1 and 2 such that $h \succ k$.⁸

⁸Requiring only $h \geq^0 f$, rather than $h \geq f$, in condition 2 is a technicality. For most applications, we can replace this condition with $h \geq f$, as Proposition 5 shows.

Define $\text{siminf}_{N,f}$ analogously. Note that for any N such that all acts in F_N are well-understood, the DM is indifferent between all acts in $\text{simsup}_{N,f}$ (similarly for $\text{siminf}_{N,f}$). When there is no risk of confusion, I will therefore abuse notation and write as if siminf and simsup are single valued (for example, $f \succsim \text{siminf}_{N,f}$ even when $\text{siminf}_{N,f}$ may contain multiple acts). It is not obvious that such bounds exist, i.e. that $\text{siminf}_{N,f}$ and $\text{simsup}_{N,g}$ are non-empty. In fact, the axioms stated will imply that both are non-empty. Conversely, Theorem 3 shows that existence of these bounds places no additional restrictions on the model parameters.

For a lottery l , I write $E_l u = E_l[u(z)]$, i.e. the expected utility given distribution l over outcomes. Although $\text{siminf}_{N,f}$ may not be single valued, I will abuse notation and write $\int_{\Omega} E_{\text{siminf}_{N,f}(\omega)} u dP(\omega)$.

Definition. Preference \succsim has a **Simple Bounds** representation if there exists an integer N , probability P on Σ , and a non-constant function $u : Z \mapsto \mathbb{R}$ such that, for every $f, g \in F$, $f \succsim g$ if and only if at least one of the following holds:

i. $f \geq g$.

ii.

$$\int_{\Omega} E_{\text{siminf}_{N,f}(\omega)} u dP(\omega) \geq \int_{\Omega} E_{\text{simsup}_{N,g}(\omega)} u dP(\omega)$$

I refer to such preferences as Simple Bounds preferences.

Theorem 2. *The following statements are equivalent:*

i. \succsim satisfies the **Basic Conditions**, **Simplicity Conditions**, **S-Independence**, and **Uniform Comparability**.

ii. \succsim has a Simple Bounds representation, with parameters P, u, N . Moreover, P is unique, u is unique up to positive affine transformations, and for all $f \in F$, $\text{siminf}_{N,f}$ and $\text{simsup}_{N,f}$ are non-empty.

Condition i. in the definition of a Simple Bounds representation requires $f \geq g$, rather than $f \geq^0 g$. This is natural if we think that it is harder to identify null events when comparing complex acts. Alternatively, we could assume that preferences obey \geq^0 dominance, and make the obvious modification to Uniform Comparability, to replace i. with $f \geq^0 g$.

For applications, it is generally without loss to assume that $\text{simsup}_{N,f} \geq f \geq \text{siminf}_{N,f}$. Roughly speaking, violations of statewise dominance only occur if f has isolated outliers in the support of P . No continuity is required.

Proposition 5. $f \geq \text{siminf}_{N,f}$ iff $f^{-1}(A)$ is non-null for every open neighborhood of $\inf E_f u(\Omega)$. Similarly, $\text{simsup}_{N,f} \geq f$ iff $f^{-1}(A)$ is non-null for every open neighborhood of $\sup E_f u(\Omega)$.

The conditions of Proposition 5 are met, for example, if P has full support on $\Omega \subseteq \mathbb{R}$ and $E_f u$ is continuous, or if its partition has no null elements, or it is a convex combination of any acts with these properties. Even if the conditions for statewise dominance fail, it is easy to see where violations will occur. The ‘‘Lebesgue approach’’, discussed below, helps clarify this point.

Theorem 2 includes the conclusion that $\text{siminf}_{N,f}$ and $\text{simsup}_{N,f}$ are non-empty for all f . A natural concern from a modeling perspective is that this may impose constraints on the other parameters of the representation. This would be the case if existence failed for some specification of u, P or N . The following theorem states that this is not the case. Let $B(\Omega)$ be the set of bounded measurable functions on Ω , and $B_N(\Omega)$ be the set of N -simple measurable functions on Ω . Abusing notation, define

$$\text{siminf}_{N,w,P} \equiv \underset{\{b \in B_N(\Omega) : b \leq^0 w\}}{\text{argmax}} \int_{\Omega} w(\omega) dP(\omega)$$

Theorem 3. For any $w \in B(\Omega)$, $\text{siminf}_{N,w,P}$ and $\text{simsup}_{N,w,P}$ are non-empty for all N, P .

A full discussion of Theorem 3 can be found in Appendix A. The proof is instructive, as it makes use of a ‘‘Lebesgue approach’’ to the problem of finding simple upper and lower bounds; rather than think of partitions of the arbitrary space Ω , I look instead at partitions of $w(\Omega)$. I show that $\text{simsup}_{N,w,P}$ and $\text{siminf}_{N,w,P}$ can be mapped to increasing functions on $w(\Omega)$, and then exploit this monotonicity and fact that $w(\Omega)$ is a bounded interval of \mathbb{R} . Put another way, Ω inherits both an order and a topology from \mathbb{R} and the measurable function w , which greatly simplifies the problem of finding simple bounds. The Lebesgue approach is also helpful in comparative statics (Appendix C). Moreover, it is easy to use this approach to show that finding $\text{simsup}_{N,w,P}$ and $\text{siminf}_{N,w,P}$ functions can be reduced to choosing a single number in $w(\Omega)$, and is thus computationally very simple.

5.0.1 Theorem 2 discussion

A natural question is how the incomplete preferences characterized above relate to those of Bewley (2002).⁹ So called ‘‘Bewley preferences’’ have the following representation: there

⁹Bewley (2002) in fact gives a representation of the strict preference. The exact unanimity representation discussed here is due to Gilboa et al. (2010).

exists a non-empty, closed, and convex set C^* of probabilities on Σ and a non-constant function $u : Z \mapsto \mathbb{R}$ such that $f \succsim g$ if and only if

$$\int_{\Omega} E_{f(\omega)} u dp(\omega) \geq \int_{\Omega} E_{g(\omega)} u dp(\omega) \quad \forall p \in C^*.$$

There is no simple relationship between the two representations. The Bewley preferences satisfy the usual independence axiom. However it is easy to see that the representation in Theorem 2 does not satisfy independence. If incompleteness à la Bewley is interpreted as reflecting complexity then it must be that mixtures do not increase the complexity of a comparison. For example, let the state space be $[0, 1]$ and for any act k identify $u(k(\omega))$ with $k(\omega)$. Let f and g be binary acts, with $f = 0, g = 1$ on $[0, 1/2)$ and $f = 1, g = 0$ on $(1/2, 1]$. In the Bewley model, if f and g are comparable then so are $1/2f + 1/2h$ and $1/2g + 1/2h$ for $h(\omega) = 10\omega^2$. Decision makers may find the later comparison, which involves acts with a greater range of values and larger partitions, more difficult. Relaxing Independence and explicitly modeling complexity allows for a model in which mixtures can increase complexity and lead to incomparability.

5.0.2 A Note on Transitivity

The Simplicity Conditions highlight the critical role of transitivity in the characterization. Transitivity is often regarded as a weak rationality condition. I maintain Transitivity for the usual reasons (no money pumps, normative desirability, etc.). It is worth pointing out, however, that it also rules out many non-separable models of pairwise choice. For example, consider the following model: acts f and g are comparable if and only if either *i*) the join (coarsest common refinement) of their partitions has less than N elements, or *ii*) one statewise dominates the other. These preferences are consistent with Monotonicity, Archimedian Continuity, Nontriviality and the Simplicity Conditions, but not Transitivity. The idea behind this model is that the decision maker must evaluate the difference between the payoffs from the two acts. Such preferences are ruled out by Transitivity however, precisely because Transitivity implies that any two acts which are comparable to all constant acts can be compared.

I extend preferences from the set of well-understood acts to complex acts using only the statewise dominance order, which satisfies Transitivity. Assuming Transitivity of \succsim rules out extensions that violate Transitivity, although some of these do have intuitive appeal. Relaxing transitivity while accommodating incompleteness arising from subjective complexity is an interesting topic for future work.

6 Completing Preferences

In many settings the decision maker is forced to make a choice. In such cases we would like to be able to make predictions about behavior even when the environment contains pairs of alternatives that are not ranked according to the incomplete preferences above. Following Gilboa, Maccheroni, Marinacci, and Schmeidler (2010), henceforth GMMS, I assume that the decision maker is characterized by a pair of binary relations (\succsim, \succsim') , interpreted as *objective rationality* and *subjective rationality* relations respectively.¹⁰ I consider decision makers with objectively rational preferences that can be represented as in Theorem 2, and consider various assumptions on the subjective relation that lead to distinct complete preference relations and representations. This approach allows me to separate the decision maker's ability to compare acts from their attitude towards choices between acts that they do not know how to compare. The first point is addressed by Theorem 2. Attitudes towards the unknown can be captured by intuitive axioms. Following GMMS, I first make the natural assumption that the subjective relation never reverses the objective.

CONSISTENCY: $f \succsim g$ implies $f \succsim' g$.

This notion of consistency allows for indifference according to the subjective relation between acts that are strictly ranked according to the objective relation. A stronger notion of consistency rules out such differences when the acts in question are well-understood. Below I discuss the reason for imposing Strong Consistency only on well-understood acts.

STRONG CONSISTENCY FOR SIMPLE ACTS: For any two well-understood acts f, g , $f \succsim g \Leftrightarrow f \succsim' g$.

Finally, I assume a cautious approach to incomparable alternatives.

CAUTION: For all $f \in F$ and $h \in F_c$, if $f \not\succeq h$ then $h \succsim' f$.

Caution is exactly the axiom used by GMMS to derive max-min expected utility (MEU) as the subjective relation given Bewley objective preferences. When the objective relation is Simple Bounds, Caution yields a representation in which acts are evaluated according to their simple lower bounds.

Definition. Preference \succsim' has a *Cautious* representation if there exists an integer N , prob-

¹⁰According to GMMS, a choice is objectively rational if the DM can convince others that they, the DM, is right in making it. It is subjectively rational if the DM cannot be convinced that they are wrong in making it.

ability P on Σ , and a non-constant function $u : Z \mapsto \mathbb{R}$ such that, for every $f, g \in F$

$$f \succsim' g \text{ iff } \int_{\Omega} E_{\text{siminf}_{N,f}(\omega)} u \, dP(\omega) \geq \int_{\Omega} E_{\text{siminf}_{N,g}(\omega)} u \, dP(\omega).$$

Call preferences that admit a Cautious representation Cautious preferences.

Theorem 4. *The following statements are equivalent:*

- i. \succsim satisfies the **Basic Conditions, Simplicity Conditions, S-Independence, and Uniform Comparability**; \succsim' satisfies **Archimedean Continuity**; and \succsim and \succsim' jointly satisfy **Consistency, Strong Consistency for Simple Acts and Caution**.*
- ii. \succsim has a Simple bounds representation and \succsim' has a Cautious representation, with common parameters P, u, N . Moreover, P is unique, u is unique up to positive affine transformations, and for all $f \in F$, $\text{siminf}_{N,f}$ and $\text{simsup}_{N,f}$ are non-empty.*

Cautious preferences capture a particular attitude towards the unknown. Faced with a difficult choice, the decision maker takes a worst case view of the set of payoffs that they consider reasonable. The relationship between Cautious preferences and models of ambiguity aversion is interesting. I discuss formally the properties of Cautious preferences, and their relationship to models of ambiguity aversion, in the Online Appendix B.1. The Cautious completion is the most ambiguity averse completion of Simple Bounds preferences, in the sense of Ghirardato and Marinacci (2002). Importantly, Cautious preferences do not satisfy the Uncertainty Aversion axiom of Gilboa and Schmeidler (1989), but do satisfy a modified version, N -Ambiguity Aversion, discussed in Hartmann and Kauffeldt (2019).

There may be situations in which decision makers take the opposite approach to what they recognize as their limited understanding of the objects of choices.

ABANDON: For all $f \in F$ and $h \in F_c$, if $h \not\prec f$ then $f \succsim' h$.

Replacing Caution with Abandon yields the obvious converse representation, which I call Reckless preferences, whereby acts are evaluated according to their simsup rather than siminf . Such preferences seem to contradict common assumptions about risk and ambiguity aversion. They have a flavor of, but are distinct from, optimism or overconfidence. The decision maker does not behave exactly as if she thought high payoff states are more likely to occur, but rather as if the underlying act is as good as possible without violating her limited understanding of the situation. Such a model may be useful for understanding the behavior of decision makers who seem to favor nebulous prospects with high potential over those that are well-understood. I will sometimes use the notation $U(f)$ or $U(f, P)$ to denote

perceived utility. If the agent is cautious $U(f, P) = E_P[E_{\text{siminf}_{N,f}(\omega)}u]$, and if the agent is reckless $U(f, P) = E_P[E_{\text{simsup}_{N,f}(\omega)}u]$.

An alternative approach to incompleteness would be to proceed as in Bewley (2002) and assume that there is a default option and the DM satisfies an inertia condition, whereby an alternative act is chosen over the default if and only if it is preferred, according to the underlying incomplete preference. While a full coverage of inertia is beyond the scope of this paper, the results presented here are helpful for understanding such a model. What matters for choice under inertia is the simsup of the default act and the siminf of the alternative. The behavior of these objects is studied in the comparative statics and applications sections.¹¹

6.1 A note on Consistency

Given the interpretation of the subjective preference as an extension of the incomplete objective preference, it would seem natural to impose Strong Consistency everywhere: for all $f, g \in F$, $f \succsim g$ implies $f \succsim' g$, and $f \succ g$ implies $f \succ' g$. Consistency allows the decision maker to be indifferent under \succsim' between some acts that are strictly ranked under \succsim .¹² It turns out that Strong Consistency everywhere is incompatible with Archimedian Continuity of \succsim' . This conflict between Caution, Strong Consistency, and Continuity is not unique to this setting. GMMS face the same trade-off obtaining MEU preferences as the Cautious completion of Bewley preference (see Section B.1), and the incompatibility holds for a broad class of incomplete preference models. Rather than relaxing consistency to retain continuity, as in Theorem 4, one can drop the continuity requirement for \succsim' and impose Strong Continuity everywhere. This yields “lexicographic” preferences: acts are first ranked by statewise dominance, and then according to their siminf if they are not ranked according to statewise dominance. The difference between the lexicographic model and that of Theorem 4 is one of continuity; the two models have nearly identical properties and predictions.

7 Consumption-Savings

In this section I study consumption-savings decisions in a two period model. Extending the model to a truly dynamic environment involves a new set of considerations. If information about the state arrives over time then the objects of choice are plans measurable with respect to information at each period. Moreover, depending on the specification of uncertainty, the

¹¹The earlier working paper version of this paper includes further discussion of inertia.

¹²In fact, under Strong Consistency for Simple Acts, this occurs for acts f, g only if a) either $f \geq g$ or $g \geq f$, and b) $\text{siminf}_{N,f} \cap \text{siminf}_{N,g} \neq \emptyset$.

dimensionality of the state space may increase in the number of periods. For example, if in each period a payoff relevant random variable realizes, then the state space is the set of all sequences of realizations. Additionally, there is the question of how aware the decision maker is of his or her own limitation. As information arrives the siminf and simsup used to evaluate a given plan may well change, leading to dynamic inconsistencies in choice. We must then consider the DM's behavior in the face of such inconsistencies. Incorporating these new subtleties into a model of decision making is beyond the scope of the current paper, and is left for future work.

I begin with a general comparative statics observation. Consider a DM with capacity N evaluating an act f . Suppose that f is real valued, and let P be the cdf of the agent's belief and u the DM's utility function, assumed to be increasing. The partitions defining the siminf and simsup will be composed of intervals, and can be described as vectors of cut-offs $\{t_i\}_{i=0}^N$. If the DM is cautious then they evaluate the uncertain payoff $u(f)$ according to the expectation $E_{P_N}[u(f)]$, where P_N is the distribution supported on the cut-offs defining $\text{siminf}_{N,f}$, $\{t_i\}_{i=0}^{N-1}$, with mass of $P(t_{i+1}) - P(t_i)$ on t_i . Similarly, the reckless agent will evaluate $u(f)$ as the expectation with respect to \tilde{P}_N , which places mass of $P(t'_i) - P(t'_{i-1})$ on the points $\{t'_i\}_{i=1}^N$ defining $\text{simsup}_{N,f}$. I will refer to P_N and \tilde{P}_N as *perceived* distributions. The key observation is that if for $M > N$ the partition defining $\text{siminf}_{M,f}$ ($\text{simsup}_{M,f}$) is finer than that defining $\text{siminf}_{N,f}$ ($\text{simsup}_{N,f}$) then the P_M will first order stochastically dominate P_N (\tilde{P}_N will FOSD \tilde{P}_M). In general partitions will not be so ordered as N increases (see section C.1.1). However for a fixed N this will hold in the limit as $M \rightarrow \infty$.

Observation 1. *Let f be real valued and u increasing. Let P be the DM's prior. Then $P \succsim_{\text{FOSD}} P_N$ and $\tilde{P}_N \succsim_{\text{FOSD}} P$. If, moreover, $u \circ f$ is integrable then $E_P[u \circ f] = \lim_{M \rightarrow \infty} E_{P_M}[u \circ f] = E_{\tilde{P}_M}[u \circ f]$.*

Observation 1 can be extended to more general environments using the Lebesgue approach of section A.5.1. This observation facilitates comparisons between constrained and unconstrained agents.

7.1 Portfolio choice with a safe and risky asset

Suppose the DM is cautious and faces the problem of allocating wealth between consumption, purchase of a risk-free asset and purchase of a risky asset. I want to compare choices under a capacity constraint to those of a fully rational agent ($N = \infty$). Let R_b be the certain return on the risk free asset, and R_s the uncertain return on the risky asset with differentiable CDF

P on the interval $[\underline{R}_s, \bar{R}_s]$. The DM with capacity N solves

$$\max_{b,s} u(w - b - s) + \beta V^N(R_b b + R_s s) \quad (7.1)$$

where

$$V^N(R_b b + R_s s) = \max_{\{\{t_i\}_{i=0}^N \in I\}} \sum_{i=1}^N [P(t_i) - P(t_{i-1})] u(R_b b + t_{i-1} s), \quad (7.2)$$

and I is the set of interval partitions of $[\underline{R}_s, \bar{R}_s]$. Clearly $\lim_{N \rightarrow \infty} V^N(R_b b + R_s s) = \int u(R_b b + r s) dP(r)$ when u is integrable. Notice that the objective in (7.2) is continuous in the cut-offs and differentiable in b, s , so that in what follows I will be able to apply the standard envelope theorem when differentiating with respect to the latter variables.

I prove two results regarding savings behavior of constrained versus unconstrained agents. As Observation 1 makes clear, for any *fixed* portfolio choice the unconstrained DM perceives a FOSD shift of the returns perceived by the constrained DM. Of course, as the constrained DM changes their portfolio allocation the perceived distribution of R_s induced by siminf will also change, so this is not the same problem as comparing portfolio choice under two different fixed yield distributions.

Proposition 6. *Let u be CRRA with a coefficient of relative risk aversion greater than or equal to 1.*¹³ *Then*

- i. For a given level of aggregate savings $x > 0$, a constrained DM allocates a greater portion of savings to the safe asset than an unconstrained DM.*
- ii. A constrained DM saves more overall than an unconstrained DM.*

As the proof illustrates, the assumption of CRRA with coefficient greater than 1, or the weaker condition in the footnote, is by no means necessary.

7.2 Equilibrium Asset Prices

Consider a representative agent model in which the agent chooses between a safe and risky asset, which must both be in zero net supply in equilibrium. Returns are as in the previous section. Normalize the price of the safe asset to 1 and let the price of the risky asset be

¹³Part (i) holds under the weaker condition that $u'''(z)u'(z)/u''(z)^2 \geq 2$ for all z . I state it for CRRA for simplicity. The assumption that the risk aversion coefficient is greater than 1 is standard. Part (ii) also holds under weaker conditions, CRRA just makes it easy to show that (B.7) in the proof holds.

given by p . I assume that the agent has an endowment of w in each period. Let $P^{b,s}$ be the perceived distribution given choices b and s . The constrained DM solves

$$\max_{b,s} u(w - b - ps) + \beta \int u(w + bR_b + sr) dP^{b,s}(r).$$

For the unconstrained DM the problem is identical, except that $P^{b,s}$ is replaced with P . It is intuitive that a cautious (reckless) constrained DM should be biased towards the safe (risky) asset, relative to the unconstrained DM. The constraint simply coarsens the DM's understanding of the stochastic payoff of the risky asset, which leads the cautious agent to undervalue it and reckless agent to overvalue. In fact, when considering assets that must be in zero net supply in equilibrium, we can extend this intuition to make comparisons between intermediate capacity levels.

Proposition 7. *The equilibrium risky asset price with a cautious (reckless) representative agent is increasing (decreasing) in the agent's capacity.*

When the agent is cautious, Proposition 7 gives an explanation for the so called "equity premium puzzle".

8 Principal-Agent

Consider a canonical principal-agent problem. The agent exerts effort a which induces a distribution of output P_a , with support on a bounded set Ω . The principal offers the agent a wage schedule $w : \Omega \mapsto \mathbb{R}$. The utility of the agent's outside option is normalized to 0. The principal receives the output minus the payment to the agent. Given a contract w , the agent chooses effort a to maximize $E_{P_a}[u(w(\omega), a)]$. The principal chooses w to maximize $E_{P_{a^*(w)}}[v(\omega, w(\omega))]$, where $a^*(w)$ is the agent's effort choice given contract w^* (assume the agent chooses the principal's preferred effort level when the agent's problem has multiple solutions). The only assumption I make on preferences is that $w \mapsto u(w, a)$ is increasing for all a and that $w \mapsto v(\omega, w)$ is decreasing for all ω .¹⁴

This formulation allows for output that lives in an arbitrary space, and for general payoffs for both the principal and agent. Nonetheless, it is relatively easy to identify features of optimal contracts for agents with different attitudes: cautious, reckless, and α -min-max (introduced below).

¹⁴The results extend immediately to the case of multi-dimensional payment spaces W , provided the monotonicity conditions on u and v hold with respect to the same order on W for all a , and that the order does not depend on a .

8.1 Cautious Agents

Let the agent be Cautious, with capacity N . Suppose that the principal offers a contract w and the agent chooses effort a . I will refer to $\langle w, a \rangle$ as the act induced by w and a (where the only uncertainty is on the w coordinate). If the principal offers the agent a contract \hat{w} such that $\langle \hat{w}, a \rangle \in \text{siminf}_{N, \langle w, a \rangle, P_a}$ then agent is no worse off; the agent can always choose the same effort level and is indifferent between the acts $\langle w, a \rangle$ and $\langle \hat{w}, a \rangle \in \text{siminf}_{N, \langle w, a \rangle, P_a}$ given output distribution P_a . In fact the nature of Cautious preferences implies that the agent's optimal effort choice will not change.

Proposition 8. *Given a contract w that induces optimal effort choice a by the agent, any contract \hat{w} such that $\langle \hat{w}, a \rangle \in \text{siminf}_{N, \langle w, a \rangle, P_a}$ induces the same effort level and has the same value for the agent.*

Proof. Suppose the agent is offered contract w and chooses effort a . Assume $\text{siminf}_{N, \langle w, a \rangle, P_a}$ is unique (the same argument applies to any selection). If the principal instead offers \hat{w} such that $\langle \hat{w}, a \rangle = \text{siminf}_{N, w, P_a}$ and effort remains unchanged then the agent's perceived payoff is unchanged, $U(\langle w, a \rangle, P_a) = U(\langle \hat{w}, a \rangle, P_a)$.

Since $w \mapsto u(w, a)$ is increasing, it must be that $w(\omega) \geq \hat{w}(\omega)$ for all ω . So for any a' , $\langle \hat{w}, a' \rangle$ is an N -simple act that satisfies $u(w(\omega), a') \geq u(\hat{w}(\omega), a')$ for all ω . In other words, for all a' , $\langle \hat{w}, a' \rangle$ is in the set of acts for which $\text{siminf}_{N, \langle w, a' \rangle, P_{a'}}$ is maximal. Then by the definition of siminf , we have $U(\langle \hat{w}, a' \rangle, P_{a'}) \leq U(\text{siminf}_{N, \langle w, a' \rangle, P_{a'}}, P_{a'}) = U(\langle w, a' \rangle, P_{a'})$ for all a' . Since the agent originally chose a it must be that $U(\langle w, a' \rangle, P_{a'}) \leq U(\langle w, a \rangle, P_a)$. Combining the inequalities in the two preceding lines, we have $U(\langle \hat{w}, a' \rangle, P_{a'}) \leq U(\langle w, a \rangle, P_a) = U(\langle \hat{w}, a \rangle, P_a)$, where the final equality follows from the definition of \hat{w} . This proves that effort a remains optimal. \square

Since w statewise dominates \hat{w} and $w \mapsto v(w, \omega)$ is decreasing we have the following immediate corollary of Proposition 8.

Corollary 1. *If the agent has capacity N and is cautious then all optimal contracts are N -simple.*

8.2 Reckless Agents

When the agent is reckless the principal is able to exploit the agent by confusing them with a complex contract and offering “prizes”, discontinuous jumps in the wage schedule. The agent sees these prizes and over-reacts, overvaluing the resulting payoffs. The results in this

section are similar in spirit to those of Vierø (2014), in that the principal exploits the limited understanding and “optimism” of the agent by offering “bait” contracts.

Formally, I show how any contract can be improved by introducing complexity and adding a discontinuous jump in a neighborhood of the highest wage. These changes can be made so as to induce the same effort level of the agent at lower cost to the principal. This implies that any optimal contract will also have these properties.

Proposition 9. *For all N , all optimal contracts for a reckless agent are N -complex and have a discrete jump in a neighborhood of the highest payoff state.*

9 Insurance Valuation

I compare the salience of various features of insurance plans for capacity constrained and fully rational ($N = \infty$) individuals. I focus on the cautious model here, although interesting results are also obtained under recklessness. I take as given a class of basic insurance contracts that are characterized by a premium p , i.e. price for the plan; a deductible d , below which the individual bears all losses; a coverage rate c specifying the fraction of losses above the deductible covered by the plan; an out-of-pocket expenditure cap m , which is the maximum amount that an individual will have to pay, excluding the premium. I do not provide a foundation for the use of these piece-wise linear contracts, but they are by far the most common form of insurance contract.

Say that a capacity-constrained individual *over-reacts* (*under-reacts*) to a change from one plan to another if the magnitude of the difference between their values for the two plans is greater (less) than that of a fully rational individual. This section relates to the large empirical literature documenting behavioral phenomena in choices of complicated contracts. The model is able to explain many observed choice patterns which differ from the predictions of standard EU theory. Abaluck and Gruber (2011) find that consumers underweight out-of-pocket spending relative to premiums. Moreover, cautious agents will respond more than the fully rational to changes in the coverage rate when there is an out-of-pocket expenditure cap and the coverage rate is high (see Proposition 10). Cutler and Zeckhauser (2004) document a bias towards low deductible plans. This is consistent with a cautious decision maker (see Proposition 10). The results of this section highlight the dependence of the qualitative nature of comparisons to the rational model on details of the insurance plans under consideration. Such variation forms an interesting basis for further empirical work, and provides a more nuanced perspective on “behavioral biases”.

The results presented here relating the siminf to plan features are the key to understanding

inertia. In Online Appendix B.2 I discuss how the predictions of the Cautious model in this setting relate to those of MEU. There is also significant evidence that inertia plays a major role in plan choice. Plan choice with Simple Bounds preferences and inertia is an interesting topic for future work.

9.1 The setting

Consider an individual facing a bounded loss distribution on $\Omega = [0, \bar{\omega}]$ with CDF P . The loss distribution will be fixed throughout. Let w be the individual's endowment wealth. In autarky the ex-post wealth is $w - \omega$, where ω is the realized loss. Assume that P is absolutely continuous. I do not consider here the choice of the contract by the insurer, only the valuation by the individual. The contracts that I consider will have the feature that the individual's ex-post wealth is decreasing in the realized loss. Recall that in this case siminf and simsup will be measurable with respect to N -element interval partitions, which I will describe by the cut-off states $t_0 = 0, t_1, \dots, t_N = \bar{\omega}$.



Figure 1: siminf with and without out-of-pocket maximum

9.1.1 Salient plan features

I will consider changes in the deductible and in the coverage rate, meaning the percentage c of losses above the deductible covered by the contract ($c = 1$ in the full insurance case). Proposition 10 states that a cautious agent over-reacts to changes in both parameters when there is no out-of-pocket expenditure cap. On the other hand, when there is a binding cap the agent's valuation is unaffected by changes in the coverage rate, provided the coverage rate is sufficiently high.

Proposition 10. *Consider marginal changes to a contract with no out-of-pocket expenditure cap. A **cautious** agent:*

1. *over-reacts to changes in the deductible.*
2. *over-reacts to changes in the coverage rate.*

Proposition 10 is silent on which of the distortions is relatively larger. If we focus on marginal changes to a baseline plan with full insurance above the deductible, then the over-reaction to the coverage rate will be relatively larger than that to the deductible if the decision maker believes that high losses are sufficiently likely. In this case the highest cut-off defining the siminf will be close to d (but always strictly below d , by Lemma 4), and so the response to a deductible change will be close to that of a fully rational DM. However caution will still cause the DM to drastically over-react to reductions in coverage rate. Similarly, when N is sufficiently high the highest cut-off defining the siminf will be close to d , so again the over-reaction to the coverage rate will be the larger of the two when N is high enough.

Restricting attention to contracts with near full insurance above the deductible, it is easy to show a stronger comparative statics result.

Proposition 11. *Consider a contract with deductible d and $c = 1$. For both the deductible and the coverage rate the magnitude of the response of a cautious agent to marginal changes is decreasing in N .*

Many of the settings in which biases towards low deductibles are observed involve full coverage above the deductible. In this case cautious individuals are always biased towards low deductible plans, and this bias is decreasing in their capacity.

Proposition 12. *With $c = 1$, the amount a cautious individual is willing to pay to lower the deductible by a given amount is decreasing in their capacity.*

Finally, I consider how a capacity constraint affects the individual's willingness to pay to decrease the out of pocket maximum.

Proposition 13. *The amount a cautious agent would be willing to pay to decrease the out-of-pocket maximum is decreasing in their capacity.*

The finding that lower capacity individuals place a greater value on the out of pocket maximum may help explain the widely documented bias towards full insurance, for example by Shapira and Venezia (2008). Full insurance plans are easy to understand, since the loss is independent of the state. In general, constrained individuals will overvalue full insurance plans relative to those for which the realized losses are a more complicated function of the state.

9.1.2 Dominated choices

Bhargava et al. (2017) find that many individuals choose dominated plans, and that the propensity to do so is positively correlated with both high expected losses and earnings levels.¹⁵ The cautious model can provide an explanation for these observations.¹⁶ The result depends on the nature of plan dominance. These results also demonstrate the usefulness of the comparative statics properties discussed in the appendix. Bhargava et al. (2017) observe individuals choosing low deductible plans even when the increase in the premium relative to a high deductible plan (holding other plan features constant) is greater than the maximum possible savings from the lower deductible. Figure 2 illustrates a situation in which the maximum possible savings from the low deductible, high premium plan is equal to the increase in the premium. Call this as a *weakly dominated low-deductible plan*.

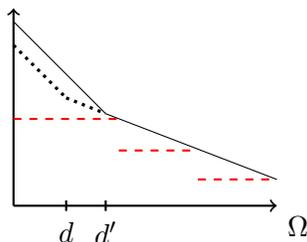


Figure 2: Indifference with a weakly dominated low-deductible plan

In Figure 2 the solid black line is the high deductible, low premium plan, and the black dashed line is a low premium, high deductible plan. Both plans have the same coverage rate. The red dotted line is the siminf, which in this case is the same for both plans. This occurs whenever the lowest cut-off for the high deductible plan is above its deductible. Let $l(\omega|d, c)$ be the amount paid by the consumer when the loss is ω given a contract with deductible d and coverage rate c (this is stated formally in Lemma 5 in the Online Appendix). More

¹⁵Interestingly, the authors find that the number of parameters needed to describe a plan, an alternative measure of complexity encountered in the literature, does not predict dominated choices. This suggest that the partitional notion of complexity may be more relevant in this setting.

¹⁶My model does not predict strictly dominated choices, but allows for indifference between pairs of plans ordered by weak dominance, even for full support beliefs. If dominated choices are related to act complexity then the empirical evidence is informative about what types of acts are perceived to be complex. Additional factors, such as difficulty understanding how plans map states to payments, may combine with complexity considerations to produce strictly dominated choices.

interesting than the fact that dominated plans can be chosen is are the conditions which are conducive to such mistakes. Bhargava et al. (2017) observe that dominated choices are correlated with both expected losses and earnings levels. Proposition 14 predicts the former. To the extent that earnings are correlated with the capacity to evaluate acts, Proposition 15 predicts that low earners will be more prone to mistaken indifference. The following corollary establishes a single-crossing property of dominated choice which implies that individuals who are more pessimistic about their losses are more prone to make dominated choices (in both cases, all other model parameters are held constant).

Proposition 14. *If an individual with belief P chooses a weakly dominated low-deductible plan then so does one with belief P' if $P' \succeq_{MLR} P$.*

Proposition 15. *If an individual with capacity N' chooses a weakly dominated low-deductible plan then so is one with capacity N if $N' \geq N$.*

10 Related Literature

Other papers have built on the idea that acts with many outcomes may be difficult for a decision maker to evaluate. Neilson (1992) proposes a model of choice under risk in which the decision maker uses a different utility function when computing expectations for lotteries with different support sizes. Puri (2020) axiomatizes a “Simplicity Representation” of choice under risk, in which a lottery p is evaluated according to $E_p[u(x)] - C(|\text{support}(p)|)$, for some increasing function C .¹⁷ While this model also relates support size to complexity, its empirical content is quite different; the Simplicity Representation and Simple bounds are very far from being “dual” in the sense one might expect at first glance. Most importantly, the Simplicity Representation makes a sharp separation between the values on which a lottery is supported and its complexity, as measured by support size. A Simple Bounds DM with capacity N will have near perfect understanding of any act that is “close” to an N -simple act, since in this case siminf and simsup will provide tight bounds on the act’s value. The same is true for Cautious and Reckless DMs. The Simplicity Representation, on the other hand, imposes the same complexity cost on an lottery, regardless of the values in its support. In this sense, the decision maker under the Simplicity Representation is more sensitive to support size than under the Simple Bounds model, or related completions. As a consequence, the

¹⁷Related ideas appear in the menu choice literature. Ortoleva (2013) axiomatizes a model of preferences over lotteries of menus in which, similar to the model of Puri (2020), the decision maker attaches a cost to lotteries with more menus in their support.

Simplicity Representation predicts potentially extreme preference reversals resulting from small changes; arbitrarily small perturbations of a lottery that increase the size of its support can dramatically change its complexity cost. This is not the case in the Simple Bounds model, in which the decision maker uses the acts they understand well to help evaluate nearby acts that they do not. The Simplicity Representation also predicts violations of first order stochastic dominance, analogous to violations of Monotonicity in my setting.

Saponara (2020) axiomatizes a Revealed Reasoning model which is close in spirit to the Cautious model. In this model a decision maker is characterized by a set of partitions \mathcal{P} , and evaluates an act f according to the best act that is uniformly below f and measurable with respect to some partition $P \in \mathcal{P}$. \mathcal{P} need not be characterized by a fixed number of elements; instead the axioms impose that \mathcal{P} satisfy a richness condition.¹⁸ Moreover, preferences over the set of acts that are measurable with respect to some $P \in \mathcal{P}$ may not have an expected utility representation. Thus the Revealed Reasoning model differs from Caution in important ways, the latter providing more choice-based structure on the DM's behavior. On a more technical note, Saponara (2020) implicitly begins with the assumption that an act's complexity is determined by its partition. One of the contributions of the current paper is to derive this conclusion from assumptions on choice behavior (primarily in Proposition 1). These differences aside, the relationship between the Simple Bounds and Revealed Reasoning models is similar to that between Bewley preferences and MEU.

There are formal similarities between this paper and Gilboa, Maccheroni, Marinacci, and Schmeidler (2010) and Lehrer and Teper (2014). As in Lehrer and Teper (2014), I begin by characterizing preferences on a small set of acts and extend these to a larger subset, although not necessarily to all acts. As in GMMS, given a characterization of incomplete preferences, I show that under additional assumptions a complete preferences relation consistent with it exists, and takes a specific form. In GMMS the incomplete and complete preference relations admit representations à la Bewley (2002) and à la Gilboa and Schmeidler (1989) respectively. The incomplete preferences arising from the partition size notion of complexity do not admit a Bewley representation. Similarly the Cautious completion is not an MEU preference, although it still captures a notion of ambiguity aversion.

One objective of the current paper is to explore the connection between Ellsberg type phenomenon (Ellsberg (1961)) arising in the presence of ambiguity and the complexity of decision making problems under uncertainty. As in Segal (1987) and Klibanoff et al. (2005), bets on ambiguous urns are viewed as a two stage act, where the first stage is subject to

¹⁸In some cases, this richness condition is in fact incompatible with \mathcal{P} being equal to the set of N -element partitions, for some N .

uncertainty. That ambiguity may arise from complexity in such an environment is not a new idea. Gilboa and Schmeidler (1989) states “One conceivable explanation of this phenomenon [Ellsberg-type preferences] which we adopt here is as follows: ... the subject has too little information to form a prior. Hence (s)he considers a set of priors as possible.” My paper is novel however in that it derives ambiguity averse preferences by explicitly characterizing subjective complexity. Bewley (2002) and Gilboa and Schmeidler (1989) relax completeness and independence respectively. I do both, but in a way that is driven by explicit assumptions about perceived complexity.

My applications relate to a number of papers studying bounded rationality and ambiguity aversion. There is a large literature on the simplicity of contracts observed in reality. Mukerji (1998) uses ambiguity aversion to explain contract incompleteness. Anderlini and Felli (1994) use a similar notion of contract complexity to my partition size definition. However they essentially impose that contracts must be simple in this sense, where as I show that such contracts are optimal from the principal’s perspective when facing a certain type of agent.

There is also a large body of experimental and observational evidence relating partition size to perceived complexity. Moffatt et al. (2015) and Sonsino et al. (2002) find experimental evidence that subjects undervalue lotteries with larger supports. Bernheim and Sprenger (2019) argue that aversion to large supports helps explain experimental data that is otherwise inconsistent with both expected utility and cumulative prospect theories.

As has been previously discussed, Cautious preferences display features of ambiguity aversion. One of the useful features of this model is that it explicitly separates the decision makers understanding of the environment from her attitude towards the unknown. This novel model of ambiguity aversion can explain some experimental findings at odds with existing models.

Chew et al. (2017) find that subjects are averse to increases in the number of possible compositions (the size of the state space in my framework) of a deck of cards on which bets are made. For example, consider an individual betting on red in a deck of red and black cards, where the number of red cards is known to be between n and $100 - n$. The authors find that subjects aversion to the ambiguous deck, measured by the difference between their certainty equivalents for an ambiguous act and the corresponding compounded lottery, is decreasing in n . The data contradict the predictions of the recursive expected utility model (Klibanoff et al. (2005), Seo (2009)), which predicts an aversion to increasing the number of possible compositions only in certain cases. Similarly, Viscusi and Magat (1992) document increasing ambiguity aversion as the range of uncertain outcomes increases.

A Omitted Proofs

A.1 Proposition 1

Proof. Suppose an act f , with partition $\{T_i\}_{i=1}^N$, has a certain equivalent. I wish to show that any other act g with the same partition also has a certain equivalent. Throughout the proof, label the partition so that $g(T_{i+1}) \succ g(T_i)$ for all i . For notational simplicity, I will identify each act f with its utility image $v \circ f$. Finally, assume that there exists a constant act \bar{c}, \underline{c} such that $\bar{c} \succ g(\omega) \succ \underline{c}$ for all ω (in the end we will establish existence of a certain equivalent when such a \bar{c}, \underline{c} do not exist). The proof will proceed by induction on N . The induction hypothesis for each $K < N$ is that all acts measurable with respect to a K -element coarsening of $\{T_i\}_{i=1}^N$ have certain equivalents. Note that, by the Basic Conditions, preferences on F_c have an expected utility representation.

Before proceeding to the induction proof, I show the following claim.

Claim 1: if $\lambda g + (1 - \lambda)c_1 \sim c_2$ for $c_1, c_2 \in F_c$ with $\bar{c} \succ c_1 \succ \underline{c}$ and $\lambda \in (0, 1)$, then g has a certain equivalent. The proof of Claim 1 is as follows. By Weak C-Independence, it suffices to show that there exists $c_3 \in F_c$ such that $\lambda c_3 + (1 - \lambda)c_1 \sim c_2$. If $c_1 \sim c_2$ then we are done. Suppose $c_1 \succ c_2$. By Monotonicity and $c_1 \succ \underline{c}$, $c_2 \succ \underline{c}$. Then there exists $\alpha \in (0, 1)$ such that $\alpha \underline{c} + (1 - \alpha)c_1 \sim c_2$. Let $c_3 = \frac{\alpha}{\lambda} \underline{c} + \frac{1-\alpha}{\lambda} c_1$. Since $g(\omega) \succ \underline{c}$, and given the expected utility representation on F_c , Monotonicity implies that $\lambda > \alpha$, so c_3 is well defined. If $c_2 \succ c_1$ replace \underline{c} with \bar{c} .

Now for the induction proof. I first show that all binary acts have certain equivalents. For any event E , let f be a bet on E that has a certain equivalent (which exists by A0), and g be another arbitrary bet on E . Let $E^c = \Omega \setminus E$. There are a few cases to consider. Suppose $f(E) \succ g(E) \succ g(E^c) \succ f(E^c)$. Then, using the expected utility representation on F_c , $\exists \lambda \in (0, 1)$ and $\hat{c} \in F_c$ such that $\lambda f(E^c) + (1 - \lambda)\hat{c} \sim g(E^c)$ and $\lambda f(E) + (1 - \lambda)\hat{c} \sim g(E)$ ($\lambda = (u(g(E)) - u(g(E^c)))/(u(f(E)) - u(f(E^c)))$). By Weak C-Independence $\lambda f + (1 - \lambda)\hat{c}$ is well-understood, and since this act is payoff equivalent to g , g is as well. Suppose instead that $g(E) \succ f(E) \succ f(E^c) \succ g(E^c)$. Then, as before, there exist λ, \hat{c} such that $\lambda g(E) + (1 - \lambda)\hat{c} \sim f(E)$ and $\lambda g(E^c) + (1 - \lambda)\hat{c} \sim f(E^c)$, so $\lambda g + (1 - \lambda)\hat{c}$ has a certain equivalent. Then g has a certain equivalent as well, by Weak C-Independence.

Suppose $g(E) \succ f(E) \succ g(E^c) \succ f(E^c)$. Then $\exists \lambda_1, \lambda_2 \in [0, 1)$ such that $\lambda_1 \bar{c} + (1 - \lambda_1)f(E) \sim g(E)$, and $\lambda_2 \bar{c} + (1 - \lambda_2)f(E^c) \sim g(E^c)$. Let $\lambda = \min\{\lambda_1, \lambda_2\}$, and suppose WLOG that this is equal to λ_1 . By Weak C-Independence, $f' := \lambda \bar{c} + (1 - \lambda)f$ has a certain equivalent. Note $f'(E) \sim g(E)$. By continuity $\exists \alpha \in [0, 1]$ such that $\alpha f'(E) + (1 - \alpha)f'(E^c) \sim g(E^c)$. Then $\alpha f'(E) + (1 - \alpha)f'$ is payoff equivalent to g and has a certain equivalent by Weak

C-Independence, so g has a certain equivalent. The remaining cases are analogous. Since by A0 every event has a well understood bet, we can conclude that any binary act has a certain equivalent. This is the first step in the induction.

Now suppose that f 's partition has $N > 2$ elements, and f has a certain equivalent. Then by A1 and the induction hypothesis, all acts g' measurable with respect to coarser partitions than f , and satisfying $\bar{c} \succ g'(\omega) \succ \underline{c}$ for all ω , have certain equivalents. I now show that it is without loss to assume that f is increasing with respect to the same order as g , i.e. $f(T_{i+1}) \succ f(T_i)$ for all i . To see this, it suffices to show that if $f(T_a) \succ f(T_b) \succ f(T_c)$ then there is an act f' with the same partition that has a certain equivalent and such that $f'(T_a) \succ f'(T_c) \succ f'(T_b)$ (a symmetric argument shows that there is f'' such that $f''(T_b) \succ f''(T_c) \succ f''(T_a)$). Define h by $h(T_a) = h(T_c) = f(T_a)$ and $h = f$ on $\Omega \setminus (T_a \cup T_c)$. Then h has a certain equivalent by the induction hypothesis. Moreover, by continuity there exists $\alpha \in (0, 1)$ such that $\alpha h(T_c) + (1 - \alpha)f(T_c) \succ f(T_b)$. Moreover, by A2, $f' = \alpha h + (1 - \alpha)f$ has a certain equivalent, as desired. So from now on, assume WLOG that f and g are comonotone.

By Claim 1, it is without loss to consider g such that $f(T_N) \succ g(T_N)$ and $g(T_1) \succ f(T_1)$. Then there exists $\lambda \in (0, 1)$ such that $\lambda f(T_N) + (1 - \lambda)f(T_1) \sim g(T_N)$. Define h_N by $h_N = f(T_1)$ on T_N and $h_N = f$ elsewhere. Then h_N has a certain equivalent by the induction hypothesis. Let $f'_N = \lambda f + (1 - \lambda)h_N$. Then f'_N has a certain equivalent by A2. Moreover, by Monotonicity, f_N defined as $f_N = g(T_N)$ on T_N and $f_N = f'_N$ elsewhere also has a certain equivalent. Then there also exists $c_{N-1} \in \{f_N(T_N), f_N(T_1)\}$ and $\lambda' \in (0, 1)$ such that $\lambda' f_N(T_{N-1}) + (1 - \lambda')c_{N-1} \sim g(T_{N-1})$. Define h_{N-1} by $h_{N-1} = c_{N-1}$ on T_{N-1} and $h_{N-1} = f_N$ elsewhere. Then by the induction hypothesis h_{N-1} has a certain equivalent, and by A2 $f'_{N-1} = \lambda' f_N + (1 - \lambda')h_{N-1}$ also has a certain equivalent. As before, define f_{N-1} by $f_{N-1} = g(T_{N-1})$ on T_{N-1} and $f_{N-1} = f'_{N-1}$ elsewhere. Then f_{N-1} also has a certain equivalent. Proceeding in this way, we arrive at an act $f_1 = g$ which has a certain equivalent, as desired.

I now need to address the assumption that there exist constant acts \underline{c}, \bar{c} such that $\bar{c} \succ g(\omega) \succ \underline{c}$ for all ω . Suppose that we have established existence of a certain equivalent for all acts f , measurable with respect to a given partition, that satisfy this interiority assumption. Suppose g is such that $g(T_N) \succ c \succ g(T_1) \forall c \in F_c$ (recall the ordering of T_i). Let $g' = \frac{1}{2}g + \frac{1}{2}(\frac{1}{2}g(T_N) + \frac{1}{2}g(T_1))$. Then g' satisfies the interiority assumption, and so has a certain equivalent by hypothesis, which can be written as $\lambda g(T_1) + (1 - \lambda)g(T_N)$, where $\lambda \in [\frac{1}{4}, \frac{3}{4}]$ by Monotonicity. The existence of a certain equivalent for g will follow from Weak C-Independence if we can show that there exists a $\kappa \in [0, 1]$ such that $\frac{1}{2}(\kappa g(T_N) + (1 - \kappa)g(T_1)) + \frac{1}{2}(\frac{1}{2}g(T_N) + \frac{1}{2}g(T_1)) \sim \lambda g(T_1) + (1 - \lambda)g(T_N)$. Equating coefficients, this holds

for $\kappa = 2\lambda - \frac{1}{2}$, which is well defined since $\lambda \in [\frac{1}{4}, \frac{3}{4}]$. □

A.2 Theorem 1

Theorem 1 is immediate from Lemma 2, A3, A4, and Proposition 1.

Lemma 2. *For any two N -element partitions $\tau = \{T_j\}_{j=1}^N$, $\tau' = \{T'_j\}_{j=1}^N$ of Ω , there is a finite sequence of N -element partitions $\{\tau^i\}_{i=1}^K$, starting with τ and ending with τ' , such that between τ^i and τ^{i+1} , $N - 2$ of the partition cells remain unchanged.*

Proof. Let $\tau^i = \{T_j^i\}_{j=1}^N$ be the last partition in the sequence constructed thus far (we begin with $\tau^1 = \tau$). Assume $\tau^i \neq \tau'$, otherwise we are done. Then there exists $T' \in \tau'$ such that $T' \cap T_1^i$ and $T' \cap T_2^i$ are both non-empty for some $T_1^i, T_2^i \in \tau^i$. There are two cases to consider. In *Case 1* we can choose T', T_1^i, T_2^i such that $T_1^i \cup T_2^i \neq T'$; in *Case 2* we cannot.

Consider first Case 1. To generate the next partition in the sequence, fix all elements of τ^i other than T_1^i, T_2^i . Choose some sets $r, r' \in \tau^i \vee \tau'$ with $r \subseteq T_1^i, r' \subseteq T_2^i$ such that $r \cup r' \subseteq T'$.¹⁹ Such sets exist since $T' \cap T_1^i$ and $T' \cap T_2^i$ are both non-empty by assumption. If $r \neq T_1^i$ then define τ^{i+1} to be the modification of τ^i in which r is merged with T_2^i , and all other elements are the same; to be precise, $\tau^{i+1} := \{T_1^i \setminus r, T_2^i \cup r, T_3^i, \dots, T_N^i\}$. If $r = T_1^i$ then define $\tau^{i+1} := \{r \cup r', (T_1^i \cup T_2^i) \setminus (r \cup r'), T_3^i, \dots, T_N^i\}$.

Consider now Case 2. We have $T_1^i \cup T_2^i = T'$. Then there exist $T_3^i \in \tau^i$ and $T'_3 \in \tau'$ such that $T'_3 \subset T_3^i$. Then define $\tau^{i+1} := \{T_1^i, T_2^i \cup T'_3, T_3^i \setminus T'_3, T_4^i, \dots, T_N^i\}$ and $\tau^{i+2} := \{T_1^i \cup T_2^i, T'_3, T_3^i \setminus T'_3, T_4^i, \dots, T_N^i\} = \{T', T'_3, T_3^i \setminus T'_3, T_4^i, \dots, T_N^i\}$.

As long as $\tau^i \neq \tau'$ the algorithm above delivers $\tau^{i+1} \neq \tau^i$ in Case 1, or $\tau^{i+2} \neq \tau^i$ in Case 2. Moreover, τ^k is a coarsening of $\tau \vee \tau'$ for all k . Since there are at most N^2 elements of $\tau \vee \tau'$, the set of possible coarsenings is finite. Thus the algorithm eventually delivers τ' as long as there are no cycles. But cycles cannot occur. To see this, first notice that if $T \in \tau^i$ and $T \in \tau'$ then $T \in \tau^{i+k}$ for all $k \geq 1$. Therefore the algorithm can only arrive at Case 2 finitely many times along any sequence, since each time it does so it delivers τ^{i+2} which has an additional 2 cells in common with τ' . Between the steps at which Case 2 is reached there can be no cycles; once r and r' are merged into the same cell in a Case 1 step they are never divided. □

A.3 Proof of Theorem 2

By Theorem 1, we know that $F_N \subseteq F_{CE}$ for some $N \geq 2$. The first part of the proof shows that preferences on F_N have an expected utility representation under the Basic Conditions

¹⁹ $\tau \vee \tau'$ is the coarsest common refinement.

and S-Independence. Then Uniform Comparability implies the desired representation.

Part 1. The restriction of \succsim on L satisfies the von Neumann-Morgenstern axioms, and so is represented by $v(l) = E_l[u]$. Without loss of generality assume that $v(L) \supset [-1, 1]$. Fix a partition $\tau = \{T_i\}_{i=1}^N$. The restriction of \succsim to acts measurable with respect to τ satisfies the standard SEU axioms (see for example Gilboa and Schmeidler (1989)). Therefore there exists a probability measure P^τ on τ such that for all τ -measurable acts f, g , $f \succsim g$ iff $E_{P^\tau}[v \circ f] \geq E_{P^\tau}[v \circ g]$.

It remains to show that there exists a probability P on Σ such that $P^\tau(A) = P(A)$ for all N -element partitions τ and $A \in \Sigma$. For any $A \in \Sigma$ and any $\tau', \tau'' \in F_N$ such that $A \in \tau', A \in \tau''$, it must be that $P^{\tau'} = P^{\tau''}$, since the certainty equivalent for the act $\mathbb{1}_A$ must be the same regardless of which of $P^{\tau'}, P^{\tau''}$ is used to represent preferences. Thus $P(A) = \{P^\tau(A) : \tau \in F_N\}$ is a well defined function. P is non-negative since each P^τ is. If $N > 2$ then it follows immediately that P is additive, since for any disjoint $A, B \in \Sigma$ there exists $\tau \in F_N$ such that $A, B \in \tau$.

When $N = 2$, additivity of P is implied by S-Independence. To see this, it is sufficient to consider 2-element partitions and $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Let $\tau^1 = \{\{\omega_1\}, \{\omega_2, \omega_3\}\}$, $\tau^2 = \{\{\omega_1, \omega_3\}, \{\omega_2\}\}$, and $\tau^3 = \{\{\omega_1, \omega_2\}, \{\omega_3\}\}$. Let $\mathbb{1}_A$ be an act such that $v \circ f(\omega) = 1$ for $\omega \in A$, and 0 otherwise. Then $P^{\tau^i}(\omega_i) = E_{P^{\tau^i}}[\mathbb{1}_{\{\omega_i\}}]$ for $i \in \{1, 2, 3\}$. We wish to show that P defined as $P(\omega_i) = P^{\tau^i}(\omega_i)$ for $i \in \{1, 2, 3\}$ is a well-defined probability on Ω that is consistent with P^{τ^1}, P^{τ^2} , and P^{τ^3} . To see this, let $f_1, f_3 \in F_2$ be acts such that $v \circ f_1(\{\omega_2, \omega_3\}) = 1$ and $v \circ f_1(\omega_1) = -1$; and $v \circ f_3(\{\omega_1, \omega_2\}) = 1$ and $v \circ f_3(\omega_3) = -1$. Then $\frac{1}{2}f_1 + \frac{1}{2}f_3 = \mathbb{1}_{\omega_2}$, so $E_{P^{\tau^2}}[\frac{1}{2}f_1 + \frac{1}{2}f_3] = P^{\tau^2}(\omega_2)$. But by S-Independence we also know that $\frac{1}{2}f_1 + \frac{1}{2}f_3 \sim \frac{1}{2}U(f_1) + \frac{1}{2}U(f_3)$, where $U(f_i)$ is a constant act such that $v(U(f_i)) = E_{P^{\tau^i}}[f_i]$. Then, we also have, with the usual abuse of notation, that $\frac{1}{2}f_1 + \frac{1}{2}f_3 \sim \frac{1}{2}(P^{\tau^1}(\{\omega_2, \omega_3\}) - P^{\tau^1}(\omega_1) + P^{\tau^3}(\{\omega_1, \omega_2\}) - P^{\tau^3}(\omega_3))$. Thus $\frac{1}{2}(P^{\tau^1}(\{\omega_2, \omega_3\}) - P^{\tau^1}(\omega_1) + P^{\tau^3}(\{\omega_1, \omega_2\}) - P^{\tau^3}(\omega_3)) = P^{\tau^2}(\omega_2)$. Using this equality it is easy to see that P is well defined and consistent; for example $P(\omega_2) = P^{\tau^2}(\omega_2) = 1 - P^{\tau^1}(\omega_1) - P^{\tau^3}(\omega_3) = 1 - P(\{\omega_1, \omega_3\})$.

By exactly the same argument, we can show that P as defined above is consistent with $P^{\tau'}$ and $P^{\tau''}$ for any 2-element partitions τ'', τ' . This gives the desired representation of \succsim for N -simple acts.

Part 2. That Uniform Comparability implies the representation in Theorem 1 is immediate; when $\text{siminf}_{N,f}$ and $\text{simsup}_{N,g}$ both exist with $\text{siminf}_{N,f} \succsim \text{simsup}_{N,g}$, they play the role of h and k in the axiom respectively.

I now show that $\text{siminf}_{N,f}$ must be non-empty for all f . The proof for $\text{simsup}_{N,f}$ is analogous. Suppose $\text{siminf}_{N,f}$ is empty. Since f is bounded, there exist constant acts $\bar{f}, \underline{f} \in F_c$

such that $\bar{f} \succsim f \succsim \underline{f}$. Let

$$s = \sup_{h \in \{h \in F_N : f \geq^0 h\}} \int_{\Omega} E_h u \, dP(\omega)$$

By Uniform Comparability, there exists an $h \in \{h \in F_N : f \geq^0 h\}$ such that $h \succsim \underline{f}$. Choose λ^* such that $\lambda^* \int_{\Omega} E_{\bar{f}} u \, dP(\omega) + (1 - \lambda^*) \int_{\Omega} E_{\underline{f}} u \, dP(\omega) = s$. Since, by hypothesis, there are no N -simple acts h with $f \geq^0 h$ that have expected utility equal to s , Uniform Comparability implies $\neg(f \succsim \lambda^* \bar{f} + (1 - \lambda^*) \underline{f})$. But this violates Archimedian Continuity.

A.4 Proof of Theorem 4

Take $f \notin F_{CE}$. I first show that $f \sim' \text{siminf}_{N,f}$. Suppose $\exists c \in F_c$ such that $c \succ f$ (otherwise the claim is trivial). By consistency $f \succsim' \text{siminf}_{N,f}$. Suppose $\text{siminf}_{N,f} \not\prec' f$. By continuity there exist constant acts $c_1 \succ c_2$ and $\lambda^* \in [0, 1)$ such that $\lambda^* c_1 + (1 - \lambda^*) c_2 \sim \text{siminf}_{N,f}$. Then for all $\lambda > \lambda^*$ it follows from Theorem 2 that $f \not\prec \lambda c_1 + (1 - \lambda) c_2$. Caution then implies that $\lambda c_1 + (1 - \lambda) c_2 \succsim' f$. By continuity of \succsim' , $\lambda^* c_1 + (1 - \lambda^*) c_2 \succsim' f$. The claim follows. Strong consistency for simple acts implies that the restrictions of \succsim and \succsim' to F_{CE} have the same representation. The theorem follows.

A.5 Existence of simsup and siminf

A.5.1 “Lebesgue Approach”

Let $\tau(\Omega)$ be the set of all partitions of Ω , and $\tau^N(\Omega)$ the set of N -element partitions. When Ω is a partially ordered set, say that $\tau = \{T_i\}_{i=1}^N$ is an *interval partition* if T_i is an interval for all i .²⁰

Fix $w \in B(\Omega)$. For any $h \in B_N(\Omega)$ and let $\tau_h = \{T_i\}_{i=1}^N$ be h 's partition. Define $T'_i := \{r \in w(\Omega) : w(\omega) = r \text{ for some } \omega \in T_i\}$. Define $\tau'_h := \{T'_i\}_{i=1}^N$ as the cover of $w(\Omega)$ induced by h . Say that h induces an interval partition of $w(\Omega)$ if τ'_h is a partition of $w(\Omega)$ and T'_i is an interval for all i .

Let Q be the law of w , defined by $Q(A) = P(w^{-1}(A))$ for any Borel set A . Let S be Q 's support. Since w is measurable, Q is a Borel measure on \mathbb{R} , and hence a Radon measure. Thus $Q(A) = 0$ for any $A \in \Omega \setminus S$ (see Parthasarathy (2005), ch 2). The idea behind the proof is to look at N -simple functions on S , rather than on Ω . To do this, I need to show that it

²⁰By an interval in a partially ordered space (J, \geq) I mean a set $I \subseteq J$ such that for all $x, y \in I$ and all $z \in J$ such that $x \geq z \geq y$, $z \in I$. I do not define an interval to be closed, as is sometimes done.

is possible to move between $B_N(\Omega)$ and $B_N(S)$. I will focus on the existence of $\text{siminf}_{N,w,P}$, as the argument for simsup is exactly symmetric.

The following lemma shows that $\text{siminf}_{N,w,P}$, if it exists, will live in the subset of $B_N(\Omega)$ that induce an interval partitions of S . The lemma has a symmetric counterpart for $\text{simsup}_{N,w,P}$.

Lemma 3. *For any $w \in B(\Omega)$, let $h \in B_N(\Omega)$ be such that $w \geq^0 h$. If h does not induce an interval partition of S then there exists an N -simple function \hat{h} that does, and such that $w \geq^0 \hat{h}$ and $E_P[\hat{h}] \geq E_P[h]$.*

Proof. Let $A \subseteq \Omega$ be the set of states ω such that $h(\omega) > w(\omega)$. We can restrict attention to functions h such that $P(A) = 0$, as this is a requirement for $\text{siminf}_{N,w,P}$. To begin, assume that $A \cap w^{-1}(S) = \emptyset$, so $w \geq_{w^{-1}(S)} h$ (i.e. $w(\omega) \geq h(\omega)$ for all ω such that $w(\omega) \in S$).

Suppose that τ'_h is not a partition of S . For any $r \in S$ and any $T_i, T_j \in \tau_h$ such that $T_i \cap w^{-1}(r) \neq \emptyset$ and $T_j \cap w^{-1}(r) \neq \emptyset$, define \hat{h} as $\hat{h}(\omega) = \max\{h(T_i), h(T_j)\} \forall \omega \in w^{-1}(r)$, and $\hat{h} = h$ elsewhere. Then $E_P[\hat{h}] \geq E_P[h]$, and τ'_h will be a partition of $w(\Omega)$. For any such r , $w \geq_{w^{-1}(S)} h$ implies $r \geq \max\{h(T_i), h(T_j)\}$. Therefore $w \geq_{w^{-1}(S)} \hat{h}$, and since $P(A) = 0$, $w \geq^0 \hat{h}$.

If there are elements of τ'_h that are not intervals then there exist states $\omega_1, \omega_2 \in \Omega$ with $w(\omega_1) < w(\omega_2)$ and $h(\omega_1) > h(\omega_2)$. Then define \hat{h} such that $\hat{h}(\omega_2) = \hat{h}(\omega_1) = h(\omega_1)$, and $h = \hat{h}$ elsewhere. Clearly $E_P[\hat{h}] \geq E_P[h]$ and $w \geq^0 \hat{h}$.

Now, I want to show that it is without loss to assume $A \cap w^{-1}(S) = \emptyset$. For any $\omega \in A$, if there exists $r \in h(\Omega \setminus A)$ such that $w(\omega) \geq r$ then we can replace $h(\omega)$ with r without altering the value of h , or the fact that it is N -simple. Assume therefore that $w(\omega) < \min h(\Omega \setminus A)$ for all $\omega \in A$. Since $w \geq_{\Omega \setminus A} h$, this implies that $w(\omega) < \inf w(\Omega \setminus A)$ for all $\omega \in A$. But then $w(A) \cap S = \emptyset$, or equivalently $A \cap w^{-1}(S) = \emptyset$. \square

Corollary 2. *When $\text{siminf}_{N,w,P}$ and $\text{simsup}_{N,w,P}$ are non-empty, they contain functions that induce interval partitions of S .*

For any function $\tilde{h} \in B_N(w(\Omega))$ we can define a function $h \in B_N(\Omega)$ by $h(\omega) = \tilde{h}(w(\omega))$. Moreover, $E_Q[\tilde{h}] = E_P[h]$ by the definition of Q . Lemma 3 and Corollary 2 are useful because they allow us to do the converse: given a function $h \in B_N(\Omega)$ that induces a partition of $w(\Omega)$ we can define a function $\tilde{h} \in B(w(\Omega))$ by $\tilde{h}(r) = h(w^{-1}(r))$. By $h(w^{-1}(r))$ I mean the value taken by h for all $\omega \in w^{-1}(r)$. For \tilde{h} to be well defined it is therefore necessary that h induce a partition of $w(\Omega)$. Interval partitions are easy to work with, as we will see, since they can be described by a vector of N cut-offs.

A.5.2 Existence proof

Proposition 16. *Suppose Ω is a closed interval in \mathbb{R} , P has full support, and $w \in B(\Omega)$ is continuous, and increasing. Then $\text{siminf}_{N,w,P}$ is non-empty.*

Proof. Since P has full support and w is continuous, it is without loss to assume that $w \geq \text{siminf}_{N,w,P}$. To find $\text{siminf}_{N,w,P}$, it will be sufficient to restrict attention to increasing functions with interval partitions. This follows since for any non-interval partition T there exist states $\omega_1, \omega_2 \in \Omega$ satisfying $w(\omega_1) \leq w(\omega_2)$ and $h_T(\omega_1) > h_T(\omega_2)$. Then adding ω_2 to the partition cell that contains ω_1 leads to a new N -simple function h' with $h' \leq w$ and $E_P[h'] \geq E_P[h_T]$. When working with interval partitions label the cells so that $T_i > T_{i-1}$ (in the obvious order). Moreover, since w is increasing we can assume that each interval in τ contains its left endpoint (since it is always optimal to assign a state to the higher interval). Any such partition can be fully described by a vector $z_\tau = (t_1, \dots, t_{N-1})$ of $N - 1$ cutoffs. Given a cutoff vector z , let h_z be the function constructed as above using the partition described by z . Since the set of such cutoff vectors is compact, it only remains to show that $z \mapsto E_P[h_z]$ is upper semi-continuous. This follows from right-continuity of the CDF and continuity of w . \square

Proof. (Theorem 3). By Corollary 2, it is without loss to look $\text{siminf}_{N,w,P}$ in the subset of $B_N(\Omega)$ that induce interval partitions of S . It is therefore without loss to look for functions $\tilde{h} \in B_N(S)$ that have interval partitions, i.e. we solve $\max_{\tilde{h} \in B_N(S): w \geq_S \tilde{h}} E_Q[\tilde{h}]$. This problem has a solution by Proposition 16. Let \tilde{h} be the solution, and define $h = \tilde{h} \circ w$ on S . Let $h = c$ on $\Omega \setminus S$, for some $c \in h(S)$. Then h will be N -simple and satisfy $w \geq^0 h$. Moreover, Corollary 2 implies that $h \in \text{siminf}_{N,w,P}$. \square

B Online Appendix

B.1 Properties of Caution

Cautious preferences can also be understood through the violations of Independence that they exhibit. A decision maker with Cautious preferences may violate independence when the mixture of two acts leads to an act that is harder to approximate from below by simple acts. In particular, the mixture of two N -simple acts between which the decision maker is indifferent will in general not be N -simple, and may thus be considered inferior to the original acts. Cautious preferences are ambiguity averse in the sense of Ghirardato and Marinacci

(2002).²¹ In particular, for fixed u and P , decision makers with a higher capacity are less ambiguity averse than those with lower capacity.

Given the discussion of Bewley (2002) above and the formal parallel between this paper and GMMS, it is natural to consider the relationship between Cautious preferences and Gilboa and Schmeidler (1989) MEU preferences. Preferences have an MEU representation if there exists a convex set C of probabilities on Σ and a non-constant function $u : Z \mapsto \mathbb{R}$ such that f is preferred to g if and only if

$$\min_{p \in C} \int_{\Omega} E_{f(\omega)} dp(\omega) \geq \min_{p \in C} \int_{\Omega} E_{g(\omega)} dp(\omega).$$

As GMMS show, such preferences are derived from Bewley preferences through the Consistency and Caution axioms, just as Cautious preferences are derived from Simple Bounds. These representations do not coincide however. This can be seen most easily by considering the Uncertainty Aversion axiom of Gilboa and Schmeidler (1989).

UNCERTAINTY AVERSION: For every $f, g \in F$, if $f \sim g$ then $(1/2)f + (1/2)g \succsim g$.

Uncertainty Aversion captures the notion that mixing acts smooths payoffs, and thus reduces exposure to uncertainty. It is easy to see that this axiom is violated by Cautious preferences. As noted above, mixing any two N -simple acts between which the decision maker is indifferent leads to an act which is no better than either of the original acts. This is because the mixed act may not be N -simple, and will thus be approximated from below, whereas the original acts were perfectly understood. More generally, whenever the two acts considered are comonotonic, Cautious preferences will satisfy the reverse Uncertainty Aversion: if $f \sim g$ then $g \succsim (1/2)f + (1/2)g$. Two acts f, g are said to be comonotonic if there is no $\omega, \omega' \in \Omega$ such that $f(\omega) \succ f(\omega')$ and $g(\omega') \succ g(\omega)$.²² The preference against the mixture will be strict unless there are elements of $\text{siminf}_{N,f}$ and $\text{siminf}_{N,g}$ which share the same partition. The intuition for this reversal is that mixing between comonotonic acts does not smooth payoffs across states in the same way that mixing between non-comonotonic acts can. On

²¹Simple Bounds preferences are also ambiguity averse, if we extend the notion to incomplete preferences.

²²This can be seen as follows. For a set $I \subseteq \Omega$ and act f define $\underline{f}(I) = \inf\{u \circ f(\omega) : \omega \in I\}$. Then the value of $\text{siminf}_{N,f}$ is given by

$$\max_{\{I_i\}_{i=1}^N \in T^N(\Omega)} \sum_{i=1}^N P(I_i) \underline{f}(I_i).$$

When f and g are comonotonic $\underline{f\alpha g}(I) = \alpha \underline{f}(I) + (1 - \alpha) \underline{g}(I)$ for all I , where $f\alpha g = \alpha f + (1 - \alpha)g$. The claim follows.

the other hand, since the two acts may have had a lot of variations on different regions of the state space, the mixed act will be more difficult to approximate from below than any of the two acts individually. This contrast illustrates that MEU and Cautious preferences capture very different notions of aversion to uncertainty.

Uncertainty Aversion is satisfied by Cautious preferences whenever the mixture of f and g is N -simple, and indeed the inequality may be strict. A modification of Uncertainty Aversion along these lines is studied by Hartmann and Kauffeldt (2019), who state the following axiom

N-AMBIGUITY AVERSION: $f_1, \dots, f_n \in F$, $\alpha_1, \dots, \alpha_n \geq 0$, $\sum_{i=1}^n \alpha_i = 1$, $\sum_{i=1}^n \alpha_i f_i = f \in F_N$ such that $f_1 \sim f_2 \cdots \sim f_n$ implies $f \succsim f_1$.

The preferences studied by Hartmann and Kauffeldt (2019) differ from Cautious preferences, primarily in that they satisfy the Comonotonic Independence axiom of Schmeidler (1989).

COMONOTONIC INDEPENDENCE: For all pairwise comonotonic acts f, g, h , and all $\alpha \in (0, 1)$, $f \succ g$ implies $\alpha f + (1 - \alpha)h \succ \alpha g + (1 - \alpha)h$.

Schmeidler’s motivation for Comonotonic Independence, as opposed to the usual Independence axiom, is similar to my motivation of S -Independence. Schmeidler (1989) notes that arbitrary mixtures may lead to acts which define a much finer (larger) algebra than the original acts, and thus violations of Independence may occur. However, if “ f, g and h are pairwise comonotonic, then the comparison of f to g is not very different from the comparison of $[\alpha f + (1 - \alpha)h]$ to $[\alpha g + (1 - \alpha)h]$ ”. In some circumstances however, comonotonicity may be too weak a notion to guarantee that mixtures do not alter the the acts under consideration in ways that lead to violations of independence. Consider the following simple example. Let the state space be the interval $[0, 1]$ and identify $u(f(\omega))$ with $f(\omega)$. Let $f = 2$ on $[0, 1/2)$ and $f = 4$ on $[1/2, 1]$. Let $g = 1$ on $[0, 2/3)$ and $g = 6$ on $[2/3, 1]$. Suppose $f \succ g$. The mixture $1/2f + 1/2g$ has both a greater spread between its highest and lowest outcomes and a larger partition than f . Both these changes may lead the DM to favor g . The Simplicity Conditions, along with S -Independence, can be seen as making precise the notion of “not very different”. Cautious preferences capture aversion to uncertainty in the sense that the decision maker, faced with an act that she does not fully understand, assigns to the act the minimum value consistent with monotonicity and her preferences over the acts that she understands well. Uncertainty here stems directly from the complexity of the acts under consideration. However, cautious preferences will always be averse to comonotonic mixtures.

COMONOTONIC MIXTURE AVERSION: For all comonotonic acts f, g with $f \sim g$ and $\alpha \in (0, 1)$, $f \succsim \alpha f + (1 - \alpha)g$.

In a related paper, I characterize a broad class of preferences satisfying Comonotonic Mixture Aversion. I show that within this class, cautious preferences are simultaneously *i*) maximally ambiguity averse, in the sense of Ghirardato and Marinacci (2002), *ii*) maximally N -Ambiguity averse, and *iii*) maximally averse to mixtures of N -simple acts.²³

B.2 MEU in insurance

The insurance plans considered here all induce payoffs that are monotone in the the state. If the set C of beliefs over which the DM with MEU preferences minimizes contains a first-order stochastically dominant belief then this will be the minimizing belief regardless of plan characteristics. In this sense the individual will behave exactly like an expected utility maximizer.

In some cases MEU and cautious preferences predict similar behavior. For example, in the MEU framework we can consider individuals who minimize over smaller sets of beliefs, which is analogous to the idea of increasing capacity. Consider comparing the relative value of two plans which induce ex-post wealth functions f_1, f_2 for two individuals, one of whom has a larger set of beliefs. If the larger set of beliefs contains an upper bound in the first-order stochastic dominance (FOSD) order, and the payoff difference $f_1 - f_2$ is increasing (decreasing) in the state then the individual with the larger belief set will value f_1 more (less) relative to f_2 than the individual with the smaller set. For example, the individual with a larger set of beliefs will value reductions in the out-of-pocket maximum more than an individual with a smaller set of beliefs.

B.3 Proof of Proposition 5

Proof. I give the proof for siminf ; simsup is symmetric. If $f^{-1}(A)$ is non-null for every open neighborhood of $\inf E_f u(\Omega)$ then there must be some ω such that $\inf E_f u(\Omega) \geq E_{\text{siminf}_{N,f}} u(\omega)$. If there was a violation of statewise dominance for some ω' then it could be removed by specifying $E_{\text{siminf}_{N,f}} u(\omega') = E_{\text{siminf}_{N,f}} u(\omega)$. If the condition doesn't hold then $\exists \omega$ such that it is strictly sub-optimal to have $E_f u(\omega) \geq E_{\text{siminf}_{N,f}} u(\omega)$. \square

²³No draft of this paper is yet available, but I am happy to provide details upon request.

B.4 Proof of Proposition 2

Proof. For any act g with partition τ and any set A define the random variable $\hat{E}_\chi[g|A]$ as follows:

$$\hat{E}_\chi[g|A] = \begin{cases} \frac{1}{\hat{P}_\chi(A)} \sum_{i=1}^N f(T_i) \frac{1}{K} \sum_{x \in \chi} \mathbb{1}\{x \in T_i \cap A\} & \text{if } \hat{P}_\chi(A) > 0 \\ \mathbb{E}[g|A] & \text{otherwise} \end{cases}$$

This definition is necessary since we will be dealing with finite samples, so $\hat{P}_\chi(A) = 0$ with positive probability even when $P(A) > 0$.

Let f_b be the act equal to b on T' and f elsewhere. In what follows $T'^c = \Omega - T'$. Notice that $\varepsilon_f(\chi)$ can be written as

$$\begin{aligned} \varepsilon_f(\chi) &= \hat{E}_\chi[f|T'](1 - \hat{P}_\chi(T'^c)) + \hat{E}_\chi[f|T'^c]\hat{P}_\chi(T'^c) - \mathbb{E}[f] \\ &= \hat{E}_\chi[f|T'] - \mathbb{E}[f|T'] + \left(\hat{E}_\chi[f|T'^c] - \hat{E}_\chi[f|T'] \right) \hat{P}_\chi(T'^c) \\ &\quad - (\mathbb{E}[f|T'^c] - \mathbb{E}[f|T']) P(T'^c) \end{aligned}$$

Similarly

$$\varepsilon_{f_b}(\chi) = \left(\hat{E}_\chi[f|T'^c] - b \right) \hat{P}_\chi(T'^c) - (\mathbb{E}[f|T'^c] - b) P(T'^c).$$

Thus

$$\begin{aligned} \varepsilon_f(\chi) &= \\ \varepsilon_{f_b}(\chi) &+ \underbrace{\hat{E}_\chi[f|T'] - \mathbb{E}[f|T'] + \left(b - \hat{E}_\chi[f|T'] \right) \hat{P}_\chi(T'^c) + (\mathbb{E}[f|T'] - b) P(T'^c)}_{\xi(\chi)} \end{aligned}$$

The weak inequality in the definition of SOSD will follow by Jensen's inequality if I can show that $\mathbb{E}[\xi | \{\chi : \varepsilon_{f_b}(\chi) = m\}] = 0$ for all m in the range of ε_{f_b} , where the expectation is taken with respect to the measure on datasets induced by P and the i.i.d. sampling procedure. The strict inequality will follow since the distribution of ξ is non-degenerate.

Notice that, because sampling is i.i.d., $\mathbb{E} \left[\hat{E}_\chi[f|T'] \mid \hat{E}_\chi[f|T'^c], \hat{P}_\chi(T'^c) \right] = \mathbb{E}[f|T']$, where we use here the specification of $\hat{E}_\chi[f|T'] = \mathbb{E}[f|T']$ when $\hat{P}_\chi(T'^c) = 1$. Independence between $\hat{E}_\chi[f|T']$, $\hat{E}_\chi[f|T'^c]$, and $\hat{P}_\chi(T'^c)$ does not hold for finite samples due to integer restrictions. Fortunately all that we need is the stated conditional mean independence condition. Given this condition

$$\begin{aligned} \mathbb{E}[\xi | \varepsilon_{f_b}] &= \mathbb{E} \left[\mathbb{E} \left[\xi | \hat{P}_\chi(T'^c) \right] | \varepsilon_{f_b} \right] \\ &= \mathbb{E} \left[(b - \mathbb{E}[f|T']) (\hat{P}_\chi(T'^c) - P(T'^c)) | \varepsilon_{f_b} \right] \end{aligned}$$

where the first equality is just the law of iterated expectations and the second follows from conditional mean independence. Clearly if $b = E[f|T']$ the expectation is zero, so we are done. \square

B.5 Proof of Proposition 4

Proof. A1 has already been shown. A2 follows immediately from Jensen's inequality. A3 follows from the fact that ϕ is strictly concave and $\hat{G}_{\bar{f}} >_{SOSD} \hat{G}_f$; for any 2-element partition τ'' of $T_1 \cup T_2$ we can choose a binary act b on τ'' that is arbitrarily close to a constant and such that $\hat{E}_{\chi}[b|\tau''] = \hat{E}_{\chi}[f|\tau'']$. So b can be chosen such that the act f'' that is equal to b on τ'' and f elsewhere will satisfy $E_{\hat{G}_{f''}}[\phi(\hat{E}[f''])] > E_{\hat{G}_f}[\phi(\hat{E}[f])]$. A0 is satisfied because for any event A we can choose a bet on A that is arbitrarily close to constant. A4 holds since null events are those for which there is not data, and when using the bootstrap the value on such sets is irrelevant. \square

B.6 Proofs for the applications

B.7 Proof of Proposition 6

Proof. Proposition 6(i). In what follows N and x , the level of savings, will be fixed. Let P^α be the perceived distribution (induced by the cutoffs in (7.2)) when the proportion of savings allocated to the risky asset is α . For a fixed level of savings x the allocation problem becomes.

$$\max_{\alpha \in [0,1]} \int u((1-\alpha)xR_b + \alpha xr) dP^\alpha(r). \quad (\text{B.1})$$

Notice that for the unconstrained DM, for whom P^α is replaced with P , the objective is concave in α . This means that the derivative of the objective crosses zero (at most once) from above. The result will follow if I can show that at any α the derivative of the objective in (B.1) with respect to α is greater for the unconstrained DM than for the constrained one. Recall that P^α is determined by a maximization problem, which satisfies the conditions of the envelope theorem. Then, by the envelope theorem, the derivative with respect to α of the objective function in (B.1) is given by

$$\int u'((1-\alpha)xR_b + \alpha xr)(xr - xR_b) dP^\alpha(r).$$

Recall that $P \succsim_{FOSD} P^\alpha$ for all α . It is therefore sufficient to show that $u'((1-\alpha)xR_b + \alpha xr)(xr - xR_b)$ is increasing in r . To do this I show that for any $b, s \geq 0$

$$\frac{d}{dr}[u'(R_b b + rs)R_b] \leq \frac{d}{dr}[u'(R_b b + rs)r],$$

or equivalently

$$(r - R_b)s \leq -\frac{u'(R_b b + rs)}{u''(R_b b + rs)} \quad (\text{B.2})$$

Notice that by assumption $R_b \in (\underline{R}_s, \bar{R}_s)$, so that the left hand side of (B.2) is negative for r low enough, while the the right hand side is always strictly positive. So it is sufficient to show that the derivative with respect to r of the LHS of (B.2) is less than that of the RHS for all r , i.e.

$$\frac{d}{dr}[(r - R_b)s] \leq \frac{d}{dr} \left[-\frac{u'(R_b b + rs)}{u''(R_b b + rs)} \right] \iff \frac{u'''(R_b b + rs)u'(R_b b + rs)}{u''(R_b b + rs)^2} \geq 2$$

This holds for all r , for example, for CRRA utility $u(x) = x^{1-\gamma}/(1-\gamma)$ when $\gamma \geq 1$. \square

Proof. Proposition 6(ii) To prove the proposition define

$$V(x, N) = u(w - x) + \beta \max_{\alpha} \int u((1 - \alpha)xR_B + \alpha xr) dP^{\alpha, x}(r).$$

and

$$V(x, \infty) = u(w - x) + \beta \max_{\alpha} \int u((1 - \alpha)xR_B + \alpha xr) dP(r).$$

I will show that V has increasing differences, in the sense that for $x'' > x'$

$$V(x', \infty) - V(x'', \infty) \geq V(x', N) - V(x'', N),$$

A sufficient condition for the above inequality is that for all $x \in [x', x'']$

$$\begin{aligned} & \frac{d}{dx} \left[\max_{\alpha} \int u((1 - \alpha)xR_B + \alpha xr) dP^{\alpha, x}(r) \right] \\ & \geq \frac{d}{dx} \left[\max_{\alpha} \int u((1 - \alpha)xR_B + \alpha xr) dP(r) \right] \end{aligned} \quad (\text{B.3})$$

Let $\alpha(x, N)$, $\alpha(x, \infty)$ be the optimal allocation proportions. Let $\xi(\alpha, r) = (1 - \alpha)xR_B + \alpha xr$.

I prove that the inequality in (B.3) holds in two parts, first I show that

$$\frac{d}{dx} \left[\max_{\alpha} \int u(\xi(\alpha, r)) dP^{\alpha, x}(r) \right] \quad (\text{B.4})$$

$$= \int u'(\xi(\alpha(x, N), r)) ((1 - \alpha(x, N))R_b + \alpha(x, N)r) dP^{\alpha, x}(r) \quad (\text{B.5})$$

$$\geq \int u'(\xi(\alpha(x, N), r)) ((1 - \alpha(x, N))R_b + \alpha(x, N)r) dP(r). \quad (\text{B.6})$$

where the first inequality follows from the envelope theorem. The inequality will follow by $P \succeq_{FOSD} P^{\alpha, x}$ if the integrand in (B.6) is decreasing in r . Taking the derivative and rearranging we can see that this is the case if and only if the coefficient of relative risk aversion $-u''(z)z/u'(z)$ is greater than or equal to 1.

To complete the proof that (B.3) holds I show that

$$\begin{aligned} & \int u'(\xi(\alpha(x, N), r))((1 - \alpha(x, N))R_b + \alpha(x, N)r)dP(r) \\ & \geq \frac{d}{dx} \left[\max_{\alpha} \int u((1 - \alpha)xR_B + \alpha xr)dP(r) \right]. \end{aligned}$$

This will follow by the Envelope Theorem and Proposition 6(i), which says that $\alpha(x, \infty) \geq \alpha(x, N)$. I need only show that for $\alpha \in [\alpha(x, N), \alpha(x, \infty)]$ the expression on the LHS of the above inequality is decreasing in α . To be precise, I need that

$$\int u'((1 - \alpha)xR_B + \alpha xr)((1 - \alpha)R_b + \alpha r)dP(r)$$

is decreasing in α . Taking the derivative, I need that

$$\int [u''(\xi(\alpha, r))((1 - \alpha)xR_b + \alpha xr) + u'(\xi(\alpha, r))](r - R_b)dP(r) \leq 0. \quad (\text{B.7})$$

Notice that the term in brackets in the integrand, $u''((1 - \alpha)xR_b + \alpha xr)((1 - \alpha)xR_b + \alpha xr) + u'((1 - \alpha)xR_b + \alpha xr)$, is less than or equal to zero by the assumption of $-u''(z)z/u'(z) \geq 1$, but $(r - R_b)$ is negative for low values of r and positive for high values. Assuming CRRA utility, (B.7) reduces to

$$\int u'((1 - \alpha)xR_B + \alpha xr)(xr - xR_b)dP(r) \geq 0 \quad (\text{B.8})$$

Notice that the LHS of (B.8) is exactly the derivative with respect to α of expected utility. Thus for all $\alpha \leq \alpha(x, \infty)$, (B.8) holds by concavity of the objective function. Then (B.3) holds, as desired. \square

B.8 Proof of Proposition 7

Proof. Let $P_N^{b,s}$ be the induced distribution for the agent with capacity N . The FONC for the constrained agent are given by

$$\begin{aligned} [b] \quad u'(w - b - ps) &= \frac{d}{db} \left[\beta \int u(w + bR_b - sr)dP_N^{b,s}(r) \right] \\ [s] \quad u'(w - b - ps)p &= \frac{d}{ds} \left[\beta \int u(w + bR_b - sr)dP_N^{b,s}(r) \right]. \end{aligned}$$

In equilibrium (with $b = s = 0$) the FONC for b implies that $R_B = 1/\beta$, regardless of capacity or attitude. Let p^N be the equilibrium risky asset price for the capacity N agent. Zero net supply requires that

$$u'(w)p^N = \frac{d}{ds} \left[\beta \int u(w + sr)dP_N^{0,s}(r) \right]_{s=0}. \quad (\text{B.9})$$

Clearly when $s = 0$ we have $\beta \int u(w + sr) dP_{N''}^{0,s}(r) = \beta \int u(w + sr) dP_{N'}^{0,s}(r)$ for all N', N'' . For any $s \neq 0$ for a cautious (reckless) DM, $\beta \int u(w + sr) dP_{N'}^{0,s}(r) < (>) \beta \int u(w + sr) dP_{N''}^{0,s}(r)$ for $N'' > N'$, since $P_N^{0,s}$ is the solution to a maximization (minimization) problem. Therefore the derivative on the RHS of (B.9) is increasing (decreasing) in N when the DM is cautious (reckless). The result follows. \square

B.9 Proofs for Section 8

B.9.1 Proof of Proposition 9

Proof. Fix a wage schedule \tilde{w} which induces effort level a . Let $\tilde{u} = u \circ \text{simsup}(\tilde{w}, P_a)$ and $\tilde{w}(\bar{\omega}) = \sup_{\omega \in \Omega} \tilde{w}(\omega)$ and $\tilde{u}(\bar{\omega}) = u(\tilde{w}(\bar{\omega}))$. Assume for simplicity that Ω is closed, u is differentiable, and P_a is absolutely continuous for all a (the results hold under weaker assumptions as well). Also assume that $\text{simsup}_{N, \tilde{w}, P_a}$ is singleton (this is without loss; we could do everything in terms of selections from $\text{simsup}_{N, \tilde{w}, P_a}$). Order the states so that \tilde{w} is increasing.

Suppose there is an N -simple contract w that induces effort level a . Then it will be possible to find a cell T_i of w 's partition and a sub-interval at the lower end of T_i such that reducing the payoff slightly on this sub-interval, leaving the contract otherwise unchanged, does not change the simsup under P_a . In other words, if w' is the contract so obtained then $\text{simsup}_{N, w, P_a} = \text{simsup}_{N, w', P_a}$. Moreover, since w' dominated statwise by w , we have $U(\text{simsup}_{N, w, P_{a'}}) \geq U(\text{simsup}_{N, w', P_{a'}})$ for all a' , so that w' also induces effort level a .

I now show that the optimal contract will have a discrete jump at the top. As before, order the states so that the original wage schedule \tilde{w} is increasing. Let $\tau = \{t_i\}_{i=0}^N$ be the set of interior cut-offs defining $\text{simsup}_{N, \tilde{w}, P_a}$, where $t_0 = \inf \Omega$, $t_N = \bar{\omega}$ so that t_{N-1} is highest cut-off such that $t_{N-1} < \bar{\omega}$. Since $\text{simsup}_{N, \tilde{w}, P_a}$ is assumed to be unique, for any other vector $\tau' = \{t'_i\}_{i=0}^N$ of $N - 1$ cut-offs we have

$$\sum_{i=1}^N u(\tilde{w}(t'_i))(P_a(t'_i) - P_a(t'_{i-1})) > U(\text{simsup}_{N, \tilde{w}, P_a}, P_a). \quad (\text{B.10})$$

In particular, this inequality holds for any cut-off vectors such that $t'_{N-1} > t_{N-1}$ (without the uniqueness assumption we would instead work with the selection from $\text{simsup}_{N, \tilde{w}, P_a}$ with the maximal t_{N-1} , the proof would be unchanged). Consider modifying \tilde{w} by lowering the payoffs for states in the interval $(t_{N-1}, \bar{\omega} - \varepsilon)$ slightly (such that monotonicity is preserved), leaving the rest unchanged. For ε small enough, $\text{simsup}_{N, w, P_a} = \text{simsup}_{N, \tilde{w}, P_a}$. Moreover, since \tilde{w} uniformly dominates w , w will also induce effort level a (as in the proof of Proposition 8. Moreover $\lim_{\omega \nearrow \bar{\omega}} w(\omega) < \tilde{w}(\bar{\omega}) = w(\bar{\omega})$. \square

B.10 Proofs for Section 9

Let $f'_-(\omega)$ and $f'_+(\omega)$ be left and right derivatives of f at ω respectively. The proof of Proposition 10 makes use of the following lemma.

Lemma 4. *Let f be a decreasing and continuous function. Suppose f has a kink at d ($f'_+(d) > f'_-(d)$). Then $\text{siminf}(f)$ is constant in a neighborhood of d .*

Proof. Suppose that there is a cut-off at d . That is, $t_n = d$ and $\text{siminf}(f)$ is discontinuous at t_n . Fix t_{n-1}, t_{n+1} . Denote the value generated by an cutoff in (t_{n-1}, t_{n+1}) by

$$V(t) = [P(t) - P(t_{n-1})]u(f(t)) + [P(t_{n+1}) - P(t)]u(f(t_{n+1}))$$

Then optimality of $t_n = d$ implies that the left derivative of V at d must be positive and the right derivative must be negative:

$$V'_-(d) = [P(d) - P(t_{n-1})]u'(f(d))f'_-(d) + p(d)[u(f(d)) - u(f(t_{n+1}))] \geq 0 \quad (\text{B.11})$$

$$V'_+(d) = [P(d) - P(t_{n-1})]u'(f(d))f'_+(d) + p(d)[u(f(d)) - u(f(t_{n+1}))] \leq 0 \quad (\text{B.12})$$

Equations (B.11) and (B.12) imply

$$f'_+(d) \leq \frac{-p(d)}{u'(f(d))[P(d) - P(t_{n-1})]}[u(f(d)) - u(f(t_{n+1}))] \leq f'_-(d)$$

which contradicts $f'_+(d) > f'_-(d)$. □

Let t_N be the highest cut-off defining $\text{siminf}(y|d)$. Clearly $t_N \leq d$, since y is flat above d for a full insurance contract. Moreover Lemma 4 implies that $t_N < d$.

Proof. Proposition 10. Let $l(\omega|d, c)$ be the amount paid by the consumer when the loss is ω given a contract with deductible d and coverage rate c . Denote the perceived value to a cautious agent of an insurance contract characterized by d, c as

$$U^N(d, c) = \max_{\hat{t}_1, \dots, \hat{t}_N} \sum_{n=1}^{N+1} [P(\hat{t}_n) - P(\hat{t}_{n-1})]u(w - l(\hat{t}_n|d, c)) \quad (\text{B.13})$$

A fully rational agent would value the contract at

$$U^\infty(d, c) = \int_0^d u(w - \omega)dP(\omega) + \int_d^{\bar{\omega}} u(w - d - (1 - c)(\omega - d))dP(\omega)$$

Let $\{t_1, \dots, t_N\}$ be the solution to the maximization problem in B.13, and let $n^* = \max\{n \in \{1, \dots, N\} : t_n \leq d\}$ be the index of the highest cut-off below d . By the envelope theorem

$$U_d^N(d, c) = -c \sum_{n=n^*+1}^{N+1} [P(t_n) - P(t_{n-1})]u'(w - d - (1 - c)(t_n - d)).$$

Moreover $U_d^\infty(d, c) = -c \int_d^{\bar{\omega}} u'(w - d - (1 - c)(\omega - d)) dP(\omega)$. Notice that

$$\begin{aligned} U_d^N(d, c) &< c[P(d) - P(t_{n^*})]u'(w - d - (1 - c)(t_{n^*+1} - d)) \\ &\quad - c \sum_{n=n^*+1}^{N+1} [P(t_n) - P(t_{n-1})]u'(w - d - (1 - c)(t_n - d)) \\ &\leq -c \int_d^{\bar{\omega}} u'(w - d - (1 - c)(\omega - d)) dP(\omega) \\ &= U_d^\infty(d, c) \end{aligned}$$

where the first inequality follows since Lemma 4 implies $t_{n^*} < d$, and the second from concavity of u . Notice that the second inequality holds with equality if and only if $c = 1$ for u strictly concave. This proves part 1 of Proposition 10.

The proof of part 2 is similar. In this case

$$\begin{aligned} U_c^N(d, c) &= \sum_{n=n^*+1}^{N+1} [P(t_n) - P(t_{n-1})]u'(w - d - (1 - c)(t_n - d))(t_n - d) \\ &> -[P(d) - P(t_{n^*})]u'(w - d - (1 - c)(t_n - d))(t_n - d) \\ &\quad + \sum_{n=n^*+1}^{N+1} [P(t_n) - P(t_{n-1})]u'(w - d - (1 - c)(t_n - d))(t_n - d) \\ &\geq \int_d^{\bar{\omega}} u'(w - d - (1 - c)(\omega - d))(\omega - d) dP(\omega) \\ &= U_c^\infty(d, c) \end{aligned}$$

where again the first inequality follows from Lemma 4 and the second from concavity of u . □

B.10.1 Proof of Proposition 11

This result follows from a similar argument as Proposition 10. The key to the proof of Proposition 10 was the result of Lemma 4 that $t_{n^*} < d$. When the baseline contract is full insurance above d the following result allows us to draw the analogous conclusion that the value of the greatest cut-off below d is increasing in N .

Let $t^N = t_0^N, \dots, t_{N+1}^N$ be the cutoffs defining $\text{siminf}(y|d)$ when the agent has capacity N and $t^{N+1} = t_0^{N+1}, \dots, t_{N+2}^{N+1}$ be the cutoffs for capacity $N + 1$. When $c = 1$ we have the immediate corollary of Lemma 8, which says that the largest cut-off below d is increasing in N . Notice that for $c = 1$, $n^* = N$ where N is the capacity of the agent. This follows since ex-post wealth is constant above d , and so it would not be optimal to have a cut-off above d .

Corollary 3. For a full insurance contract above a deductible, $t_N^N < t_{N+1}^{N+1}$.

We can now prove Proposition 11.

Proof. (Proposition 11) For $c = 1$ an envelope theorem implies that

$$\begin{aligned} U_d^N(d, c) &= -[1 - P(t_N^N)]u'(w - d) \\ &\leq [P(t_{N+1}^{N+1}) - P(t_N^N)]u'(w - d) - [1 - P(t_N^N)]u'(w - d) \\ &= U_d^{N+1}(d, c) \end{aligned}$$

The proof for decreasing c (recall that c is bounded above by 1), is similar. \square

B.10.2 Propositions 14 and 15

Propositions 14 and 15 follow immediately from Corollary 8 and 7 respectively.

B.10.3 Proof of Propositions 13 and 12

I give the proof Proposition 13 here. The argument for Proposition 12 is essentially identical.

Proof. Let m be the out of pocket maximum, and $U^N(m)$ be the perceived value, holding c and d fixed. Let $\underline{l}(m) = \min\{\omega \in \Omega : l(\omega|d, c, m) = m\}$. Let $C^*(N, [0, \underline{l}(m)])$ be the cut-offs corresponding to the elements of the plan siminf. Since P is absolutely continuous and $\underline{l}(\cdot)$ is continuous, $C^*(N, [0, \cdot])$ is upper-hemicontinuous by Berge's maximum theorem.

Recall that for a decreasing and continuous function f on an interval of the reals the cell function corresponding to simple lower bounds is given by

$$v([a, b]) = f(b)P(\{\omega \in [a, b]\}). \quad (\text{B.14})$$

Since utility is strictly decreasing on $[0, \underline{l}(m)]$ and P is full support the cell function generating $U^N(m)$ is strictly submodular by Lemma 7. Moreover, it satisfies the conditions for regularity in Tian (2015), so by Tian (2015) Theorem 3, C'' and C' are sandwiched for all $C'' \in C^*(N + 1, [0, \underline{l}(m)])$ and $C' \in C^*(N, [0, \underline{l}(m)])$, so upper-hemicontinuity implies that for ε small enough there exist sandwiched selections from $C^*(N, m)$ and $C^*(N + 1, m - \varepsilon)$.

For an interval $I \subseteq \Omega$ and $C \in C^N(I)$ let $V(I, C)$ be the coarse value corresponding to the cell function in (B.14). Since utility is constant above $\underline{l}(m)$, we can write $U^N(m) = V([0, \underline{l}(m)], C^*(N, [0, \underline{l}(m)])) + (1 - P(\underline{l}(m)))u(w - m)$ (with an abuse of notation when $C^*(N, [0, \underline{l}(m)])$ is non-singleton). The result follows from Lemma 9. \square

B.10.4 Lemma 5

Lemma 5. *Consider two plans with no out-of-pocket maximum, coverage rate c , deductibles d and d' , with $d < d'$, and premiums p, p' such that $p' - p = (1 - c)(d' - d)$. Then a cautious agent is indifferent between the two plans if the lowest cut-off defining siminf of the high deductible plan is (weakly) greater than d' .²⁴*

Proof. Let t be the lowest cut-off for the siminf of the high deductible plan. Since the high deductible dominates the low deductible plan, any set of cut-offs defines a less preferred lower bound for the latter than for the former. Since the siminf of the high deductible plan is also dominated statwise by the low deductible plan it is an element of the siminf for the low deductible plan as well. \square

C Comparative Statics

Given the centrality of the siminf and simsup in the representations of incomplete and complete preferences discussed above, it helpful in applications to understand how these functions vary with the parameters of the problem. This section contains a number of comparative statics results that are helpful in this regard. In addition, some of the results apply to a more general class of problems and are of independent interest. I characterize the responses siminf and simsup to changes in both the capacity and the beliefs. I also show that the marginal returns to additional capacity are higher when the relevant state space is larger. These results are used in the applications.

I will discuss real valued acts, with the understanding that all conclusions apply to the utility images of any acts. Moreover, I will make statements about *every* element of siminf and simsup , with the understanding that these apply “up to sets of measure zero under P ”. I will first focus on properties of $\text{siminf}_{N,w,P}$ when $\Omega = [\omega, \bar{\omega}] \subset \mathbb{R}$ and w is increasing. Results in this setting can be extended in two ways. First, since $\text{siminf}_{N,w,P} = -\text{simsup}_{N,-w,P}$ all results regarding siminf can be translated directly to simsup . Second, using the “Lebesgue approach” of Section A.5.1, results for increasing functions on an interval can be translated to results for arbitrary bounded functions on an arbitrary state space. Finally, for simplicity I will assume throughout that $f \geq \text{siminf}_{N,w,P}$. Using Proposition 5 we know when this will hold. Moreover, Proposition 5 tells us where violations of statewise dominance can occur. The

²⁴The siminf for the high deductible plan need not be unique. I state the result in this way for simplicity, but it holds as long as the condition is satisfied for a selection from the siminf . In fact, if siminf for the high deductible plan is single valued then the converse holds as well.

results presented below will apply with only minor modifications when $\neg(f \geq \text{siminf}_{N,w,P})$, as we can just ignore the zero measure set on which violations of statewise dominance occur.

C.1 Submodular cell functions

I first present some results in a general setting. The framework, as well as many of the general results mentioned here, are from Tian (2015) and Tian (2016). Let $\Omega = [\underline{\omega}, \bar{\omega}] \subset \mathbb{R}$.²⁵ Let $I(\Omega)$ be the set of interval subsets of Ω . A *cell function* is a function $v : I(\Omega) \mapsto \mathbb{R}$. An *interval partition* is a partition consisting only of intervals. Given an interval partition $\tau = \{T_i\}_{i=1}^N$ of Ω , the *coarse value* associated with cell function v is defined as

$$V(\tau) = \sum_{i=1}^N v(T_i).$$

I will be interested in cell functions that are submodular.

Definition. A cell function v is (strictly) **submodular** if for all intervals I, I' with $I \cap I' \neq \emptyset$, we have

$$v(I \cap I') + v(I \cup I') (<) \leq v(I) + v(I').$$

The following Lemma says that elements of siminf and simsup will have interval partitions. The proof follows from that given for Lemma 3.

Lemma 6. *If w is increasing then every element of $\text{siminf}_{N,w,P}$ and $\text{simsup}_{N,w,P}$ has an interval partition.*

For an interval I let $\underline{w}_a(I) = \inf_{\omega \in I} \{w_a(\omega)\}$. Given this Lemma, we know that when $f \geq \text{siminf}_{N,u(a),P}$, $\text{siminf}_{N,u(a),P}$ will be defined by the partition maximizing the coarse value associated with the cell function, which I will call the *siminf cell function* $v(I) = \underline{w}_a(I)P(I)$. Moreover, I will show that this coarse value is submodular. I turn now to general properties of submodular cell functions.

Submodularity can be thought of as diminishing returns to larger intervals. Indeed, this intuition is made concrete by the following observation from Tian (2015). For each cell $I = [a, b]$ and c in I , let $\Delta v(c, I) = [v([a, c]) + v([c, b])] - v(I)$. Notice that for submodular cell functions $v(c, I)$ is always non-negative. We can make the following observation.

Proposition 17. (Tian (2015), Observation 1) *A cell function is (strictly) submodular if and only if for all intervals I, I' with $(I' \subsetneq I) \implies I' \subseteq I$ and any $c \in I'$ we have*

$$\Delta v(c, I) (>) \geq \Delta v(c, I').$$

²⁵It matters that Ω is bounded, but the fact that it is closed here is irrelevant

Suppose $\tau \in T^N(X)$ is an interval partition of the sub-interval $X \in I(\Omega)$. Then τ can be described by a set of cut-offs $C = \{t_i\}_{i=1}^{N-1} \in \mathbb{R}^{N-1}$ giving the interior endpoints of the partition intervals. I will sometimes use the convention $t_0 = \inf\{X\}$ and $t_N = \sup\{X\}$ when referring to a partition of the interval X . Cut-off states can be assigned to partition cells in any way.²⁶ Since for finding the siminf it will always be optimal to assign cut-off states to the higher interval I will assume this assignment from now on. Let $C^N(\Omega)$ be the set of cut-offs defining interval partitions of Ω . Abusing notation, I will write $V(\Omega, C)$ to denote the coarse value of the partition induced on Ω by cut-off vector $C \in C^N(\Omega)$.

Endow $C^N(\Omega)$ with the pointwise partial order. Denote the least upper bound and greatest lower bound of two cut-off vectors $C, C' \in C^N(\Omega)$ by $C \vee C'$ and $C \wedge C'$. Define the union of two sets of cut-offs in the obvious way, as the ordered union of the two sets.²⁷

Definition. *A coarse value is (strictly) **supermodular** if it is (strictly) supermodular as a function of cut-off vectors.*

The following result also relates submodularity of cell functions to diminishing returns of cell division. It is extremely useful in the comparative statics results.

Proposition 18. (Tian (2015), Observation 3) *The coarse value is (strictly) supermodular if and only if the cell function is (strictly) submodular.*

Propositions 17 and 18 are from Tian (2015), Observations 1 and 3. Although Tian (2015) does not consider the strict version of either result, this follows by the same argument.

Let $W(N, \Omega) = \sup_{C \in C^N(\Omega)} V(\Omega, C)$. Tian (2015) also proves that supermodular coarse values exhibit diminishing marginal returns to capacity.

Proposition 19. (Tian (2015), Theorem 1). *For a supermodular coarse value $W(N + 1, \Omega) - W(N, \Omega) \leq W(N, \Omega) - W(N - 1, \Omega)$.*

²⁶Technically, a full description of the partition should include, in addition to the cut-offs, a vector in $\{0, 1\}^{N+1}$ identifying whether each cut-off state is assigned to the interval immediately above or below it. A cut-off assigned to the lower interval is considered to be lower than the same cut-off assigned to the higher interval, otherwise the usual order on \mathbb{R}^N is used to order cut-off vectors.

²⁷By this definition the union of two cut-off vectors defines a partition which is the join, in the refinement sense, of the partitions defined by each of the individual cut-off vectors. This is different than the join of the cut-off vectors, which I define as the coordinate-wise maximum. The later notion is restricted to cut-off vectors of the same length, whereas the union can be taken of any two cut-off vectors.

It is easy to verify that the siminf cell function satisfies the condition of Proposition 17, and is thus submodular, so that the associated coarse value is supermodular by Proposition 18.

Lemma 7. *The siminf cell function $v(\cdot)$ is submodular. It is strictly submodular if P has full support and w_a is strictly increasing.*

Proof. For any interval $[a, b]$ and $c \in [a, b]$ we have

$$\begin{aligned}\Delta v(I, c) &= \underline{w}_a([a, c])P([a, c]) + \underline{w}_a(c, b)P([c, b]) - \underline{w}_a([a, b])P([a, b]) \\ &= (\underline{w}_a([c, b]) - \underline{w}_a([a, c]))P([c, b])\end{aligned}$$

where the last equality follows since $\underline{w}_a([a, c]) = \underline{w}_a([a, b])$ when w_a is increasing, as it is here by assumption. Suppose $[a, b] \subseteq [l, m]$. Then $\underline{w}_a([c, b]) = \underline{w}_a([c, m])$, $\underline{w}_a([l, c]) \leq \underline{w}_a([a, c])$ and $P([c, b]) \leq P([c, m])$, so $\Delta v([a, b], c) \leq \Delta v([l, m], c)$. The claim follows from Proposition 17. \square

Corollary 4. *The coarse value for the siminf cell function is supermodular. It is strictly supermodular if w_a is strictly increasing and P has full support.*

Proposition 17 relates submodularity of the cell function to the returns to making a subdivision of a cell at a given place. The coarse values of submodular cell functions also satisfy the following increasing differences property.

Lemma 8. *If the cell function is submodular then*

- For all $a \in \Omega$, $b'' \geq b' \geq a$ and $C'' = \{t''_i\}_{i=1}^{N-1}$, $C' = \{t'_i\}_{i=1}^{N-1} \in C^N([a, b'])$ with $t''_{N-1} \geq t'_{N-1}$,

$$V([a, b''], C'') - V([a, b''], C') \geq V([a, b'], C'') - V([a, b'], C'). \quad (\text{C.1})$$

- For all $b \in \Omega$, $b \geq a'' \geq a'$ and $C'' = \{t''_i\}_{i=1}^{N-1}$, $C' = \{t'_i\}_{i=1}^{N-1} \in C^N([a'', b])$ with $t''_1 \geq t'_1$,

$$V([a', b], C') - V([a', b], C'') \geq V([a'', b], C') - V([a'', b], C''). \quad (\text{C.2})$$

Moreover, if the cell function is strictly submodular then the inequality in (C.1) is strict when $t''_{N-1} > t'_{N-1}$ and $b'' > b'$, and that in (C.2) is strict when $t''_1 > t'_1$ and $a'' > a'$.

Proof. I prove the first claim. The proof of the second is analogous. The key is that the partitions induced on $[a, b'']$ and $[a, b']$ by a cut-off vector C differ only in the highest cell of the partition. This implies that $V([a, b''], C'') - V([a, b''], C'' \cup C') = V([a, b'], C'') -$

$V([a, b'], C'' \cup C')$, since $t''_{N-1} \geq t'_{N-1}$ implies that for a given interval the highest cells of the partitions induced by C'' and $C'' \cup C'$ are the same. It also means that $V([a, b''], C'' \cup C') - V([a, b''], C') \geq V([a, b'], C'' \cup C') - V([a, b'], C')$ by the characterization of submodular cell functions in Proposition 17. Therefore

$$\begin{aligned}
& V([a, b''], C'') - V([a, b''], C') \\
&= V([a, b''], C'') - V([a, b''], C'' \cup C') + V([a, b''], C'' \cup C') - V([a, b''], C') \\
&= V([a, b'], C'') - V([a, b'], C'' \cup C') + V([a, b''], C'' \cup C') - V([a, b''], C') \\
&\geq V([a, b'], C'') - V([a, b'], C'' \cup C') + V([a, b'], C'' \cup C') - V([a, b'], C') \\
&= V([a, b'], C'') - V([a, b'], C')
\end{aligned}$$

□

The following standard comparative statics result follows from Lemma 8 and Proposition 18. For sets A, B I write $A \geq_{SSO} B$ when A is larger than B in the strong set order.

Corollary 5. *With a submodular cell function, for intervals $I'', I' \subseteq \Omega$ with $I'' \geq_{SSO} I'$, $C^*(N, I'') \geq_{SSO} C^*(N, I')$.*

Corollary 6. *With a strictly submodular cell function the elements of $C^*(N, X)$ are ordered for all N and all $X \subseteq \Omega$.*

Related to the first part of Corollary 6, Tian (2015) provides alternative conditions under which the elements of $C^*(N, X)$ are ordered.

C.1.1 Changing N

It will often be useful to understand how the cut-offs defining optimal partitions change with capacity. The following definition is from Tian (2015).

Definition. *Two sets of cut-offs $C \in \mathbb{R}^N$ and $C' \in \mathbb{R}^{N+1}$ are **sandwiched** if $t'_i < t_{i+1} < t'_{i+1}$ for all i .*

Say that cut-offs are weakly sandwiched if the strict inequalities in the above definition are replaced with weak inequalities. The following result is related to Proposition 2 in Tian (2015), which shows that the maximal elements of $C^*(N, [a, b])$ and $C^*(N+1, [a, b])$, when they exist, are sandwiched (and similarly for the minimal elements). While Tian (2015) uses continuity of the coarse value to guarantee that minimal and maximal elements of the sets of optimal cut-off vectors exist, we can conclude that sandwiched selections exist without this assumption.

Corollary 7. *With a sub-modular cell function, if $C^*(N, [a, b])$ and $C^*(N+1, [a, b])$ are non-empty then for any $C' \in C^*(N, [a, b])$ there exists $C'' \in C^*(N+1, [a, b])$ such that C' and C'' are weakly sandwiched.*

Proof. Start with $C'' = \{t''_i\}_{i=1}^N \in C^*(N+1, [a, b])$ and $C' = \{t'_i\}_{i=1}^{N-1} \in C^*(N, [a, b])$. Define $\tilde{C} = \{\tilde{t}_i\}_{i=1}^N \equiv (\{t''_i\}_{i=1}^{N-1} \wedge C') \cup t''_N$ and $\hat{C} = \{\hat{t}_i\}_{i=1}^N \equiv (\{\tilde{t}_i\}_{i=2}^N \vee C') \cup \tilde{t}_1$. By Tian (2015), $\hat{C} \in C^*(N+1, [a, b])$.

If $\{\hat{t}_i\}_{i=1}^{N-1} \leq C'$ then it will follow that \hat{C} and C' are weakly sandwiched. To see that this holds, note that $\hat{t}_1 = \tilde{t}_1 = \min\{t''_1, t'_1\} \leq t'_1$ and for $i > 1$, $\hat{t}_i = \max\{\tilde{t}_i, t'_{i-1}\}$. Moreover, $\tilde{t}_N = t''_N$ and $\tilde{t}_i = \min\{t''_i, t'_i\}$ for $i < N$. Thus for $1 < i < N$, we have $\hat{t}_i = \max\{\min\{t''_i, t'_i\}, t'_{i-1}\} \leq t'_i$. □

Under an additional condition on the cell function, Tian (2015) shows that all cut-offs are sandwiched.

Tian (2015), Theorem 3. *With a regular cell function, C'' and C' are sandwiched for all $C'' \in C^*(N+1, [a, b])$ and $C' \in C^*(N, [a, b])$.*

The definition of a regular cell function can be found in Tian (2015).

C.1.2 Marginal returns to capacity

Consider an interval state space. Submodularity of the cell function means that the benefit of dividing an interval at a certain point is higher for larger (in the inclusion order) intervals. The following proposition shows that with a submodular cell function there will also be lower returns to increasing capacity when dividing a sub-interval. Assume the state space is an interval, and let $S \subseteq S'$ be two interval subsets of the state space.

Let $C = \{t_i\}_{i=1}^{N-1}$ be a cut-off vector defining an N -element interval partition of the interval $X \subseteq \Omega$, where $t_0 = \inf\{X\}$ and $t_N = \sup\{X\}$. I follow the convention that cut-off states are assigned to the higher interval, although this has no bearing on the results. If X is closed at the bottom and open at the top then

$$V(X, C) = \sum_{i=0}^{N-1} v([t_i, t_{i+1})) + v([t_N, t_{N+1}]).$$

Let the set of cut-offs defining N -element interval partitions of X be $C^N(X) \subseteq \mathbb{R}^{N-1}$. Let $C^*(N, X) = \operatorname{argmax}_{C \in C^N(X)} V(X, C)$, and assume $C^*(N, X)$ is non-empty for all $X \subseteq \Omega$ (as is the case when the coarse value is the expectation of a simple lower or upper bound of a

bounded function). Let $W(N, X) = \max_{C \in \mathcal{C}^N(X)} V(C)$. The following lemma is similar to the characterization of submodularity given by Proposition 17. It states that the marginal returns to capacity are higher when dividing larger intervals.

Lemma 9. *Let the cell function be submodular, and let S, S' with $S \subseteq S'$ be two intervals. Assume that there exist selections from $\mathcal{C}^*(N, S')$ and $\mathcal{C}^*(N + 1, S)$ that are sandwiched. Then $W(N + 1, S') - W(N, S') \geq W(N + 1, S) - W(N, S)$.*

Remark. Recall that there exist sandwiched selections from $\mathcal{C}^*(N, X)$ and $\mathcal{C}^*(N + 1, X)$ for any X by 7 or Tian (2015), Theorem 3. Thus the condition of Lemma 9 will be satisfied if $v(\cdot)$ is regular, the optimal cut-offs are continuous in the endpoints of the intervals, and S, S' are sufficiently close.

Proof. Let $C' = \{t'_i\}_{i=1}^{N-1}$ and $C = \{t_i\}_{i=1}^N$ be the sandwiched selections from $\mathcal{C}^*(N, S')$ and $\mathcal{C}^*(N + 1, S)$ respectively. Since the cut-offs are sandwiched it must be that $\inf\{S\} = t_0 \leq t_1 \leq t'_1$ and $\sup\{S\} = t_{N+1} \geq t_N \geq t'_{N-1}$. Therefore C' also defines an N -element partition of S . The partition induced by C' on S and that induced by C' on S' differ only on the highest and lowest partition cells, which are larger for the latter. Similarly for the partitions induced by C on S and S' . By optimality of C' and C we have

$$\begin{aligned} W(N + 1, S') - W(N, S') &= W(N + 1, S') - V(S', C') \\ &\geq V(S', C) - V(S', C'), \end{aligned}$$

and

$$\begin{aligned} W(N + 1, S) - W(N, S) &= V(S, C) - W(N, S) \\ &\leq V(S, C) - V(S, C') \end{aligned}$$

so it suffices to show $V(S', C) - V(S', C') \geq V(S, C) - V(S, C')$. We have

$$V(S', C) - V(S', C') = V(S', C') - V(S', C \cup C') + V(S', C \cup C') - V(S', C')$$

and

$$V(S, C) - V(S, C') = V(S, C) - V(S, C \cup C') + V(S, C \cup C') - V(S, C').$$

By the sandwiched property of C and C' , $V(S', C) - V(S', C \cup C') = V(S, C) - V(S, C \cup C')$. Since the partition induced by C' on S and that induced by C' on S' differ only on the highest and lowest partition cells, which are larger for the latter, submodularity of the cell function implies that $V(S', C \cup C') - V(S', C') \geq V(S, C \cup C') - V(S, C')$. \square

C.1.3 Changing the distribution

I look here at how the functions in simsup and siminf change as beliefs change. As before, I will discuss the siminf case, but all the results hold without alteration for simsup . I remain in the one dimensional setting. The results will be extended in the next section. As I will be varying beliefs while holding capacity fixed I will make explicit the dependence on P of the, cell function, coarse value, and optimal cut-offs by writing these as $v(\cdot|P)$, $V(\Omega, \cdot|P)$, and $C^*(N, \Omega|P)$.

Let \hat{P} and P be distributions, with densities \hat{p} and p respectively. Recall that distribution \hat{P} that dominates P according to the *monotone likelihood ratio* (MLR) property if there exists a non-negative increasing function α such that $\hat{p}(\omega) = \alpha(\omega)p(\omega)$ for all $\omega \in \Omega$.

Given two real valued functions g and h defined on a partially ordered set Z , say that g dominates h by the *interval dominance order* ($g \succeq_I h$) if $f(z'') \geq f(z') \implies g(z'') \geq g(z')$ holds for all $z'' \geq z'$ such that $g(z) \geq g(z') \forall z \in [z', z''] = \{z \in Z : z' \leq z \leq z''\}$. The following is implied by Tian (2016), Proposition 3.

Proposition 20. *Let Ω be an interval of \mathbb{R} , and $G \succeq_{MLR} P$. Assume that w is absolutely continuous (and thus differentiable a.e.). Then $V(\Omega, \cdot|G) \succeq_I V(\Omega, \cdot|P)$ on $C^N(\Omega)$ for all N .*

If w is decreasing then the same conclusion holds; $V^N(\cdot|G) \succeq_I V^N(\cdot|P)$.

By Proposition 18 and Lemma 7, V is supermodular in C . By Theorem 1 in Quah and Strulovici (2009) and Proposition 20 we obtain the following.

Corollary 8. *If $G \succeq_{MLR} P$ then $C^*(N, \Omega|G) \geq_{SSO} C^*(N, \Omega|P)$.*

References

- J. Abaluck and J. Gruber. Choice inconsistencies among the elderly. *American Economic Review*, 101(4):1180–1210, 2011.
- L. Anderlini and L. Felli. Incomplete written contracts: Undescribable states of nature. *The Quarterly Journal of Economics*, 109(4):1085–1124, 1994.
- F. J. Anscombe and R. J. Aumann. A definition of subjective probability. *Annals of mathematical statistics*, 34(1):199–205, 1963.
- R. J. Aumann. Utility theory without the completeness axiom. *Econometrica: Journal of the Econometric Society*, pages 445–462, 1962.
- B. D. Bernheim and C. Sprenger. Direct tests of cumulative prospect theory. *Available at SSRN 3350196*, 2019.
- T. F. Bewley. Knightian decision theory. part i. *Decisions in economics and finance*, 25(2): 79–110, 2002.
- S. Bhargava, G. Loewenstein, and J. Sydnor. Choose to lose: Health plan choices from a menu with dominated option. *The Quarterly Journal of Economics*, 132(3):1319–1372, 2017.
- S. H. Chew, B. Miao, and S. Zhong. Partial ambiguity. *Econometrica*, 85(4):1239–1260, 2017.
- D. M. Cutler and R. Zeckhauser. Extending the theory to meet the practice of insurance. *Brookings-Wharton Papers on Financial Services*, 2004(1):1–53, 2004.
- D. Ellsberg. Risk, ambiguity, and the savage axioms. *The quarterly journal of economics*, pages 643–669, 1961.
- T. Galaabaatar and E. Karni. Subjective expected utility with incomplete preferences. *Econometrica*, 81(1):255–284, 2013.
- P. Ghirardato and M. Marinacci. Ambiguity made precise: A comparative foundation. *Journal of Economic Theory*, 102(2):251–289, 2002.
- I. Gilboa and D. Schmeidler. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141–153, 1989.
- I. Gilboa, F. Maccheroni, M. Marinacci, and D. Schmeidler. Objective and subjective rationality in a multiple prior model. *Econometrica*, 78(2):755–770, 2010.

- L. Hartmann and T. F. Kauffeldt. A hierarchy of ambiguity aversion. Working Paper, 2019.
- J. L. Horowitz. The bootstrap. In *Handbook of econometrics*, volume 5, pages 3159–3228. Elsevier, 2001.
- P. Klibanoff, M. Marinacci, and S. Mukerji. A smooth model of decision making under ambiguity. *Econometrica*, 73(6):1849–1892, 2005.
- E. Lehrer and R. Teper. Extension rules or what would the sage do? *American Economic Journal: Microeconomics*, 6(1):5–22, 2014.
- M. Mandler. Incomplete preferences and rational intransitivity of choice. *Games and Economic Behavior*, 50(2):255–277, 2005.
- R. Mehra and E. C. Prescott. The equity premium: A puzzle. *Journal of monetary Economics*, 15(2):145–161, 1985.
- P. G. Moffatt, S. Sitzia, and D. J. Zizzo. Heterogeneity in preferences towards complexity. *Journal of Risk and Uncertainty*, 51(2):147–170, 2015.
- S. Mukerji. Ambiguity aversion and incompleteness of contractual form. *American Economic Review*, pages 1207–1231, 1998.
- W. S. Neilson. Some mixed results on boundary effects. *Economics Letters*, 39(3):275–278, 1992.
- P. Ortoleva. The price of flexibility: Towards a theory of thinking aversion. *Journal of Economic Theory*, 148(3):903–934, 2013.
- K. R. Parthasarathy. *Probability measures on metric spaces*, volume 352. American Mathematical Soc., 2005.
- I. Puri. Simplicity theory of choice under risk. Working Paper, 2020.
- J. K.-H. Quah and B. Strulovici. Comparative statics, informativeness, and the interval dominance order. *Econometrica*, 77(6):1949–1992, 2009.
- N. Saponara. Revealed reasoning. Working Paper, 2020.
- D. Schmeidler. Subjective probability and expected utility without additivity. *Econometrica*, pages 571–587, 1989.

- U. Segal. The ellisberg paradox and risk aversion: An anticipated utility approach. *International Economic Review*, pages 175–202, 1987.
- K. Seo. Ambiguity and second-order belief. *Econometrica*, 77(5):1575–1605, 2009.
- Z. Shapira and I. Venezia. On the preference for full-coverage policies: Why do people buy too much insurance? *Journal of Economic Psychology*, 29(5):747–761, 2008.
- D. Sonsino, U. Benzion, and G. Mador. The complexity effects on choice with uncertainty—experimental evidence. *The Economic Journal*, 112(482):936–965, 2002.
- J. Tian. Optimal interval division. Working Paper, 2015.
- J. Tian. Monotone comparative statics for cut-offs. Working Paper, 2016.
- M.-L. Vierø. Bait contracts. *Oxford Economic Papers*, 66(3):702–720, 2014.
- W. K. Viscusi and W. A. Magat. Bayesian decisions with ambiguous belief aversion. *Journal of Risk and Uncertainty*, 5(4):371–387, 1992.
- J. Von Neumann and O. Morgenstern. Game theory and economic behavior, 1944.