Cases and Scenarios in Decisions under Uncertainty[*]

Itzhak Gilboa[†], Stefania Minardi[‡] and Larry Samuelson[§]

February 6, 2017

Abstract

We offer a model that combines and generalizes case-based decision theory and expected utility maximization. It is based on the premise that an agent looks ahead and assesses possible future scenarios, but may not know how to evaluate their likelihood and may not be sure that the set of scenarios is exhaustive. Consequently, she also looks back at her memory for past cases, and makes decisions so as to maximize a combined function, taking into account both scenarios and cases. We allow for non-additive set functions, both over future scenarios and over past cases, to capture (i) incompletely specified or unforeseen scenarios, (ii) ambiguity, (iii) the absence of information about counterfactuals, and (iv) some forms of case-to-rule induction ("abduction") and statistical inference. We axiomatize this model. Learning in this model takes several forms, and, in particular, changes the relative weights of the two forms of reasoning.

[†]HEC, Paris-Saclay, and Tel-Aviv University. tzachigilboa@gmail.com
[‡]HEC, Paris-Saclay. minardi@hec.fr
[§]Yale University. larry.samuelson@yale.edu

Cases and Scenarios in Decisions under Uncertainty

Contents

Cases and Scenarios in Decisions under Uncertainty

1 Introduction

1.1 Background

Expected utility theory rests on intuitive axiomatic foundations (Ramsey 1926, de Finetti 1931, 1937, von Neumann and Morgenstern 1944, Savage 1954, Anscombe and Aumann 1963). These axiomatic foundations have been powerful rhetorical devices in the argument that subjective expected utility theory is the only "rational" way of making decisions in the face of uncertainty, despite objections by prominent dissenters such as Knight (1921) and Keynes (1921).

Expected utility theory has been criticized on both positive and normative grounds. As a description of people's behavior, expected utility theory has been challenged by the classical "paradoxes" of Allais (1953) and Ellsberg (1961), as well as by the works of Kahneman and Tversky (1979) and their followers. As a normative theory, the expected-utility paradigm, and especially the application of Savage's axioms to large state spaces, has been criticized by (among others) Shafer (1986) and Gilboa, Postlewaite, and Schmeidler (2008, 2009, 2012).

These critiques have given rise to the development of new decision models, for the most part generalizing expected utility theory by relaxing the underlying axioms. Examples include Prospect Theory (Kahneman and Tversky, 1979), Rank-Dependent Utility (Quiggin, 1982, Chew, 1983, Yaari, 1987), Cumulative Prospect Theory (Tversky and Kahneman, 1992, Wakker, 2010), Choquet Expected Utility (Schmeidler, 1989), Maxmin Expected Utility (Gilboa and Schmeidler, 1989), the "Smooth" model (Klibanoff, Marinacci, and Mukerji, 2005), and Variational Preferences (Maccheroni, Marinacci, and Rustichini, 2006a,b). For surveys, see Camerer and Weber (1992), Harless and Camerer (1994), Gilboa (2009), Wakker (2010), and Gilboa and Marinacci (2013).

The generalizations of expected utility theory vary in their cognitive interpretations and in the degree to which they capture intuition about how decisions are made. Some have relatively clear interpretations that could be viewed as descriptions of the mental processes a person (or organization) goes through while reaching a decision. Others are typically viewed as "as-if" descriptions of behavior. However, even the more intuitive theories do not describe the origin of beliefs, regardless of whether these beliefs are represented by probabilities or otherwise. This lacuna has implications for

both normative and positive applications of the theory. On the normative side, axiomatic derivations that consider the behavior of an agent as a "black box" may well convince the agent that she would like to behave according to a certain model, but they do not provide the self-help tools to achieve this goal. In particular, the axioms provide no clues as to which probability measure(s) the agent should adopt as her beliefs. Similarly, when the theory is interpreted descriptively and applied to models in economics, finance, political science, and other fields, the modelers may be convinced that a specific decision model is a reasonable building block to be used in their analysis, but the axiomatic foundations provide no guidelines as to what beliefs the agents in their models are likely to hold.

It has accordingly been suggested that decision theory should say more about the way people reason about uncertainty and the scope of problems for which beliefs can be modeled by probabilities. One obvious source of information about the likelihood of future events is the past. The importance of past experience has been the primary motivation for the development of "case-based decision theory" by Gilboa and Schmeidler (1995, 2001), dealing with decision making that is guided by analogies to past cases: acts that performed well in similar problems in the past tend to be chosen again; acts that performed poorly tend to be avoided.[1]

Case-based decision theory captures the importance of the past in shaping an agent's behavior, but still does not deal with beliefs. Instead of linking past events to beliefs about the future, case-based decision theory focusses exclusively on the past—its agents do not engage in any explicit predictions for the future. Realizing the potential of linking past events to beliefs about the future requires a model exhibiting features of both case-based decision theory and expected utility theory.

1.2   Cases and Scenarios in Decisions under Uncertainty

This paper presents a model of decision making that combines and generalizes case-based decision and expected-utility theory. When evaluating an act, our agent looks back at past cases to see how well the act has performed, but also evaluates scenarios describing future events.

In Section 2 we illustrate how the model is intended to generalize case-based decision theory and expected utility theory and to deal with their limitations. Section 3 provides an axiomatic foundation and an attendant representation theorem, showing that an agent characterized by these axioms

---

[1]While the basic idea of reasoning by analogies goes back to Hume (1748) at the latest, the term "case-based reasoning" was coined by Schank (1986).

chooses acts so as to maximize a function that (not necessarily additively) aggregates past cases as well as scenarios.

The axiomatization offered in Section 3 presents a static model of an agent characterized by a fixed assessment of past cases and fixed beliefs about future scenarios. Section 4 introduces learning, examining some simple ways in which the agent may alter her assessments and beliefs in response to experience.

2 Motivation

This section describes the considerations that motivate various aspects of our model, and explains how the model is designed to address these considerations.

2.1 Combining Case-Based and Expected-Utility Reasoning

2.1.1 Black Swans

Following the attack on the Twin Towers on September 11, 2001, the New York Stock Exchange was closed for five days. A day before it was reopened, a prominent market analyst was asked what the Dow Jones Industrial Average would do on the following day. His answer was based on the drop in the Dow following similar attacks on the US, most notably Pearl Harbor. This answer (which proved to be quite accurate, perhaps because other analysts were focusing on the same past cases) was fully case-based.

If we had asked the same analyst for his predictions a week earlier, it is likely that his answer would have been based on scenarios describing future events, with accompanying probabilities. Indeed, his answer could well have been fully Bayesian. However, this Bayesian analysis would almost certainly have given no weight to the possibility of an attack like that of September 11, which would not have been among the scenarios considered by the analyst. Once the attack had occurred, all of the familiar scenarios that had previously provided the core of the analyst's reasoning were inapplicable. At this point, the analyst turned to past experience, looking for similar cases that would allow the analyst to generate predictions.

We would like our model to capture this ebb and flow of case-based and scenario-based reasoning. The ebb and flow arises out of the differing characteristics of these two modes of reasoning. Scenario-based reasoning takes on a relatively ambitious challenge: imagine what will happen in the future, compile a list of scenarios capturing all the important possibilities,

and attach probabilities to each scenario. Such scenarios are powerful aids for making predictions, and so are often useful, but are also vulnerable to being easily falsified, and so may often leave the agent without guidance. By contrast, case-based reasoning is only based on past experience, and makes no explicit predictions. Cases thus often appear to be less useful than do scenarios, but cases cannot be "falsified" and so always provide some guidance to the agent.

After a long period with few surprises, people tend to believe that they have figured out the basic processes that govern the consequences they experience. Consequently, they decrease the weight placed on past cases and become more reliant on future scenarios, possibly up to a point where they become fully Bayesian, effectively taking the scenarios as sufficient statistics for the past cases. However, when a "black swan" appears, the agent's world view is shaken. Very few scenarios, if any, are then compatible with observations and remain in the game. Until new scenarios are developed and tested, the agent falls back on past cases, which have always been there in memory, but which now take on new relevance as the best available tools for making predictions.

2.1.2 A Preliminary Model

Our first step is thus to combine case-based and expected-utility reasoning. We begin with some notation. A case is a quadruple (p, a, o, r) that specifies a problem p, an act a that was chosen in the problem, an eventuality o, and a consequence r. These are drawn (respectively) from the sets P, A, O, and R.

We are imposing more structure on cases here than in the original development of Gilboa and Schmeidler (1995)—we could build case-based decision theory on simply a set of problems P, a set of acts A, and a set of consequences R. We think of an eventuality o and the consequence r as both being realized after an act is chosen, but as capturing different aspects of the agent's experience. The eventuality o is to be thought of as a description of the external world, reflecting an underlying process that the agent thinks can be learned and is "objective". That is, the agent implicitly believes that other agents would describe cases with the same eventualities and would learn the associations between problem-act pairs and eventualities in the same way. By contrast, consequences describe the payoff-relevant outcomes that may have been associated with given eventualities in the past. They are thus inherently subjective. For example, the problem p may be a financial investment problem. The act a taken by the agent may be the signing of a

futures contract to supply oil at a specified price in one year's time. The eventuality o may describe conditions in the oil market in a year, and in particular may specify that there is a worldwide shortage of oil. The consequence r may specify that the combination of committing to supply oil at a fixed price and an oil shortage cause the agent to lose a prodigious amount of money.

Eventualities and consequences will play different roles in learning from past cases. Eventualities can be thought of as carrying that part of the mechanism that the agent believes she has understood. She learns how problem-act pairs relate to eventualities, and uses this knowledge for generating scenarios and reasoning about their likelihood. For this type of learning the consequences that *happened* to be associated with eventualities in the past do not matter; all regularities that are expected to be repeated are already encapsulated in the eventualities. However, if the agent is not sure that she has figured out the causal mechanisms involved, she will also be affected by past cases directly, that is, not mediated by the scenarios she can imagine. Past consequences will then affect the desirability of an act, reflecting the fact that the agent is uncertain of the underlying causal mechanisms. Note that the distinction between eventuality and consequence resides, ultimately, in the agent's mind, and she may well be wrong about drawing the line between them.

An agent considering an act a in a problem $\tilde{p}$ recalls the collection of past cases M_a in which act a was chosen. The influence of a past case (p, a, o, r) on the evaluation of act a in problem $\tilde{p}$ is determined by the similarity of the past problem, p, to the current problem $\tilde{p}$, captured by the function $s(\tilde{p}, p)$. This "similarity" value, which is behaviorally defined, will typically reflect not only intrinsic similarity judgments, but also salience considerations that enhance the prominence of some cases, judgments as to relevance, and so on.

The agent also imagines a collection of scenarios Ω, with each act a inducing a function $a : \Omega \to R$, with $a(\omega)$ interpreted as identifying the consequence that will be realized if act a is chosen and scenario ω is realized. Intuitively, scenarios are the equivalent of states of the world—we could build expected utility theory on simply a set of scenarios Ω, a set of acts A, and a set of consequences R. Scenarios also share many of the mathematical properties of states of the world: they are interpreted as being mutually exclusive, and each scenario defines a consequence for each possible act. Section 2.2.1 explains how scenarios differ from states. The influence of the scenario ω on the evaluation of act a in decision problem $\tilde{p}$ is given by the weight $l(\omega)$.

A preliminary formulation, that allows the agent to take into account both the backward-looking evaluation of past cases and the forward-looking evaluation of future scenarios, calls for the agent to evaluate act a in problem $\tilde{p}$ according to the following:

$$U(a) = \sum_{(p,a,o,r) \in M_a} s(\tilde{p},p)\, u(r) + \sum_{\omega \in \Omega} l(\omega)\, u(a(\omega)), \tag{1}$$

where $u : R \to \mathbb{R}$ is a utility function that maps consequences into utilities. We then imagine the agent choosing that act a which maximizes $U(a)$. We can thus view this as the sum of an additive model of case-based decision theory and an expected utility function. The relative weight placed on expected utility when evaluating act a in problem $\tilde{p}$ is decreasing in the sum of similarities $\sum_{(p,a,o,r) \in M_a} s(\tilde{p},p)$, with case-based decision theory and expected utility maximization as limiting special cases. Notice that the agent's memory is assumed to be the disjoint union of sets of the form M_a for different acts a. By contrast, the set of scenarios, Ω, is the same for all acts—for each scenario $\omega \in \Omega$ and each possible act a, the function $a(\omega)$ describes the consequence of act a in scenario ω.

It is noteworthy that eventualities o make no appearance in (1). Indeed, if the agent has a fixed memory of cases and similarities and a fixed set of future scenarios and likelihoods, eventualities need not be considered—expected utility maximization requires only the scenarios and their likelihoods, and case-based reasoning requires only that the agent associate consequences with the various problems in her memory. The addition of eventualities to the model is necessary in order to capture some forms of learning. In particular, we would like to think of the agent as using her experience with past cases to assess the likelihood of her scenarios and generate new scenarios. This form of induction, often referred to as "abduction", requires that the agent be able to reason about eventualities that have occurred in the world independently of the payoff-relevant consequences that happened to be associated with them, because different consequences might be associated with a given eventuality in the future. For example, the agent needs to consider the possibility of a shortage in oil, while recognizing that this need not entail losing a prodigious amount of money. We first touch on this type of learning in 2.2.2.

2.2 Generalization

We generalize this preliminary model in three ways.

2.2.1 Unforeseen Scenarios

First, despite their formal resemblance to states of the world, scenarios do not share all of the properties of states. Instead, the agent realizes that the list of scenarios need not be exhaustive, nor does each scenario necessarily resolve all uncertainty. Indeed, these lacunae provide part of the motivation for also considering past cases as an additional source of information.[2] Thus, the set of scenarios is a possibly nonexhaustive collection of pairwise disjoint events. We capture this by assuming that the weights attached to scenarios need not sum to unity. The relative weight put on scenarios, out of the total weight on scenarios and cases, now offers a measure of the agent's confidence that she has come up with the important scenarios, relative to the information she has.

2.2.2 Evaluating New Scenarios

Second, the weighting function by which scenarios are evaluated need not be additive. Additivity may fail for two reasons.

The first is familiar—the agent might feel unsure about the probabilities of future scenarios and might accordingly follow a decision model that allows for, say, ambiguity aversion.

Second, we view the likelihoods attached to the various scenarios as reflecting the agent's experience with past cases. In particular, we note that the eventuality o in a case is reminiscent of a scenario ω (a link we develop more carefully in Section 4), and we expect the likelihoods attached to scenarios to reflect the eventualities that have appeared in past cases.

This link from cases to scenarios introduces some basic tension with additivity. Each case tells the story of a single act, but each scenario makes predictions for *all* acts. A case (p, a, o, r) tells us the consequence of act a in problem p and given eventuality o, not what would have been the consequence of other acts had they been chosen in the context of problem p and eventuality o. But once a scenario is added to Ω, the agent must stand ready to identify the consequence of *every* act at this scenario. How does the agent use her observations of cases in which particular acts were chosen to evaluate scenarios that make predictions about what would happen if other acts are chosen? Equivalently, how does she "lump together" cases,

---

[2]The set of scenarios can be guaranteed to be exhaustive (and all scenarios mutually exclusive) if it is defined as all the conceivable truth assignments to a set of propositions. But the agent should be aware of the fact that the propositions she came up with may not exhaust the set of all relevant ones.

in each of a which a single act was chosen, into scenarios, in each of which the consequence of each act is specified?

In some cases one can fill in the required counterfactual information about what would have happened had other actions been chosen to generate an additive weighting function over scenarios. For instance, suppose that every morning Jill has to choose which route to use to get to her office, a or b. She decides to experiment and tries both of them many times to obtain a large enough sample for each. Suppose that a seems to be faster than b on average, as shown by statistical analysis. However, Jill does not know that her selection of days for a and for b was independent of some global events that affect both. For all she knows, it is possible that whenever a was chosen, b would have been faster than a and, whenever b was chosen, a would have been faster. Still, if Jill's choice was random, she can hope that this is not the case. Or, more simply, she can judge all past cases to be sufficiently similar to each other and take the actual consequences experienced in cases in which one act was chosen also as proxies for the counterfactual consequences that act would have yielded in the problems in which it was not chosen.

Sometimes a rational agent can go further and fill in counterfactual consequences even when nearly identical repetitions are not available. Suppose, for example, that Jack is a small investor in the stock market. Every day he holds a particular portfolio and gets information about the relevant consequence, namely, the monetary value of this portfolio. However, with the same ease he can also figure out what would have been the value of any other portfolio he could have held. Even though no two trading days are identical, the mere assumption that Jack is too small to affect market prices implies that every day he obtains an observation of many cases, one of which is actual and the others of which are counterfactual. Augmenting his memory with counterfactual cases, Jack can use empirical frequencies to come up with an additive likelihood function on scenarios.

In other cases, these sorts of exercises would be highly speculative, and counterfactuals are difficult to reason about based on hard evidence. We do not know how World War II would have ended had Hitler crossed the channel. We do not know how the 1929 financial crisis would have evolved had governments assumed a more active role in dealing with it. Experts and historians might surely have educated guesses about these, but the "hard evidence" memory provides does not suffice to evaluate all acts at a given scenario. Even if all past problems are equally similar to the present one, if act a resulted in consequence r we can only say that the resulting case is equivalent to a *set* of scenarios: all those compatible with a yielding r (and

8

with all other acts yielding some consequences).

It follows that, even if we assume that the current problem is equally similar to all past ones, observations of past cases should in general be viewed as providing evidence on *sets* of scenarios rather than on specific scenarios. The data provided by past cases can therefore be thought of as the Möbius transform of a belief function a la Dempster (1967) and Shafer (1976). Such belief functions are monotone set functions that are not necessarily (and typically are not) additive, and we accordingly do not require the weighting function attached to scenarios to be additive.

2.2.3 Rule-Based Similarity

Third, the similarity function by which past cases are evaluated need not be additive. For example, standard techniques in statistics take into account the size of the database in conducting statistical inference (such as forming confidence sets), but the impact of the size of the database is not additive.

Case-based decision theory (Gilboa and Schmeidler, 1995) offered two extreme ways in which past cases can be summarized for the evaluation of an act in problem $\tilde{p}$: one by addition, and the other by averaging. The additive formula is the first term in (1):

$$\sum_{(p,a,o,r)\in M_a} s\left(\tilde{p}, p\right) u\left(r\right) \tag{2}$$

whereas the averaged one is:

$$\sum_{(p,a,o,r)\in M_a} \frac{s\left(\tilde{p}, p\right)}{\sum_{(p,a,o,r)\in M_a} s\left(\tilde{p}, p\right)} u\left(r\right) \tag{3}$$

assuming that the denominator does not vanish.

Both formulas can only serve as rough approximations to the way agents learn from past cases. In (2) the simple addition has a strong flavor of habit formation. For example, assume that all past problems are equally similar to the present one, and act a has been chosen 100 times, yielding a consequence with utility 2 each time, while act b has been chosen 1000 times, yielding a consequence with utility 1 each time. The latter (act b) would be the maximizer of (2), while the former (act a) seems like the obvious choice of any agent who would be deemed rational.

Alternatively, the formula (3) deals with the above challenge simply and intuitively by taking the (similarity-weighted) average utility for each act. However, the formula (3) does not behave smoothly, nor even intuitively,

when the sum of past similarities is close to 0. Given the fact that case-based decision theory was developed in order to deal with novel situations, this is a major difficulty.

The present proposal can deal with these issues more elegantly than either (2) or (3). In place of the similarity function $s : P \times P \to \mathbb{R}$ defined on $P \times P$, we assume that for each action a there is a potentially non-additive monotone function v_a that assigns weights to sets of cases in which action a is chosen.[3] This allows us to say that 1000 cases weigh more than do 100 cases, though not necessarily by a factor of 10. For example, the function v_a can be the lower envelope of a set of measures, where the set is defined by a confidence set as in classical statistics.

A nonadditive set function can also capture our intuitive reasoning when generalizing cases into rules. Suppose, for example, that act a can result in a success or failure. Assume further that it has been used 100 times in seemingly-identical problems, and resulted in success in all of them. It seems natural that the agent feels rather confident that a would yield a success next time as well. This can be captured by statistical inference, as described above, but also by simple induction: at some point the agent formulates in her mind the general rule "a always yields success" and starts having growing confidence in this rule. If, however, a fails on the 101th trial, the general rule is proved wrong. Act a still has a very impressive track record, and the confidence set for its consequence will still be tightly focused on success. But the possibility that a always succeeds, that it is a rule of nature that it would succeed, is lost forever. The agent is forced to accept the fact that there might be exceptions.

In such a situation, we can allow the set function v_a to reflect both the apparent certainty obtained by 100 cases out of 100, as well as the shadow of doubt cast by a single counterexample. Importantly, to this end we have to make v_a dependent on the set of problems in memory (as the same 100 cases would be assigned a different weight if there aren't any additional ones than if there are).[4] The resulting function v_a will in general not be additive: fix the 100 past cases in which a was chosen, and vary only its consequences. With any 99 successes out of the 100, a should be reasonably trusted to succeed. But with 100 successes out of 100, there is the additional support

---

[3]We work with one set function v_a for each act $a \in A$. Analogously, though the set function s is defined on $P \times P$, when evaluating act a the only relevant aspect of s is its specification on pairs of problems in which a was chosen.

[4]Indeed, the axiomatic derivation in Section 3 will be provided for a given set of problem-act-eventuality triples, considering different potential consequences but not different sets of problems (or of acts chosen in them, or of resulting eventualities).

deriving from the universal quantifier. Indeed, it might not even occur to the agent that a might fail, just as it doesn't occur to her that the sun might not rise tomorrow.

More generally, the support that a set of cases yields to a certain conclusion is obtained from two forms of reasoning: case-to-case induction, which is captured in the additive formula (2), but also case-to-rule induction. The latter cannot be captured by simple addition of weights assigned to cases, as it involves a different space, namely the space of general rules. We do not model this space explicitly here, but capture its implications in the form of the non-additivity of the set function on past cases.

2.3 The Decision Making Model

To capture the considerations discussed in the preceding subsections, we generalize (1) as follows. Given a problem p, a memory M_a specifying the cases in which each act $a \in A$ was chosen, and a set of scenarios Ω, we assume there exists a monotonic set function v_a on 2^{M_a} (for each a) and a monotonic set function ν on 2^Ω such that act a is evaluated by

$$U\left(a\right) = \int_{M_a} u\left(r\right) dv_a + \int_\Omega u\left(a\left(\omega\right)\right) d\nu \tag{4}$$

where u is the utility function defined on consequences. Once again we then imagine the agent choosing that act a which maximizes $U(a)$. This is the representation that is axiomatized in Section 3.

Notice that our model allows different acts a to have different values for $v_a\left(M_a\right)$. In particular, we need not have $v_a\left(M_a\right) = 1$, and may even have $v_a\left(M_a\right) = 0$ for some or all acts—the agent may have no experience with some acts. Hence, v_a is not necessarily a capacity. Similarly, the model allows for the possibility that $\nu(\Omega) \neq 1$ or that Ω is empty or that ν vanishes—the agent may have no inkling of relevant scenarios. Hence, ν is also not necessarily a capacity.[5] In addition, we expect the functions v_a to depend on the problem at hand. The agent may have many cases relevant to choosing act a in some problem p, in which case $v_a(M_a)$ will be large, but may view none of these cases as informative when choosing act a in another problem $\tilde{p}$, in which case $v_a(M_a)$ may be zero. The similarity function in (1) makes this dependence explicit, while it appears only implicitly in

---

[5]Note that the Choquet integral relative to such a set function is well-defined (and equals zero). When $\nu(\Omega) = 0$, we have a familiar model of case-based reasoning, and so we simplify the presentation in Section 3 by focussing on the interesting case in which $\nu(\Omega) > 0$.

(4). The latter convention follows the example of expected utility theory. In the latter, the specification of an act $a(\omega)$ depends on the problem at hand—selling oil short may have salutary effects in the scenario that the price of oil collapses if the problem is a portfolio investment problem, but detrimental effects if the problem is one of securing fuel for a shipping line. Our combination of case-based and expected-utility theory thus follows the notational conventions of the latter.

2.4 Learning

The representation given by (4) combines case-based and expected-utility theory, but does not yet allow us to examine the ebb-and-flow between them. To do this, Section 4 introduces learning. It is here that the eventualities in our model pull their weight. There are three ways in which learning takes place in our model.

First, the agent may try an act a in a problem p, observe an eventuality o and consequence r, and add a case (p, a, o, r) to her memory. The function v_a would then have to be reassessed, as its domain M_a has expanded. In the simplest version of the model nothing else would change—neither the counterparts of the function v_a attached to other acts nor the evaluation ν of future scenarios.

Second, the agent may learn of a new fact, which rules out certain scenarios. In this case she would be expected to have a smaller set of scenarios, consisting only of those that were not ruled out by evidence. This type of learning is akin to Bayesian updating, with the important distinction that the function ν need not be re-normalized to satisfy $\nu(\Omega) = 1$. Rather, in this type of learning the relative importance of scenarios is expected to decrease.

Third, the agent may use past cases as a motivation for attaching weight to new future scenarios. Here, past eventualities play the role of manifestations of actual scenarios realized in the past. This process requires inductive reasoning, abstraction from details, and imagination in combining parts of past cases into new sequences of occurrences that form the inspiration for considering scenarios that have previously be ignored or even unimagined. In our model, it will be reflected in an updating of the function ν: new scenarios ω are generated and added to Ω. We can then expect $\nu(\Omega)$ to increase at the expense of the importance of past cases. Even if the intrinsic similarity between a past case and a present problem is unchanged, the relevance of the past case would be reduced if the agent views the new scenario as adequately capturing the information contained in that case.

3   The Model

3.1   Setup

We adapt the framework of Gilboa and Schmeidler (1995). Let P and A be finite and nonempty sets of *problems* and *acts*, respectively. For each $p \in P$, there is a nonempty set O_p of possible *eventualities* associated with p. Denote by $O = \cup_{p \in P} O_p$ the set of all eventualities. Let R stand for the set of *consequences*.

The set of conceivable *cases* is given by

$$C = P \times A \times O \times R.$$

Thus, a generic case $c = (p, a, o, r)$ identifies a problem p faced by the agent, an action a chosen in p, an observed eventuality o, and a resulting consequence r. Importantly, the eventuality-consequence pair (o, r) realizes after the action has been chosen and represents a comprehensive description of the outcome arising from case c.

A *memory* is a finite subset $M \subseteq C$ containing all past cases that actually occurred. For each $p \in P$, there is at most one triple (a, o, r) such that $(p, a, o, r) \in M$. Equivalently, we view every case in the memory M as involving a distinct problem. This can be assured by making the description of problems sufficiently rich. It imposes no restrictions on our agent, as she is always free to view two distinct problems as subjectively identical.

For each $a \in A$, $M_a \subseteq M$ denotes the set of past cases in which $a \in A$ was chosen. We allow for the possibility that $M_a = \varnothing$ for some or all a. For every $a \in A$, let

$$H_a = \{(p, o) \in P \times O \,|\, \exists r \in R, \text{ such that } (p, a, o, r) \in M_a\}$$

be the set of problem-eventuality pairs in which a was chosen and let $H = \cup_{a \in A} H_a$ be the set of all problem-eventuality pairs that appear in the agent's memory.

Let Ω be a finite and nonempty set of future *scenarios*. Each scenario specifies the predicted consequence in R for each act, as is commonly assumed about states of the world. We refer to functions $y \colon \Omega \to R$ as *future profiles* induced by acts and assume that all future profiles are conceivable by the agent.

An act a defines a *past profile* by the consequences it has yielded: $x \colon H_a \to R$. We will assume that the agent has preferences not only over acts with actual past profiles, but also with hypothetical ones. We denote by

$$\mathcal{H}_a = \{x \,|\, x \colon H_a \to R\}$$

the set of all *hypothetical past profiles* defined over the experienced problem-eventuality pairs in which a was chosen. A hypothetical past profile is generated from an actual profile by keeping fixed the pairs of problems and eventualities experienced (i.e., the set H_a) and letting vary the resulting consequences.

We can draw an analogy between future scenarios and past eventualities that helps motivate our interest in hypothetical past profiles. We view a past eventuality $o \in O$ as a possible scenario describing the payoff-relevant uncertainty at the time in which the agent faced problem p. That is, if the agent looks back at a past case (p, a, o, r) and asks herself how she viewed the world at problem p, then the eventuality o would have been a possible scenario, and r a possible consequence associated to act a.[6] Then, in the same spirit as standard decision-theoretic models require that any future scenario can be associated to any consequence, giving rise to all conceivable future profiles, we also assume that any past eventuality can be associated to any consequence, not just the realized one, giving rise to all hypothetical past profiles. This amounts to saying that the agent is able to reason about counterfactual consequences in terms of both past and future reasoning. Subsection 3.2 will show that hypothetical profiles are needed in our axiomatic derivation for the same reason Savage's model maintains that any consequence is compatible with any state of the world.

We will assume that the agent has preferences over all pairs of past-and-future-profiles. That is, the agent is assumed to have meaningful preferences between vectors of the form

$$
\begin{aligned}
f &= (x, y) \colon H_a \cup \Omega \to R \\
g &= (z, w) \colon H_b \cup \Omega \to R
\end{aligned}
$$

interpreted as follows: "Assume that act a, when chosen, resulted in the past profile x, and that it is guaranteed to yield profile y in the future scenarios. Assume further that act b, when *it* was chosen, resulted in the past profile z, and that it is guaranteed to yield profile w in the future scenarios. Under these circumstances, would you rather choose a or b?"

Thus, for each act $a \in A$, we consider *profiles* of the form $f = (x, y) \colon H_a \cup \Omega \to R$ where $x \in \mathcal{H}_a$ is a hypothetical past profile and $y \in R^\Omega$ is a future profile. Denote all such profiles by

$$
\mathcal{F}_a = \{f \mid f \colon H_a \cup \Omega \to R\} = R^{(H_a \cup \Omega)}
$$

---

[6]Differently from scenarios, the agent may have failed to envisage an eventuality at the time of her choice. As shown later, only the realized eventualities will play a role in the axiomatic derivation.

for $a \in A$. Finally, define

$$\mathcal{F} = \cup_{a \in A} \mathcal{F}_a.$$

Our primitive is a binary relation $\succsim_{H,\Omega}$ on $\mathcal{F}$ which depends on the history H of experienced problem-eventuality pairs and the set Ω of envisaged scenarios.

We do not assume that this binary relation is complete. In particular, we do not assume the agent ranks two profiles such as $f = (x, y), f' = (x', y') : H_a \cup \Omega \to R$ that are defined on the *same* set of problems. In essence, the agent is not assumed to imagine that, over the very same set H_a the results obtained were simultaneously x and x'. However, under some richness conditions, preferences between such profiles f and f' will follow from preferences between profiles defined over different acts and transitivity.

The set of future profiles R^Ω is a subset of all profiles, and it represents the profiles of acts with empty histories. Because preferences are defined directly on profiles, we implicitly assume that two acts with empty histories and the same predicted consequences are identical. The restriction of $\succsim_{H,\Omega}$ to R^Ω is denoted by $\succsim_\Omega$. Moreover, we will use the notation $\succsim$ to refer to a binary relation defined on R as usual: For every $\alpha, \beta \in R$, $\alpha \succsim \beta$ if and only if $(\alpha, \alpha, \ldots, \alpha) \succsim_\Omega (\beta, \beta, \ldots, \beta)$. For $\alpha \in R$, the element $\boldsymbol{\alpha}$ stands for the constant profile which yields α on the appropriate domain.

For $a \in A, f \in \mathcal{F}_a, \alpha \in R$, and $s' \in H_a \cup \Omega$, we denote by $\alpha\{s'\}f$ the profile in $\mathcal{F}_a$ which yields α in s' and $f(s)$ for all $s \neq s'$. This definition can be used recursively, so that, for $\alpha, \beta \in R$ and $s', s'' \in H_a \cup \Omega, s' \neq s''$, the profile $\alpha\{s'\}\beta\{s''\}f$ in $\mathcal{F}_a$ yields α in s' and $(\beta\{s''\}f)(s)$ for all $s \neq s'$.

For any $a \in A$, we say that $s \in H_a \cup \Omega$ is *null* on $F \subseteq \mathcal{F}_a$ if $\alpha\{s\}f \sim_{H,\Omega} f$ for all $\alpha\{s\}f$ and f in F.

A few mathematical notions are needed. First, we recall that a capacity is a monotone and normalized set function that is not necessarily additive. That is, a *capacity* on a finite set S is a set function $v : 2^S \to [0, 1]$ such that: (i) $v(\varnothing) = 0$; (ii) $A \subseteq B$ implies $v(A) \leq v(B)$; and, (iii) $v(S) = 1$. We say that a set function $v : 2^S \to \mathbb{R}_+$ is a *pseudo-capacity* if it satisfies conditions (i) and (*ii*). Second, we assume throughout the following:

Assumption 1 *The set R is a connected topological space and the set $\mathcal{F}$ is endowed with the product topology.*

3.2 Axioms on Preferences

We assume that the set of acts is nontrivial and that there is at least one act that was never chosen in the past:

Assumption 2 (Richness) *There exist at least two acts, and at least one $a' \in A$ such that $H_{a'} = \varnothing$.*

We impose the following axioms on $\succsim_{H,\Omega}$. The first is the weak order axiom, restricted to profiles that belong to different acts. The second is a monotonicity assumption, stated in a way that takes into account the possibility that preferences may not be defined between profiles that belong to the same act. The continuity axiom is standard.

Axiom 1 (Restricted Weak Order) *The binary relation $\succsim_{H,\Omega}$ on $\mathcal{F}$ is reflexive and transitive. For every $a, b \in A, a \neq b$, every $f \in \mathcal{F}_a$ and $g \in \mathcal{F}_b$, $f \succsim_{H,\Omega} g$ or $g \succsim_{H,\Omega} f$.*

Axiom 2 (Monotonicity) *For every $a, b \in A, a \neq b$, $f, f' \in \mathcal{F}_a$ with $f(s) \succsim f'(s)$ for all $s \in H_a \cup \Omega$, and every $g \in \mathcal{F}_b$,*

- *$f' \succsim_{H,\Omega} g$ implies $f \succsim_{H,\Omega} g$,*

- *$g \succsim_{H,\Omega} f$ implies $g \succsim_{H,\Omega} f'$.*

Axiom 3 (Continuity) *For every $a, b \in A, a \neq b$, and every $f \in \mathcal{F}_a$, the sets $\{g \in \mathcal{F}_b \colon f \succ_{H,\Omega} g\}$ and $\{g \in \mathcal{F}_b \colon g \succ_{H,\Omega} f\}$ are nonempty and open in $\mathcal{F}_b$.*

Observe that, given an act a' with an empty history (whose existence is ensured by Assumption 2), Axiom 3 implies that there exists at least one non-null scenario, and hence $\nu(\Omega) > 0$. This can be relaxed by weakening Assumption 2 to allow all scenarios to be null, though we would then need an alternative condition to ensure uniqueness of our representation.

The following definition will be used to state the next axiom. It is the standard definition of pairwise comonotonic sets of profiles, apart from the fact that, in our case, comonotonicity will not involve comparisons of pairs of past problems and realized eventualities with future scenarios. Thus, two profiles are comonotonic if they do not rank any two problem-eventuality pairs differently, nor any two future scenarios. (See Appendix A for details.)

Definition 1 *For any $a \in A$, a set of profiles in $\mathcal{F}_a$ is pairwise comonotonic if there are no two profiles f and g in the set such that*

$$f(p, o) \succ f(p', o') \text{ and } g(p, o) \prec g(p', o')$$

for some $(p, o), (p', o') \in H_a$, or

$$f(\omega) \succ f(\omega') \text{ and } g(\omega) \prec g(\omega')$$

16

for some $\omega, \omega' \in \Omega$.

We can now state the version of the tradeoff consistency axiom we will need for our representation. This condition strengthens the one used in Köbberling and Wakker (2003) by considering pairwise comonotonic sets.

Axiom 4 (Pairwise Comonotonic Tradeoff Consistency) *For every $a \in A, f, f', g, g' \in \mathcal{F}_a, \alpha, \beta, \gamma, \delta, \delta^* \in R,$ and $s, s' \in H_a \cup \Omega$, if*

$$\alpha\{s\}f \sim_{H,\Omega} \beta\{s\}f', \quad \gamma\{s\}f \sim_{H,\Omega} \delta\{s\}f', \quad \alpha\{s'\}g \sim_{H,\Omega} \beta\{s'\}g', \quad then$$
$$\gamma\{s'\}g \sim_{H,\Omega} \delta^*\{s'\}g' \quad \Leftrightarrow \quad \delta \sim \delta^*$$

whenever $\{\alpha\{s\}f, \beta\{s\}f', \gamma\{s\}f, \delta\{s\}f'\}$ and $\{\alpha\{s'\}g, \beta\{s'\}g', \gamma\{s'\}g, \delta^*\{s'\}g'\}$ are pairwise comonotonic sets, and s and s' are non-null on the first and second set, respectively.

The next axiom guarantees the existence of a neutral consequence that is independent of the choice of act $a \in A$. To understand its meaning, notice that when we compare an act $y \in R^{\Omega}$, with no history, to acts that have the same future profile but different histories, it would make sense that one of the two will hold: either (i) past profiles do not affect the act's desirability (that is, pairs in H_a are null), which will be the case if agent believes that all the information in past cases is already incorporated into the likelihood of future scenarios; or (ii) some past histories make the act more attractive and some make it less so. In either case, one would expect there to be a consequence $\alpha^* \in R$ such that the constant past profile $\boldsymbol{\alpha}^*$ makes y just as desirable as it would be with no history at all, that is, $(\boldsymbol{\alpha^*}, y) \sim_{H,\Omega} y$. (In case (i) this would be the case for any α^*, and in case (ii) it would follow for some α^* by continuity.) Such a consequence α^* is reminiscent of the aspiration level in the satisficing model of Simon (1957): having better consequences throughout the past would make the act look desirable, and the agent would tend to keep choosing it; having worse consequences by having used this act would make the agent try to explore new paths.

One could think of models in which this "aspiration level" depends on the act a under discussion and/or on the future profile y. Having different aspiration levels for different acts a might occur if the agent has a certain intrinsic preferences for some acts over others. For example, an agent might have preferences for the labels associated with some acts. However, in our model we assume that the agent is consequentialist, in the sense that only past consequences matter. Thus, we wish to require that the consequence α^* be independent of the act under consideration. Moreover, in line with

the problem-scenario separability, we also require that this "neutral" consequence be independent of the future profile y:

Axiom 5 (Act Independent Aspirations) *There exists $\alpha^* \in R$ such that, for every $a \in A$ and $y \in R^\Omega$, the vector $(\boldsymbol{\alpha^*}, y) \in \mathcal{F}_a$ satisfies $(\boldsymbol{\alpha^*}, y) \sim_{H,\Omega} y$.*

3.3 The Representation Result

We can now state our representation result, which combines past-based and future-based reasoning in a single criterion.[7]

Theorem 1 *Let $\succsim_{H,\Omega}$ be a binary relation on $\mathcal{F}$. Assume that Richness holds and that there exists a comonotonic set in R^Ω with at least two non-null scenarios. The following statements are equivalent:*

1. *$\succsim_{H,\Omega}$ satisfies Restricted Weak Order, Monotonicity, Continuity, Pairwise Comonotonic Tradeoff Consistency, and Act Independent Aspirations;*

2. *There exist a continuous function $u \colon R \to \mathbb{R}$ such that $0 \in int(u(R))$, a pseudo-capacity v_a on 2^{H_a} for each $a \in A$, and a capacity $\nu \colon 2^\Omega \to [0,1]$, such that, for all $a, b \in A, f \in \mathcal{F}_a, g \in \mathcal{F}_b$,*

$$f \succsim_{H,\Omega} g \iff \int_{H_a} u(f)dv_a + \int_\Omega u(f)d\nu \geq \int_{H_b} u(g)dv_b + \int_\Omega u(g)d\nu. \tag{5}$$

Notice that we require ν be a capacity, despite stressing in our discussion of (4) that ν need only be a pseudo-capacity. However, as we noted in our discussion of Axiom 3, we are focussing on the case in which $\nu(\Omega) > 0$, so that we are not simply reproducing case-based reasoning. Once we do this, it is only a convenient normalization to assume that ν (or indeed, v_a for any $a \in A$) is a capacity.

A cognitive interpretation of the underlying decision process is that the agent uses past eventualities as a source of information to assess the future. If the agent believes that the informational content in the past eventualities lends support to an exhaustive theory of the structure of the world, then she will transfer all the weight from cases to scenarios and evaluate profiles

---

[7]Notice that we could just as well define the pseudo-capacity v_a on M_a, as we did in (4) in order to avoid prematurely introducing the notation H_a.

only by means of her future-based reasoning. However, this is an extreme case. In many decision problems, one can expect that past cases will not be sufficient to fully understand all possible causal relationships. In these situations, the agent still reasons in terms of scenarios, and transfers some weight from past cases to future scenarios, reflecting what she understood from the past, but also leaves some weight on past cases. The latter weight is a measure of the component that she could not explain yet—that is, it reflects the uncertainty about how the future is likely to be similar to the past.

Our utility representation has uniqueness properties that are almost precisely the standard ones. The only small difference is that the utility function cannot generally be shifted by any constant: as long as there are some acts a for which H_a isn't a null set, one has to make sure that the utility function assigns the value 0 to these consequences α^* that satisfy the condition of Act Independent Aspirations. Thus, one would typically expect there to be a constraint $u(\alpha^*) = 0$, which makes the utility function unique only up to a unit of measurement (without freedom in setting the value 0). Only in the degenerate (but important) case in which all the information is incorporated in ν, and where all of H_a are null, does one regain the freedom to shift the utility function as in the classical models. More explicitly:

Proposition 1 Assume that Richness holds and that there exists a comonotonic set in R^Ω with at least two non-null scenarios. Two triples $(u, \{v_a\}_{a\in A}, \nu)$ and $(\hat{u}, \{\hat{v}_a\}_{a\in A}, \hat{\nu})$ represent the same binary relation $\succsim_{H,\Omega}$ as in Theorem 1 if and only if (i) $\hat{v}_a = v_a$ for all $a \in A$, and $\hat{\nu} = \nu$; and (ii-a) there exists $a \in A$ such that H_a isn't null, and there exists $\lambda > 0$ such that $\hat{u} = \lambda u$; or (ii-b) for all $a \in A$, H_a is null, and there exists $\lambda > 0$ and $d \in \mathbb{R}$ such that $\hat{u} = \lambda u + d$.

The proofs are given in Appendix B.

4   Learning

We mentioned in Section 2.4 that there are three types of learning that may occur in this setting. First, new cases are continually added to the memory. Second, new evidence may exclude some scenarios from the support of the pseudo-capacity ν (no longer insisting on the normalization with which we simplified Theorem 1). Third, the agent may use past cases to motivate including additional scenarios with the support of ν, a process that we will refer to as induction. This section introduces and explores a model of

learning. We begin with three simple examples of the connection between scenarios and past cases that illustrate key aspects of the learning process, and present a more general learning model.

4.1 Example 1: Counting

We first consider a case in which the agent continually revises her weighting function ν, in the process remaining confident enough of this function that she need devote *no* weight to cases. Suppose that in each period $\tau \in \{1, 2, \ldots\}$, the agent must choose an act $a \in A$ where A is finite. The agent views an act a, chosen in problem p, as inducing a distribution over eventualities $o \in O$ and consequences $r \in R$, where O and R are also finite. Suppose that the agent believes that a given act a induces the same distribution over eventualities in every problem, but does now know these distributions. The agent does know how the combination of an act and an eventuality map into a consequence, and hence is concerned only with learning the distribution over eventualities induced by each act. For example, an act may be a choice of how to travel to work, an eventuality may include traffic conditions and other factors that determine whether the agent is late to work, the problem may specify whether the agent has an early-morning performance review or product presentation to her boss, and the consequence may be that the agent gets demoted or invited to head a new division of the firm.

The agent looks at past cases with the purpose of learning the distribution over O that each act a induces. The agent can summarize her beliefs about this distribution in the form of scenarios. In particular, let a scenario ω be represented by a function $h_\omega : A \to O$ associating eventualities with actions, with the set of scenarios given by

$$\Omega = \{\, \omega \mid h_\omega : A \to O \,\} \equiv O^A.$$

Each scenario ω associates an eventuality with each act a, according to the obvious relationship $a(\omega) = h_\omega(a)$. Each case (p, a, o, r) indicates that a particular act, a, resulted in a particular eventuality, o. Such a case says nothing about the eventualities that might have followed from other acts. Thus, it does not provide support for any particular scenario $\omega \in \Omega$ but only to the event

$$[a, o] \equiv \{\, \omega \in \Omega \mid h_\omega(a) = o \,\}.$$

Because the weights are assigned to such subsets rather than to singletons, the agent's beliefs about her scenarios are summarized not by a probability

measure, but (as noted in Section 2.2.2) by a belief function a la Dempster (1967) and Shafer (1976). Specifically, if act a was chosen n_a times and out of these $n_{a,o}$ times resulted in eventuality o, one may define two candidates for the belief function ν over Ω: the un-normalized one, whose Möbius transform is given by

$$m\left([a, o]\right) = n_{a,o}$$

for all $a \in A$, $o \in O$ (and zero on all other events), and the normalized one, defined by

$$m'\left([a, o]\right) = \frac{n_{a,o}}{n_a}$$

whenever $n_a > 0$ (and, again, zero for all other events).

Let ν_m be the belief function defined by a Möbius transform m, that is

$$\nu_m\left(S\right) = \sum_{T \subset S} m\left(T\right)$$

for any $S \subset \Omega$. Then, the resulting capacities, ν_m and $\nu_{m'}$ both reflect the distribution of the eventualities for each act. Neither attempts to divide the mass of evidence among scenarios within events of the form $[a, o]$.[8]

The difference between the normalized and un-normalized capacities will be reflected in the impact of the number of times each act has been chosen: the un-normalized version, ν_m, would take this number into account, exhibiting habit formation for positive utilities and variety seeking for negative utilities. By contrast, the normalized version, $\nu_{m'}$, would ignore the numbers n_a and follow only relative frequencies.

Importantly, in both versions the agent loses nothing by putting zero weight on the original cases. As she believes that the environment is stationary and the past is only a source of information for the future, once the cases are counted there is no need to recall them specifically. If, for example, $n_{a,o} = 10$, then the agent considers the mass $m\left([a, o]\right) = 10$ (or $m'\left([a, o]\right) = 10/n_a$) as a sufficient statistic. Once the states are conceived of and the statistical information is encapsulated in the capacity over them, there is no need to retain the specific cases (observations) that gave rise to these statistics.

---

[8]Indeed, this is also not needed for decision making in this context: if the agent only wishes to maximize her expected utility, she should choose an act with the highest expected utility relative to the act's marginal distribution, irrespective of the joint distribution of all acts. Thus, maximizing the Choquet expected utility relative to ν_m or $\nu_{m'}$ would be tantamount to choosing the act with the "best" marginal expected utility, and the non-additivity of the capacity would not be observable.

4.2 Example 2: Theorizing

We now present a contrasting example in which cases retain their influence. Assume that stationarity is discarded completely, but the set-up is simplified to be a prediction problem. In each period $\tau \in \{1, 2, \ldots\}$, the agent predicts an eventuality in O and then observes a case. She believes that her actions have no effect on eventualities, and she obtains a consequence that reflects the accuracy of her prediction. Hence, she cares only about the value of o in each period, and we can think of her as observing a process $\{o_\tau\}_{\tau=1,2,\ldots}$ and attempting to predict o_t based on $(o_\tau)_{\tau<t}$. When thinking about the past, there is no need for her to recall her past actions (guesses), or anything else besides the sequence of observations $\{o_\tau\}_{\tau<t}$.

Imagine now that scenarios consist of theories that the agent can come up with. For simplicity, assume that the agent can conceive of k theories about the process, where k is intended to capture some cognitive limitation. Each theory provides a deterministic prediction for each τ. Theories that have been refuted by the data are assumed to be discarded, leaving room for new ones. For concreteness, the theories can be thought of as Turing machines, where the agent entertains the k simplest hitherto-unrefuted theories, i.e., the k unrefuted machines whose minimal description length is the lowest. Among these k theories, we may assume that each theory is assigned a weight that is decreasing in its (Kolmogorov) complexity, and that these weights are used additively.

In such a model the scenarios part may well behave in an additive way, along the lines of expected utility theory. However, there is no compelling reason to put all the weight on scenarios and discard past cases completely. Rather, the agent may realize that there exist other theories are yet to be conjured up, and that the process she is observing may not follow any theory (see Gayer and Gilboa, 2013). Hence, she may find it safer to leave some weight also on the case-based mode of reasoning, which may complement any rule-based reasoning that she can embed in scenarios.

4.3 Example 3: Parametric Statistical Inference

Assume that the agent is faced with a statistical inference problem. She is faced with a prediction problem as in Section 4.2, but she now assumes that past cases are a result of an i.i.d. process, characterized by some vector of parameters θ. That is, to the extent that a given observation affects future one, all that matters for the latter is incorporated in the agent's estimate of θ.

In this application, scenarios are possible values of the parameters θ characterizing the unknown distribution. A very special case would be the estimation of a parameter of a Bernoulli random variable, where past cases are realizations of 0's and 1's. This would also be a special case of the counting example above, where the act is suppressed.[9]

Point estimation in classical statistics—as, say, maximum likelihood estimators—would yield a learning algorithm that uses past cases and singles out one of the parameter values.[10] A confidence set, by contrast, could be viewed as a capacity that puts weight 1 on every set of parameters that contains the confidence set, and 0 on others.[11] Bayesian statistics would have a prior over the scenarios (i.e., over the values of the parameters), and update it by Bayes rule given the past cases.

Our model can thus encompass familiar statistical techniques as specifications of how beliefs over scenarios are generated based on past cases. When we add decision making to the problem, mimicking Bayesian statistics would lead to expected utility maximization, whereas adopting confidence intervals would lead to a maxmin expected utility model. Importantly, however, our model does more than simply suggest that we should see decision making based on expected utility or maxmin expected utility. In this setting as in Section 4.2, we expect the agent to retain some weight put on past cases, reflecting the agent's doubt in her own statistical reasoning.

4.4   A Learning Model

As before, we let P, O, and A be finite sets of problems, eventualities, and actions. We imagine an agent who is characterized by a memory M_τ and faces a problem p_τ. In general we need place no structure on the memory M_τ,

---

[9]This class of problems will typically involve an uncountable set of scenarios, each specifying the value of all relevant parameters. While this assumption is in conflict with our assumption of a finite set of scenarios, it is rather straightforward to extend our axiomatic derivation to this case. The axiomatic derivation of Section 3 can easily be extended to this case, relying on the tools developed in Köbberling and Wakker (2003).

When we think of scenarios as descriptions of the way an economic or political process may evolve, it seems reasonable that the agent can only conceive of finitely many scenarios. However, when the scenarios are values of a continuous variable, such as the expectation of a random variable, there is no difficulty in imagining a continuum of scenarios.

[10]Here, again, there is a slight deviation from our axiomatic model, which assumed that there are at least two non-null states. This, however, is a technical detail that can easily be dealt with using alternative richness conditions.

[11]Maximizing Choquet expected utility with respect to such a capacity is equivalent to maximizing the minimal expected utility with respect to all parameters values in the confidence set.

but it helps to fix ideas to work with a particular case, for which the notation is well suited. Consider a sequence of time periods $\{1, 2, \ldots\}$, indexed by τ. In each period τ the agent encounters a problem $p_\tau \in P$, chooses an act $a_\tau \in A$, and then observes the eventuality $o_\tau \in O$ and the consequence $r_\tau \in R$.

As in Sections 4.2 and 4.3, we focus on a prediction problem, in which the agent attempts to learn the nature of a process that she assumes is independent of her actions. While acts and their associated consequences can be designed to generate bets that may be useful eliciting the agent's belief, both the acts (predictions or bets) and their consequences are immaterial for learning. Hence all that matters for learning are the problems and the eventualities that accompanied them. We can thus think of the period-τ memory as a sequence of observations $((p_1, o_1), \ldots, (p_{\tau-1}, o_{\tau-1}))$.

Each period, the agent encounters a new case to be added to her memory, and she also has some new evidence that is potentially useful in revising the capacity ν she attaches to her scenarios. We will normalize so that $\max_{a \in A} v_a = 1$, and focus on the revision of ν.

A *theory* is a capacity $\mathfrak{v}$ on the set $\Omega = P \times O$ of pairs of problems and eventualities. We let V be the set of such theories. We could imagine a more complicated setting in which the agent does not know the sets P and O, and indeed cannot even imagine some elements of these sets. We view such considerations as quite realistic, but also believe that they can be captured in the current setting by allowing capacities $\mathfrak{v}$ that do not have full support.

The agent now proceeds as follows. In each period τ, she observes a current problem p_τ and recalls her memory M_τ. The agent then applies an evaluation rule to the theories in V. The evaluation rule collects some data from the memory M_τ, and then applies an evaluation to each theory. Given the environment we are working in, an obvious and simple evaluation rule is to chose some subset of the memory and then to assign to each theory the likelihood of that subset. For example, an evaluation method may choose (only) the previous period's problem $p_{\tau-1}$ and eventuality $o_{\tau-1}$ and then attach to each theory $\mathfrak{v}$ the evaluation

$$\mathfrak{v}(p_{\tau-1}, o_{\tau-1}).$$

Alternatively, a theory may consult the last n periods, evaluations of the form

$$\prod_{k=\tau-n}^{\tau-1} \mathfrak{v}(p_k, o_k).$$

24

Why might an agent not use all the data in the memory? We have made no assumptions about the process that generates problems and eventualities. The agent may be concerned that this process is not stationary, and hence that more recent cases are more informative than older cases. Alternatively, this may reflect limitations on the agent's ability to remember or process information.

Once the agent has evaluated the theories, the agent must construct a pseudo-capacity ν to use in the criterion (5). Doing so requires balancing two sorts of considerations. First, the relative magnitudes of the weights attached to the various scenarios by the function ν will reflect the relative importance of the various theories. We will assume a simple rule here, namely that the agent selects the theory $\mathfrak{v}$ with the highest evaluation and uses as probabilities the marginal of this theory on the problem p_τ, denoted by $\mathfrak{v}|_{p_\tau}$, to fix the relative weights attached to the various scenarios. Second, the agent must choose the relative weights to place on backward-looking and forward-looking reasoning. We let $z(s, \tau)$ be a function, where s is the evaluation score achieved by the theory currently in vogue and τ is the period, and assume that the pseudo-capacity ν is obtained by scaling $\mathfrak{v}|_{p_\tau}$ so that $\mathfrak{v}|_{p_\tau}(\Omega) = z(s, \tau)$.

The underlying intuition is that having chosen the "best" theory, the agent may still be relatively confident of this theory or relatively skeptical. More confidence in a theory will be reflected in larger values of z. The period τ appears in the function $z(s, \tau)$ for multiple reasons. If, for example, the evaluation rule is the likelihood and all of the data is used, then likelihoods will necessarily decline as the number of periods grow, and the function z must be adjusted accordingly. The agent may encounter a theory in period 2 that has a high likelihood, but the paucity of evidence may nonetheless cause the agent to be relatively skeptical of this theory. In period 100, the agent may be highly confident of a theory with a tiny likelihood. Second, the number of cases in the memory is expanding as does the number of periods, and this may cause the importance placed on cases to grow. Given our normalization of the total weight attached to cases, this must also be reflected in the function z. We assume that $z(s, \tau)$ is strictly increasing in s, so that a more highly evaluated theory receives more weight in the agent's decision making.

If the process generating cases is sufficiently maniacal, the agent will have little hope of learning anything useful about the process. Let us consider the best case for learning, that in which problems and eventualities are drawn independently across periods from identical distributions. We then have enough structure to give the following:

Proposition 2 Suppose that problems and eventualities are drawn independently and identically across periods from distributions with full support, and the evaluation rule is the likelihood rule. Then:

[2.1] Suppose the evaluation rule takes account of only the previous n periods. Then for any period τ in which the current theory is some theory $\mathfrak{v}$, with probability one there will be a subsequent period in which the current theory is not $\mathfrak{v}$.

[2.2] Suppose the evaluation rule takes account of the entire history. Then generically, with probability one, there will be a time T and a theory $\mathfrak{v}$ such that for every $\tau > T$, the theory chosen by the data generating process is $\mathfrak{v}$.

The argument behind the first of these is straightforward. In each period, the evaluation rule takes in a finite amount of data, produced by a generating process with full support. Different sequences of problems and eventualities will have the highest likelihood under different theories, and hence the agent will be destined to continually switch between theories. For the second result, we similarly note that, given independence, the strong law of large numbers ensures that the likelihoods (under the various theories) of the realized string of data must converge, and hence (generically so that there are no ties between theories) the agent must settle on a particular theory. These results could obviously be generalized beyond evaluation rules given by the likelihood.

One reaction to the first result is that the agent is ill-advised to use only part of the data in a stationary world. If the agent is convinced the world is stationary and faces no constraints, this is a quite reasonable assessment. However, an agent unconvinced of the stationarity of the world or facing limitations on the amount of data she can process may have no other choice.

Let λ_τ be the relative weight placed on scenarios at time τ. Then we have:

Corollary 1 *Suppose that problems and eventualities are drawn independently and identically across periods from distributions with full support, the evaluation rule is the likelihood rule, and the evaluation rule takes account of only the previous n periods. Then*

$$\liminf_{\tau \to \infty} \lambda_\tau \neq \limsup_{\tau \to \infty} \lambda_\tau.$$

This indicates that the relative weights placed on cases and scenarios will be constantly shifting, giving us our ebb and flow.

Interestingly, the determination of when the weight on scenarios is relatively high will depend upon finer details of the environment. To illustrate, suppose there are only two eventualities, 0 and 1. Suppose the evaluation rule takes account of the eventualities in the last n periods. Then we can view the input into the evaluation rule as a birth-death process on the set $\{0, 1/n, 2/n, \ldots, n/n\}$, recording the proportion of 1 observations in the previous n periods, and each period causing this indicator to either remain constant, move one up, or move one down. Now let there be two theories, one that predicts nearly all 0 values and one that predicts nearly all 1 values. The theory currently in use will switch whenever the current summary of the data crosses the midpoint, i.e., the proportion of 1 values in the data crosses $1/2$. This is precisely when the likelihood is lowest. Hence, theories will be particularly uninfluential as the process switches from one theory to another. Here, the agent will be convinced that theories are unreliable at the same time that she has little idea which theory is best. In contrast, suppose that one theory predicts $1/2 - \varepsilon$ and the other theory predicts $1/2 + \varepsilon$. Again, we will switch from theory to theory as the summary statistic of the last n periods crosses $1/2$. In this case, however, the likelihood of both theories will be relatively large, and so scenarios will be particularly important when switching from one theory to another. Scenarios will be unimportant when the summary statistic is near the ends of the unit interval, in which case the agent will be relatively convinced of which theory to use, conditional on using a theory, but will think all theories are of little use.

4.5   Shifting Modes of Reasoning

In this section we consider a more concrete example of a combination of case-based and rule-based reasoning for the prediction of a single-dimensional real-valued variable.

4.5.1   An Investment Problem

Consider a process $x_t \in \mathbb{R}$ for periods $t = 1, 2, \ldots$. The variable x_t can be interpreted as the value of a financial asset at time t. Each period t is viewed as a decision problem, where the agent can buy or sell different quantities of the asset. Each such choice results in a monetary payoff. Thus, a past case is a triple (t, a, x, r) where t, the problem, is simply the time period; a is the financial decision; x is the eventuality, namely the market value of the asset, and r is the monetary payoff of the agent (depending on the portfolio she held).

27

As we noted in Section 2.2.2, it simplified the learning problem to assume that the investor is small, in the sense that her decisions have no causal effect on the process x_t. In particular, such investors have no difficulty in computing the counterfactual consequences they would have experienced had they made other choices. Thus, each past problem t can be used to evaluate both the act a that has been actually chosen at time t and any other act b that was not. In this sense, this decision problem is basically a prediction problem.

We can thus simplify notation as in Subsection 4.4 and suppress the acts chosen and consequences experienced in past cases. As there is no causal relationship between the agent's past choice and the uncertainty she faces in the new problem, and given that the performance of each possible act can be easily computed given x_t, the evaluation of acts a, b at present is independent of past choices. We thus assume that the set of cases at time t is $\{(\tau, x_\tau)\}_{\tau < t}$.

The agent conceives of theories that exhibit linear trends. Each such theory is given by two real-valued parameters, β_0, β_1, and thus a scenario is a pair $(\beta_0, \beta_1) \in \mathbb{R}^2$. We interpret the scenario (β_0, β_1) as suggesting the underlying relationship

$$x_t = \beta_0 + \beta_1 t + \varepsilon_t$$

where ε_τ is an error term implicitly interpreted as in standard OLS models.

For simplicity, assume that the agent focuses only on the (OLS, maximum likelihood) point estimate of the two parameters. The OLS learning procedure assumes as input past observations and computes the estimates, $\left(\hat{\beta}_0, \hat{\beta}_1\right)$, of the regression line coefficients. We assume here that only the past k observations are taken into account, so that the observations $\{(\tau, x_\tau)\}_{t-k \leq \tau < t}$ are used to select a single scenario, $\left(\hat{\beta}_0, \hat{\beta}_1\right)$. The capacity defined by this learning procedure puts all the weight on a single scenario, and it can therefore be viewed as a Dirichlet, additive measure.[12] Let the point estimate of the next observation, x_t, based on this regression equation be denoted by $\hat{x}_t^S$ (where S stands for "scenarios"). Denote by R_t^2 the coefficient of determination of this regression line.

---

[12] Alternatively, one can imagine that the agent considers a confidence set for the unknown parameters, centered around $\left(\hat{\beta}_0, \hat{\beta}_1\right)$, and chooses a capacity over the scenarios that assigns weight 1 to this confidence set, and perhaps weight 0 to its proper subsets. Observe that, as in Example 3, this model deviates from our axiomatic derivation in two ways: first, it involves a continuum of scenarios; second, it allows the agent to have only one non-null scenario. As mentioned above, both deviations can be taken care of in a more elaborate axiomatic derivation.

We assume that the agent also engages in case-based reasoning, which takes the simple form of computing the simple average of the past k periods. This is captured by the similarity function, defined for problems $\tau < t$,

$$s(t, \tau) = \left\{ \begin{array}{ll} \frac{1}{k} & \tau \geq t - k \\ 0 & \tau < t - k \end{array} \right.$$

and resulting in the point estimate of x_t given by

$$\hat{x}_t^C = \frac{1}{k} \sum_{\tau = t-k}^{\tau = t-1} x_\tau.$$

We assume that the relative weight of the scenario-based estimate, $\hat{x}_t^S$, and the case-based one, $\hat{x}_t^C$, in producing the final estimate depends on the success of the regression analysis. Specifically, define

$$\hat{x}_t = \left(1 - R_t^2 \right) \hat{x}_t^C + R_t^2 \hat{x}_t^S$$

This specification suggests that the agent has a preference for theories, when these seem to work without being too complex. If, among the linear functions of the past k observations there is one which fits the data well, the agent would tend to adopt it. In the extreme case of $R_t^2 = 1$, reasoning would be completely rule-based, putting all the weight on the scenarios and their likelihood. However, if R_t^2 is relatively low, the agent realizes that the (linear) theories she can come up with are not doing too well in terms of explaining the (past k) observations, and she resorts to case-based reasoning. In the extreme cases of $R_t^2 = 0$, reasoning would be completely case-based. In this case, as the agent can find no trend in the data she resorts to simple averaging of past observations in order to generate predictions.

The preference for rule-based reasoning assumed here is based on the intuition that people prefer to have general, simple theories that explain the data rather than to be looking at large databases and attempt to draw inferences from them. There are a few related reasons for this supposition. Simple theories that fit the data are efficient. They serve as "sufficient statistics" for large databases, saving on memory and on computation time in generating predictions. Importantly, they provide a feeling of having understood the process, of cutting Nature at its joint, as it were. Correspondingly, having figured out a simple rule that explains the data is pleasurable. This warm feeling one obtains from understanding might be explained from an evolutionary viewpoint: it is a way of implanting in us preferences for simple

theories that, as long as they fit the data, have a chance of uncovering some truth about the reality we live in.[13]

The simplest model of a financial market would assume that the realization of x_t is distributed around the estimate $\hat{x}_t$. For example, if all agents have the same estimate $\hat{x}_t$ (namely, if they all follow the assumptions of our model with the same k), then this estimate will be expected to be the equilibrium price. Various external factors can be expected to introduce noise into the system—from economic shocks and political crises to fundamentalists trading policies that are based on analysis of economic fundamentals rather than on the past behavior of x_t. For the sake of our discussion it will suffice to assume that

$$x_t = \hat{x}_t + \varepsilon_t$$

where ε_t is normally distributed around 0.

We focus on two characteristics of this model. First, we again have an ebb and flow between case-based and scenario-based reasoning: Agents are likely to be switching between case-based and rule-based reasoning. Moreover, it will generally take longer for a theory to be established than to be refuted. Second, predictions may not behave monotonically as a function of past values: higher values of the asset in previous periods may lead the agent to predict a lower current value.

4.5.2   Switching between Modes of Reasoning

In light of Corollary 1, it is not surprising that, with case-based and rule-based reasoning being determined by a bounded memory, their relative weights need not converge. Under mild assumptions on the data generating process, the relative weight of each mode of reasoning will move into any open neighborhood of 0 and of 1 infinitely often with probability 1. Thus, we expect to observe patterns by which theories get established, and then dethroned. But, importantly, they will not only be replaced by other theories: infinitely often the rule-based mode of reasoning will give way to the case-based one. Likewise, the opposite will also happen infinitely often. Moreover, this pattern need not be symmetric: the belief in a theory will grow gradually, while discarding it will happen faster. That is, we expect that it will typically take longer to establish a theory than to destroy it. To see why this is the case, we first state the following.

---

[13]See Gilboa and Samuelson (2012) for the selection of theories and for subjective preference over theories such as the preference for simplicity. See also Gayer and Gilboa (2014), who argue that, if reality is simple, reasoning is likely to converge to rule-based reasoning.

Proposition 3 For $k \geq 3$, let $\{(\tau, x_\tau)\}_{t-k \leq \tau < t}$ be a sequence of k observations at time t. Assume that the last three observations $x_{t-3}, x_{t-2}, x_{t-1}$ are not collinear. Then, there exists a $\hat{c} \in (0, 1)$ such that, for all $c \in (0, \hat{c})$,

1. the time length necessary to achieve $R^2 \geq 1 - c$ is $(k - 2)$;

2. the time length necessary to achieve $R^2 \leq c$ is 1.

The proof of this proposition is quite simple. Clearly, if the last three points are not collinear, a very high R^2 would require all but the last two to be forgotten, that is, not to participate in the regression equation. By contrast, a single observation suffices to set R^2 to zero. As will be clear from the proof, such an observation might be rather extreme. One may therefore be interested in similar results, where future observations are limited to the range of past ones. In this case, the number of observations needed to get $R^2 = 0$ is still no more than $\frac{k}{2}$.

The main point of Proposition 3 and its possible variants is that the number of observations needed to increase R_t^2 from a value in the neighborhood of 0 to a value in the neighborhood of 1 is larger than vice versa. If we wish R_t^2 to be within a small neighborhood of 1, the future $(k - 2)$ observations would have to be within suitably small neighborhoods of the regression line. By contrast, fewer observations would have to be within small neighborhoods for R_t^2 to be close to 0. Under various assumptions about the data generating process, the probability of the former event will be much smaller than of the latter event.

4.5.3   Non-monotonicity in Predictions

How does prediction behave as a function of the past? Will higher values of the most recent k observations result in a higher prediction? The obvious answer is positive for the case-based, but negative for the rule-based mode of reasoning: the former is a mere average, so that it is a monotone function of each of its components; by contrast, the latter seeks trends, and higher values earlier on may tilt the regression line downward, identifying a negative trend.

Let us, therefore ask a more refined question: suppose that we compare two possible sequences of the past k observations, $\{(\tau, x_\tau)\}_{t-k \leq \tau < t}$ and $\{(\tau, y_\tau)\}_{t-k \leq \tau < t}$ where (i) one dominates the other pointwise, say, $y_\tau \geq x_\tau$ for all $t - k \leq \tau < t$; and (ii) the dominating sequence also generates a higher prediction according to the regression line, $\hat{y}_t^S > \hat{x}_t^S$. Will it then be necessarily true that the higher sequence results in a higher prediction?

It turns out that the answer is negative. Consider the following two sequences, for $k = 3$:

	x	y
1	0	0
2	100	175
3	200	200
Average	100	125
Regression Line	$-100 + 100\tau$	$-75 + 100\tau$
R^2	1.00	0.84
Regression Prediction	300	325
Overall Prediction	300	293

In this example, the y sequence dominates the x sequence pointwise (hence also in the average), but also in its regression prediction. If the relative weights of the two modes of reasoning were held constant, the weighted average for y would have been higher than that for x. However, this isn't the case: the R^2 of the y sequence is lower than that of the x. As a result the weight put in the overall prediction on the regression line in y is lower than the corresponding one for x. As in the famous Simpson's Paradox (Simpson, 1951), when relative weights change, averages can change in direction that is opposite to that of all the components involved.

Intuitively, the sequence y seems to suggest better news than does x, not only in terms of the actual values observed but also in terms of the forecasts they induce for the future. However, while the sequence x reassures the agent that she observes a steady growth, and that her linear model works perfectly, the sequence y leaves some doubt. As the points are not exactly aligned, an agent who thinks in terms of linear theories may be wondering whether she had indeed cut Nature at its joints. She has to admit that her theory does well, but not perfectly well, and that it might be safer to leave room for alternatives, which are here the cases. Putting some weights on their average can decrease the overall prediction.

While this example may seem anomalous, it is important to point out that it is not an artifact of the small number of observations involved. The following proposition establishes that it can happen for any k.

Proposition 4 For each $k \geq 3$, there exist two sequences, $\{(\tau, x_\tau)\}_{t-k \leq \tau < t}$ and $\{(\tau, y_\tau)\}_{t-k \leq \tau < t}$, of k observations with $y_\tau \geq x_\tau$ for all $t - k \leq \tau < t$, $\hat{y}_t^C > \hat{x}_t^C$, and $\hat{y}_t^S > \hat{x}_t^S$ such that $\hat{y}_t < \hat{x}_t$.

When applied to financial markets, we note that good news about past performance can be double-edged. In our model, investors would like to get good news about their investments, but when these are not easily explained, they might become edgy and might pull out of a market despite recent successes.

5 Appendix

5.1 Appendix A: Pairwise Comonotonicity

We start with some simple observations regarding the notion of pairwise comonotonic profiles in Definition 1. Observe that pairwise comonotonicity in our model is less restrictive than the standard comonotonicity condition, in that comparisons are only made between two components within a set of problem-eventuality pairs and within a set of scenarios, but not across these. The standard characterizations of comonotonicity in terms of monotonicity with respect to a permutation have natural counterparts in our case, as explained below.

Let $a \in A$, $H_a = \{(p_i, o_i)\}_{i=1}^{n(a)}$ and $\Omega = \{\omega_j\}_{j=1}^{m}$. For the sake of brevity, let $e_i = (p_i, o_i)$. For permutations π on H_a and ρ on Ω, consider the sets $C^\pi = \{f \in \mathcal{F}_a \colon f(\pi(e_1)) \succsim \ldots \succsim f(\pi(e_{n(a)}))\}$ and $C^\rho = \{f \in \mathcal{F}_a \colon f(\rho(\omega_1)) \succsim \ldots \succsim f(\rho(\omega_m))\}$.

Then, $C^\pi \cap C^\rho$ is a maximal pairwise comonotonic set (*pairwise comoncone*). Note that (standard) comonotonic sets are subsets of pairwise comonotonic sets (relative to the appropriate permutations).

For $f \in \mathcal{F}_a$, define the binary relations

- $\succeq_{H_a}^{f}$ on H_a as $e_i \succeq_{H_a}^{f} e_j$ if and only if $f(e_i) \succsim f(e_j)$;

- $\succeq_{\Omega}^{f}$ on Ω as $\omega_i \succeq_{\Omega}^{f} \omega_j$ if and only if $f(\omega_i) \succsim f(\omega_j)$.

For a set $F \subseteq \mathcal{F}_a$, define the binary relations $\succeq_{H_a}^{F} = \bigcap_{f \in F} \succeq_{H_a}^{f}$ and $\succeq_{\Omega}^{F} = \bigcap_{f \in F} \succeq_{\Omega}^{f}$.

Then, the following observation adapts a basic result in the literature to our notion of pairwise comonotonicity.

Lemma 1 Let $a \in A$ and $F \subset \mathcal{F}_a$. Assume that $\succsim$ on R is a weak order. Then, the following statements are equivalent:

1. F is a pairwise comonotonic set;

2. $\succeq^F_{H_a}$ and $\succeq^F_\Omega$ are weak orders;

3. $F \subset C^\pi \cap C^\rho$ for some permutations π on H_a and ρ on Ω.

The proof of the above lemma follows easily by adapting well-known results (see, e.g., Lemma 3.1 in Wakker, 1989).

   We mention in passing that the above analysis can be easily extended to any partition of a finite coordinate set. Similarly, our pairwise comonotonic tradeoff consistency, as stated (with "comonotonicity" understood to hold only within each element of the partition) would be equivalent to the capacity being additive over elements of the partition. This will follow from adaptation of Steps 1-5 in the proof of our main result.

5.2   Appendix B: Proofs

Throughout the appendix, for every $a \in A$, the binary relation $\succsim_{H_a,\Omega}$ stands for the restriction of $\succsim_{H,\Omega}$ to $\mathcal{F}_a$. If $H_a = \varnothing$, then $\succsim_{\varnothing,\Omega}$ coincides with the restriction of $\succsim_{H,\Omega}$ to R^Ω, and is simply denoted by $\succsim_\Omega$.

   We recall that, for $a \in A$, a binary relation $\succsim_{H_a,\Omega}$ on $\mathcal{F}_a$ is:

- *monotone* if $f(s) \succsim g(s)$ for all $s \in H_a \cup \Omega$ implies $f \succsim_{H_a,\Omega} g$;

- *continuous* if, for every $f \in \mathcal{F}_a$, the sets $\{g \in \mathcal{F}_a \colon f \succsim_{H_a,\Omega} g\}$ and $\{g \in \mathcal{F}_a \colon g \succsim_{H_a,\Omega} f\}$ are closed.

   We start with some preliminary results which will be useful to prove Theorem 1.

Lemma 2 For every $a \in A$, let the binary relation $\succsim_{H_a,\Omega}$ on $\mathcal{F}_a$ be a monotone and continuous weak order that satisfies Pairwise Comonotonic Tradeoff Consistency. Then, for every $f, g \in \mathcal{F}_a, \alpha, \gamma \in R$, and $s \in H_a \cup \Omega$,

$$\alpha\{s\}f \succsim_{H_a,\Omega} \alpha\{s\}g \quad \Longleftrightarrow \quad \gamma\{s\}f \succsim_{H_a,\Omega} \gamma\{s\}g \tag{6}$$

whenever the set $\{\alpha\{s\}f, \alpha\{s\}g, \gamma\{s\}f, \gamma\{s\}g\}$ is pairwise comonotonic.

Observe that (6) is a stronger version of the standard Comonotonic Coordinate Independence axiom.

Proof. For short notation, set $e_i = (p_i, o_i)$. Let $a \in A$ and F be a pairwise comoncone in $\mathcal{F}_a$ which contains $\alpha\{s\}f, \alpha\{s\}g, \gamma\{s\}f$, and $\gamma\{s\}g$. Without loss of generality, assume that the profiles in F are ordered from best to worst using identity permutations on both H_a and Ω.

In the formulation of Pairwise Comonotonic Tradeoff Consistency, set $\alpha = \beta, \gamma = \delta$ and $f = f'$. Then, this axiom implies that

$$\alpha\{s\}f \sim_{H_a,\Omega} \alpha\{s\}g \iff \gamma\{s\}f \sim_{H_a,\Omega} \gamma\{s\}g. \tag{7}$$

For the strict part of the statement, suppose, by contradiction, that $\alpha\{s\}f \succ_{H_a,\Omega} \alpha\{s\}g$ and $\gamma\{s\}f \prec_{H_a,\Omega} \gamma\{s\}g$. Then, there exists $s' \in H_a \cup \Omega$ such that $s' \neq s$ and $f(s') \succ g(s')$.

To ease notation, set $f = \alpha\{s\}f$ and $g = \alpha\{s\}g$. The following arguments are analogous to the steps of the proof of Lemma 31 in Köbberling and Wakker (2003).

Step 1.1: Consider the set $S = \{\omega_i \in \Omega \colon f(\omega_i) \succ g(\omega_i)\}$ and suppose it is nonempty. Let j be the first rank-ordered index in S. Consider the profile $f(\omega_j)\{\omega_j\}g \in \mathcal{F}_a$ and note that it belongs to F because $g(\omega_{j-1}) \succsim f(\omega_{j-1}) \succsim f(\omega_j) \succ g(\omega_j) \succsim g(\omega_{j+1})$.

If $f \succ_{H_a,\Omega} f(\omega_j)\{\omega_j\}g$, replace the original profile g with $g = f(\omega_j)\{\omega_j\}g$ and repeat Step 1.1 by taking the next rank-ordered index in S.

If $f(\omega_j)\{\omega_j\}g \succsim_{H_a,\Omega} f \succ_{H_a,\Omega} g$, we can find $\beta \in R$ such that $\beta\{\omega_j\}g \sim_{H_a,\Omega} f$ because $\succsim_{H_a,\Omega}$ is continuous. By monotonicity, $f(\omega_j) \succsim \beta \succ g(\omega_j)$, hence $\beta\{\omega_j\}g \in F$. In this case, replace the original profile g with $g = \beta\{\omega_j\}g$ and proceed with Step 2.

Observe that, if H_a is a null set (or simply $H_a = \varnothing$), then there is at least one $\omega_i \in \Omega$ such that $f(\omega_i) \succ g(\omega_i)$ and the procedure described in Step 1.1 can be implemented. If H_a is not null, then Step 1.2 may also be needed to reach the desired conclusion.

Step 1.2: Suppose that $S = \varnothing$ or $S \neq \varnothing$ but, after applying Step 1.1 iteratively using all elements in S, we still have $f \succ_{H_a,\Omega} fSg$. Then, it must be that H_a is not null, and $f(e_i) \succ g(e_i)$ for at least one $e_i \in H_a$. Let $Q = \{e_i \in H_a \colon f(e_i) \succ g(e_i)\}$ and e_j stand for the first rank-ordered index in Q. Consider the profile $f(e_j)\{e_j\}g \in \mathcal{F}_a$ (where g could be the original profile or the profile resulting from the transformations in Step 1.1) and note that it belongs to F, too. By proceeding analogously to Step 1.1, we can find some $\delta \in R$ such that $\delta\{e_j\}g \sim_{H_a,\Omega} f$.

Step 2: Denote by $\bar{g} \in \mathcal{F}_a$ the profile constructed from the original profile g in Step 1. Observe that $\alpha\{s\}f \sim_{H_a,\Omega} \alpha\{s\}\bar{g}$, which implies, by (7), that

$\gamma\{s\}f \sim_{H_a,\Omega} \gamma\{s\}\bar{g}$. However, note that $\bar{g}(s) \succsim g(s)$ for all $s \in H_a \cup \Omega$. Hence, by monotonicity, $\gamma\{s\}\bar{g} \succsim_{H_a,\Omega} \gamma\{s\}g \succ_{H_a,\Omega} \gamma\{s\}f$, contradiction. ∎

Corollary 2 *For every $a \in A$, let the binary relation $\succsim_{H_a,\Omega}$ on $\mathcal{F}_a$ be a monotone and continuous weak order that satisfies Pairwise Comonotonic Tradeoff Consistency. Then, for every $x, x' \in \mathcal{H}_a$, and $y, y' \in R^\Omega$,*

$$(x,y) \succsim_{H_a,\Omega} (x,y') \quad \Longleftrightarrow \quad (x',y) \succsim_{H_a,\Omega} (x',y').$$

whenever the set $\{(x,y),(x,y'),(x',y),(x',y')\}$ is pairwise comonotonic.

Proof. The result follows from Lemma 2 using an inductive argument. ∎

For two utility functions, $u, u' \colon R \to \mathbb{R}$ the notation $u \approx u'$ implies that they are positive affine transformations of each other.

Proof of Theorem 1. We prove the sufficiency of the axioms. The necessity part follows by standard arguments.

Step 1 Let $a, b \in A, a \neq b$ and $f \in \mathcal{F}_a$. Consider $U = \{g \in \mathcal{F}_b \colon g \succsim_{H,\Omega} f\}$ and $V = \{g \in \mathcal{F}_b \colon f \succsim_{H,\Omega} g\}$. By Continuity, the sets U and V are nonempty and closed; by Restricted Weak Order, $U \cup V = \mathcal{F}_b$. Since $\mathcal{F}_b$ is connected, $U \cap V \neq \varnothing$. Hence, for every $f \in \mathcal{F}_a$, there exists $g \in \mathcal{F}_b$ such that $g \sim_{H,\Omega} f$. □

Step 2 Let $a \in A$. We show that the binary relation $\succsim_{H_a,\Omega}$ satisfies all axioms of Corollary 10 in Köbberling and Wakker (2003). Clearly, $\succsim_{H_a,\Omega}$ is a preorder by Restricted Weak Order. Let $f, f' \in \mathcal{F}_a$. By Step 1, there exists $g \in \mathcal{F}_b$, for $b \in A, b \neq a$, such that $f \sim_{H,\Omega} g$. Then, Restricted Weak Order implies that $\succsim_{H_a,\Omega}$ is complete and, therefore, is a weak order. Moreover, $\succsim_{H_a,\Omega}$ is monotone: Let $f, f' \in \mathcal{F}_a$ be such that $f(s) \succsim f'(s)$ for all $s \in H_a \cup \Omega$. From Step 1 and Monotonicity, it follows that $f \succsim_{H_a,\Omega} f'$. The binary relation $\succsim_{H_a,\Omega}$ is also continuous: Indeed, let $f \in \mathcal{F}_a$ and consider the set $\{f' \in \mathcal{F}_a \colon f' \succ_{H_a,\Omega} f\}$. Step 1 and Continuity directly imply that this set is open in $\mathcal{F}_a$. Similarly, it can be shown that $\{f' \in \mathcal{F}_a \colon f \succ_{H_a,\Omega} f'\}$ is open, too. Finally, Pairwise Comonotonic Tradeoff Consistency implies that $\succsim_{H_a,\Omega}$ satisfies the comonotonic tradeoff consistency axiom of Köbberling and Wakker (2003).

Hence, $\succsim_{H_a, \Omega}$ satisfies all the axioms of Corollary 10 in Köbberling and Wakker (2003) and, therefore, admits a Choquet expected utility representation: there exist a continuous function $u_a \colon R \to \mathbb{R}$ and a capacity $v_a \colon 2^{H_a \cup \Omega} \to [0, 1]$ such that $V_a(f) = \int_{H_a \cup \Omega} u_a(f) dv_a$ represents $\succsim_{H_a, \Omega}$. In particular, for all $b \in A$ such that H_b is null, there exist a continuous function $u \colon R \to \mathbb{R}$ and a capacity $\nu \colon 2^{\Omega} \to [0, 1]$ such that $V_\Omega(f) = \int_\Omega u(f) d\nu$ represents $\succsim_\Omega$. $\qquad\square$

Step 3 Let $a \in A$ and, for a given $x \in \mathcal{H}_a$, consider the set

$$T = \{f = (x, y) \in \mathcal{F}_a \mid y(\omega) \succsim x(p, o) \text{ for all } (p, o) \in H_a \text{ and } \omega \in \Omega\} .$$

Using the representation of $\succsim_{H_a, \Omega}$, we observe that $\succsim_{H_a, \Omega}$ restricted to T coincides with $\succsim_\Omega$ and, therefore, we have that $(x, y) \succsim_{H_a, \Omega} (x, y')$ if and only if $y \succsim_\Omega y'$ for all $(x, y), (x, y') \in T$. By applying Corollary 1, it follows that

$$(x', y) \succsim_{H_a, \Omega} (x', y') \quad \Leftrightarrow \quad y \succsim_\Omega y' \tag{8}$$

for all $x' \in \mathcal{H}_a$ such that the set $\{(x, y), (x, y'), (x', y), (x', y')\}$ is pairwise comonotonic. Hence, by the uniqueness properties of the representation of Köbberling and Wakker (see their Observation 9), we have that $u_a \approx u$ and $\frac{v_a(A)}{v_a(\Omega)} = \nu(A)$ for all $A \subseteq \Omega$. Shift u so that $u(\alpha^*) = 0$ for the consequence α^* of the Act Independent Aspirations axiom. (Note that the remaining freedom in selecting u is only multiplication by a positive number.) $\qquad\square$

Step 4 We claim that, for every $a \in A$ and $f \in \mathcal{F}_a$,

$$V_a(f) = \int_{H_a} u(f) dv_a + \int_\Omega u(f) dv_a.$$

To this end, it is sufficient to show that $v_a(E \cup F) = v_a(E) + v_a(F)$ for all $E \subseteq H_a$ and $F \subseteq \Omega$.

Let $y = (\alpha, D; \beta, \Omega \backslash D)$ and $y' = (\gamma, \Omega)$, where $D \subsetneq \Omega$ and $\alpha, \beta, \gamma \in R$ are such that $\alpha \succ \gamma \succ \beta$ and $\nu(D) = \frac{u(\gamma) - u(\beta)}{u(\alpha) - u(\beta)}$. Note that we can find such set D because, by assumption, there exists a comonotonic set in R^Ω with at least two non-null scenarios. Then, $y \sim_\Omega y'$. Now, let $x = (\beta, H_a) \in \mathcal{H}_a$. By (8), we have $(x, y) \sim_{H_a, \Omega} (x, y')$ which, using the representation of $\succsim_{H_a, \Omega}$, is equivalent to

$$u(\alpha) v_a(D) + u(\beta) \left[v_a(H_a \cup \Omega) - v_a(D) \right] = u(\gamma) v_a(\Omega) + u(\beta) \left[v_a(H_a \cup \Omega) - v_a(\Omega) \right].$$

Replace x with $x' = (\theta, B; \beta, H_a \backslash B) \in \mathcal{H}_a$, where $B \subseteq H_a$ and $\theta \in R$ such that $\gamma \succ \theta \succ \beta$. Then, by Corollary 2, $(x', y) \sim_{H_a, \Omega} (x', y')$ and, using the representation, we have

$$u(\alpha)v_a(D) + u(\theta)\left[v_a(B \cup D) - v_a(D)\right] + u(\beta)\left[v_a(H_a \cup \Omega) - v_a(B \cup D)\right] =$$
$$u(\gamma)v_a(\Omega) + u(\theta)\left[v_a(B \cup \Omega) - v_a(\Omega)\right] + u(\beta)\left[v_a(H_a \cup \Omega) - v_a(B \cup \Omega)\right].$$

Then, subtracting the previous equality from this last one, we get

$$\left[u(\theta) - u(\beta)\right]\left[v_a(B \cup D) - v_a(D)\right] = \left[u(\theta) - u(\beta)\right]\left[v_a(B \cup \Omega) - v_a(\Omega)\right],$$

and $u(\theta) - u(\beta) > 0$ delivers

$$v_a(B \cup D) - v_a(D) = v_a(B \cup \Omega) - v_a(\Omega) \tag{9}$$

for all $B \subseteq H_a$ and $D \subsetneq \Omega$.

It remains to show that $v_a(B \cup D) - v_a(D) = v_a(B)$, which can be proved by following a similar argument. Specifically: let $z = (\gamma, H_a) \in \mathcal{H}_a$. Then, $(z, y) \sim_{H_a, \Omega} (z, y')$ if and only if

$$u(\alpha)v_a(D) + u(\gamma)\left[v_a(H_a \cup D) - v_a(D)\right] + u(\beta)\left[v_a(H_a \cup \Omega) - v_a(H_a \cup D)\right]$$
$$= u(\gamma)v_a(H_a \cup \Omega) \qquad .$$

Now, replace z with $z' = (\zeta, B; \gamma, H_a \backslash B) \in \mathcal{H}_a$, where $B \subseteq H_a$ and $\zeta \in R$ with $\alpha \succ \zeta \succ \gamma$. Then, $(z', y) \sim_{H_a, \Omega} (z', y')$, which is equivalent to

$$u(\alpha)v_a(D) + u(\zeta)\left[v_a(B \cup D) - v_a(D)\right] +$$
$$u(\gamma)\left[v_a(H_a \cup D) - v_a(B \cup D)\right] + u(\beta)\left[v_a(H_a \cup \Omega) - v_a(H_a \cup D)\right]$$
$$= u(\zeta)v_a(B) + u(\gamma)\left[v_a(H_a \cup \Omega) - v_a(B)\right].$$

A similar subtraction yields

$$\left[u(\zeta) - u(\gamma)\right]\left[v_a(B \cup D) - v_a(D)\right] = \left[u(\zeta) - u(\gamma)\right]v_a(B)$$

and $u(\zeta) - u(\gamma) > 0$ delivers

$$v_a(B \cup D) - v_a(D) = v_a(B). \tag{10}$$

By combining conditions (9) and (10), we have $v_a(E \cup F) = v_a(E) + v_a(F)$ for all $E \subseteq H_a$ and $F \subseteq \Omega$. $\qquad \square$

Hence, for every $a \in A$, the binary relation $\succsim_{H_a,\Omega}$ is represented by $V_a(f) = \int_{H_a} u(f) dv_a + \int_\Omega u(f) dv_a$. Moreover, by the uniqueness properties of u and v_a discussed in Step 3, we can apply the normalization $V'_a = \frac{V_a}{v_a(\Omega)}$ and obtain, with little abuse of notation, that

$$V'_a(f) = \int_{H_a} u(f) dv_a + \int_\Omega u(f) d\nu$$

represents $\succsim_{H_a,\Omega}$, too.

Step 5 It remains to derive the representation of $\succsim_{H,\Omega}$ on $\mathcal{F}$ — i.e., when comparing profiles induced by distinct acts a and b in A.

Fix $a \in A$ and $f \in \mathcal{F}_a$. Step 1 implies that there exists $y \in R^\Omega$ such that $f \sim_{H,\Omega} y$. By Act Independent Aspirations, $y \sim_{H,\Omega} (\boldsymbol{\alpha}^*, y)$ (where $\boldsymbol{\alpha}^* \in \mathcal{H}_a$), and by transitivity we also get $f \sim_{H_a,\Omega} (\boldsymbol{\alpha}^*, y)$ (that is, $f \sim_{H,\Omega} (\boldsymbol{\alpha}^*, y)$ and both these profiles are in $\mathcal{F}_a$). Using the representation of $\succsim_{H_a,\Omega}$ from Step 4, we have that $f \sim_{H_a,\Omega} (\boldsymbol{\alpha}^*, y)$ if and only if

$$
\begin{aligned}
&\int_{H_a} u(f) dv_a + \int_\Omega u(f) d\nu &&\text{(11)}\\
={}& \int_{H_a} u(\boldsymbol{\alpha}^*) dv_a + \int_\Omega u(y) d\nu \\
={}& \int_\Omega u(y) d\nu.
\end{aligned}
$$

Next, consider $a, b \in A$, and a pair of profiles, $f \in \mathcal{F}_a$ and $g \in \mathcal{F}_b$. Choose $y, y' \in R^\Omega$ so that $f \sim_{H,\Omega} y$ and $g \sim_{H,\Omega} y'$. Representation of $\succsim_{H,\Omega}$ by the sum of the integrals over the entire space follows from its representation on R^Ω and (11). More explicitly,

$$
\begin{aligned}
f \succsim_{H,\Omega} g \quad\Leftrightarrow\quad & y \succsim_\Omega y' \quad\Leftrightarrow\quad \int_\Omega u(y) d\nu \geq \int_\Omega u(y') d\nu \quad\Leftrightarrow \\
& \int_{H_a} u(f) dv_a + \int_\Omega u(f) d\nu \geq \int_{H_b} u(g) dv_b + \int_\Omega u(g) d\nu.
\end{aligned}
$$

$\square$ ∎

Proof of Proposition 1. Assume first that, for all $a \in A$, H_a is null. Then any representation of $\succsim_{H_a,\Omega}$, $(u, \{v_a\}_{a \in A}, \nu)$, satisfies $v_a \equiv 0$ for all $a \in A$. In this case u is unique up to an affine transformation, and ν is

39

unique, as in Observation 9 of Köbberling and Wakker (2003), while the identically-zero pseudo-capacities v_a are clearly unique.

Next, assume that, for some $a \in A$, H_a isn't null. Assume first that both $(u, \{v_a\}_{a \in A}, \nu)$ and $(\hat{u}, \{\hat{v}_a\}_{a \in A}, \hat{\nu})$ represent $\succsim_{H,\Omega}$ as in Theorem 1 By the aforementioned Observation 9, we have (i) $\hat{v}_a = v_a$, for all $a \in A$, and $\hat{\nu} = \nu$; (ii) there exist $\lambda, d \in \mathbb{R}$ with $\lambda > 0$ such that $\hat{u} = \lambda u + d$. Choose a consequence α^* such that $(\boldsymbol{\alpha^*}, y) \sim_{H,\Omega} y$ where $(\boldsymbol{\alpha^*}, y) \in \mathcal{F}_a$, whose existence is guaranteed by Act Independent Aspirations. As $v_a(H_a) = \hat{v}_a(H_a) > 0$, it has to be the case that $\hat{u}(\alpha^*) = 0 = u(\alpha^*)$. Hence, $d = 0$ and $\hat{u} = \lambda u$.

Conversely, it is easy to verify that, if the triple $(u, \{v_a\}_{a \in A}, \nu)$ represents $\succsim_{H,\Omega}$ as in Theorem 1, so will any triple $(\hat{u}, \{v_a\}_{a \in A}, \nu)$ where $\hat{u} = \lambda u$ for any $\lambda > 0$. ∎

Proof of Proposition 3. Let $k \geq 3$ and $\{(\tau, x_\tau)\}_{t-k \leq \tau < t}$ be a sequence of observations such that $(x_{t-3}, x_{t-2}, x_{t-1})$ are not collinear.

Given the last two observations, x_{t-2}, x_{t-1}, there clearly exists a unique sequence of $(k-2)$ observations that makes the coefficient of determination R^2_{t+k-3} of the regression line computed using the last k observations equal to 1. Denote such sequence by $x^* = \{(\tau, x^*_\tau)\}_{t \leq \tau < t+k-2}$. Note also that, as a function defined over the set of sequences of length k, the coefficient R^2_τ is continuous except at points $x \in \mathbb{R}^k$ such that $x_i = x_j$ for all $i, j = 1, \ldots, k$. Hence, for every $c > 0$, there exists an $\varepsilon > 0$ such that, for all sequences $x' = \{(\tau, x'_\tau)\}_{t \leq \tau < t+k-2}$, if $|x^*_\tau - x'_\tau| < \varepsilon$ for all $t \leq \tau < t + k - 2$, then $|R^2_{t+k-3}(x^*) - R^2_{t+k-3}(x')| \leq c$, which implies that $R^2_{t+k-3}(x') \geq 1 - c$.

Given the last three observations, $(x_{t-3}, x_{t-2}, x_{t-1})$, denote by $\bar{R}^2$ the highest value of the coefficient of determination that can be obtained with just $(k-3)$ additional observations. Then, it must be that $0 < \bar{R}^2 < 1$. It cannot be 1 because, by assumption, the last three points are not collinear, and, by construction, cannot be 0 either. Set $\bar{R}^2 = 1 - \hat{c}$ for some $\hat{c} \in (0, 1)$. It follows that, for all $c < \hat{c}$, we need at least $(k-2)$ observations to have $R^2 \geq 1 - c$.

As standard, recall that for simple linear regressions, $R^2 = \rho^2$, where, ρ is the coefficient of correlation between the dependent and the independent variables. Observe also, that, given $\{(\tau, x_\tau)\}_{t-k \leq \tau < t}$, the sign of ρ equals the sign of the covariance between the time periods and the corresponding past observations. Moreover, for any additional observation x_t, the covariance

between the last k observations and the last k time periods is given by

$$\sum_{\tau=t-k+1}^{t} (\tau - \bar{\tau})(x_\tau - \bar{x}) = \sum_{\tau=t-k+1}^{t-1} (\tau - \bar{\tau})(x_\tau - \bar{x}) + (t - \bar{\tau})(x_t - \bar{x}), \quad (12)$$

where $\bar{\tau} = \frac{\sum_{\tau=t-k+1}^{t} \tau}{k}$ and $\bar{x} = \frac{\sum_{\tau=t-k+1}^{t} x_\tau}{k}$. Since $(t - \bar{\tau}) > 0$, it follows that there exists a point $x_t \in \mathbb{R}$ such that the covariance in (12) is 0. We conclude that, given such point, $R_t^2 = 0$, as well. ∎

Proof of Proposition 4. Let $k \geq 3$ and consider a sequence $\{(\tau, x_\tau)\}_{1 \leq \tau \leq k}$ of k observations such that $x_\tau = \tau$ for all $\tau = 1, \ldots, k$. Since $\sum_{\tau=1}^{k} \tau = \frac{k(k+1)}{2}$, we have that the case-based estimate is $\hat{x}_{k+1}^C = \frac{\sum_{\tau=1}^{k} \tau}{k} = \frac{k+1}{2}$. The scenario-based estimate is $\hat{x}_{k+1}^S = \hat{\beta}_0 + (k+1)\hat{\beta}_1 = k+1$ because $\hat{\beta}_0 = 0$ and $\hat{\beta}_1 = 1$. (Clearly, the coefficient of determination of this regression line is 1.)

Next, we show that there exists another sequence, $\{(\tau, y_\tau)\}_{1 \leq \tau \leq k}$, of k observations with $\hat{y}_{k+1}^C > \hat{x}_{k+1}^C$ and $\hat{y}_{k+1}^S > \hat{x}_{k+1}^S$ and such that $\hat{y}_{k+1} < \hat{x}_{k+1}$. We distinguish between two cases, depending on whether k is odd or even.

Case 1: Let k be odd and define a sequence $\{(\tau, y_\tau)\}_{1 \leq \tau \leq k}$ as

$$y_\tau = \begin{cases} \tau & \text{for } \tau \neq \frac{k+1}{2} \\ \frac{k+1}{2} + \Delta & \text{for } \tau = \frac{k+1}{2} \end{cases}$$

for some $\Delta \in \mathbb{R}_{++}$. Then, the case-based estimate is $\hat{y}_{k+1}^C = \frac{\sum_{\tau=1}^{k} y_\tau}{k} = \hat{x}_{k+1}^C + \frac{\Delta}{k}$. Moreover, as only the value of y_τ for the midpoint $\tau = \frac{k+1}{2}$ changes, the slope $\hat{\beta}_1$ of the regression line does not change, and its intercept $\hat{\beta}_0$ increases from 0 to $\frac{\Delta}{k}$. Hence, the scenario-based estimate is $\hat{y}_{k+1}^S = \hat{x}_{k+1}^S + \frac{\Delta}{k}$.

The overall estimate is

$$\hat{y}_{k+1} = (1 - R_k^2)\hat{x}_{k+1}^C + R_k^2 \hat{x}_{k+1}^S + \frac{\Delta}{k},$$

where R_k^2 is the coefficient of determination of the regression line used to compute the estimate $\hat{y}_{k+1}^S$. It follows that $\hat{y}_{k+1} < \hat{x}_{k+1}$ if and only if Δ satisfies $(1 - R_k^2)(\hat{x}_{k+1}^S - \hat{x}_{k+1}^C) > \frac{\Delta}{k}$, that is,

$$1 - R_k^2 > \frac{2\Delta}{k(k+1)}. \quad (13)$$

41

Express R_k^2 as ρ_k^2 where $\rho_k = \frac{Cov(T,Y)}{\sigma_T \sigma_Y}$ is the coefficient of correlation between the past realizations $Y = \{y_\tau\}_{1 \le \tau \le k}$ and the corresponding time periods $T = \{\tau\}_{1 \le \tau \le k}$. We have that

$$Cov(T,Y) = \sum_{\tau=1}^{k} \tau y_\tau - k\bar{\tau}\bar{y} = \sum_{\tau=1}^{k} \tau^2 + \left(\frac{k+1}{2}\right)\Delta - k\bar{\tau}\bar{y}$$

$$= \frac{k^3}{3} + \frac{k^2}{2} + \frac{k}{6} + \left(\frac{k+1}{2}\right)\Delta - k\left(\frac{k+1}{2}\right)\left(\frac{k+1}{2} + \frac{\Delta}{k}\right)$$

$$= \frac{k(k^2-1)}{12},$$

$$\sigma_T = \sqrt{\sum_{\tau=1}^{k} \tau^2 - k\bar{\tau}^2} = \sqrt{\frac{k^3}{3} + \frac{k^2}{2} + \frac{k}{6} - k\left(\frac{k+1}{2}\right)^2} = \sqrt{\frac{k(k^2-1)}{12}},$$

$$\sigma_Y = \sqrt{\sum_{\tau=1}^{k} y_\tau^2 - k\bar{y}^2} = \sqrt{\sum_{\tau=1}^{k} \tau^2 + (k+1)\Delta + \Delta^2 - k\left(\frac{k+1}{2} + \frac{\Delta}{k}\right)^2}$$

$$= \sqrt{\frac{(k-1)(k^3 + k^2 + 12\Delta^2)}{12k}},$$

where $\bar{\tau} = \frac{\sum_{\tau=1}^{k} \tau}{k} = \hat{x}_{k+1}^C$ and $\bar{y} = \frac{\sum_{\tau=1}^{k} y_\tau}{k} = \hat{y}_{k+1}^C$.[14]

It follows that we can express R_k^2 as a function of k and Δ as $R_k^2 = \frac{k^2(k+1)}{k^3+k^2+12\Delta^2}$. Plugging this formula for R^2 in (13), we obtain that the condition to be satisfied is $\Delta > \frac{k}{6} + \frac{2\Delta^2}{k(k+1)}$ or, equivalently, $\Delta\left(1 - \frac{2\Delta}{k^2+k}\right) > \frac{k}{6}$. Observe that the left-hand side of the last inequality, as a function of Δ, is maximized at $\Delta^* = \frac{k^2+k}{4}$. Finally, it is easily seen that $\Delta^*\left(1 - \frac{2\Delta^*}{k^2+k}\right) = \frac{k(k+1)}{8} > \frac{k}{6}$ holds for all odd $k \ge 3$.

Case 2: Let k be even and define a sequence $\{(\tau, y_\tau)\}_{1 \le \tau \le k}$ as

$$y_\tau = \begin{cases} \tau & \text{for } \tau \ne \lfloor \frac{k+1}{2} \rfloor, \lceil \frac{k+1}{2} \rceil \\ \lfloor \frac{k+1}{2} \rfloor + \Delta & \text{for } \tau = \lfloor \frac{k+1}{2} \rfloor \\ \lceil \frac{k+1}{2} \rceil + \Delta & \text{for } \tau = \lceil \frac{k+1}{2} \rceil \end{cases}$$

for some $\Delta \in \mathbb{R}_{++}$. Note that $\hat{y}_{k+1}^C = \hat{x}_{k+1}^C + \frac{2\Delta}{k}$. Since only the values of y_τ for the midpoints $\tau = \lfloor \frac{k+1}{2} \rfloor, \lceil \frac{k+1}{2} \rceil$ change, we have that the covariance

---

[14]Recall that $\sum_{\tau=1}^{k} \tau^2 = \frac{k^3}{3} + \frac{k^2}{2} + \frac{k}{6}$ by Faulhaber's formula (Conway and Guy, 1991, p. 106).

$Cov(T, Y)$ for k even coincides with the one found in Case 1. Thus, the slope $\hat{\beta}_1$ of the regression line does not change (i.e., $\hat{\beta}_1 = 1$), whereas its intercept $\hat{\beta}_0$ increases from 0 to $\frac{2\Delta}{k}$. Therefore, $\hat{y}_{k+1}^S = \hat{x}_{k+1}^S + \frac{2\Delta}{k}$, and the overall estimate is

$$\hat{y}_{k+1} = (1 - R_k^2)\hat{x}_{k+1}^C + R_k^2 \hat{x}_{k+1}^S + \frac{2\Delta}{k}.$$

Analogously to Case 1, it follows that $\hat{y}_{k+1} < \hat{x}_{k+1}$ if and only if Δ satisfies $(1 - R_k^2)(\hat{x}_{k+1}^S - \hat{x}_{k+1}^C) > \frac{2\Delta}{k}$, that is,

$$1 - R_k^2 > \frac{4\Delta}{k(k+1)}. \tag{14}$$

Express R_k^2 as ρ_k^2. As argued above, the covariance does not change compared to Case 1. We only need to compute the standard deviation of Y, which is given by

$$\sigma_Y = \sqrt{\sum_{\tau=1}^{k} y_\tau^2 - k\bar{y}^2} = \sqrt{\frac{k^3}{3} + \frac{k^2}{2} + \frac{k}{6} + 2\Delta(k+1) + 2\Delta^2 - k\left(\frac{k+1}{2} + \frac{2\Delta}{k}\right)^2}$$

$$= \sqrt{\frac{k^2(k^2-1) + 24\Delta^2(k-2)}{12k}}.$$

Hence, for k even, we have $R_k^2 = \frac{k^2(k^2-1)}{k^2(k^2-1) + 24\Delta^2(k-2)}$. Using this expression for R^2, we can rewrite condition (14) as $\Delta > \frac{4\Delta^2}{k^2+k} + \frac{k(k-1)}{6(k-2)}$ or, equivalently, $\Delta\left(1 - \frac{4\Delta^2}{k^2+k}\right) > \frac{k(k-1)}{6(k-2)}$. The left-hand side of the last inequality, as a function of Δ, is maximized at $\Delta^* = \frac{k^2+k}{8}$. Note that $\Delta^*\left(1 - \frac{4\Delta^*}{k^2+k}\right) > \frac{k(k-1)}{6(k-2)}$ holds for all even $k \geq 4$.

We conclude that, for each $k \geq 3$ and given $\{(\tau, x_\tau)\}_{1 \leq \tau \leq k}$, there exists a $\Delta \in \mathbb{R}_{++}$ such that the corresponding sequence $\{(\tau, y_\tau)\}_{1 \leq \tau \leq k}$ defined above satisfies the properties described in Proposition 4. ∎

6   Bibliography

Allais, M. (1953), "Le Comportement de L'Homme Rationnel devant le Risque: critique des Postulats et Axiomes de l'Ecole Americaine", *Econometrica*, **21**: 503-546.

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

Camerer, C. and M. Weber (1992), "Recent Developments in Modeling Preferences: Uncertainty and Ambiguity", *Journal of Risk and Uncertainty*, **5**: 325-370.

Chew, S. H. (1983), "A Generalization of the Quasilinear Mean with Applications to the Measurement of Income Inequality and Decision Theory Resolving the Allais Paradox" , *Econometrica,* **51**: 1065-1092.

Choquet, G. (1953-4), "Theory of Capacities" , *Annales de l'Institut Fourier* (Grenoble), **5**: 131-295.

Conway, J. H. and R. K. Guy (1996), *The Book of Numbers*, New York: Springer-Verlag.

de Finetti, B. (1931), Sul Significato Soggettivo della Probabilità, *Fundamenta Mathematicae*, **17**: 298-329.

——— (1937), "La Prevision: ses Lois Logiques, ses Sources Subjectives", Annales de l'Institut Henri Poincare, **7**: 1-68.

Dempster, A. P. (1967), "Upper and Lower Probabilities Induced by a Multivalued Mapping", *The Annals of Mathematical Statistics* **38**: 325-339.

Di Tillio, A., I. Gilboa, and L. Samuelson (2013), "The Predictive Role of Counterfactuals", *Theory and Decision*, **74**: 167-182.

Eichberger. J. and A. Guerdjikova (2013), "Ambiguity, Data and Preferences for Information – A Case-Based Approach", *Journal of Economic Theory*, **148**: 1433-1462.

Ellsberg, D. (1961), "Risk, Ambiguity and the Savage Axioms", *Quarterly Journal of Economics*, **75**: 643-669.

Fix, E. and J. Hodges (1951), "Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties". Technical Report 4, Project Number 21-49-004, USAF School of Aviation Medicine, Randolph Field, TX.

——— (1952), "Discriminatory Analysis: Small Sample Performance". Technical Report 21-49-004, USAF School of Aviation Medicine, Randolph Field, TX.

Gilboa, I. and M. Marinacci (2013), "Ambiguity and the Bayesian Approach", in *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress of the Econometric Society*, Daron Acemoglu, Manuel Arellano, and Eddie Dekel (Eds.). New York: Cambridge University Press.

Gilboa, I., A. Postlewaite, and D. Schmeidler (2008), "Probabilities in Economic Modeling", *Journal of Economic Perspectives*, **22**: 173-188.

——— (2009), "Is It Always Rational to Satisfy Savage's Axioms?", *Economics and Philosophy*, **25**: 285-296.

——— (2012), "Rationality of Belief", *Synthese*, **187**: 11-31.

Gilboa, I., L. Samuelson, and D. Schmeidler (2013), "Dynamics of Inductive Inference in a Unified Model", *Journal of Economic Theory*, **148**: 1399-1432.

——— (2015), *Analogies and Theories: Formal Models of Reasoning.* (The Lipsey Lectures, 2015) Oxford: Oxford University Press.

Gilboa, I. and D. Schmeidler (1989), "Maxmin Expected Utility with a Non-Unique Prior", *Journal of Mathematical Economics*, **18**: 141-153.

——— (1995), "Case-Based Decision Theory", *The Quarterly Journal of Economics*, **110**: 605-639.

——— (2001), *A Theory of Case-Based Decisions*, Cambridge: Cambridge University Press.

——— (2012), *Case-Based Predictions.* World Scientific Publishers, Economic Theory Series (Eric Maskin, Ed.), 2012.

Gilboa, I., D. Schmeidler, and P. P. Wakker (2002), "Utility in Case-Based Decision Theory", *Journal of Economic Theory*, **105**: 483-502.

Harless, D. and C. Camerer (1994), "The Utility of Generalized Expected Utility Theories", *Econometrica*, **62**: 1251-1289.

Hume, D. (1748), *An Enquiry Concerning Human Understanding.* Oxford: Clarendon Press.

Kahneman, D. and A. Tversky (1979), "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, **47**: 263-291.

Keynes, J. M. (1921), *A Treatise on Probability.* London: MacMillan and Co.

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

Knight, F. H. (1921), *Risk, Uncertainty, and Profit.* Boston, New York: Houghton Mifflin.

Köbberling, V. and Wakker, P. P. (2003), "Preference Foundations for Non-expected Utility: A Generalized and Simplified Technique", *Mathematics of Operations Research*, **28**: 395-423.

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

——— (2006b), "Dynamic Variational Preference", *Journal of Economic Theory* **128**: 4-44.

Quiggin, J. (1982), "A Theory of Anticipated Utility", *Journal of Economic Behaviorand Organization*, **3**: 225-243.

Ramsey, F. P. (1926a), "Truth and Probability", in R. Braithwaite (ed.), (1931), *The Foundation of Mathematics and Other Logical Essays.* London: Routledge and Kegan.

Ramsey, F. P. (1926b), "Mathematical Logic", *Mathematical Gazette*, **13**: 185-194.

Savage, L. J. (1954), *The Foundations of Statistics.* New York: John Wiley and Sons. (Second addition in 1972, Dover)

Schank, R. C. (1986), *Explanation Patterns: Understanding Mechanically and Creatively.* Hillsdale, NJ, Lawrence Erlbaum Associates.

Schmeidler, D. (1989), "Subjective Probability and Expected Utility without Additivity", *Econometrica*, **57**: 571-587. (Working paper, 1982).

Shafer, G. (1976), *A Mathematical Theory of Evidence.* Princeton: Princeton University Press.

——— (1986), "Savage Revisited", *Statistical Science*, **1**: 463-486.

Simon, H. A. (1957), *Models of Man.* New York: John Wiley and Sons.

Simpson, E. H. (1951), "The Interpretation of Interaction in Contingency Tables", *Journal of the Royal Statistical Society*, Series B. **13**: 238–241.

Tversky, A., and D. Kahneman (1992), "Advances in Prospect Theory: Cumulative Representation of Uncertainty", *Journal of Risk and Uncertainty*, **5**: 297-323.

von Neumann, J. and O. Morgenstern (1944), *Theory of Games and Economic Behavior.* Princeton: Princeton University Press .

Wakker, P. P. (1989), "Continuous subjective expected utility with non-additive probabilities", *Journal of Mathematical Economics*, **18**: 1-27.

Wakker, P. P. (2010), *Prospect Theory.* Cambrdige: Cambridge University Press.

Yaari, M. E. (1987), "The Dual Theory of Choice under Risk", *Econometrica*, **55**: 95-115.