Games and Economic Behavior 123 (2020) 22–40

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Theories and cases in decisions under uncertainty ☆

Itzhak Gilboa [a,b,*], Stefania Minardi [a], Larry Samuelson [c]

[a] HEC, Paris, France
[b] Tel-Aviv University, Israel
[c] Yale University, United States of America

ARTICLE INFO

Article history:
Received 22 January 2020
Available online 17 June 2020

Keywords:
Decision under uncertainty
Case-based reasoning
Rule-based reasoning
Theories

ABSTRACT

We present and axiomatize a model combining and generalizing theory-based and analogy-based reasoning in decision under uncertainty. An agent has beliefs over a set of theories describing the data generating process, given by decision weights. She also puts weight on similarity to past cases. When a case is added to her memory and a new problem is encountered, two types of learning take place. First, the decision weight assigned to each theory is multiplied by its conditional probability. Second, subsequent problems are assessed for their similarity to past cases, including the newly-added case. If no weight is put on past cases, the model is equivalent to Bayesian reasoning over the theories. However, when this weight is positive, the learning process continually adjusts the balance between case-based and theory-based reasoning. In particular, a "black swan" which is considered a surprise by all theories would shift the weight to case-based reasoning.

1. Introduction

1.1. Motivation

Savage's (1954) derivation of expected utility maximization is a powerful argument for the supposition that people behave as if they have subjective probabilities, but it says little about the nature of such beliefs. This paper models agents who form beliefs based on general theories as well as on specific past cases.[1] The two approaches to belief formation depict complementary modes of reasoning and have been studied independently of one another. By contrast, introspection, as well as some evidence, suggest that individuals may take into account both criteria in their belief formation process.

To illustrate, consider the attack on the World Trade Center on September 11, 2001; the New York Stock Exchange remained closed for the following 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 (a seven percent decline) was based on the

☆ We thank David Schmeidler for comments and discussions. We thank Xiangliang Li, a coeditor, an associate editor, and two referees for comments and suggestions. Gilboa gratefully acknowledges ISF Grant 1077/17. Gilboa and Minardi gratefully acknowledge the Investissements d'Avenir ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX-0047. Samuelson gratefully acknowledges NSF Grant 1459158.

* Corresponding author.
E-mail addresses: tzachigilboa@gmail.com (I. Gilboa), minardi@hec.fr (S. Minardi), larry.samuelson@yale.edu (L. Samuelson).

[1] There have been attempts to complement the general paradigm by a model of the way that beliefs are formed. The works closest to ours are Gilboa and Samuelson (2012), who address various ways in which similarity to past-cases can be used to assess probabilities, and Gilboa et al. (2013), who consider a unified model of belief formation. However, the former ignores general rules, and the latter only models deterministic ones. The present paper focuses on the combination of probabilistic theories with case-based reasoning.

https://doi.org/10.1016/j.geb.2020.06.001
0899-8256/

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 the day before the attack, it is likely that his answer would have been based on theories describing distributions of random variables, perhaps with Bayesian beliefs over these theories. However, each of these theories would almost certainly have given very low probability to an attack like that of September 11. An agent who uses only theories would have had to make predictions by appealing to those theories that were least wrong, as it were. By contrast, the expert in question switched to a completely different mode of reasoning. It seemed that, given an event that was surprising to all theories, his weight of reasoning shifted from the theories to cases. He wasn't only learning which theory is more trustworthy than others, but also how all theories taken together should be weighed in relation to the less ambitious mode of reasoning by analogies.[2]

This paper develops a model of decision making that combines theory-based and case-based reasoning and deals with some of their limitations. When evaluating the probability of an outcome, our agent looks at the probabilities of this outcome according to each theory she entertains, but also at similar past cases in which this outcome has occurred. The relative weight placed on theory-based, as compared to case-based reasoning depends on several factors. First, it depends on past success of the theories the agent entertains; if they all seem to do poorly, the relative weight of case-based reasoning will be higher. Second, the similarity of past cases would also matter; a very novel problem would be evaluated mostly by general theories, in the absence of concrete experience. Finally, the balance between the two modes of reasoning may also be a personal trait, which might depend on cognitive style, education, and so forth.[3]

Section 2 presents the general model of belief formation and discusses the way it captures learning. Section 3 provides an axiomatic foundation and an attendant representation theorem. It assumes that, for the purposes of elicitation, we may collect data on preferences over acts, each having a specific history (associating payoffs to the problems in which the act has indeed been chosen) and a specific payoff given each theory. Axioms on preferences over such acts identify the agent's (i) similarity function, between the current problem and sets of past problems; (ii) beliefs over theories; (iii) relative weight of the two modes of reasoning, and (iv) utility function. Section 4 presents some implications of the model. Section 4.1 explores the comparative statics of the model, explaining how we can characterize people as being relatively prone to rely on either theory-based or case-based reasoning. Section 4.2 formulates the classic Likelihood Principle within our framework by describing the revision of beliefs over theories as new cases are added to the history. As an illustration of the learning process, we discuss how the interplay between cases and theories may amplify the recency effect. Section 5 presents an example of how different modes of reasoning might affect play in finitely repeated games. Finally, Section 6 contains some concluding remarks and directions for future research. All proofs are collected in the Appendix.

2. The model

2.1. Belief formation

Let P be a set of *problems*, A a set of *acts*, and R a set of *outcomes*. A case is a triple (p, a, r), interpreted as follows: p denotes the circumstances of a decision problem; a is the act chosen in it; and r is the outcome that resulted. The set of all cases is $C \equiv P \times A \times R$. *Histories* are finite sequences of cases, and the set of all histories is $\mathcal{H} \equiv \cup_{n \geq 0} C^n$. We will assume without loss of generality that every problem can be encountered at most once in every history $H \in \mathcal{H}$. Formally, we are only interested in histories $H = (c_i = (p_i, a_i, r_i))_{i \leq n}$ where $p_i \neq p_{i'}$ for $i \neq i'$. It is therefore convenient to think of P as an infinite set. However, in each history only finitely many problems have been encountered. For simplicity, we also assume that A and R are finite.

Let T be a finite set of theories, where each theory provides a probabilistic prediction of the outcome $r \in R$ given (i) a history $H \in \mathcal{H}$; (ii) a new problem $p \in P$; and (iii) an act $a \in A$. Formally, a *theory* is a function

$$t : \mathcal{H} \times P \times A \to \Delta(R).$$

We will only be interested in theories $t(H, p, a)$ for which p does not appear in history H (that is, if $H = (c_i = (q_i, a_i, r_i))_{i \leq n}$, the notation (H, p, a) implicitly assumes that $p \neq q_i$ for all $i \leq n$).[4]

The agent faces a decision problem p and a history H. For each act a, she would like to generate beliefs over the possible outcomes in R. In this paper we consider probabilistic beliefs. The agent thus asks herself, given history $H = ((q_i, a_i, r_i))_{i \leq n}$ and the theories contained in T, how likely is an outcome r to result from act a in the current problem p? She finds support for the beliefs that outcome r would indeed transpire from two sources: the similarity of the current problem to past cases, and the likelihood of the theories. More precisely, the agent has:

[2] In a similar vein, Giacomini et al. (2020) report that professional inflation forecasters typically act as if they are Bayesian, until forced by a financial crisis to reconsider their models. Again, they may come up with probabilistic beliefs, but these are not always obtained just by updating the relative probabilities of the theories; at times of crisis the very possibility of theorizing is questioned, and the agent may rely on other modes of reasoning.

[3] In this sense the preference for general theories (vs. specific analogies) is similar to risk aversion, but it pertains to beliefs rather than to tastes.

[4] We exclude from consideration a host of objects that intuitively lie between cases and theories. For example, one might think of incorporating ambiguity by examining functions that map from $\mathcal{H} \times P \times A$ into capacities (rather than probability measures) over R.

(i) A similarity function σ, which is defined over sets of past problems and the current problem p, formally

$$\sigma : 2^P \times P \to \mathbb{R}_+, \tag{1}$$

where σ is non-decreasing in its first argument with respect to set inclusion; and[5]

(ii) A probability measure ν, defined over theories:

$$\nu : 2^T \to [0, 1].$$

These are used as follows. For a given history $H = ((q_i, a_i, r_i))_{i \le n}$ and act $a \in A$ and for each $r \in R$, the agent considers the weight

$$\sigma(\{q \in P | (q, a, r) \in H\}, p),$$

which is the similarity of all past problems in H in which a was chosen and r resulted. We can think of the agent as taking into account all these cases as supporting the prediction of r at the current problem p, should a be chosen, whereas all other cases—in which a was chosen but a different outcome ($r_i \neq r$) transpired, or in which a has not been chosen to begin with—do not provide support for this prediction.

Next the agent considers all theories, where each theory $t \in T$ predicts that r will occur with probability $t(H, p, a)(r)$. The agent's belief in the theory is captured by $\nu(t | H) \ge 0$ and the weighted sum of the supports is thus $\sum_{t \in T} \nu(t | H)\, t(H, p, a)(r)$. Taken together, the support for the prediction of r is

$$W(r | H, p, a) \equiv \sigma(\{q \in P | (q, a, r) \in H\}, p) + \sum_{t \in T} \nu(t | H)\, t(H, p, a)(r). \tag{2}$$

Assuming that (2) is positive for at least one r, we may also normalize the vector of weights to obtain a probability vector over R:

$$w(r | H, p, a) = \frac{W(r | H, p, a)}{\sum_{r \in R} W(r | H, p, a)}.$$

Two comments are in order. First, at this point we have no insight into the relative magnitudes of the similarity function σ and the probability measure ν, and hence of the weights $\sigma(\{q \in P | (q, a, r) \in H\}, p)$ and $\sum_{t \in T} \nu(t | H)\, t(H, p, a)(r)$ attached to the various outcomes by cases and theories. These relative magnitudes will be uniquely determined in our representation theorem, and can be interpreted as capturing the decision maker's cognitive style, i.e., her tendency to rely relatively heavily on either cases or theories in her reasoning.

Second, the way past cases are used in this model is more general than the case-based model of Gilboa and Schmeidler (1995). In the latter, similarity is evaluated between pairs of problems, and is aggregated additively over problems in history. That is, it is assumed that

$$\sigma(P_0, p) = \sum_{q \in P_0} s(q, p)$$

for some similarity function $s : P \times P \to \mathbb{R}_+$.[6] The main reason to consider non-additive aggregation is that learning in general, and statistical learning in particular, doesn't follow linear formulas. For example, if a specific outcome r resulted in all cases in history in which act a was chosen, the agent would have support for the supposition that a would also lead to r in the present problem. But this support would not grow linearly in the number of cases in which a was chosen. The first 100 cases would reach near certainty, while the following 100 would add little to the belief in this prediction. This is captured by a set function $\sigma(\cdot, p)$ which need not be additive with respect to disjoint unions of its first argument.

2.2. Learning

Learning is performed by the addition of cases to history, and the attendant updating of the probability weights of the theories. Consider a history $H = ((q_i, a_i, r_i))_{i \le n}$, leading to a probability assessment $\nu \in \Delta(T)$, and a new problem p. Assume that, in this problem, act a was chosen and outcome r resulted. The first component of learning is simply the addition of the new case (p, a, r) to history, so that the new history is

$$H' = ((q_i, a_i, r_i))_{i \le n+1}$$

[5] In fact, we will only use the function σ for pairs (P_0, p) such that P_0 is finite and does not include p.

[6] In fact, the term "similarity" is more intuitive in the original usage. A new problem is similar to each of past problems, not to a set of problems. However, in the context of the general model it seems clearer to retain the familiar term "similarity" rather than to introduce a new one.

with $(q_{n+1}, a_{n+1}, r_{n+1}) = (p, a, r)$. The second component is the updating of the agent's beliefs on theories:

$$\nu\left(t \mid H'\right) = \frac{\nu(t|H)\, t(H, p, a)(r)}{\sum_{t \in T} \nu(t|H)\, t(H, p, a)(r)}.$$

This second component is akin to Bayesian updating: given the realized outcome r at the observed (p, a), the agent revises her assessment of theory t by multiplying her initial belief of t by the conditional probability that the theory used to assign to outcome r (at problem p following history H) should act a be chosen. If $\varnothing$ denotes the empty history, then $\nu(t|\varnothing)$ is the agent's prior belief in theory t, whereas $\nu\left(t \mid H'\right)$ is her posterior belief. Note that our updating rule allows for the possibility of experiencing a new case which rules out particular theories. In this context, an agent would be expected to have a smaller set of theories, consisting only of those that were not ruled out by evidence. In turn, the relative importance of theories may well decrease.

Observe that in our formulation theories predict neither the problem p which may arise nor the agent's choice a. The former limitation may be relaxed with no difficulty: we may let the theories predict also the circumstances that the agent would encounter, and penalize them for "wrong" predictions as we do for "wrong" predictions of outcomes. The latter limitation is inherent to the model: as the agent has beliefs over the theories, if these suggest probabilities over acts, the agent would have beliefs over the act she is about to be taking. This is a type of circularity that we would like to avoid, as is standard in decision theory.

If the agent ignores similarity to (sets of) past cases, so that $\sigma \equiv 0$, our model boils down to a Bayesian model in which beliefs are generated by general theories, as is standard in Bayesian statistics. The agent starts with prior beliefs over theories and updates them according to Bayes's law. The introduction of the similarity function makes the model non-Bayesian. The similarity function $\sigma(\cdot, p)$ is defined for each problem p independently, and we have not imposed any constraints on its magnitude relative to ν. Thus, it is possible that the agent encounters a very novel situation p, so that $\sigma(\cdot, p) \equiv 0$ and she behaves at p as if she had Bayesian beliefs generated by the theories. It is also possible that σ takes very large values relative to those of ν, so that case-based reasoning is much more important at p than is theory-based Bayesian reasoning.

3. Elicitation

Given a history $H = ((q_i, a_i, r_i))_{i \le n}$ and a new problem p, we wish to elicit the agent's subjective parameters that feed into the belief formation process *and* the agent's utility function. These former include (i) the similarity function over sets of past problems (to the current problem p) and (ii) the subjective probability of theories. Observe that, by eliciting their values we also implicitly elicit the relative weight of case-based and theory-based reasoning.

3.1. Payoffs

Savage (1954) begins with the primitive concepts of a set of states and a set of outcomes. Acts are then functions that map from the set of states to the set of outcomes. Savage's elicitation result requires that the agent's preferences rank the set of all possible acts, i.e., the set of all possible functions from states to outcomes. This requirement poses no formal difficulties, though tension can arise when intuitive interpretations are attached to some of the acts.[7] Our elicitation result will similarly require preferences over a rich set of alternatives, with tensions again arising out of the fact that we think of each case specifying an intuitively interpreted act.

We introduce a set of *payoffs* G. Our basic requirement (Assumption 1) is that G be a connected topological space. This will be the case if G is a convex subset of $\mathbb{R}$, allowing us to interpret elements of G as monetary payoffs, or a convex subset of $\mathbb{R}^d$ (for a natural number d), allowing us to think of elements of G as commodity bundles.

The set of outcomes R and the set of payoffs G are distinct. We interpret the elements of R as material outcomes. For example, an outcome r_i may specify that a politician was caught in scandalous circumstances or that a financial asset increases in value. An element of G identifies the payoff to the agent associated with this outcome, such as the loss of one's income if the scandal is ruinous, or the financial windfall if the scandal facilitates a lucrative book deal, or the financial gain or loss if one has either purchased or shorted the asset. Importantly, the agent can imagine different payoffs in G associated to any outcome in R. Just as with the umbrella in Savage's world, once we allow the possibility that an agent may earn a positive payoff from the outcome increased-asset-value and a negative payoff from the outcome decreased-asset-value, we must also consider possibilities such as a negative payoff from the outcome increased-asset-value and a positive payoff from the outcome decreased-asset-value. The description of the case specifies the material outcome, but leaves the agent free to (perhaps counterfactually) consider different payoffs that could be associated with this outcome.

If, like Savage (1954), we were concerned only with characterizing preference orders without mentioning learning, we could dispense with the outcome set R and work only with the set of payoffs G. If we were concerned only with beliefs and their updating without mentioning preferences, we would dispense with the payoffs in G and work only with the set

[7] To echo a well-known example, the state space {rain, sun} and outcome space {wet, dry-and-encumbered, dry-and-unencumbered} allows one to examine actions interpreted as carrying an umbrella (rain→dry-and-encumbered, sun→dry-and-encumbered) and not carrying an umbrella (rain→wet, sun→dry-and-unencumbered), but then also requires the agent to evaluate the act (rain→dry-and-unencumbered, sun→dry).

of outcomes R. Given our aspiration to not only characterize preferences but examine the nature and updating of the beliefs behind these preferences, we need both sets. Intuitively, one may think of R as identifying features of the external world about which the agent may learn, and G as identifying the implications of these features for the agent's payoff.

For each act a we consider the set of cases in which it has been chosen (in the given history H),

$$H_a = ((q_i, a_i, r_i))_{i \le n, a_i = a}$$

and attach a payoff vector $x_a : H_a \to G$ to these cases. We also attach a payoff vector to the theories, $y_a : T \to G$. As Section 3.2 explains, we then ask the agent to rank *profiles* of the form (x_a, y_a) that are payoff (G)-valued vectors in the appropriate space. (Explicitly, $(x_a, y_a) \in G^m$ where $m = |H_a| + |T|$.)

The interpretation of a vector of payoffs x_a is as follows. We ask the agent to imagine an act that was chosen precisely in the cases H_a, and yielded payoffs $x_a \in G^{|H_a|}$ in those cases. The agent is capable of assessing various vectors x_a. For example, we may consider a financial crisis as the outcome r_i, but ask the agent to consider either a high payoff (from an appropriately contrarian strategy that flourishes in a crisis) or a low payoff (from a risky strategy and subsequent bankruptcy). Notice that when the agent is asked to consider the payoffs in x_a that might be attached to the cases H_a, she is *not* asked to imagine different outcomes r_i. The outcomes r_i are determined by the cases actually experienced, and are relevant for updating the agent's beliefs over her theories. However, the agent is capable of evaluating arbitrary payoff vectors that might be attached to these outcomes, giving us the sufficiently rich preference domain needed for a representation result.

The vector of payoffs y_a is to be interpreted as specifying, for each theory $t \in T$, the payoff $y_a(t)$ that will materialize should theory t be correct and action a taken. Observe that, as opposed to pulling balls out of urns, the resolution of this uncertainty is not immediate. It may only be in the limit that we can tell which theory is true. Yet, there is no conceptual difficulty in imagining contracts that will guarantee given payoffs in the limit. For example, uncertainty about global warming or the success of autonomous cars will only be resolved years hence. Economic agents buy and sell assets—ranging from real estate to equity—that are contractual agreements over payoffs in the long run. Observing whether someone purchases a coastal house below sea level may give us a hint about one's belief in the extent of global warming. This economic decision can be made at present, and reflect beliefs in theories that will only be proven true or false in the long run.

3.2. Profiles

Given history $H = ((q_i, a_i, r_i))_{i \le n}$ and the set T of theories, for each act $a \in A$ we consider the set of all a-profiles

$$\mathcal{F}_a \equiv G^{H_a \cup T} = \{ f \mid f = (x_a, y_a) \colon H_a \cup T \to G \}.$$

For any two distinct acts $a, b \in A$, the agent is assumed to have preferences between any two profiles, $f \in \mathcal{F}_a$, $g \in \mathcal{F}_b$. We also define

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

The agent's preferences thus allow her to answer questions of the following form: "Assume that act a resulted in the past profile x_a in those problems in which it was chosen, and that it will yield future profile y_a for the theories. Assume further that act b resulted in the past profile x_b in those problems in which it was chosen, and that it will yield profile y_b for the theories. Would you rather choose a or b?" The agent thus calls upon both her past experience and her future expectations in evaluating the two acts. The agent is capable of reasoning about the desirability of a boxing match with the reigning boxing champion under the presumption (or the future profile) that the agent will thereby win a great deal of money, even if she has no box-office drawing power, and the agent is similarly capable of reasoning about such an act under the presumption (or the past profile) that she has won a great deal of money in each of her previous boxing matches, even if she did not.

For any act $a \in A$, the set of all possible past profiles is given by G^{H_a}. A typical case $(q_i, a, r_i) \in H_a$ has actually occurred, in which the agent chose act a, the outcome was r_i, and the agent received some payoff. However, we ask the agent to imagine the desirability of the act had the payoff in this case been different. We can interpret this as reflecting the agent's ability to engage in counterfactual reasoning.

In presenting classical axiomatic results such as Savage's (1954), one thinks of elicitation of beliefs given preferences between acts, i.e., functions from states of the world to outcomes, which are equivalent to our payoffs. It is usually assumed that the attendant payoffs will soon be known. Thus, in Savage's omelette example, one will learn whether the sixth egg is rotten just after the agent makes the decision. By contrast, our payoff profiles allow a more nuanced relationship with time: past profiles involve replacing the payoffs, which are known to have occurred, by counterfactual ones, whereas theory-dependent, future profiles will have their payoffs determined only in the long run.

Both of these may seem to be problematic for tests of our axioms, and they may suggest that the axiomatization is more of a gedankenexperiment than an actual elicitation procedure. Notice, however, that similar problems arise with applying Savage's model to many set-ups that go beyond the omelette example. When Savage (1954) is cited as a reason to believe that agents have a common prior over their types (as in Harsanyi (1967–1968)), which they update to posteriors once

each knows her own type, the Savage questionnaire required to elicit preferences also involves hypothetical questions. And, similarly, when Savage's model is used to justify prior beliefs over outcomes such as global warming, one needs again to imagine uncertainty that will not be resolved in the near future. More importantly, these limitations do not render the preferences we deal with inherently unobservable. Acts with different past payoffs than those actually experienced can be used in experiments.[8] For example, an agent may be asked to make investment decisions in assets, where all the information provided about them is their past performance in given cases. And betting about long-run payoffs can be observed when agents purchase assets such as real estate.

3.3. Preferences

Our primitive is a binary relation $\succsim_{H,T}$ on $\mathcal{F}$ describing preferences over profiles for a fixed problem p, which depends on the history H of experienced cases and the set T of envisaged theories.

We do not assume that this binary relation is complete. Since the agent is asked to imagine different payoffs of acts in problems in which they were actually chosen, we will not ask her to compare pairs of vectors of payoffs defined over the history of the *same* act. In particular, we do not assume the agent ranks two profiles such as $f = (x_a, y_a)$, $f' = (x'_a, y'_a) : H_a \cup T \to G$ that are defined over the history of the same act. Beyond this restriction, we will assume that preferences are complete, so that, for any distinct acts a, b, the vectors (x_a, y_a), (x_b, y_b) can be compared for any x_a, x_b, y_a, y_b.

Returning to our example, the agent is allowed to consider whether she would rather fight the reigning boxing champion (act a) or not (act b) under the presumption that she has won great sums of money in the previous problems H_a in which she chose to fight. She is also allowed to consider whether she would rather fight the reigning boxing champion (act a) or not (act b) under the presumption that she has lost disastrously in the previous problems H_a in which she chose to fight. However, she is not asked to choose between (i) fighting under the presumption that she has been earned great sums in the previous problems H_a and (ii) fighting under the presumption that she has lost disastrously in the previous problems H_a. Under some richness conditions, preferences between such profiles f and f' will follow from the preferences between profiles defined over different acts and transitivity.

The set of future profiles G^T 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 $\succsim_{H,T}$ to G^T is denoted by $\succsim_T$. Moreover, we will use the notation $\succsim$ to refer to a binary relation defined on G as usual: For every $\alpha, \beta \in G$, $\alpha \succsim \beta$ if and only if $(\alpha, \alpha, \ldots, \alpha) \succsim_T (\beta, \beta, \ldots, \beta)$. For $\alpha \in G$, the element $\boldsymbol{\alpha}$ stands for the constant profile which yields α on the appropriate domain.

For $a \in A$, $f \in \mathcal{F}_a, \alpha \in G$, and $s' \in H_a \cup T$, 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 $\alpha, \beta \in G$ and $s', s'' \in H_a \cup T, s' \neq s''$, 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 $s \in H_a \cup T$ is *null* on $F \subseteq \mathcal{F}_a$ if $\alpha\{s\}f \sim_{H,T} f$ for all $\alpha\{s\}f$ and f in F.

3.4. Assumptions

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 $\sigma : 2^S \to [0, 1]$ such that: (i) $\sigma(\varnothing) = 0$; (ii) $A \subseteq B$ implies $\sigma(A) \leq \sigma(B)$; and, (iii) $\sigma(S) = 1$. We say that a set function $\sigma : 2^S \to \mathbb{R}_+$ is a *pseudo-capacity* if it satisfies conditions (i) and (ii) and $\sigma(S) \in [0, \infty)$. It is convenient to condition on the fixed problem p and note that the function σ in (1) induces a pseudo capacity $\sigma : 2^P \to \mathbb{R}_+$ (for which we retain the name σ) and for a history H to adopt the shorthand notation $\sigma(H)$ for $\sigma(\{q \in P : (q, a_i, r_i) \in H\})$.

Second, we assume throughout the following:

Assumption 1. The set G is a connected topological space (Simmons 1983, p. 143) and, for every $a \in A$, the set $\mathcal{F}_a$ is endowed with the product topology.

In our leading interpretations of G as the set $\mathbb{R}$ of monetary payoffs or the set $\mathbb{R}^d$ of commodity bundles, G is a connected topological space.

Finally, 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 $H_{a'} = \varnothing$.

[8] Bleichrodt et al. (2017) show how payoffs from past cases can be engineered so as to elicit the preferences associated with case-based decision theory.

3.5. Axioms on preferences

We impose the following axioms on $\succsim_{H,T}$. 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 $\succsim_{H,T}$ on $\mathcal{F}$ is reflexive and transitive. For every $a, b \in A, a \neq b$, every $f \in \mathcal{F}_a$ and $g \in \mathcal{F}_b$, $f \succsim_{H,T} g$ or $g \succsim_{H,T} f$.

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 $s \in H_a \cup T$, and every $g \in \mathcal{F}_b$,

- $f' \succsim_{H,T} g$ implies $f \succsim_{H,T} g$,
- $g \succsim_{H,T} f$ implies $g \succsim_{H,T} f'$.

Axiom 3 *(Continuity).* For every $a, b \in A, a \neq b$, and every $f \in \mathcal{F}_a$, the sets $\{g \in \mathcal{F}_b : f \succ_{H,T} g\}$ and $\{g \in \mathcal{F}_b : g \succ_{H,T} f\}$ 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 theory (because for any $a \in A$ and $f \in \mathcal{F}_a$, the sets $\{g \in \mathcal{F}_{a'} : f \succ_{H,T} g\}$ and $\{g \in \mathcal{F}_{a'} : g \succ_{H,T} f\}$ are both nonempty). Hence, $\nu(T) > 0$. This can be relaxed by weakening Assumption 2 to allow all theories 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 only apply to past problems. Thus, two profiles are *historically-comonotonic* if they do not rank any two problems differently. Appendix A explains how pairwise historical comonotonicity is a weaker requirement than standard comonotonicity notions (e.g., Köbberling and Wakker (2003, p. 400)) and shows that pairwise-historically-comonotonic sets share some familiar properties of comonotonic sets.

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

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

for some $p, p' \in H_a$.

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 imposing its implication on pairwise historically-comonotonic rather than simply comonotonic sets. A familiar version of comonotonicity would suffice to characterize decisions based only on theories or based only on problems, but this somewhat stronger notion is needed to connect the two.

Axiom 4 *(Pairwise comonotonic ("BiCo") tradeoff consistency).* For every $a \in A, f, f', g, g' \in \mathcal{F}_a, \alpha, \beta, \gamma, \delta, \delta^* \in G$, and $s, s' \in H_a \cup T$, if

$$\alpha\{s\}f \sim_{H,T} \beta\{s\}f', \quad \gamma\{s\}f \sim_{H,T} \delta\{s\}f', \quad \alpha\{s'\}g \sim_{H,T} \beta\{s'\}g',$$

then

$$\gamma\{s'\}g \sim_{H,T} \delta^*\{s'\}g' \iff \delta \sim \delta^*,$$

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 *historically-comonotonic* sets, and s and s' are non-null on the first and second set, respectively.

The interpretation of this axiom is familiar. Suppose (i) a switch from α to β in (problem or theory) s just balances a switch from f to f' elsewhere, (ii) a switch from γ to δ in s similarly just balances a switch from f to f' elsewhere, and (iii) a switch from α to β in (problem or theory) s' just balances a switch from g to g' elsewhere. Then the agent is consistent in how she evaluates tradeoffs, in that a switch from γ to δ in s' similarly just balances a switch from g to g' elsewhere. Note that a set of profiles that have the same past profile is trivially a pairwise historically-comonotonic set. This means that if $s, s' \in T$ then the above axiom is equivalent to imposing the full tradeoff consistency property over preferences restricted to future profiles. Köbberling and Wakker (2003), in the process of using tradeoff consistency axioms to characterize expected utility, Choquet expected utility, and prospect theory, argue that such axioms have the advantage of depending only on indifferences.

The next axiom guarantees the existence of a "neutral payoff" that is independent of the choice of act $a \in A$. Suppose that the agent compares an act $y \in G^T$, with no history, to an act with the same future profile but also with the history H_a. We expect that either (i) past profiles do not affect the ranking of the two acts (that is, problems in H_a are null), which will be the case if the agent believes that all the information in past cases is already incorporated into the likelihood of theories; or (ii) some past profiles make the act with the history more attractive and some make it less so. In either case, one would expect there to be a payoff $\alpha^* \in G$ such that the constant past profile $\boldsymbol{\alpha}^*$ makes y just as desirable as it would be with no history at all, that is, $(\boldsymbol{\alpha}^*, y) \sim_{H,T} y$. (In case (i) this would be the case for any α^*, and in case (ii) it would follow for some α^* by continuity.)

One could think of models in which this neutral payoff $\boldsymbol{\alpha}^*$ depends on the act a under discussion and/or on the future profile y. Having different neutral payoffs for different acts a might occur if the agent has a certain intrinsic preference 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 payoffs matter. Thus, we wish to require that the payoff α^* be independent of the act under consideration. Moreover, in line with the problem-theory separability, we also require that this "neutral" payoff be independent of the future profile y:

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

3.6. The representation result

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

Theorem 1. *Let $\succsim_{H,T}$ be a binary relation on $\mathcal{F}$. Assume that Richness holds and that there exist at least two non-null theories. The following statements are equivalent:*

(i) *$\succsim_{H,T}$ satisfies Restricted Weak Order, Monotonicity, Continuity, BiCo Tradeoff Consistency, and Consequentialism;*

(ii) *There exist a continuous function $u\colon G \to \mathbb{R}$ such that $0 \in int(u(G))$, a pseudo-capacity σ_a on 2^{H_a} for each $a \in A$, and a probability ν on T, such that, for all $a, b \in A$, $f \in \mathcal{F}_a$, $g \in \mathcal{F}_b$,*

$$f \succsim_{H,T} g \iff \int_{H_a} u(f)d\sigma_a + \int_T u(f)d\nu \geq \int_{H_b} u(g)d\sigma_b + \int_T u(g)d\nu. \tag{3}$$

According to representation (3), our agent compares acts on the basis of both their performance in past problems and their expected future performance. The agent's beliefs about the relevance of past problems to the present one are captured by a collection $(\sigma_a)_{a\in A}$ of pseudo-capacities. These are monotone set functions that are not normalized to 1, because the past problems in which different acts were chosen may be of different relevance. On the other hand, the agent's beliefs about the likelihood of theories are reflected by a single probability measure ν. The representation depicts a formal link between these two criteria. Indeed, the overall utility of an act a depends not only on how the two criteria evaluate it independently, but also on the relative weight placed on each criterion.

In the extreme situations in which the agent puts all the weight on either past problems or on theories, we recover the special cases of case-based decision theory or of Subjective Expected Utility, respectively. The relative weight of the two modes of reasoning (say, the ratio of $\sigma_a(H_a)$ to $\nu(T)$) is derived from preferences given a specific history, and it is expected to change from one history to another depending both on the similarity of past cases to the new problem encountered, and on the performance of theories in these past cases. Consider our motivating example again. The day before September 11, the market analyst might have entertained a collection of theories T, each with some probability weight $\nu(t) > 0$. The history $H = ((q_i, a_i, r_i))_{i\leq n}$ he had was pretty rich, but we can assume, for simplicity, that the similarity function σ satisfied $\sigma\left(\{q_i\}_{i\leq n}, q_{n+1}\right) \approx 0$ for a typical new problem q_{n+1}. That is, there is a sense that all the information in past cases has already been gleaned and processed, crystallized into theories, and no additional weight should be put on past cases above and beyond their impact on future beliefs. However, the new case (q_i, a_i, r_i), involving the attacks on the World Trade Center and the Pentagon, was hardly predicted by any of these. We may thus assume that, for a small $\varepsilon > 0$, $\nu(t) < \varepsilon$. By contrast, the new problem was perceived as similar to previous attacks (such as Pearl Harbor), and similarity-based prediction became dominant.

We will say that a triple $\left(u, (\sigma_a)_{a\in A}, \nu\right)$ represents a binary relation $\succsim_{H,T}$ if it satisfies the properties of Theorem 1. In the degenerate (but important) case in which all the information is incorporated in ν, and hence all of H_a are null, one has the standard result that $(\sigma_a)_{a\in A}$ and ν are unique and u is unique up to increasing affine transformations, i.e., u is an interval scale. In general, $(\sigma_a)_{a\in A}$ and ν are unique and u is unique up to increasing linear transformations, i.e., u is a ratio scale. In the latter case, the utility function cannot be shifted by a constant: as long as there are some acts a for which H_a is not a null set, one has to make sure that the utility function assigns the value 0 to these payoffs α^* that satisfy the condition of Consequentialism. Finally, note that the uniqueness properties would be weaker if one does not require the existence of at least two non-null theories. More explicitly:

Proposition 1. *Assume that Richness holds and that there exists at least two non-null theories. Two triples* $\left(u, (\sigma_a)_{a\in A}, \nu\right)$ *and* $\left(\hat{u}, \left(\hat{\sigma}_a\right)_{a\in A}, \hat{\nu}\right)$ *represent the same binary relation* $\succsim_{H,T}$ *if and only if*

- *(i)* $\hat{\sigma}_a = \sigma_a$ *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* $\hat{u} = \lambda u + d$*.*

As anticipated earlier on, this strong uniqueness property allows us to relate the representations of any two preferences $\succsim_{H,T}$ and $\succsim_{H',T}$, where $H = ((q_i, a_i, r_i))_{i\le n}$ and $H' = ((q_i, a_i, r_i))_{i\le n+1}$, in terms of learning. First, we identify the impact of the additional case on the Bayesian updating of the probability over theories: in particular, the prior probability assigned to each theory is multiplied by its conditional probability (given the realized case). Second, the next problem is assessed for its similarity to the larger set of past cases. As a result of this learning process, the relative weight between theory-based and analogy-based reasoning may well shift.

4. Implications

4.1. Cases vs. theories

The decision criterion given by (3) combines information from past cases and information about theories. How much weight will be placed on past cases and how much on theories?

We expect the answer to depend on a variety of factors, relating both to the problem and to the agent. It is no surprise that the mix between cases and theories should differ across problems. If asked to predict the outcome of a coin toss, we expect most people to rely primarily on reasoning about theories. When asked to reason about the implications of autonomous cars, we expect most people to rely on cases involving previous technological innovations. Even the best chess players are unable to enumerate a complete set of theories, and instead rely heavily on analogies to past cases, reflected in their intensive study of previous games. Checkers has been solved, and the best checker players are more likely to rely on theories.[9]

We are interested here in differences across people in proclivities to rely on cases or theories. Our goal is to offer a behavioral definition of such differences, and then to link this definition to our representation of preferences.

4.1.1. Cognitive styles

We conjecture that people differ in their "cognitive style" in terms of their propensity to rely relatively heavily on the case-based or the theory-based modes of reasoning. Similar to the way in which economic agents may differ in terms of their risk aversion or their ambiguity aversion, they will differ in the degree to which they believe that they can glean all the information in past cases and mold it into a probability on theories. Some people may be prone to believe that past cases are still relevant more than others.

In any attempt to capture such differences, we wish to assume away other differences, such as concavity of the utility function, or any feature of the beliefs that may capture some notion of ambiguity aversion. Further, we do not wish to compare individuals who differ in their memory of cases or ability to conceive of theories. Hence, we limit attention to the comparison of agents who are identical in terms of (i) the memory of past cases and the set of theories; (ii) their attitude towards risk and uncertainty; (iii) their similarity judgment of past cases, or sets thereof.

Intuitively, we expect differing proclivities to engage in case-based vs. theory-based reasoning to be reflected in the relative weights in (3) captured by the pseudo-capacity σ_a on cases and the probability ν on theories. To make this precise, we offer a behavioral definition of what it means for the agents to be identical in the terms described in the previous paragraph, and then a behavioral definition of what it means for one agent to be more prone to theory-based reasoning than another. We then confirm that these translate into statements about the weights captured by the agents' pseudo-capacities on cases.

We begin with the following behavioral notion of comparability.

Definition 2. We say that $\succsim^1_{H,T}$ is *comparable* to $\succsim^2_{H,T}$ if the following conditions hold:

(a) For all $a, b \in A$, $x \in G^{H_a}$, $x' \in G^{H_b}$, and $y \in G^T$,

$$(x, y) \succsim^1_{H,T} \left(x', y\right) \iff (x, y) \succsim^2_{H,T} \left(x', y\right); \tag{4}$$

(b) For all $y, y' \in G^T$,

$$y \succsim^1_T y' \iff y \succsim^2_T y'. \tag{5}$$

[9] Madrigal (2017) reports that checkers legend Marion Tinsley, at the tenth move of a game, was able to see a path to victory by working through the game tree to the end, as much as 64 moves away (though the opponent resigned after only 26 more moves).

The comparisons in (4) and (5) involve profiles that differ only in terms of past profiles or only in terms of future profiles, so that no tradeoff between cases and theories arises. The requirement of comparability is that the two agents agree in their ranking of such profiles. Differences between the two agents will then arise out of the differing weights agents put on past cases and theories.

We can provide a behavioral notion of what it means for one agent to be more prone to theory-based reasoning than another.

Definition 3. Assume that $\succsim^1_{H,T}$ is comparable to $\succsim^2_{H,T}$. We say that $\succsim^1_{H,T}$ is more prone to theory-based reasoning than $\succsim^2_{H,T}$ if, for all $a, b \in A$ and $\alpha, \beta, \theta \in G$ such that $\beta \succ \alpha \succ \theta$, the vectors $(\boldsymbol{\theta}, \boldsymbol{\beta}) \in \mathcal{F}_a$ and $(\boldsymbol{\alpha}, \boldsymbol{\alpha}) \in \mathcal{F}_b$ satisfy

$$(\boldsymbol{\theta}, \boldsymbol{\beta}) \succsim^2_{H,T} (\boldsymbol{\alpha}, \boldsymbol{\alpha}) \implies \left(\boldsymbol{\theta}, \boldsymbol{\beta}\right) \succsim^1_{H,T} \left(\boldsymbol{\alpha}, \boldsymbol{\alpha}\right).$$

To interpret this definition, notice that $(\boldsymbol{\alpha}, \boldsymbol{\alpha})$ is a constant profile, yielding some payoff α in all past problems and all theories. Compare this to a profile $(\boldsymbol{\theta}, \boldsymbol{\beta})$ that is expected to deliver a payoff β better than α in all theories, but also yielded a payoff θ worse than α in all past problems. Suppose that agent 2 prefers $(\boldsymbol{\theta}, \boldsymbol{\beta})$ to $(\boldsymbol{\alpha}, \boldsymbol{\alpha})$. That is, receiving a more attractive payoff in the future compensates the gloomy memory of the past, making the overall profile at least as desirable as the constant one yielding simply α. Definition 3 states that, if agent 1 is more prone to theory-based reasoning than agent 2, then agent 1 should find the profile $(\boldsymbol{\theta}, \boldsymbol{\beta})$ better than the profile $(\boldsymbol{\alpha}, \boldsymbol{\alpha})$.

Note that the limit case of maximal proneness corresponds to full reliance on future-based reasoning. In this case, an agent will always find $(\boldsymbol{\theta}, \boldsymbol{\beta})$ strictly better than $(\boldsymbol{\alpha}, \boldsymbol{\alpha})$ as long as $\beta \succ \alpha$.[10] In particular, a simple implication of Definition 3 is that, given a maximally prone agent, any agent with comparable preferences will be less prone to future-based reasoning (or, equivalently, more prone to case-based reasoning).

We first characterize comparability of two agents. The notation $u^1 \approx u^2$ stands for $u^1 = \lambda u^2 + d$ for some $\lambda > 0$ and $d \in \mathbb{R}$.

Proposition 2. *Let $\succsim^1_{H,T}$ and $\succsim^2_{H,T}$ be two binary relations on $\mathcal{F}$ that are represented by $\left(u^1, \left(\sigma^1_a\right)_{a\in A}, \nu^1\right)$ and $\left(u^2, \left(\sigma^2_a\right)_{a\in A}, \nu^2\right)$, respectively. Then, the following conditions are equivalent:*

(i) *$\succsim^1_{H,T}$ is comparable to $\succsim^2_{H,T}$;*
(ii) *$u^1 \approx u^2$, $\nu^1 = \nu^2$, and, for all $a \in A$, $\sigma^1_a = \lambda_H \sigma^2_a$ for some $\lambda_H > 0$.*

If $\succsim^1_{H,T}$ and $\succsim^2_{H,T}$ are comparable, we will denote by $\lambda^{1,2}_H$ the coefficient λ_H provided by the above proposition. (Observe that it is unique, and clearly, $\lambda^{1,2}_H \lambda^{2,1}_H = 1$.)

The implication of this characterization is that, given two individuals, 1 and 2, such that $\succsim^1_{H,T}$ and $\succsim^2_{H,T}$ are comparable, we attribute any behavioral differences between them to a single factor, captured by $\lambda^{1,2}_H$ or $\lambda^{2,1}_H$, reflecting their "cognitive style": the degree to which they tend to trust their analysis of past cases. Note that the individuals are assumed not to differ in other cognitive capacities such as the prowess of their memory or imagination. They have the same set of cases in mind, recalled to the same (relative) degree, and the same set of theories in mind, judged to have the same (relative) likelihood. Further, their tastes are identical. Thus, the differences we might observe can only be attributed to the degree they are prone to case-based vs. theory-based reasoning.

We can now characterize proneness to theory-based reasoning, building on our characterization of comparability:

Proposition 3. *Let $\succsim^1_{H,T}$ and $\succsim^2_{H,T}$ be two binary relations on $\mathcal{F}$ that are represented by $\left(u^1, \left(\sigma^1_a\right)_{a\in A}, \nu^1\right)$ and $\left(u^2, \left(\sigma^2_a\right)_{a\in A}, \nu^2\right)$, respectively. Assume that $\succsim^1_{H,T}$ is comparable to $\succsim^2_{H,T}$. Then, the following conditions are equivalent:*

(i) *$\succsim^1_{H,T}$ is more prone to theory-based reasoning than $\succsim^2_{H,T}$;*
(ii) *$\lambda^{1,2}_H \leq 1$.*

According to Proposition 3, the value of the coefficient $\lambda^{1,2}_H$ provides a succinct comparative measure of proneness to future-based reasoning. Namely, being more prone translates into a coefficient smaller than 1. In turn, this means that agent 1 places a smaller relative weight on case-based reasoning than agent 2. The limit case of maximal proneness is characterized by $\lambda^{1,2}_H$ approaching 0.

[10] More precisely, observe that we can capture maximal proneness only as a limit case. The reason is that our definition of comparability implies that two comparable agents share the same set of non-null past problems.

4.1.2. Illustration

To illustrate the tradeoffs between case-based and theory-based reasoning, suppose a team of entrepreneurs seeks funding from a venture capital firm. The team describes their business plan in terms of a set of theories T dealing with the efficacy of their new cancer treatment, the appearance of possible competitors, the possible delays and costs involved in clinical trials, the amount insurers will pay for the treatment, and so forth. According to their estimates, the expected profits on the initial investment are very enticing.

John finds the description of the theories T and their attendant probabilities compelling, and has little experience with past cases. He adopts the probability $\nu \in \Delta(T)$ presented by the entrepreneurs and concludes that the investment looks quite promising.

Sarah, more experienced, responds skeptically to his enthusiastic description of the project, commenting that she sees nothing wrong with John's analysis, but "I've seen a parade of entrepreneurs in my time. All of them seemed convincing, but only a fraction of these projects actually made it. I can't challenge your construction of T or the weights you put on its elements, but I put considerable weight on all these past cases."

Rachel, with little experience with past cases, is also skeptical of the project, despite the fact that she objects to no details of his analysis. Rachel explains, "We can't be sure that even the most careful of calculations have captured all the relevant possibilities and weighted them appropriately. The only thing we can do is rely on our experience, scant as it might be, which forces us to be cautious about claims involving fantastic new technologies."

In this example, each new investment problem gives rise to a set of theories T and a probability ν describing beliefs about the theories. Sarah's reasoning differs from John's in that Sarah has a larger database of past cases upon which to call. Past cases are evaluated according to the $(\sigma_a)_{a\in A}$, which are only assumed to be pseudo-capacities, allowing a larger database to have larger values of $\sigma_a(H_a)$. As a result, John and Sarah may well agree on the probabilities of success of the present enterprise, and may even have the same cognitive style ($\lambda_H^{J,S} = 1$), but the very fact that Sarah has seen more cases can make her more skeptical. In contrast, we imagine John and Rachel as having the same memory of past cases, but having different cognitive styles. Rachel is less sanguine about the ability to imagine and evaluate theories, and as a matter of course places more weight on cases—we have $\lambda_H^{J,R} < 1$.

There are many other problems in which people reason with a combination of specific theories and cases. No two economic booms are precisely identical, and one can find theories explaining why the current economy will keep growing despite the experiences of past recessions. Indeed, Reinhart and Rogoff (2009) point to a wealth of theories claiming that "this time is different". However, it is also both natural and rational for an agent to reason that "While I cannot find any flaw with the arguments that underlie ν, I do not understand fully how the economy works, and I also cannot help seeing some similarities to past cases in which recessions did occur. Hence, it might be wise to take them into account alongside the theories listed in T." And we can well imagine different people resolving these conflicting forces by placing different emphases on case-based and theory-based reasoning.

4.2. Learning

4.2.1. The likelihood principle

The agent learns as new cases are added to her memory. The entry of these new cases into memory gives a direct learning effect, altering the realized values of the function σ in (1) and (2). They also prompt a revision in the probabilities attached to the theories in T in (2). We assume the agent updates her probability ν over theories, upon the addition of a new case to her history, by applying the classic Likelihood Principle: For every $H = ((q_i, a_i, r_i))_{i\le n}$, $H' = ((q_i, a_i, r_i))_{i\le n+1}$ and for any two theories t, t' with $\nu\left(t'|H\right), t'(H, p, a)(r_{n+1}) > 0$,

$$\frac{\nu\left(t|H'\right)}{\nu\left(t'|H'\right)} = \frac{\nu(t|H)\, t(H, p, a)(r_{n+1})}{\nu(t'|H)\, t'(H, p, a)(r_{n+1})}$$

This allows the model to be consistent with Bayesian reasoning when the agent relies entirely on theories.

Observe that the principle is stated in the language of conditional probabilities given the entire history, and it therefore does not assume statistical independence of consecutive cases.

4.2.2. Recency effects

Following a disaster, people seem to react in ways that are extreme in the short run, especially in comparison to the long run reactions. For example, following a plane crash, people may avoid flying for a while (or avoid flying with the same airline or from the same airport), but go back to normal behavior pretty fast. (See, e.g., Gigerenzer (2004).) We explain here how learning in our model may give rise to such recency effects, which have been recognized since the early works of Ebbinghaus (1913).

Theories alone may give rise to a recency effect. One theory, say t_0, posits that plane crashes are *iid* events occurring with probability $\varepsilon > 0$. An agent whose reasoning is guided solely by this theory would learn nothing from observing an adverse outcome, and would make no adjustments in behavior. However, the agent may also entertain other theories involving "regime changes" having to do with airport security, equipment safety, and so on. For example, for every period τ there exists a theory t_τ, which posits that the probability of a crash is the aforementioned (small) $\varepsilon > 0$ up to period τ,

but takes the higher value ρ in period τ and thereafter—it may be that airport security has become lax, or that the airline's fleet is growing older, and so on. The likelihood principle would increase the value assigned to t_τ as compared to t_0 after an adverse outcome in period τ, as the former assigns a higher probability to the crash than did the latter, prompting the agent to shun air travel. The agent will become more amenable to air travel as subsequent periods τ' pass without incident, decreasing the weight placed on theory τ'.

The presence of cases can also give rise to recency effect. When an new airline accident joins the history, theories such as t_0 suffer an adverse likelihood shock, whereas case-based reasoning does not. The value $\sigma(\{q_n\}, q_{n+1})$ (measuring the similarity of the current problem to the crash case) may be sufficiently high as to generate a powerful recency effect. For instance, let $H = ((q_i, a, r))_{i \le n}$ be the history of cases in which act a was chosen and outcome r resulted, and consider a simple additive specification:

$$\sigma(\{q \in P | (q, a, r) \in H\}, q_{n+1}) = \sum_{i \le n} e^{-(n+1-i)}. \tag{6}$$

To interpret this specification, we may think of each past problem as inherently equally similar to the current one, but with similarity decreasing with time. An agent may forget past problems. Alternatively, an agent may have perfect memory, and yet decide that, if two cases are equally similar to the current one, it makes more sense to rely on a more recent case in making predictions. This can be justified by the assumption that data generating process might change (as in "regime change" models), and a more recent past case is less likely to belong to a different (pre-change) data generating process.

The combination of theories and cases can amplify recency effects. If the agent were solely case-based in her reasoning, any function that weighs more distant cases less would explain some aspects of the recency effect. But in our model the effect of the single event, as well as of the events following it, would be magnified because of the existence of theories: when the plane crash occurs it is not only the most recent case; it also represents a surprise that shakes the agent's belief in the theories she used to employ for prediction. As a mirror image, when normal, non-crash events start accumulating, the effect is not only the exponential decay of the weight assigned to the crash; rather, there is a rise in the relative weights assigned to (at least some) theories as these start looking better despite their failure to predict the crash.

5. Reasoning in games

This section illustrates how our model might be used to study players' reasoning in games. Specifically, reasoning by cases and by theories is reminiscent of Selten's (1978, Section 5) discussion of three levels of decision making—routine, imagination, and reasoning. The routine level can be captured by case-based reasoning, where no strategic sophistication is assumed: the player expects the others to keep playing as they used to. Theories in our model can reflect strategic reasoning, and can thus capture some aspects of Selten's "imagination" and "reasoning".

The interaction between different modes of reasoning can be particularly fruitful in rationalizing well-known experimental departures from the game-theoretic predictions based on backward-induction. For instance, different cognitive styles may help explain the abundant evidence on the Centipede game: ordinary players tend to play the game for many stages before defecting, suggesting a predominance of case-based reasoning. On the contrary, more experienced players, such as professional chess players, engage into backward induction more easily—suggesting a dominance of theory-based reasoning.[11] More generally, our model provides insights into the role of experience in the emergence of cooperation and, in turn, into the learning process guiding the play of some popular finitely repeated games.

Consider a finitely repeated Prisoner's Dilemma (PD) in which Player I and Player II play $L > 1$ repetitions of the stage game:

	C	D
C	3, 3	0, 4
D	4, 0	1, 1

.

At stage $l \in \{1, \ldots, L\}$ each player has a memory consisting of $(l-1)$ cases, where case $i < l$ corresponds to stage i in the repeated game, and specifies the stage number, the player's choice and the outcome. Thus, if at $i < l$ the players played, say, (C, C), each would have a case $(i, C, (C, C))$ in her memory, identifying the problem i, her choice C, and the outcome $(C, C) \in R$. The payoff $u(r) \in G(\equiv \mathbb{R})$ is given by the matrix above.[12] We assume that both players share the same similarity function.

A "theory" is a repeated game strategy of the other player. Thus, each theory is sufficiently detailed to allow for the computation of the player's payoff in the repeated game given each repeated game strategy she may choose. Note that in

[11] See, e.g., Palacios-Huerta and Volij (2009).

[12] The nature of the strategic interaction is such that each player's payoff depends on the outcome emerging from the acts chosen by both players. In this context, there is a one-to-one map between outcomes and payoffs, which allows us to define utilities directly over outcomes and leave aside the distinction between outcomes and payoffs. Furthermore, we do not need hypothetical payoffs, as these are required only for parameters' elicitation in our representation, whereas we can focus on actual payoffs in most of applications.

this model the objects of choice in the "routine" and the "reasoning" levels differ: the former, case-based mode of thinking applies to the selection of a move in the stage game (C or D); the latter, theoretical one, selects repeated game strategies (in the subgame that remains to be played). Case-based reasoning cannot apply to entire repeated game strategies in this model, as there is no history of such games. By contrast, strategic reasoning should allow players to think of their entire strategies, not only the stage move they select. If, for example, she can imagine her opponent playing C until the last stage and then defecting, we would like her to be able to do the same thing herself. Hence, theories rank entire strategies, while cases rank only moves in specific stages. Thus, we assume that a player's choice of a move in a stage game is an act that yields the highest payoff, when summing up the case-based payoff for this act in past stages with the expected payoff of the best repeated game strategy that starts with this act.

When L is large it is unlikely to assume that all repeated game strategies are considered by a player. We will restrict attention to strategies that play cooperatively (C) until a certain stage $l \leq L$ and defect (D) thereafter, where the strategy switches to D either as a response to the other player's choice of D or of its own initiative when the period is late enough. In other words, for $l = 1, ..., L+1$, define a strategy s_l that plays D at period $i \geq l$ independently of history, and plays C at period $i < l$ if and only if the other player has played C for all $j < i$. Otherwise, at $i < l$, if the other player has defected at least once before i, s_l plays D. This assumption makes the model less cognitively demanding for the players. It also simplifies our analysis, because it offers a one-to-one correspondence between the strategies and the game stages, so that at each stage the player basically has to decide whether to switch to D assuming neither player has done it so far. In other words, the game resembles a Centipede game, where choosing D for the first time is equivalent to stopping the game.

Note that s_{L+1} corresponds to a "grim trigger" strategy, which never switches to D on its own. Consequently, if it is matched with itself, it will choose a dominated move at period L. Similarly, s_1 is the always-defect strategy, and the only Nash equilibrium of the repeated game is (s_1, s_1). The strategies of each player correspond to theories of levels of reasoning: s_L can be thought of as a variant of Level-0 reasoning: it plays C until the last stage, where C is dominated and D is chosen. The best response to s_L is s_{L-1}, which corresponds to Level-1 reasoning. More generally, s_{L-k} is the best response to s_{L-k+1} (for $k \geq 1$) and s_{L-k} corresponds Level-k reasoning. Let the set of theories therefore be $T = \{s_1, ..., s_L\}$. Importantly, all of these theories but s_1 would be ruled out by iterated domination.

In general, each theory in T specifies a probability distribution over outcomes in R given a problem, a history, and an act. In this model these probabilistic predictions are degenerate, as the theories describe pure strategies of the other player. We let the belief ν on T be an additive probability, and assume that it is identical for both players. For the sake of the example, assume that $\nu(\{s_l\})$ is increasing in l for $l > 1$. That is, we allow the players to have some belief on the equilibrium, backward induction theory s_1, and do not compare this probability to those assigned to other theories. Many people can conceive of the common knowledge argument (corresponding to $k = \infty$ and strategy s_1) without climbing up the belief hierarchy step by step, while lower values of k (strategies s_l, $l > 1$) are typically arrived at by inductive steps. Among these remaining finite levels of reasoning we assume that higher levels of k have lower probabilities. While this may be too simplistic when $k \leq 3$, we find this assumption reasonable for higher levels of k. We point out the following.

Observation 1. *Under the assumptions above, the game will end with* (s_1, s_1) *or* (s_L, s_L).

The logic behind this observation is straightforward: in period 1, the players have empty memories. All theories s_l for $l > 1$ suggest a higher payoff for act C over D, while s_1 does the opposite. If its relative weight, $\nu(s_1)$, is high enough, the players will play the equilibrium (s_1, s_1), that is, will both choose D. At the next stage all the other theories will have been refuted, and the players will be "certain" that they play the equilibrium. Further, with the payoffs being all non-negative, the outcome (D, D) makes the choice D more attractive to each player, and they will indeed, play (D, D) throughout the game.[13] If, however, the relative weight of the equilibrium theory, $\nu(s_1)$, isn't high enough, the players will both play C. In this situation two things will occur: the equilibrium theory s_1 will be refuted; and a case will be added with act C and payoff 3. Both have the same effect of making C even more attractive in the next stage: the only theory that favored D at the present stage is now out of the game, and C has a nicer record. If C was chosen over D in the presence of s_1 and with no record, it will certainly be chosen now. The argument continues in the same way until the last stage.

This analysis is rather simplistic. Yet, we find that it captures some intuitive reasoning. If we find two players cooperating at stage 10 out of 20 repetitions of the prisoners' dilemma, and ask one of them why, she might say, "Look, I don't know exactly what my opponent thinks; I guess we don't have common knowledge of rationality between us, or else we wouldn't be here, but I don't know exactly what are the higher order beliefs, and for now I don't really care. I've been playing C for ten stages and got a nice payoff. Whatever the reason, it may well apply at stage 11 as well." Of course, this would break down at the last stage.

More realistic models can be constructed along similar lines. First, as already mentioned, ν need not be monotone throughout the range $l > 1$. When the game reaches stages $L-3$ or $L-2$, the players might suspect that the backward induction solution is going to be played, and therefore play it. Second, the reasoning of players in stages of the repeated

[13] Notice that if we were to change the neutral payoff to be above 1 this result would no longer hold. Intuitively, if the players look at the outcome (D, D) and think to themselves, "We must be able to do better than that", they might choose C even in the absence of a theory that explains how C would result in better long-run payoffs.

game might rely on (i) past stage games from other repeated games, and (ii) "cases" that describe entire games. Starting with (i), we may assume that each of C and D has been played in other, similar games of different lengths. Some of these might have been a priori unbounded in length, some may still be played at present. Thus, cooperation might look more attractive thanks to its payoffs in other games. As for (ii), players may reason about the entire repeated game strategy, and can say, "Oh, I know from experience that my opponent will cooperate until the last 10% of the periods". Incorporating these types of reasoning is compatible with our general framework, which can therefore be viewed as a theoretical basis to investigate the role of experience in learning how to play games.[14]

6. Discussion

We propose an axiomatic model that gives us a snapshot of an agent who combines analogy-based and theory-based reasonings. We study how the relative weight between these two modes may depend on (i) the current problem faced; and (ii) the agent's "cognitive style" to reason about uncertainty. Learning takes place in the form of addition of new cases to history that may affect the revision of beliefs over theories—in a Bayesian updating fashion—as well as the relative weight between case-based and theory-based reasoning.

The similarity function in our model is best interpreted not as a measure of "pure" similarity of cases, but similarity *given* the theories. As our elicitation procedure measures the similarity values in tandem with the probability weights, we should interpret it as similarity-in-context. Part of the learning procedure that we do not model in this paper is the extent to which "pure similarity" is reduced given the fact that the theories already capture some of the lessons that a given case can provide. More generally, one may wish to have a theory of the way past cases are summarized by new theories, transferring weight from these cases' similarity values to the theories' likelihoods.

We intend our model both to provide a representation of preferences and to characterize how the beliefs behind these preferences are formed and updated. There is still work to be done on beliefs. In our current formulation of the model, the set of theories can only shrink, as cases appear that falsify theories. However, we expect an agent who accumulates sufficient experience with cases to identify regularities that are then incorporated into new theories. We can thus view theories as summaries of coherent collections of past cases. An agent may initially assign high weight to cases and low weight to theories, indicating that she does not trust the theories she has available as sufficiently informative. However, the agent may use past cases as a motivation for modifying existing theories or forming and attaching weight to new theories. In the process, she may diminish or eliminate the weight placed on cases, as she comes to believe that her theories provide an effective explanation of her environment. 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 new theories.

This process would be reflected in our model in an updating of the function ν: new theories t will be generated and added to T. We can then expect the relative importance of past cases to decrease. 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 theory as adequately capturing the information contained in that case. We view a formal development of this process as an important area for further work.

Appendix A. Pairwise comonotonicity

We start with some simple observations regarding the notion of pairwise historically-comonotonic profiles in Definition 1. Observe that pairwise historic-comonotonicity in our model is less restrictive than the standard comonotonicity condition (e.g., Köbberling and Wakker (2003, p. 400)), in that comparisons are only made between two components within a set of problems. The standard characterizations of comonotonicity in terms of monotonicity with respect to a permutation have natural counterparts in our case, as explained below.

Let $a \in A$ and $H_a = \{p_i\}_{i=1}^{n(a)}$. Given a permutation π on H_a, consider the set $C^\pi = \{f \in \mathcal{F}_a \colon f(\pi(p_1)) \gtrsim \ldots \gtrsim f(\pi(p_{n(a)}))\}$. Then, C^π is a maximal pairwise historically-comonotonic set (*pairwise comoncone*). Note that (standard) comonotonic sets are subsets of pairwise historically-comonotonic sets (relative to the appropriate permutations).

For $f \in \mathcal{F}_a$, define the binary relation $\succeq^f_{H_a}$ on H_a as

$$p_i \succeq^f_{H_a} p_j \quad \Leftrightarrow \quad f(p_i) \gtrsim f(p_j).$$

For a set $F \subseteq \mathcal{F}_a$, define the binary relation $\succeq^F_{H_a} = \bigcap_{f \in F} \succeq^f_{H_a}$. The following lemma adapts a basic result in the literature to our notion of pairwise historic-comonotonicity.

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

[14] Among others, Selten and Stoecker (1986) and, more recently, Embrey et al. (2018) provide experimental studies of the impact of experience on cooperative behavior in the finitely repeated Prisoner's Dilemma. See Fudenberg and Levine (2016) for a more general discussion on learning in strategic interactions.

1. *F is a pairwise historically-comonotonic set;*
2. *$\succeq^F_{H_a}$ is a weak order;*
3. *$F \subset C^\pi$ for some permutation π on H_a.*

The proof of the above lemma follows easily by adapting well-known results (see, e.g., Wakker (1989, Lemma 3)).

Appendix B. Proofs

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

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

- *monotone* if $f(s) \succsim g(s)$ for all $s \in H_a \cup T$ implies $f \succsim_{H_a,T} g$;
- *continuous* if, for every $f \in \mathcal{F}_a$, the sets $\{g \in \mathcal{F}_a : f \succsim_{H_a,T} g\}$ and $\{g \in \mathcal{F}_a : g \succsim_{H_a,T} f\}$ 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 $\succsim_{H_a,T}$ on $\mathcal{F}_a$ be a monotone and continuous weak order that satisfies BiCo Tradeoff Consistency. Then, for every $f, g \in \mathcal{F}_a, \alpha, \gamma \in G$, and $s \in H_a \cup T$,*

$$\alpha\{s\}f \succsim_{H_a,T} \alpha\{s\}g \iff \gamma\{s\}f \succsim_{H_a,T} \gamma\{s\}g \tag{7}$$

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

Observe that (7) is a stronger version of the standard Comonotonic Coordinate Independence axiom. In particular, for all $s \in T$, (7) is equivalent to Coordinate Independence imposed on preferences restricted to future profiles.

Proof. 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 the identity permutation on H_a.

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

$$\alpha\{s\}f \sim_{H_a,T} \alpha\{s\}g \iff \gamma\{s\}f \sim_{H_a,T} \gamma\{s\}g. \tag{8}$$

For the strict part of the statement, suppose, by contradiction, that $\alpha\{s\}f \succ_{H_a,T} \alpha\{s\}g$ and $\gamma\{s\}f \prec_{H_a,T} \gamma\{s\}g$. Then, there exists $s' \in H_a \cup T$ 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 $S = \{t_i \in T : f(t_i) \succ g(t_i)\}$ and suppose it is nonempty. Pick some $t_j \in S$ and consider the profile $f(t_j)\{t_j\}g \in \mathcal{F}_a$. Note that it belongs to F because $g \in F$.

If $f \succ_{H_a,T} f(t_j)\{t_j\}g$, replace the original profile g with $g = f(t_j)\{t_j\}g$ and repeat Step 1.1 by taking another element in S.

If $f(t_j)\{t_j\}g \succsim_{H_a,T} f \succ_{H_a,T} g$, we can find $\beta \in G$ such that $\beta\{t_j\}g \sim_{H_a,T} f$ because $\succsim_{H_a,T}$ is continuous. Note that $\beta\{t_j\}g \in F$. In this case, replace the original profile g with $g = \beta\{t_j\}g$ 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 $t_i \in T$ such that $f(t_i) \succ g(t_i)$ 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 $f \succ_{H_a,T} fSg$. Then, it must be that H_a is not null, and $f(p_i) \succ g(p_i)$ for at least one $p_i \in H_a$. Let $Q = \{p_i \in H_a : f(p_i) \succ g(p_i)\}$ and p_j stand for the first rank-ordered index in Q. Consider the profile $f(p_j)\{p_j\}g \in \mathcal{F}_a$ (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 $\delta \in G$ such that $\delta\{p_j\}g \sim_{H_a,T} f$.

Step 2: Denote by $\bar{g} \in \mathcal{F}_a$ the profile constructed from the original profile g in Step 1. Observe that $\alpha\{s\}f \sim_{H_a,T} \alpha\{s\}\bar{g}$, which implies, by (8), that $\gamma\{s\}f \sim_{H_a,T} \gamma\{s\}\bar{g}$. However, note that $\bar{g}(s) \succsim g(s)$ for all $s \in H_a \cup T$. Hence, by monotonicity, $\gamma\{s\}\bar{g} \succsim_{H_a,T} \gamma\{s\}g \succ_{H_a,T} \gamma\{s\}f$, contradiction. ■

Corollary 1. *For every $a \in A$, let the binary relation $\succsim_{H_a,T}$ on $\mathcal{F}_a$ be a monotone and continuous weak order that satisfies BiCo Tradeoff Consistency. Then, for every $x, x' \in G^{H_a}$, and $y, y' \in G^T$,*

$$(x, y) \succsim_{H_a,T} (x, y') \iff (x', y) \succsim_{H_a,T} (x', y'),$$

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

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

For two utility functions, $u, u' : G \to \mathbb{R}$ 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 $U = \{g \in \mathcal{F}_b : g \succsim_{H,T} f\}$ and $V = \{g \in \mathcal{F}_b : f \succsim_{H,T} g\}$. 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 $g \sim_{H,T} f$. □

Step 2. Let $a \in A$. We show that the binary relation $\succsim_{H_a,T}$ satisfies all axioms of Corollary 10 in Köbberling and Wakker (2003). Clearly, $\succsim_{H_a,T}$ 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 $f \sim_{H,T} g$. Then, Restricted Weak Order implies that $\succsim_{H_a,T}$ is complete and, therefore, is a weak order. Moreover, $\succsim_{H_a,T}$ is monotone: Let $f, f' \in \mathcal{F}_a$ be such that $f(s) \succsim f'(s)$ for all $s \in H_a \cup T$. From Step 1 and Monotonicity, it follows that $f \succsim_{H_a,T} f'$. The binary relation $\succsim_{H_a,T}$ is also continuous: Indeed, let $f \in \mathcal{F}_a$ and consider the set $\{f' \in \mathcal{F}_a : f' \succ_{H_a,T} f\}$. Step 1 and Continuity directly imply that this set is open in $\mathcal{F}_a$. Similarly, it can be shown that $\{f' \in \mathcal{F}_a : f \succ_{H_a,T} f'\}$ is open, too. Finally, BiCo Tradeoff Consistency implies that $\succsim_{H_a,T}$ satisfies the Comonotonic Tradeoff Consistency axiom of Köbberling and Wakker (2003).

Hence, $\succsim_{H_a,T}$ 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 $u_a : G \to \mathbb{R}$ and a capacity $\sigma_a : 2^{H_a \cup T} \to [0,1]$ such that $V_a(f) = \int_{H_a \cup T} u_a(f) d\sigma_a$ represents $\succsim_{H_a,T}$. Moreover, for all $b \in A$ such that H_b is null, there exist a continuous function $u : G \to \mathbb{R}$ and a capacity $\nu : 2^T \to [0,1]$ such that $V_T(f) = \int_T u(f) d\nu$ represents $\succsim_T$. □

Step 3. Let $a \in A$ and, for a given $x \in G^{H_a}$, consider the set

$$Z = \left\{ f = (x, y) \in \mathcal{F}_a \mid y(t) \succsim x(p) \text{ for all } p \in H_a \text{ and } t \in T \right\}.$$

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

$$(x', y) \succsim_{H_a,T} (x', y') \quad \Leftrightarrow \quad y \succsim_T y' \tag{9}$$

for all $x' \in G^{H_a}$ such that the set $\left\{(x, y), (x, y'), (x', y), (x', y')\right\}$ is pairwise historically-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 $\frac{\sigma_a(A)}{\sigma_a(T)} = \nu(A)$ for all $A \subseteq T$. Moreover, Pairwise Comonotonic Tradeoff Consistency directly implies that $\succsim_T$ satisfies the standard Tradeoff Consistency axiom on G^T (see, e.g., Definition 2 of Köbberling and Wakker (2003)). Thus, ν is a probability measure, and the pseudo-capacity σ_a restricted to T is additive. Shift u so that $u(\alpha^*) = 0$ for the consequence α^* of the Consequentialism axiom. (Note that the remaining freedom in selecting u is only multiplication by a positive number.) □

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

$$V_a(f) = \int_{H_a} u(f) d\sigma_a + \int_T u(f) d\sigma_a.$$

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

Let $y = (\alpha, D; \beta, T \backslash D)$ and $y' = (\gamma, T)$, where $D \subsetneq T$ and $\alpha, \beta, \gamma \in G$ 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 a set D because, by assumption, there exists at least two non-null theories. Then, $y \sim_T y'$. Now, let $x = (\beta, H_a) \in G^{H_a}$. By (9), we have $(x, y) \sim_{H_a,T} (x, y')$ which, using the representation of $\succsim_{H_a,T}$, is equivalent to

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

Replace x with $x' = (\theta, B; \beta, H_a \backslash B) \in G^{H_a}$, where $B \subseteq H_a$ and $\theta \in G$ such that $\gamma \succ \theta \succ \beta$. Then, by Corollary 1, $(x', y) \sim_{H_a,T} (x', y')$ and, using the representation, we have

$$\begin{aligned} u(\alpha)\sigma_a(D) + u(\theta)\left[\sigma_a(B \cup D) - \sigma_a(D)\right] + u(\beta)\left[\sigma_a(H_a \cup T) - \sigma_a(B \cup D)\right] = \\ u(\gamma)\sigma_a(T) + u(\theta)\left[\sigma_a(B \cup T) - \sigma_a(T)\right] + u(\beta)\left[\sigma_a(H_a \cup T) - \sigma_a(B \cup T)\right]. \end{aligned}$$

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

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

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

$$\sigma_a(B \cup D) - \sigma_a(D) = \sigma_a(B \cup T) - \sigma_a(T) \tag{10}$$

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

It remains to show that $\sigma_a(B \cup D) - \sigma_a(D) = \sigma_a(B)$, which can be proved by following a similar argument. Specifically: let $z = (\gamma, H_a) \in G^{H_a}$. Then, $(z, y) \sim_{H_a,T} (z, y')$ if and only if

$$u(\alpha)\sigma_a(D) + u(\gamma)\left[\sigma_a(H_a \cup D) - \sigma_a(D)\right] + u(\beta)\left[\sigma_a(H_a \cup T) - \sigma_a(H_a \cup D)\right] = u(\gamma)\sigma_a(H_a \cup T) \quad .$$

Now, replace z with $z' = (\zeta, B; \gamma, H_a \backslash B) \in G^{H_a}$, where $B \subseteq H_a$ and $\zeta \in G$ with $\alpha \succ \zeta \succ \gamma$. Then, $(z', y) \sim_{H_a,T} (z', y')$, which is equivalent to

$$\begin{aligned} u(\alpha)\sigma_a(D) + u(\zeta)\left[\sigma_a(B \cup D) - \sigma_a(D)\right] + \\ u(\gamma)\left[\sigma_a(H_a \cup D) - \sigma_a(B \cup D)\right] + u(\beta)\left[\sigma_a(H_a \cup T) - \sigma_a(H_a \cup D)\right] \\ = u(\zeta)\sigma_a(B) + u(\gamma)\left[\sigma_a(H_a \cup T) - \sigma_a(B)\right]. \end{aligned}$$

A similar subtraction yields

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

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

$$\sigma_a(B \cup D) - \sigma_a(D) = \sigma_a(B). \tag{11}$$

By combining conditions (10) and (11), we have $\sigma_a(E \cup F) = \sigma_a(E) + \sigma_a(F)$ for all $E \subseteq H_a$ and $F \subseteq T$. □

Hence, for every $a \in A$, the binary relation $\succsim_{H_a,T}$ is represented by $V_a(f) = \int_{H_a} u(f)d\sigma_a + \int_T u(f)d\sigma_a$. Moreover, by the uniqueness properties of u and σ_a discussed in Step 3, we can apply the normalization $V'_a = \frac{V_a}{\sigma_a(T)}$ and obtain, with little abuse of notation, that

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

represents $\succsim_{H_a,T}$, too. As shown in Step 3, recall that $\nu \in \Delta(T)$.

Step 5. It remains to derive the representation of $\succsim_{H,T}$ 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 $y \in G^T$ such that $f \sim_{H,T} y$. By Consequentialism, $y \sim_{H,T} (\boldsymbol{\alpha}^*, y)$ (where $\boldsymbol{\alpha}^* \in G^{H_a}$), and by transitivity we also get $f \sim_{H_a,T} (\boldsymbol{\alpha}^*, y)$ (that is, $f \sim_{H,T} (\boldsymbol{\alpha}^*, y)$ and both these profiles are in $\mathcal{F}_a$). Using the representation of $\succsim_{H_a,T}$ from Step 4, we have that $f \sim_{H_a,T} (\boldsymbol{\alpha}^*, y)$ if and only if

$$\begin{aligned} &\int_{H_a} u(f)\, d\sigma_a + \int_T u(f)\, d\nu \\ &= \int_{H_a} u(\boldsymbol{\alpha}^*) d\sigma_a + \int_T u(y)\, d\nu \\ &= \int_T u(y)\, d\nu. \end{aligned} \tag{12}$$

Next, consider $a, b \in A$, and a pair of profiles, $f \in \mathcal{F}_a$ and $g \in \mathcal{F}_b$. Choose $y, y' \in G^T$ so that $f \sim_{H,T} y$ and $g \sim_{H,T} y'$. Representation of $\succsim_{H,T}$ by the sum of the integrals over the entire space follows from its representation on G^T and (12). More explicitly,

$$f \succsim_{H,T} g \quad \Leftrightarrow \quad y \succsim_T y' \quad \Leftrightarrow \quad \int_T u(y)\, d\nu \geq \int_T u\left(y'\right) d\nu \quad \Leftrightarrow$$

$$\int_{H_a} u(f)\, d\sigma_a + \int_T u(f)\, d\nu \geq \int_{H_b} u(g)\, d\sigma_b + \int_T u(g)\, d\nu. \quad \square$$

Proof of Proposition 1. Assume first that, for all $a \in A$, H_a is null. Then any representation of $\succsim_{H_a,T}$, $\left(u, (\sigma_a)_{a\in A}, \nu\right)$, satisfies $\sigma_a \equiv 0$ for all $a \in A$. In this case u is unique up to an affine transformation, and ν is unique, as in Observation 6 of Köbberling and Wakker (2003), while the identically-zero pseudo-capacities σ_a are clearly unique.

Next, assume that, for some $a \in A$, H_a isn't null. Assume first that both $\left(u, (\sigma_a)_{a\in A}, \nu\right)$ and $\left(\hat{u}, \left(\hat{\sigma}_a\right)_{a\in A}, \hat{\nu}\right)$ represent $\succsim_{H,T}$ as in Theorem 1. By Observation 9 of Köbberling and Wakker (2003), we have (i) $\hat{\sigma}_a = \sigma_a$, 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 $(\boldsymbol{\alpha^*}, y) \sim_{H,T} y$ where $(\boldsymbol{\alpha^*}, y) \in \mathcal{F}_a$, whose existence is guaranteed by Consequentialism. As $\sigma_a(H_a) = \hat{\sigma}_a(H_a) > 0$, 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 $\left(u, (\sigma_a)_{a\in A}, \nu\right)$ represents $\succsim_{H,T}$ as in Theorem 1, so will any triple $\left(\hat{u}, (\sigma_a)_{a\in A}, \nu\right)$ where $\hat{u} = \lambda u$ for any $\lambda > 0$. ■

Proof of Proposition 2. We only prove that (i) implies (ii), the converse being routine. First, note that condition (5) of Definition 2 implies that $\succsim^1_T$ coincides with $\succsim^2_T$. Thus, by Observation 6(c) of Köbberling and Wakker (2003), we have that $\nu_1 = \nu_2$ and $u^1 = \gamma u^2 + d$ for some $\gamma > 0$ and $d \in \mathbb{R}$. For the remaining, set $u := u^1 \approx u^2$.

If H_a is null for all $a \in A$, condition (ii) trivially holds. So, assume that H_a is not null for some $a \in A$. For any such $a \in A$, define $\succsim^i_{H_a}$ on G^{H_a}, for $i = 1, 2$, as $x \succsim^i_{H_a} x'$ if and only if $(x, y) \succsim^i_{H_a,T} (x', y)$ for some $y \in G^T$. Using the representation of $\succsim^i_{H_a,T}$, we have that

$$x \succsim^i_{H_a} x' \iff \int_{H_a} u(x) d\sigma^i_a \geq \int_{H_a} u(x') d\sigma^i_a.$$

Hence, $\succsim^i_{H_a}$ is independent of the choice of y and, therefore, is well defined. Moreover, condition (4) implies that

$$x \succsim^1_{H_a} x' \iff x \succsim^2_{H_a} x' \quad \forall x, x' \in G^{H_a}.$$

That is, $\succsim^1_{H_a}$ coincides with $\succsim^2_{H_a}$. Set $\succsim_{H_a} := \succsim^1_{H_a} = \succsim^2_{H_a}$. Clearly, $\succsim_{H_a}$ satisfies all axioms of Corollary 10 in Köbberling and Wakker (2003). Thus, there exists a unique capacity $\sigma_a : 2^{H_a} \to [0, 1]$ that represents $\succsim_{H_a}$. It follows that the pseudo-capacity σ^1_a must be proportional to the pseudo-capacity σ^2_a, i.e., there exists $\lambda_{H_a} > 0$ such that $\sigma^1_a = \lambda_{H_a}\sigma^2_a$.

It remains to show that $\lambda_{H_a} = \lambda_{H_b}$ for all $a, b \in A$. Let $x \in G^{H_a}$, $x' \in G^{H_b}$, and $y \in G^T$. By condition (4), we have

$$(x, y) \succsim^1_{H,T} (x', y) \iff (x, y) \succsim^2_{H,T} (x', y)$$

which, given the representations, is equivalent to

$$\lambda_{H_a} \int_{H_a} u(x) d\sigma^2_a \geq \lambda_{H_b} \int_{H_b} u(x') d\sigma^2_b \iff \int_{H_a} u(x) d\sigma^2_a \geq \int_{H_b} u(x') d\sigma^2_b.$$

Since $u(G)$ is an open interval, the last equivalence can hold only if $\lambda_{H_a} = \lambda_{H_b}$. ■

Proof of Proposition 3. Since $\succsim^1_{H,T}$ is comparable to $\succsim^2_{H,T}$, we have $u := u^1 \approx u^2$, $\nu^1 = \nu^2$, and, for all $a \in A$, $\sigma^1_a = \lambda_H \sigma^2_a$ for some $\lambda_H > 0$.

(i) implies (ii). Choose $\alpha^* \in G$ such that $u(\alpha^*) = 0$. Recall that this can be done because $u(G)$ is an open interval containing 0. Let $\beta, \theta \in G$ such that $\beta \succ \alpha^* \succ \theta$ and $(\boldsymbol{\theta}, \boldsymbol{\beta}) \sim^2_{H,T} (\boldsymbol{\alpha^*}, \boldsymbol{\alpha^*})$. Using the representation of $\succsim^2_{H,T}$, we have

$$u(\beta) = -u(\theta)\sigma^2_a(H_a).$$

Since $\succsim^1_{H,T}$ is more prone to theory-based reasoning than $\succsim^2_{H,T}$, we have $(\boldsymbol{\theta}, \boldsymbol{\beta}) \succsim^1_{H,T} (\boldsymbol{\alpha^*}, \boldsymbol{\alpha^*})$ which is equivalent to

$$u(\beta) \geq -\lambda_H u(\theta)\sigma^2_a(H_a)$$

by the representation of $\succsim^1_{H,T}$. Hence, $-u(\theta)\sigma^2_a(H_a) \geq -\lambda_H u(\theta)\sigma^2_a(H_a)$, implying that $\lambda_H \leq 1$.

(ii) implies (i). Assume that $\lambda_H \leq 1$ and that $(\boldsymbol{\theta}, \boldsymbol{\beta}) \succsim^2_{H,T} (\boldsymbol{\alpha}, \boldsymbol{\alpha})$ for some $\beta \succ \alpha \succ \theta$. By the representation, we have $u(\theta)\sigma^2_a(H_a) + u(\beta) \geq u(\alpha)\sigma^2_a(H_a) + u(\alpha)$. By applying comparability, we obtain

$$\frac{u(\beta) - u(\alpha)}{u(\alpha) - u(\theta)} \geq \sigma^2_a(H_a) \geq \lambda_H \sigma^2_a(H_a),$$

which implies that $(\boldsymbol{\theta}, \boldsymbol{\beta}) \succsim^1_{H,T} (\boldsymbol{\alpha}, \boldsymbol{\alpha})$. ■

References

Bleichrodt, Han, Filko, Martin, Kothiyal, Amit, Wakker, Peter, 2017. Making case-based decision theory directly observable. American Economic Journal: Microeconomics 9 (1), 123–151.

Ebbinghaus, Hermann, 1913. On Memory: A Contribution to Experimental Psychology. Teachers College, New York.

Embrey, Matthew, Fréchette, Guillaume R., Yuksel, Sevgi, 2018. Cooperation in the finitely repeated prisoner's dilemma. The Quarterly Journal of Economics 133 (1), 509–551.

Fudenberg, Drew, Levine, David K., 2016. Whither game theory? Towards a theory of learning in games. The Journal of Economic Perspectives 30, 151–170.

Giacomini, Raffaella, Skreta, Vasiliki, Turén, Javier, 2020. Heterogeneity, inattention and Bayesian updates. American Economic Journal: Macroeconomics 12 (1), 282–309.

Gigerenzer, Gerd, 2004. Dread risk, September 11, and fatal traffic accidents. Psychological Science 15 (4), 286–287.

Gilboa, Itzhak, Samuelson, Larry, 2012. Subjectivity in inductive inference. Theoretical Economics 7 (2), 183–216.

Gilboa, Itzhak, Schmeidler, David, 1995. Case-based decision theory. The Quarterly Journal of Economics 110, 605–640.

Gilboa, Itzhak, Samuelson, Larry, Schmeidler, David, 2013. Dynamics of inductive inference in a unified model. Journal of Economic Theory 148 (4), 1399–1432.

Harsanyi, John C., 1967–1968. Games with incomplete information played by Bayesian players, parts I-III. Management Science 14, 159–182, 320–334, 486–502.

Köbberling, Veronica, Wakker, Peter P., 2003. Preference foundations for nonexpected utility: a generalized and simplified technique. Mathematics of Operations Research 28, 395–423.

Madrigal, Alexis C., 2017. How checkers was solved. The Atlantic. July 19.

Palacios-Huerta, Ignacio, Volij, Oscar, 2009. Field centipedes. American Economic Review 99 (4), 1619–1635.

Reinhart, Carmen M., Rogoff, Kenneth S., 2009. This Time Is Different: Eight Centuries of Financial Folly. Princeton University Press, Princeton.

Savage, Leonard J., 1954. The Foundations of Statistics. Dover Publications, New York. Second edition, 1972.

Savage, Leonard J., 1972. The Foundations of Statistics. Dover Publications, New York. Originally 1954.

Selten, Reinhard, 1978. The chain store paradox. Theory and Decision 9 (2), 127–159. https://doi.org/10.1007/BF00131770.

Selten, Reinhard, Stoecker, Rolf, 1986. End behavior in sequences of finite prisoner's dilemma supergames: a learning theory approach. Journal of Economic Behavior & Organization 7, 47–70.

Simmons, George F., 1983. Introduction to Topology and Modern Analysis. R. E. Krieger Publishing Company, Malabar, Florida.

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