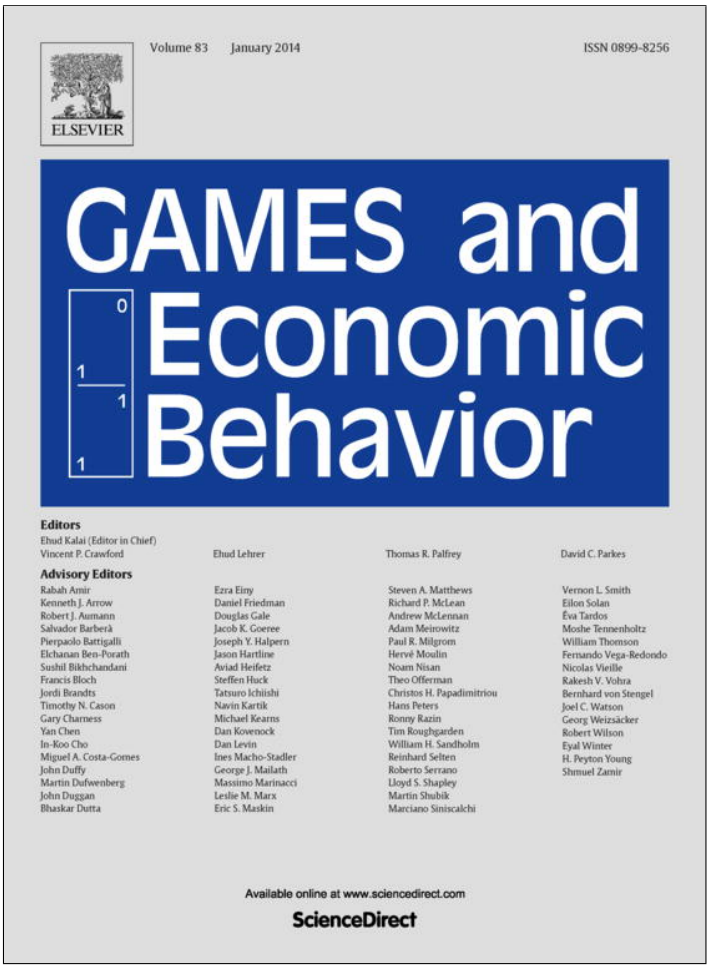

Games and Economic Behavior 83 (2014) 267–283

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Analogies and theories: The role of simplicity and the emergence of norms ☆

Gabrielle Gayer [a,*], Itzhak Gilboa [b,c]

[a] Bar-Ilan University, Israel
[b] HEC, Paris, France
[c] Tel-Aviv University, Israel

ARTICLE INFO

Article history:
Received 27 September 2012
Available online 3 December 2013

JEL classification:
D8
C7

Keywords:
Case-based reasoning
Rule-based reasoning
Model selection
Social norms
Equilibrium selection

ABSTRACT

We consider the dynamics of reasoning by general rules (theories) and by specific cases (analogies). When an agent faces an exogenous process, we show that, under mild conditions, if reality happens to be simple, the agent will converge to adopt a theory and discard analogical thinking. If, however, reality is complex, analogical reasoning is unlikely to disappear. By contrast, when the agent is a player in a large population coordination game, and the process is generated by all players' predictions, convergence to a theory is much more likely. This may explain how a large population of players selects an equilibrium in such a game, and how social norms emerge. Mixed cases, involving noisy endogenous processes are likely to give rise to complex dynamics of reasoning, switching between theories and analogies.

1. Introduction

Consider the prediction problem of agents who live in an uncertain environment. They might be facing a process that is *exogenous*, that is, independent of the agents' predictions, or *endogenous*, that is, fully determined by these predictions. For example, natural processes such as the weather, earthquakes, or hurricanes are exogenous. On the other hand, social processes such as the adoption of a social norm are largely endogenous, as they are fundamentally determined by the agents' predictions thereof. Many other processes are combinations of exogenous and endogenous processes. These include, for example, prices in markets such as real estate, commodities and stock exchanges, which respond both to exogenous news and to speculative trade.

How do agents reason about such processes? Do they think about exogenous and endogenous processes in the same way? This paper attempts to address these questions in a formal way. We consider a dynamic model in which, at each period t, an agent tries to predict the value of a variable y_t, based on a set of observable variables, x_t, as well as the history of both x and y (that is, $(x_i, y_i)_{i<t}$). One common mode of reasoning is regression analysis, whereby y_t might be regressed on x_t, on its own past values, $(y_i)_{i<t}$, or some combination of these. This process belongs to a general category known as *rule-based* learning that involves a selection of theories based on observations. In philosophy, this mode of reasoning is referred to as (case-to-rule) induction, and it is based on the belief that a rule that has been valid in the past will remain valid in the future. Hume (1748) famously pointed out that this belief requires justification, thereby stating the

☆ We are grateful to two referees and an associate editor for comments and references. Gilboa gratefully acknowledges ISF Grant 396/10 and ERC Grant 269754.

* Corresponding author.
E-mail addresses: gabi.gayer@gmail.com (G. Gayer), tzachigilboa@gmail.com (I. Gilboa).

0899-8256/$ – see front matter
http://dx.doi.org/10.1016/j.geb.2013.11.003

problem of induction. Wittgenstein (1922) suggested that the process of induction consists in finding the simplest theory that conforms to the observations, while Goodman (1955) claimed that the notion of simplicity is language-dependent.[1] The basic mechanism of using unrefuted theories for prediction has remained a fundamental method of inference in science, statistics, and everyday life.

Another, perhaps simpler mode of reasoning involves analogical thinking. In its simplest manifestation, when the variable x_t is ignored, y_t is predicted to be the most frequently encountered value in the past.[2] If, however, different periods $i < t$ are characterized by different values of x_i, one may wish to rely more heavily on more similar periods.[3] In artificial intelligence, this mode of reasoning is referred to as *case-based* (see Schank, 1986; Riesbeck and Schank, 1989), and it has been axiomatized in Gilboa and Schmeidler (2001, 2003). Slade (1991) and Kolodner (1992) pointed out some advantages of case-based systems over rule-based systems.[4]

Rule-based reasoning requires agents to speculate about the data generating process that governs the phenomenon of interest. It is akin to problems of model selection, and to the classical problem of statistical inference: the data generating process is assumed to be in a certain class (i.e., a set of rules), and the problem is to find which process, within this class, is the actual one. Case-based reasoning, by contrast, is closer to statistical techniques such as kernel estimation (Akaike, 1954; Parzen, 1962, and see also Silverman, 1986), which keep on using the entire database for predictions, and do not attempt to summarize the database by a concise rule. Thus, rule-based reasoning is more ambitious than case-based reasoning. Correspondingly, rule-based reasoning requires more a priori knowledge about the structure of the problem (say, the class of possible data generating processes) than does case-based reasoning. Indeed, the former is closer to parametric statistical methods, whereas the latter – to non-parametric ones. In a sense, rule-based reasoning is future-oriented, whereas case-based reasoning – past-oriented: engaging in rule-based reasoning, one attempts to come up with general theories that can describe the entire process in the future (and as a by-product, to make predictions for the next period). By contrast, case-based reasoning only tries to predict one period ahead using past observations (where the accumulation of these predictions generates a prediction for the entire process).

It appears that both case-based reasoning and rule-based reasoning are common in everyday life, as well as in formal statistical analysis. In the artificial intelligence literature there are attempts to combine the two modes of reasoning in order to exploit their respective advantages (see for example Rissland and Skalak, 1989, and Domingosu, 1996). However, we are unaware of theoretical work that analyzes such combinations, especially as models of human reasoning, dealing with questions such as, when do agents tend to use analogies, and when – theories? Do they converge to one such mode of reasoning in the long run, and if so, which? Or, under which conditions will case-based reasoning be asymptotically dominant, and under which conditions will long-run behavior be governed by rule-based reasoning? Specifically, are there differences between exogenous and endogenous processes in this respect?

We start with an adaptation of the model of Gilboa, Samuelson, and Schmeidler (GSS, 2013), which provides a unified framework for case-based, rule-based, and Bayesian reasoning. In this model, weights are assigned to "conjectures", which are subsets of states of the world. A prediction is made by comparing the total weights of conjectures that support each possible outcome. Learning in this model is done by elimination of refuted conjectures. In the simplest case, when each conjecture is a single state of the world, the model turns out to be equivalent to Bayesian reasoning, where the elimination of refuted conjectures is the counterpart of Bayesian updating. But when conjectures consist of more than one state of the world, other modes of reasoning may be captured by the model. For instance, a conjecture might reflect the belief that "period t will have the same outcome as period $i < t$", in which case the aggregation over such conjectures captures case-based reasoning.

The focus of GSS (2013) is the robustness of Bayesian reasoning versus other modes of reasoning, where its main message is conveyed by comparing the Bayesian and the case-based conjectures. The paper provides examples of rules, for which it is difficult to provide a universal definition. For instance, rules can have different domains of applicability, allowing them, in particular, to vary in their starting periods. Thus, a rule in GSS (2013) might be "Starting at time $t = t_0$, y_t will equal y_0". In this paper, by contrast, we limit attention to particular rules that can be thought of as general theories. All theories are required to share the same domain of applicability, and, in particular, to make predictions at each and every period t, of y_t given x_t and the history that preceded it. We assume that each theory can be explicitly contemplated, described, and referred to by the agents. It is therefore natural to assume that the set of theories is countable, as only countably many theories can have a concrete, finite description.[5] We also assume that the set of theories is rich enough so that for each history it contains a theory that can describe it. An example of such a set of theories is the functions that are *computable*, that is, those that can be described by Turing machines (or computer programs) that, for each history, compute a prediction in finite time. However, the set of computable theories is but an example, and computability plays no role in the formal model. These theories are contrasted with case-based reasoning by examining the long-run behavior of the relative weights of these two modes of reasoning.

[1] Solomonoff (1964) showed that, in an appropriate model, the dependence of simplicity judgments on language can be bounded.
[2] Or, in a continuous model, the average value of past observations.
[3] As suggested by Hume (1748, Section IV), "From causes which appear similar we expect similar effects."
[4] Their definition of rule-based systems is, however, different from the definition we use in this paper.
[5] See, however, Section 5.1, discussing the relaxation of this assumption.

We consider deterministic conjectures, which, given each possible history, divide outcomes dichotomously into "possible" and "impossible". Correspondingly, given a history, the conjectures are dichotomously classified as "refuted" or "unrefuted" by the given history. Probabilistic predictions can be captured by aggregation over deterministic ones.[6] However, our preferred interpretation of the model is that the agents attempt to foresee long-run trends that are beyond short-term noise variables. Thus, the variables discussed – x_t, y_t – are better viewed as ranges of values for averages taken over relatively long periods, rather than instantaneous values. Correspondingly, a "theory" would be expected to predict whether the average price of oil will go up or down over the next year, but not necessarily tomorrow.

The analysis of the relative weight of rule-based vs. case-based reasoning turns out to critically depend on the process that generates the variable y_t. When the process is *exogenous*, namely when y_t is completely independent of the agent's reasoning process, we show that, under mild assumptions, rule-based reasoning will prevail if one of the theories that the agents conceive of, and assign positive weight to, happens to be correct. Moreover, rule-based reasoning may be dominant even if no single theory is correct, and the agent indefinitely replaces one general theory by another. However, the states of the world in which case-based reasoning asymptotically vanishes is small. Using topological and measure-theoretic notions, we show that, in general, one should expect case-based reasoning to remain prominent, alongside rule-based reasoning. For example, predicting the change of seasons is purely rule-based, because it is a simple enough process to be described by one of the theories that agents can imagine ex ante. By contrast, the occurrence of tropical storms does not seem to follow any simple pattern, and their prediction is, at least partly, based on similarity to past cases.

Next, consider a process that is perfectly *endogenous*, as in the case where y_t is the mode of the predictions of the agents. In contrast to the exogenous case, several results indicate that in this situation rule-based reasoning would be more often encountered, relative to case-based reasoning. Specifically, in the exogenous case we prove that, in a well-defined sense, "most" states of the world are such that case-based reasoning becomes asymptotically dominant, and rule-based reasoning vanishes. In the endogenous case, by contrast, under mild richness assumptions this phenomenon does not occur at *any* state of the world. The reason is, roughly, that case-based reasoning can be described by a theory. Assuming that a majority of the agents follow (a particular) case-based reasoning, they will soon figure out that the theory that makes this very prediction does indeed hold, and they will therefore use it for prediction. More generally, rule-based reasoning is likely to be dominant in the long run, because the agents' shared prediction agrees with a certain theory that becomes the theory of choice for their predictions. Thus, when we consider equilibrium selection in a game among many agents, it is more likely to find the agents converging to simple rules than in the case where these agents predict, say, the weather. This convergence to rules may explain the emergence of social norms as the selection of equilibria in coordination games.

As mentioned above, there are many economic phenomena involving intermediate cases, where the process y_t is determined partly by agents' predictions, and partly by exogenous factors. Speculative trade is one such example. In these cases, due to the external "noise" factors, no single theory can remain valid in the long run (unless the noise factors diminish over time). Nevertheless, when the noise factors are relatively weak, it may take a very long time for the process to converge, and in the meantime the agents' reasoning will fluctuate between rule-based and case-based reasoning. In particular, the agents' reasoning may select theories that become the equilibrium prediction for a certain period, until they are refuted, and then replaced by new theories, or by periods of case-based reasoning.

The rest of the paper is organized as follows. Section 2 describes the basic framework. It uses the framework of GSS (2013) and defines rule-based and case-based reasoning. Section 3 deals with a purely exogenous process, showing that rule-based reasoning is likely to emerge in simple states of the world, but not in all states. Section 4 then deals with a purely endogenous process, showing that rule-based reasoning is more likely to emerge as the asymptotic mode of reasoning than case-based reasoning. Section 5 concludes with comments on some variations of these models. It discusses the robustness of the results with respect to various assumptions and highlights the main conclusions of the analysis.

2. Framework

2.1. The unified model

We adapt the unified model of induction of GSS (2013). An agent makes predictions about the value of a variable y based on some observations x. She has a history of observations of past x and y values to rely on. We make no assumptions about independence or conditional independence of the variables across periods, or any other assumption about the data generating process.

Let the set of periods be $\mathbb{T} \equiv \{0, 1, 2, \ldots, t, \ldots\}$. At each period $t \in \mathbb{T}$ there is a *characteristic* $x_t \in X$ and an *outcome* $y_t \in Y$. The sets X and Y are finite and non-empty.[7] We exclude the trivial case where $|Y| = 1$. The set of all *states of the world* is

$$\Omega = \{\omega : \mathbb{T} \to X \times Y\}.$$

[6] See the discussion in Section 5.2.

[7] Clearly, x_t may be a vector of characteristics, or explanatory variables, each of which assumes only finitely many values.

For a state ω and a period t, let $\omega(t) = (\omega_X(t), \omega_Y(t))$ denote the element of $X \times Y$ appearing in period t given state ω. Let

$$h_t(\omega) = \big(\omega(0), \ldots, \omega(t-1), \omega_X(t)\big)$$

denote the history of characteristics and outcomes in periods 0 through $t-1$, along with the period-t characteristic, given state ω. Let H_t denote all possible histories at period t, i.e., $H_t = \{h_t(\omega) \mid \omega \in \Omega\}$ and let H denote all histories, i.e., $H \equiv \bigcup_{t \geqslant 0} H_t$. We let (h_t, y) denote the concatenation of the history h_t with the outcome y.

For a history $h_t \in H_t$, define

$$[h_t] = \big\{\omega \in \Omega \;\big|\; \big(\omega(0), \ldots, \omega(t-1), \omega_X(t)\big) = h_t\big\}.$$

Thus, $[h_t]$ is the event consisting of all states that are compatible with history h_t. Similarly, for $h_t \in H_t$ and a subset of outcomes $Y' \subset Y$, we define the event

$$\big[h_t, Y'\big] = \big\{\omega \in [h_t] \;\big|\; \omega_Y(t) \subset Y'\big\},$$

consisting of all states that are compatible with history h_t and with the next outcome being in the set Y'.

We endow Ω with the topology generated by the sets $\{[h_t] \mid h_t \in H_t\}$. We also denote by Σ the σ-algebra generated by these sets,[8] and define λ to be the uniform measure on Σ, that is, the unique probability measure according to which

$$\lambda\big([h_t]\big) = \lambda\big(\big[h'_t\big]\big) \quad \forall t,\ \forall h_t, h'_t \in H_t.$$

In each period $t \in \mathbb{T}$, the agent observes a history, h_t, and makes a prediction about the period-t outcome, $\omega_Y(t) \in Y$. A *prediction* is a ranking of subsets in Y given h_t.

Predictions are made with the help of conjectures. A *conjecture* is an event $A \subset \Omega$. It can represent a theory, an analogy, or in general any reasoning aid one may employ in predicting y_t. Indeed, any such reasoning tool can be described extensively, by the set of states that are compatible with it. However, not every subset of Ω may be considered by the agent. Rather, we assume that the agent only conceives of a *countable* subset $\mathcal{A}$ of 2^{Ω}, referred to as the set of *conjectures*. We think of a conjecture as an idea that the agent can think of in a concrete way, reason about, describe in words, and perhaps relate to others. With this interpretation in mind, it appears natural to assume that there are only countably many conjectures.[9] As a result, summation over sets of conjectures will be well-defined.[10]

GSS (2013) show that the notion of conjectures is general enough to capture Bayesian, rule-based, as well as case-based reasoning. Specifically, they assume that the agent has a *model*, which is a function $\phi : \mathcal{A} \to \mathbb{R}_+$, where $\phi(A)$ is interpreted as the weight attached to conjecture A for the purpose of prediction. For a subset of conjectures $\mathcal{D} \subset \mathcal{A}$, ϕ is defined additively, that is,

$$\phi(\mathcal{D}) = \sum_{A \in \mathcal{D}} \phi(A).$$

We assume that the total weight of all conjectures is finite, and, without loss of generality, that $\phi(\mathcal{A}) = 1$.[11]

The agent learns by ruling out conjectures that have been refuted by evidence. Specifically, given a history $h_t \in H_t$, a conjecture A that is disjoint from $[h_t]$ should not be taken into consideration in future predictions.[12]

Fixing a subset of conjectures $\mathcal{D} \subset \mathcal{A}$, a history $h_t \in H_t$ and a subset of outcomes $Y' \subset Y$, consider the set of conjectures in $\mathcal{D}$ that have not been refuted by h_t and that predict that the outcome will be in Y':

$$\mathcal{D}\big(h_t, Y'\big) = \big\{A \in \mathcal{D} \;\big|\; \varnothing \neq A \cap [h_t] \subset \big[h_t, Y'\big]\big\}.$$

Observe that the conjectures in $\mathcal{D}(h_t, Y')$ are various events, many pairs of which may not be disjoint. This is important to bear in mind in the following definitions, where we sum over the weights assigned to different conjectures. As GSS (2013) point out, a model ϕ is a belief function as defined by Dempster (1967) and Shafer (1976). Further, discarding refuted conjectures is equivalent to the Dempster–Shafer updating rule of belief functions.

Given a model $\phi : \mathcal{A} \to \mathbb{R}_+$, the weight assigned to Y' by the unrefuted conjectures in $\mathcal{D}$ is

$$\phi\big(\mathcal{D}\big(h_t, Y'\big)\big).$$

[8] Observe that, because Y is finite, this σ-algebra coincides with that generated by the sets $[h_t, \{y\}]$ for history h_t and outcome $y \in Y$.

[9] For example, if conjectures are assumed to be computable (by finite algorithms), the set of conjectures is countable.

[10] See also Section 5.1 for a discussion of relaxation of the countability assumption.

[11] In GSS (2013), the set of conjectures is uncountable, and ϕ is defined as a measure over sets of conjectures, that is, sets of subsets of states of the world. This complication is obviated thanks to the countability assumption.

[12] One may allow theories to have a few wrong predictions before being considered "refuted". For example, for $k > 1$, we may consider a theory to consist of all the states along which the theory's prediction with the potential exception of k periods. It will be clear from the following analysis that such a modification of the notion of "refutation" does not alter our results in any fundamental way.

The total weight assigned to a subset $Y' \subset Y$ by all unrefuted conjectures is thus given by

$$\phi\big(\mathcal{A}(h_t, Y')\big).$$

The agent's prediction is a ranking of the subsets of Y, with Y' considered more likely than Y'' iff

$$\phi\big(\mathcal{A}(h_t, Y')\big) > \phi\big(\mathcal{A}(h_t, Y'')\big).$$

It will be useful to have notation for the set of conjectures, in a class $\mathcal{D} \subset \mathcal{A}$, that are relevant for prediction at history h_t:

$$\mathcal{D}(h_t) = \bigcup_{Y' \subsetneq Y} \mathcal{D}\big(h_t, Y'\big).$$

Observe that $\mathcal{D}(h_t)$ is the set of conjectures in $\mathcal{D}$ that have not been refuted and that could lend their weight to *some* nontautological prediction after history h_t (and hence $\mathcal{D}(h_t) \subset \mathcal{D}(h_t, Y)$).

2.2. Rule-based reasoning: theories

The notion of a rule is rather general. There are *association rules*, which, conditional on the value of x_t, restrict the possible values of y_t. For example, the rule "if the Democratic candidate wins the election, taxes will rise" says something about the rate of taxation, y_t, if the president is a Democrat (i.e., if x_t assumes a certain value). Such a rule does not restrict prediction if its antecedent does not hold. By contrast, there are *functional rules*, which predict that y_t be equal to $f(x_t)$ for a certain function f, and thus provide non-trivial predictions for every value of x_t. Other rules may be time-dependent, and allow y_t to be a function of x_t as well as of t itself. Further, rules may differ in their domain. In particular, GSS (2013) provide an example of rule-based reasoning in which the rules predict a certain constant y value beginning with a given period t, and making no predictions prior to that t.

In this paper we restrict attention to rules that can be viewed as general theories. Such theories are constrained to make a specific prediction (i.e., a single y_t) *at each and every* t, and for any possible value of x_t. It is important to highlight that our results depend on this assumption. In particular, our results do not apply to more general definitions of "theories", whereby a theory may be restricted in its domain, say, by making a prediction at only a finite number of periods, starting at a particular t, or limiting its bite to particular values of x_t. Similarly, our notion of a theory does not include paradigms, which can be viewed as general frameworks for generating specific theories. For the purposes of this paper, a "theory" can be tested, and can possibly be refuted, at each history h_t. Observe however that the definition is general in the sense that the functions are allowed to depend on the entire history h_t, and thus on previous values (x_i, y_i) for $i < t$.

Let $\mathcal{R} \subset Y^H$ be the set of functions considered by the agent. For a model ϕ and a theory $f \in \mathcal{R}$, we will use $\phi(f)$ to denote the weight assigned by ϕ to the conjecture consisting of all the states that do not contradict f, that is, $\phi(f) = \phi([f])$ where

$$[f] = \big\{\omega \in \Omega \;\big|\; \omega_Y(t) = f\big(h_t(\omega)\big)\ \forall t\big\}.$$

We assume that the agent satisfies the following two conditions:

(i) For every history h_t, there exists $f \in \mathcal{R}$ such that $h_t \cap [f] \neq \varnothing$;
(ii) $\phi(f) > 0$ for every $f \in \mathcal{R}$.

The first condition states that no history will ever find the agent at a loss for theories. Whatever is the history observed, h_t, the agent will be able to conceive of a theory that would have predicted precisely the observed realizations of y_i for each $i < t$. The second condition further insists that the agent gives some positive weight to each such theory, that is, that she doesn't conceive of some theories but then arbitrarily decides to rule them out. Recall that we assume that the set of all conjectures, $\mathcal{A}$, is countable, and hence so is $\mathcal{R} \subset \mathcal{A}$. Indeed, condition (ii) directly implies that $\mathcal{R}$ is countable, as the weight of all conjectures, $\phi(\mathcal{A})$, is assumed to be finite (and normalized to 1).

We find these conditions rather intuitive. If $\mathcal{R}$ is assumed to be countable, condition (ii) may be viewed as a notational convention: $\mathcal{R}$ is defined, without loss of generality, as the support of ϕ. Ignoring theories in $\mathcal{R}$ with zero ϕ weight can be viewed as a failure to distinguish between theories that the agent has conceived of and dismissed and theories that the agent has not been aware of to begin with. Because the analysis will not make this distinction, condition (ii) is basically equivalent to the assumption that the set of theories that the agent conceives of is countable. Condition (i) states that, for every history, the agent can conceive of, and assign a positive weight to at least one theory that is consistent with this history.

A natural class of theories that satisfies these conditions is the set of computable theories. A theory $f : H \to Y$ is *computable* if there exists a Turing machine (or, equivalently, a computer program in a higher-level language such as PASCAL, C++, etc.), which, for every $h_t \in H$, halts in finite time and computes $y_t = f(h_t) \in Y$. Each computable theory can thus be described by an algorithm, that is, a finite set of instructions that define the value of $f(h_t)$ for every h_t. Since each such theory can be defined by a finite description, it seems reasonable that the agent would have to conceive of it. At the

same time the agent may not be able to think about other, non-computable theories, as they cannot be described in a well-defined way. Since there are only countably many computable functions, the set of computable functions appears as a natural candidate for the set of theories $\mathcal{R}$ satisfying conditions (i) and (ii) above. However, nothing in the ensuing analysis depends on the assumption of computability.

Observe that the definition assumes that a theory $f \in \mathcal{R}$ assigns a prediction for every history h_t, including histories that are inconsistent with f itself. This is reminiscent of the definition of a strategy in extensive form games. Alternatively, one may restrict the domain of a theory f only to the histories that do not contradict it.[13]

If there are no x values to be observed (that is, $|X| = 1$), then for every $f \in \mathcal{R}$, there exists a unique state of the world compatible with it. In this case, a model ϕ that puts positive weight only on theories in $\mathcal{R}$ can also be viewed as a Bayesian model (as defined in GSS, 2013), namely as a model assigning probabilities to single states.[14] However, in the more general case, a theory $f \in \mathcal{R}$ is compatible with a non-singleton conjecture, because such a theory, as opposed to a Bayesian conjecture, need not predict the values of the x_t's. Thus, a theory f can predict the outcomes of a Hurricane, should one occur, but it need not commit to a prediction about its occurrence.

A model ϕ_R is (a priori) *purely rule-based* if $\phi_R(\mathcal{R}) = 1$, equivalently, $\phi_R(\mathcal{A}\backslash\mathcal{R}) = 0$ or $\sum_{f\in\mathcal{R}} \phi_R(f) = 1$. Such a model can also be viewed as a probability distribution over $\mathcal{R}$.

2.3. Case-based reasoning: analogies

Case-based conjectures are defined as in GSS (2013): for every $i < t$, $x, z \in X$, let

$$A_{i,t,x,z} = \left\{\omega \in \Omega \;\middle|\; \omega_X(i) = x, \omega_X(t) = z, \omega_Y(i) = \omega_Y(t)\right\}.$$

We can interpret this conjecture as indicating that, *if* the input data in period i are given by x and in period t by z, *then* periods i and t will produce the same outcome (value of y). Notice that a single case-based conjecture consists of many states: $A_{i,t,x,z}$ does not restrict the values of $\omega_X(k)$ or $\omega_Y(k)$ for $k \neq i, t$.

Let the set of all conjectures of this type be denoted by

$$\mathcal{CB} = \{A_{i,t,x,z} \mid i < t, x, z \in X\} \subset \mathcal{A}. \tag{1}$$

A model ϕ_{CB} is a priori *purely case-based* if all weight is put on the case-based conjectures. Note that the number of case-based conjectures is countable, and it is thus possible to assign a positive weight to each and every one of them.

While there is no restriction on how the weights should be divided among the different conjectures in $\mathcal{CB}$, it seems more natural that the agent assigns higher weights to cases that are more similar to each other than to cases with no resemblance. For example, the agent might have a similarity function over the characteristics,

$$s : X \times X \to \mathbb{R}_+,$$

and a memory decay factor $\beta \leqslant 1$. Given history $h_t = h_t(\omega) \in H_t$, a possible outcome $y \in Y$ is assigned a weight proportional to

$$S(h_t, y) = \sum_{i=0}^{t-1} \beta^{t-i} s\big(\omega_X(i), \omega_X(t)\big)\mathbf{1}_{\{\omega_Y(i)=y\}},$$

where $\mathbf{1}$ is the indicator function of the subscripted event. Hence, the agent may be described as if she considered past cases in the history h_t, chose all those that resulted in some period i with the outcome y, and considered the aggregate similarity of the respective characteristic $\omega_X(i)$ to the current characteristic $\omega_X(t)$. The resulting sums $S(h_t, y)$ can then be used to rank the possible outcomes y. If $\beta = 1$ and in addition the similarity function is constant, the resulting number $S(h_t, y)$ is proportional to the relative empirical frequency of y's in the history h_t.

As noted by GSS (2013), for every similarity function s and decay factor β one may define a model $\phi_{s,\beta}$ by setting $\phi_{s,\beta}(A_{i,t,x,z})$, for each t, to be proportional to $\beta^{(t-i)}s(x, z)$, and $\phi_{s,\beta}(\mathcal{A}\backslash\mathcal{CB}) = 0$.[15] In this case, for every history h_t and every $y \in Y$, $\phi_{s,\beta}(\mathcal{A}(h_t, \{y\}))$ is proportional to $S(h_t, y)$. Such a model $\phi_{s,\beta}$ will be equivalent to case-based prediction according to the function S.

It is worthy of note that a particular case-based conjecture is not intended to capture an entire world view. A conjecture $A_{i,t,x,z}$ only says that, conditional on $x_i = x$ and $x_t = z$, the observed values of y in these periods will be the same ($y_t = y_i$). This can hardly be viewed as a theory about the world, and as a result case-based conjectures are not meant to be tested as

[13] Such a restriction would not make a major difference, because the definition of a theory at histories incompatible with it will be immaterial for our purposes.

[14] The resulting Bayesian prior, however, is restricted to have a countable support.

[15] Observe that only the conjectures $\{A_{i,t,x,z}\}_{i<t,x,z\in X}$ are used for prediction at time t. This implies that, should the weights of these conjectures be multiplied by a positive constant (for a given t), prediction would be unaffected. Hence, embedding a similarity function s in a model ϕ one has a degree of freedom for each period t.

are rule-based ones. In fact, when the mechanism of "refutation" of conjectures in our framework is applied to case-based conjectures, it is used to *select* the relevant conjectures, focusing on the conjectures $A_{i,t,x,z}$ for which $x_i = x$ and $x_t = z$. When the reasoner knows whether such a conjecture predicted the correct y_t (i.e., whether $y_t = y_i$), the conjecture is anyway irrelevant, as it does not constrain $y_{t'}$ for $t' > t$.

A case-base conjecture can be thought of as a building block that yields a meaningful prediction mainly when aggregated with many other building blocks according to the weight function ϕ. This is similar to the status of conjectures in Bayesian reasoning: each conjecture is a single state of the world, $\{\omega\}$, and it should not necessarily be interpreted as the statement that ω will obtain. Rather, when one aggregates over all such states, using the prior as a weighting function, one obtains a meaningful prediction tool. We tend to view rule-based reasoning along similar lines: while a single rule-based conjecture, such as "$y_t = x_t$ for every history h_t", can be viewed as a theory about the way the world functions, we allow the rule-based agent to aggregate over several (or even many) such specific conjectures.

Finally, observe that there may be other modes of case-based reasoning that go beyond the conjectures in $\mathcal{CB}$. For example, an agent seeking patterns in the data may observe that the past three periods are similar to other triples of consecutive periods in the past. To capture these analogies, one would need to define more elaborate conjectures than those in $\mathcal{CB}$. Our results would not change if one expands the set of conjectures to include these more involved analogies.

2.4. Open-mindedness

We restrict attention to rule-based reasoning and case-based reasoning of the types described above. Formally, we assume that the set of conjectures is $\mathcal{A} = \mathcal{R} \cup \mathcal{CB}$. Within this constraint, we wish to guarantee that the agent is open-minded. Thus, we will henceforth assume that the agent assigns a positive weight $\phi(A) > 0$ to each conjecture in $\mathcal{A} = \mathcal{R} \cup \mathcal{CB}$. We denote this set of *open-minded* models by Φ_+.

3. Exogenous process

3.1. Learning rules

We now turn to study the dynamics of rule-based versus case-based reasoning. We first show that, if a particular theory (in $\mathcal{R}$) happens to describe the data generating process, the agent in our model will learn this fact. This result is not too surprising: in many different set-ups one may show that, should and agent conceive of the true model and assign some credence to it, then, in the limit, the agent will converge to believing in the true model. (See the discussion following the proposition below.) Our main interest, however, is not in the agent's belief in the true model relative to other models, but in the mode of reasoning the agent employs. The following result is therefore an important benchmark: it provides conditions under which, at such states of the world, rule-based reasoning will become the dominant mode of reasoning, and case-based reasoning will become negligible. This benchmark is to be contrasted with the analysis that follows.

For each theory $f \in \mathcal{R}$, recall that $[f]$ is the event in which f is never refuted. A state $\omega \in [f]$ is simple in the sense that a given theory, f, is always valid in it. In particular, if the set $\mathcal{R}$ consists of computable theories, for such a state ω the computation of y_t given h_t can be done in finite time, employing a program that is independent of t, justifying the adjective "simple".[16]

We define the *set of simple states* to be

$$\mathbb{S} = \bigcup_{f \in \mathcal{R}} [f].$$

We can now state

Proposition 1. *For every $\phi \in \Phi_+$ and every $\omega \in \mathbb{S}$,*

$$\frac{\phi(\mathcal{CB}(h_t(\omega)))}{\phi(\mathcal{R}(h_t(\omega)))} \to 0 \quad \text{as } t \to \infty.$$

That is, in all simple states, the agent will converge to reason by theories and will gradually discard case-based reasoning.

The logic of this proposition is straightforward: if we consider a simple state ω, where a certain theory f holds, the initial weight assigned to this theory will serve as a lower bound on $\phi(\mathcal{R}(h_t(\omega)))$ for all t, because the theory will never be refuted at ω. By contrast, the total weight of the set of all case-based conjectures that are relevant for prediction at time t converges to zero because it is an element in a convergent series. Intuitively, because at ω theory f is correct, it retains its original weight of credence. By contrast, case-based conjectures concern only pairs of periods, $i < t$, and thus, for each new

[16] Observe that even in a simple state ω, the pattern of x_t's (in ω) may be rather complicated. Theories in $\mathcal{R}$ attach y values to histories h_t, but they are not supposed to predict the x values.

value of t, a new set of case-based conjectures is being considered. It is inevitable that the total weight of this set (which is disjoint from sets considered in previous periods) converge to zero.

Proposition 1 may bring to mind "merging" results as in the literature that started with Blackwell and Dubins (1962). (See, for instance, Kalai and Lehrer, 1993.) Indeed, the positive weight assigned to a theory f, at a state $\omega \in [f]$, is reminiscent of an absolute continuity condition, implying that a state that has a positive probability (according to the real data generating process) also has a positive belief (according to the agent's subjective prior). However, Proposition 1 is different from the merging results: whereas the latter compare the true theory to other theories, Proposition 1 compares the true theory to the case-based conjectures, which do not have a Bayesian counterpart. The driving force behind the merging results is that, if at ω theory f is true, other theories, which are not equivalent to f, will be proven false. By contrast, the driving force behind Proposition 1 is not that the case-based conjectures are found to be false; rather, it is that the case-based conjectures relevant at time t have a total weight that converges to zero – a fact that is known a priori. Relatedly, the Bayesian theories that are compared to the true f in a merging results are the same theories at each period t (and each history h_t), as opposed to the case-based conjectures of time t that are different from those at time $t' \neq t$.

3.2. The insufficiency of rule-based reasoning

Next we turn to study states that are not simple. In such states, by definition, no single theory is correct, and thus rule-based reasoning has to evolve over time: the theories that the agent finds most credible at a given point of time are bound to be refuted in due course, and leave the stage to other theories, which were considered less credible at the outset. But, beyond the relative importance of different theories, the entire mode of rule-based reasoning may change its weight relative to that of case-based reasoning. Will rule-based reasoning be more dominant than case-based reasoning, or rather will the agent discard reasoning by theories and converge to analogical thinking?

The answer will depend on the specification of the model ϕ. In particular, the main question is how the total weight of the case-based conjectures changes over time. Since the sets of such conjectures that are relevant at two distinct periods ($t' \neq t$) are disjoint, it may appear intuitive to assume that their weight is "uniform" over the different periods. However, this is impossible as the total weight of the relevant case-based conjectures at period t must converge to zero (as explained above). The main question is, therefore, how fast should this convergence be?

In this paper we assume that the convergence is not too fast. A rate of decay that approximates the uniform weights, yet guarantees convergence would be polynomial: assume that ϕ satisfies the following conditions: there exist $\gamma < -1$ and $c > 0$, such that, for every t, and every $x, z \in X$,

$$\sum_{i<t} \phi(A_{i,t,x,z}) \geqslant ct^{\gamma}. \tag{2}$$

The set of these models will be denoted by $\Phi_+^p \subset \Phi_+$.

It is important to emphasize that the assumption $\phi \in \Phi_+^p$ is crucial for the analysis that follows. In particular, if one were to replace it by the assumption

$$\sum_{i<t} \phi(A_{i,t,x,z}) \leqslant ca^{-t} \tag{3}$$

for $a > 1$ and $c > 0$, the next two propositions will not hold.

We find assumption (2) more reasonable than (3), since that latter reduces the weight of case-based reasoning in an artificial way. The key point is that as t grows, the number of analogies grows polynomially (as opposed to the number of theories that grows exponentially). To illustrate this point, consider a version of our model with a finite time horizon, T. Define two conjectures to be T-equivalent if they make the same predictions for every h_t with $t \leqslant T$. In this case we would be able to divide the weight of analogies, as well as the weight of theories, uniformly within each set of conjectures, and it would follow immediately that the weight of the case-based conjectures decreases in a polynomial rate, which is in line with assumption (2). By comparison, in this case the weight of theories would decrease exponentially. However, should one prefer an assumption such as (3) to (2), the ensuing analysis would have to be revised.

We now turn to distinguish between states according to their asymptotic behavior. Define

$$\Omega_{R\phi} = \left\{\omega \in \Omega \;\middle|\; \exists T,\ \phi\left(\mathcal{R}\left(h_t(\omega)\right)\right) > \phi\left(\mathcal{CB}\left(h_t(\omega)\right)\right)\ \forall t \geqslant T\right\},$$

$$\Omega_{C\phi} = \left\{\omega \in \Omega \;\middle|\; \exists T, \phi\left(\mathcal{R}\left(h_t(\omega)\right)\right) < \phi\left(\mathcal{CB}\left(h_t(\omega)\right)\right)\ \forall t \geqslant T\right\},$$

$$\Omega_{M\phi} = \left\{ \omega \in \Omega \;\middle|\; \begin{array}{l} \forall T,\ \exists t, t' \geqslant T,\ \text{such that} \\ \phi\left(\mathcal{R}\left(h_t(\omega)\right)\right) \geqslant \phi\left(\mathcal{CB}\left(h_t(\omega)\right)\right) \\ \phi\left(\mathcal{R}\left(h_{t'}(\omega)\right)\right) \leqslant \phi\left(\mathcal{CB}\left(h_{t'}(\omega)\right)\right) \end{array} \right\}.$$

Thus, $\Omega_{R\phi}$ is the set of states where rule-based reasoning becomes more important than case-based reasoning, from some point on. Similarly, $\Omega_{C\phi}$ is the set of states where case-based reasoning becomes, from a certain point on, weightier than rule-based reasoning. Their complement is the set $\Omega_{M\phi}$, where reasoning is bound to remain asymptotically mixed:

in these states there are infinitely many periods where rule-based reasoning is at least as weighty as case-based reasoning, but also infinitely many periods where the opposite is true.

We can now state

Proposition 2. *Let there be given a model $\phi \in \Phi_+^p$. Then $\Omega_{R\phi}$, $\Omega_{C\phi}$, and $\Omega_{M\phi}$ are dense in Ω.*

Thus, every open set of the state space contains both states where rule-based becomes weightier forever, and states where case-based reasoning becomes weightier forever, as well as states where neither is true. In particular, after each history h_t (where $[h_t]$ defines an open set), there are continuations in which the reasoner will be mostly rule-based, or mostly case-based, and there are continuations where she will keep switching between these two modes of reasoning.

While the topological notion of denseness suggests that rule-based reasoning and case-based reasoning will be encountered just as frequently, a measure-theoretic notion indicates that the most common situation – when $\phi \in \Phi_+^p$ – is that case-based reasoning takes over. To be precise, a conclusion which is the opposite of that of Proposition 1 holds almost everywhere:

Proposition 3. *For every $\phi \in \Phi_+^p$,*

$$\lambda\left(\frac{\phi(\mathcal{R}(h_t(\omega)))}{\phi(\mathcal{CB}(h_t(\omega)))} \to_{t\to\infty} 0\right) = 1.$$

To conclude, there are states, dubbed "simple", in each of which a particular theory holds. In these states, case-based reasoning will vanish. However, these are by no means the "majority" of states. Overall, the sets of states where rule-based reasoning is asymptotically weightier; where case-based reasoning is asymptotically weightier; and where neither of the above holds are all dense. In particular, no finite set of observations may determine whether the agent's reasoning will converge to be mostly rule-based, mostly case-based, or neither. Furthermore, a simple count of histories of a given length, as does the uniform measure λ, sustains that in most histories case-based reasoning is weightier than rule-based reasoning asymptotically. Admittedly, the uniform measure may not be the only way to aggregate over states. Yet, in summary of the topological and measure-theoretic results, we find that there is no reason to assume that rule-based reasoning suffices in order to describe the way agents think about exogenous processes.

3.3. Discussion

The driving force behind Proposition 3 is that theories make predictions at each and every history, and therefore each new observation partitions the set of theories that are still unrefuted. Thus, at time t there are exponentially many disjoint sets of theories, and only one of them will contain unrefuted theories once history h_t unfolds. This implies that the set of theories that are relevant for prediction at h_t carries a weight that, when added to the weights of exponentially many competitor sets, sums up to a given constant (the a priori weight of rule-based conjectures). If all histories of length t are to have equal weight, it follows that the weight decreases exponentially fast in t. If not, it is certainly possible that some such sets would retain a relatively high weight, but this cannot be true of most of these sets.

This result is a consequence of the conception of a "theory" as a *general* rule, that is, as a statement that has a universal quantifier ("for all history $h_t \ldots$"). By contrast, analogies do not make universal statements, and thus there can be much fewer distinct analogies than distinct theories. Differently put, the fact that the number of T-equivalence classes of theories grows exponentially in T is inherent to the definition of a theory as making a prediction – and risking refutation – at each history. By contrast, analogies are silent most of the time, and thus the number of possible analogies up to time T can be only polynomially large. Hence the weight of the relevant case-based conjectures need not decrease at an exponential rate.

The results that the set of case-based conjectures does not disappear in many (Proposition 2) or most (Proposition 3) states of the world does not depend on the case-based conjectures being analogical in nature. But rather it depends on the class of case-based conjectures being small (of polynomial growth) and relevant (offering a predictions at every period). Consequently, any other class of conjectures that has these two features, which, provided it obtains a priori credence (satisfying a counterpart of assumption (2)), may also not disappear. Our focus on case-based conjectures is motivated by the fact that analogical reasoning is a simple and basic mode of reasoning which has been extensively studied in psychology and in philosophy. It so happens that when this mode of reasoning is embedded in our model, it generates a set of conjectures that does not grow too fast.

Observe that, under our assumptions, both the set of theories and of analogies are countable. If we follow a state of the world, we may find that the agent uses a sequence of theories and a sequence of analogies for making predictions along the (histories that contain) the state. Clearly, while there are only countably many conjectures of each type, the set of sequences of such conjectures is uncountable. However, the agent is not assumed to be aware of this sequence; at no point in time is she assumed to envision the entire path of her reasoning in the future. Rather, at each t the agent is aware of countably many conjectures – be they rule-based or case-based – and only an outside observer who keeps track of the agent's reasoning would have to consider uncountably many sequences of (sets of) conjectures. Thus, we find the treatment

of rule-based and case-based conjecture symmetric. Importantly, the set of conjectures the agent is presumed to be aware of is countable for each class.

Having said that, it is also worth noting that the basic logic of Proposition 3 and the conclusion that case-based reasoning is, generally speaking, unlikely to vanish do *not* depend on the fact that there are only countably many theories or that theories are deterministic. (See Sections 5.1 and 5.2.) Indeed, the reasoning described above applies to sets of theories that are consistent with a history h_t, whether this set is countable or not.

4. Endogenous process

In this section we consider a process that is governed by the reasoning of a set of agents. For example, consider the behavior of agents involved in a coordination game, where each agent tries to predict the social norm that will govern the behavior of others, and to match that norm in her choice of strategy.

In this section we analyze the case in which all agents share the same weight function $\phi \in \Phi$. This extreme case attempts to capture the intuition that, while people vary in their a priori judgment of theories, these judgments are correlated. Specifically, people tend to prefer simpler theories to more complex ones, and similarity judgments are also correlated across people. For example, people might disagree whether the pattern 011111... is simpler than the pattern 010101..., but practically everyone would agree that 000000... is simpler than 011001.... Along similar lines, people would tend to concur that, other things being equal, a more recent period is more relevant for prediction than is a less recent one. Hence the assumption that all agents share the same ϕ, while extreme, is an acceptable benchmark. For such a function ϕ, define

$$\Omega_\phi = \left\{\omega \in \Omega \;\middle|\; \omega_Y(t) \in \arg\max_{y\in Y} \phi\big(\mathcal{A}\big(h_t, \{y\}\big)\big)\ \forall t \geqslant 0\right\}.$$

Thus, it is assumed that the agents' predictions determine the actual outcome. As in the case of the exogenous process, the agents are not assumed to predict the values of x_t, nor to affect them.

We first note that every state of the world may unfold in an endogenous process:

Proposition 4. *For every $\omega \in \Omega$, there exists $\phi \in \Phi_+^p$ such that $\omega \in \Omega_\phi$.*

The proof of Proposition 4 is constructive: given a state $\omega \in \Omega$ the proof describes an algorithm that generates $\phi \in \Phi_+^p$ such that $\omega \in \Omega_\phi$.

Our interest is in the dynamics of reasoning of the agents along states in Ω_ϕ for $\phi \in \Phi_+^p$. To this end, we introduce the following definitions. *Rule-based reasoning is dominant* at state $\omega \in \Omega_\phi$ at period t if

(i) $\phi\big(\mathcal{R}\big(h_t(\omega)\big)\big) > \phi\big(\mathcal{CB}\big(h_t(\omega)\big)\big)$ and
(ii) $\omega_Y(t) \in \arg\max_{y\in Y} \phi\big(\mathcal{R}\big(h_t, \{y\}\big)\big)$.

Thus, rule-based reasoning is dominant if there is more weight put on rule-based reasoning than on case-based reasoning, and if the prediction of the rule-based reasoning is indeed the prediction that the agents make (and that defines the next observation y_t). Similarly, we say that *case-based reasoning is dominant* at state $\omega \in \Omega_\phi$ at period t if

(i) $\phi\big(\mathcal{R}\big(h_t(\omega)\big)\big) < \phi\big(\mathcal{CB}\big(h_t(\omega)\big)\big)$ and
(ii) $\omega_Y(t) \in \arg\max_{y\in Y} \phi\big(\mathcal{CB}\big(h_t, \{y\}\big)\big)$.

Observe that, at $\omega \in \Omega_\phi$ at period t we may have neither mode of reasoning dominating either if they happen to be equally weighty, that is, if, $\phi(\mathcal{R}(h_t(\omega))) = \phi(\mathcal{CB}(h_t(\omega)))$, or if the weightier mode of reasoning does not correctly predict the outcome. This may happen, for instance, if the weight of these conjectures is split between the different predictions, so as to make the other mode of reasoning pivotal. Define $\Omega_{RB\phi}$ to be the set of states $\omega \in \Omega_\phi$ such that, for some T, rule-based reasoning is dominant at state $\omega \in \Omega_\phi$ at all $t \geqslant T$. Define $\Omega_{CB\phi}$ accordingly to be the states at which case-based reasoning dominates from some period on.

Proposition 5. *For every $\phi \in \Phi_+^p$ we have $\mathbb{S} \cap \Omega_\phi \subset \Omega_{RB\phi}$.*

Thus, for every weight function that satisfies our assumptions, the set of states in which rule-based reasoning is eventually dominant contains all the simple states that may emerge from the process. One might wonder whether in complex states case-based reasoning might be dominant in the long run. It turns out that this possibility is precluded if the set of theories satisfies a mild richness condition, namely, that one of the theories describes case-based reasoning. Formally, we say that $\phi \in \Phi_+^p$ is *theoretically closed* if the following holds: for every h_t there exists a theory $f \in \mathcal{R}$ (and thus $\phi(f) > 0$) such that $[h_t] \subset [f]$ and, for every $t' \geqslant t$, and every continuation $h_{t'}$ of h_t ($[h_{t'}] \subset [h_t]$), $f(h_{t'}) \in \arg\max_{y\in Y} \phi(\mathcal{CB}(h_t, \{y\}))$.

In essence, the condition says that if an external observer can predict the result of case-based reasoning according to ϕ, then the agents involved should also be able to conceive of the theory the external observer formulated. The condition is slightly stronger, in that it requires that it is possible for such a theory to begin with any finite history, and proceed according to (ϕ-) case-based reasoning after that history. For example, this condition will be satisfied if the function ϕ itself is computable, and the set $\mathcal{R}$ contains all computable functions.

We can now state:

Observation 1. *Assume that $\phi \in \Phi_+^p$ is theoretically closed and that, for some T, $\arg\max_{y\in Y} \phi(\mathcal{A}(h_t, \{y\}))$ is a singleton for every h_t with $t \geqslant T$. Then we have $\Omega_{CB\phi} = \varnothing$.*

The reasoning behind Observation 1 is very simple (as stated in the introduction): if ω were a state that is, in the long run, governed by case-based reasoning, then, after a certain history h_t, it can be described by the theory, f, that y_t is a maximizer of the case-based part of ϕ. Theory f is therefore identical to ω, apart from histories where case-based reasoning has a non-unique maximizer, and thus the choice of y_t according to ω and according to f need not coincide. However, if case-based reasoning provides a unique prediction from some point on, this prediction can be captured by a theory.

Observation 1 stands in stark contrast to Proposition 3: in the latter, we saw that the set of states in which case-based reasoning becomes overwhelmingly important relative to rule-based reasoning is large: not only is it a dense set, it has measure 1 according to the uniform measure over all histories. By contrast, Observation 1 implies that, under a mild assumption, this set is empty. Thus, exogenous processes are more likely to give rise to case-based reasoning than are endogenous processes.

5. Variants

5.1. Uncountably many theories

As stated above, if a "theory" is supposed to be conceived of, stated, and conveyed by humans, it should be describable in words, and thus there can only be countably many theories to consider. However, one might wonder how our results would change if one were to adopt a more abstract notion of a theory, according to which any function from histories to predictions could be a "theory", resulting in uncountably many of these. Will this additional freedom make rule-based reasoning more powerful? Will it be, for instance, more likely to be observed in exogenous processes?

The answer is negative. The main weakness of rule-based reasoning, exemplified in Lemma 1 and Proposition 3, remains valid: these results rely on the overall weight that rule-based reasoning might muster, and they do not depend on the set of theories being countable, or on reference to specific theories. Rather, the very fact that the overall weight of rule-based conjectures has to be divided among exponentially many histories, implies that in "most" of these histories this weight has to decay exponentially. This argument (used in the proof of Lemma 1) can be repeated for any set of theories, as long as each theory has to make a prediction given each and every history.[17] By contrast, the strength of rule-based reasoning, stated in Proposition 1, does not immediately follow in the uncountable case. For example, if the agent's belief in theories defines a uniform prediction at each history, then at each and every state of the world, including the simple ones, the weight of rule-based reasoning will decay exponentially. In order to guarantee that a result akin to Proposition 1 holds, one would have to add an assumption that, among the uncountably many theories, a countable subset are assigned positive weights. Similarly, replicating the results in the endogenous process case would also require that the agents (or a majority thereof) assign positive weight to specific theories, for these theories to emerge as equilibria.

5.2. Probabilistic theories

One may wish to generalize the model to deal with theories whose predictions are probabilistic. For example, a theory may assign to each history h_t a distribution over outcomes $f(h_t) \in \Delta(Y)$, rather than a specific outcome. As mentioned above, such theories can be captured by the deterministic theories in their support. Specifically, a weight $\phi(f)$ assigned to a theory f can be split among a class of deterministic theories $D(f)$ so that the total weight of these theories at history h_t equals $\phi(f)\Pr(h_t|f)$ that is, the original weight of f, $\phi(f)$, multiplied by f's likelihood at h_t. The predictions generated by such a class of deterministic theories will be the same as those generated by the original probabilistic f, and, furthermore, the same would apply to aggregation over such theories.

Observe, however, that a single probabilistic theory will typically result, in this construction, in uncountably many deterministic ones. For example, if f states that the distribution over Y is uniform for every history h_t, all states of the world would be in the support of f. As just noted (in Section 5.1), this would require the more general set-up of GSS (2013), where ϕ is a measure over subsets of conjectures.

[17] The formal statement of this result, as well as the entire discussion, require a more general model, in which ϕ is not defined for individual conjectures, but is, rather, a measure of a σ-algebra of conjectures. That is, ϕ is defined on set of subsets of states of the world, as in GSS (2013).

As observed in Section 5.1, results such as Lemma 1 and Proposition 3 would still hold in such a model: the driving force behind them is the fact that the total weight of all theories, probabilistic or not, is divided among exponentially many disjoint histories, and only a fraction of these can have a weight that is not decreasing at an exponential rate. However, one may argue that it is inappropriate to multiply the weight of a probabilistic theory by its likelihood function: when comparing probabilistic theories, all of whom make predictions at each and at every period, the likelihood function seems to be an obvious tool. But when such theories are compared with case-based predictions, which have the luxury of making predictions only at very specific periods, such a comparison might not give theories a fair chance, as it were. Consider the following example.

Suppose that $|X| = 1$ and $Y = \{0, 1\}$. A probabilistic theory states that y_t are i.i.d. with $y_t \sim B(0.5)$. After a history of length t, the likelihood function of this theory is 2^{-t}. This theory will therefore be discredited in comparison with case-based conjectures. However, a proponent of this theory might say, "It's unfair to penalize me for not being able to predict the pattern of 0's and 1's in the data. I argued that the variables are i.i.d., and when I see a sequence with no obvious pattern, I feel vindicated. You may penalize me if the average of the y_t's is far from 0.5; you may also discredit my theory if there is an obvious pattern in the data; but you shouldn't ask me to predict a particular sequence of 0's and 1's to begin with. My probabilistic theory precisely states that there is no point in making such predictions."

Thus, when probabilistic theories are concerned, it may be more reasonable to lump periods together, and apply our model to averages of relatively long chunks of periods, predicting that these averages lie in certain positive-length intervals, as suggested in the introduction.

5.3. Hybrid models

Consider the case of trade in financial markets. Financial assets are affected by various economic variables that are exogenous to the market, ranging from weather conditions to technological innovation, from demand shocks to political revolutions. At the same time, financial assets are worth what the market "thinks" they are worth. In other words, such markets have a strong endogenous factor as well. It seems natural to assume that such processes (y_t) are governed partly by the predictions $(\hat{y}_t)$ as in Section 4 and partly by random shocks as in Section 3. For instance, assume that $\alpha(h_t)$ is the probability that agents' reasoning determines y_t, and with the complement probability y_t is determined by a random shock. That is,

$$y_t = \begin{cases} \hat{y}_t, & \alpha(h_t), \\ \tilde{y}_t, & 1 - \alpha(h_t) \end{cases}$$

where $\hat{y}_t \in \arg\max_{y \in Y} \phi(\mathcal{A}(h_t, \{y\}))$ and $\tilde{y}_t$ is uniformly distributed over Y. Thus, if $\alpha(h_t) \equiv 1$ we consider a model as in Section 4, which is likely to converge to a single dominant theory, and when $\alpha(h_t) \equiv 0$ we consider a model as in Section 3, coupled with a non-degenerate i.i.d. measure that guarantees asymptotic case-based reasoning. Obviously, the interesting case is where $\alpha(h_t) \in (0, 1)$ (for most if not all histories h_t).

If $\alpha(h_t)$ is independent of history, so that $\alpha(h_t) \equiv \alpha \in (0, 1)$, no theory can be dominant asymptotically. Indeed, every theory that correctly predicts $\hat{y}_t$ has a fixed positive probability $(1 - \alpha)$ of being refuted at each period, and will thus be refuted at some point with probability 1. Moreover, when t is large, we know that with very high probability the number of "noise" periods is approximately $(1 - \alpha)t$. Over these periods we are likely to observe a complex pattern of y_t's, and thus a result similar to Proposition 3 holds: the total weight of rule-based conjectures decreases, on average, exponentially fast in the number of noise periods. Because the number of noise periods increases linearly in t (as it is roughly $(1 - \alpha)t$), this weight is also an exponentially decreasing function of t and thus it decays faster than do the case-based conjectures. Thus, case-based reasoning will be asymptotically dominant in "most" states of the world even if $\alpha(h_t) \equiv \alpha$ is very close to 1.

However, the probability of noise in an endogenous process is likely to be endogenous as well. For example, consider the choice of driving on the right or on the left in a large population. When agents are not quite sure which equilibrium is being played, it is easier for a random shock to switch equilibria. But when all the agents are rather certain that everyone is going to drive, say, on the right, it is highly unlikely that at least half of them would behave differently from what they would find optimal based on their predictions. Thus, it stands to reason that $\alpha(h_t)$ depends on h_t, and, moreover, that it converges to 1 as t grows, if a simple theory fits the data h_t. Such convergence would allow the process to be asymptotically dominated by rule-based conjectures with positive probability.

5.4. Heterogeneous beliefs

The analysis in Section 4 assumes that all agents share the function ϕ, which is the natural counterpart of the common prior assumption in economics. Clearly, this assumption is not entirely realistic; people vary in their similarity judgments, in their prior beliefs in theories, as well as in their tendency to reason by theories vs. by analogies. Hence one may consider an endogenous process in which the population is distributed among different credence functions ϕ.

If we consider the case of computable theories, we find that subjectivity has limits: the distinction between computable and incomputable states is an objective one. Agents may vary in the language they use to describe theories, and, correspondingly, in their judgment of simplicity. However, any two languages that are equivalent to the computational model of a Turing machine can be translated to each other. Thus, if the process follows a simple (computable) path, all agents will

notice this regularity. Different agents may discard case-based reasoning in favor of the unrefuted theory at different times, but (under the assumption of open-mindedness) all of them will eventually realize that this unrefuted theory is indeed "correct". Interesting dynamics might emerge if the agents who are slow to switch to prediction by the correct theory are sufficiently numerous to refute that theory, thereby changing the reasoning of those agents who were the first to adopt the theory.

5.5. Main messages

To conclude, it might be useful to summarize the main messages of this paper. The first is that rule-based reasoning is generally not sufficient to describe the way agents think about the world – neither at the outset nor asymptotically. The second is that, other things being equal, an endogenous process is more likely to give rise to convergence to rule-based reasoning than is an exogenous one.

We believe that these messages are relatively robust to the assumptions of our model. For example, for an exogenous process, if one relaxes the assumption that the set for rules is countable, one can no longer guarantee that rule-based reasoning becomes dominant even in a state where a particular rule holds. Thus, the assumption of countability of the set of rules is relatively favorable to rule-based reasoning, in terms of becoming dominant when valid. Conversely, the general message that rule-based reasoning does not suffice is strengthened when the assumption of countability is relaxed. At the same time, when an endogenous process is concerned, it is still possible that some rules have a positive weight (that is, that the measure over the set of rules has atoms), and these rules would be more likely to emerge as equilibria than to be a priori selected by an exogenous random process.

Appendix A. Proofs

A.1. Proof of Proposition 1

Assume that $\omega \in [f_r]$ for some r. In this case the denominator is bounded from below by the weight assigned to the correct theory f_r. In fact,

$$\mathcal{R}\big(h_t(\omega)\big) \searrow \phi(f_r) > 0$$

as $t \to \infty$.

By contrast, $\mathcal{CB}(h_t(\omega))$ includes the ϕ-weight only of those case-based conjectures that are relevant at t, that is

$$\phi\big(\mathcal{CB}\big(h_t(\omega)\big)\big) = \sum_{\{(i,x,z)|i<t,\omega_X(i)=x,\omega_X(t)=z\}} \phi(A_{i,t,x,z}).$$

Clearly,

$$\phi\big(\mathcal{CB}\big(h_t(\omega)\big)\big) \leqslant \sum_{\{(i,x,z)|i<t\}} \phi(A_{i,t,x,z}).$$

Defining

$$\alpha_t = \sum_{\{(i,x,z)|i<t\}} \phi(A_{i,t,x,z})$$

and observing that

$$\sum_t \alpha_t = \phi(\mathcal{CB}) < 1$$

we must have

$$\alpha_t \to 0$$

as $t \to \infty$. Hence $\phi(\mathcal{CB}(h_t(\omega)))$ also converges to zero as $t \to \infty$, and this completes the proof.

A.2. Proof of Proposition 2

First consider $\Omega_{R\phi}$. By Proposition 1, for $\omega \in \mathbb{S}$ we have $\frac{\phi(\mathcal{CB}(h_t(\omega)))}{\phi(\mathcal{R}(h_t(\omega)))} \to 0$ and thus $\mathbb{S} \subset \Omega_{R\phi}$. However, $\mathbb{S}$ is dense in Ω, because we assumed that the set of theories is rich enough to intersect any open set $[h_t]$. Hence $\Omega_{R\phi}$ is dense.

To see that $\Omega_{C\phi}$ is dense, we use Proposition 3 proven below. (The proof of Proposition 3 does not make use of the present result.) That proposition states that

$$\lambda\left(\frac{\phi(\mathcal{R}(h_t(\omega)))}{\phi(\mathcal{CB}(h_t(\omega)))} \to_{t\to\infty} 0\right) = 1$$

and that implies that the set $\{\omega \in \Omega \mid \frac{\phi(\mathcal{R}(h_t(\omega)))}{\phi(\mathcal{CB}(h_t(\omega)))} \to_{t\to\infty} 0\}$ is dense in Ω. However, this set is a subset of $\Omega_{C\phi}$. Hence, $\Omega_{C\phi}$ is also dense.

Finally, we wish to prove that $\Omega_{M\phi}$ is dense. It suffices to show that for every h_t there exists $\omega \in [h_t] \cap \Omega_{M\phi}$. Given such a history h_{t_0}, choose $\omega_1 \in [h_{t_0}] \cap \Omega_{R\phi}$. Let $t_1 > t_0$ be such that, at ω_1, for $t \geqslant t_1$, $\phi(\mathcal{R}(h_t(\omega))) \geqslant \phi(\mathcal{CB}(h_t(\omega)))$. Then consider the history h_{t_1} defined by ω_1. Find $\omega_2 \in [h_{t_1}] \cap \Omega_{C\phi}$ and let $t_2 > t_1$ be such that, at ω_2, for $t \geqslant t_2$, $\phi(\mathcal{R}(h_t(\omega))) \leqslant \phi(\mathcal{CB}(h_t(\omega)))$. Continuing by induction, one constructs a state $\omega \in [h_t] \cap \Omega_{M\phi}$. □

A.3. Proof of Proposition 3

We first prove the following lemma, which suggests that, in the "vast majority" of states, the weight of rule-based reasoning decays at a semi-exponential rate. This lemma does not depend on the allocation of weights among the case-based conjectures, and it is stated for all $\phi \in \Phi_+$ (and not only for $\phi \in \Phi_+^p$).

Lemma 1. *Let there be given $\phi \in \Phi_+$ and let λ be the uniform measure defined on Σ. There exists $0 < \delta < 1$ such that for every $\varepsilon > 0$ there exists T_0 such that*

$$\lambda\left(\left\{\omega \,\middle|\, \phi\left(\mathcal{R}\left(h_t(\omega)\right)\right) \leqslant \delta^{t/2} \ \forall t \geqslant T_0\right\}\right) > 1 - \varepsilon.$$

A.3.1. Proof of Lemma 1

Let there be given an open-minded model ϕ. For a period t and a sequence $x_{(t)} = (x_0, \ldots, x_{t-1}) \in X^t$, consider the state space $\Omega_{x_{(t)}}$ defined by the corresponding $y_{(t)} = (y_0, \ldots, y_{t-1}) \in Y^t$ and containing $|Y|^t$ states. Thus $\Omega_{x_{(t)}}$ is a replica of Y^t and when no confusion is likely to arise we will refer to elements of $\Omega_{x_{(t)}}$ as $y_{(t)}$. Let $\lambda_{x_{(t)}}$ be the corresponding (uniform) measure on $\Omega_{x_{(t)}}$.

Choose $\frac{1}{|X||Y|} < \delta < 1$. Observe that $\lambda_{x_{(t)}}$ attaches a probability not exceeding δ^t to each element in the space $\Omega_{x_{(t)}}$.

Let W be a random variable defined on $\Omega_{x_{(t)}}$, and measuring the total weight of rule-based conjectures that are compatible with history. That is, for $y_{(t)} = (y_0, \ldots, y_{t-1})$ choose an arbitrary $x_t \in X$ and define h_t by $h_t = ((x_0, y_0), \ldots, (x_{t-1}, y_{t-1}), x_t)$. Choose ω such that $h_t(\omega) = h_t$ and define

$$W(y_{(t)}) = \phi\left(\mathcal{R}\left(h_t(\omega)\right)\right).$$

Clearly, such states ω exist.

Observe that $\{\mathcal{R}(h_t(\omega))\}_\omega$ defines a partition of $\mathcal{R}$: each theory $f \in \mathcal{R}$ is compatible with precisely one state $y_{(t)} \in \Omega_{x_{(t)}}$. Hence

$$\sum_{y_{(t)} \in \Omega_{x_{(t)}}} \phi\left(\mathcal{R}\left(h_t(\omega)\right)\right) = r < 1$$

and therefore

$$\begin{aligned} E(W) &= \sum_{y_{(t)} \in \Omega_{x_{(t)}}} \lambda_{x_{(t)}}(y_{(t)}) W(y_{(t)}) \\ &= \sum_{y_{(t)} \in \Omega_{x_{(t)}}} \lambda_{x_{(t)}}(y_{(t)}) \phi\left(\mathcal{R}\left(h_t(\omega)\right)\right) \\ &< \delta^t r < \delta^t. \end{aligned}$$

Denoting by B_t the event $W > \delta^{t/2}$, and using Markov's inequality, we get

$$\lambda_{x_{(t)}}(B_t) = \lambda_{x_{(t)}}\left(W > \delta^{t/2}\right) < \frac{E(W)}{\delta^{t/2}} < \frac{\delta^t}{\delta^{t/2}} = \delta^{t/2}.$$

We will also use B_t to denote the corresponding event in Ω. Since we have shown that $\lambda_{x_{(t)}}(B_t) = \lambda(B_t | x_{(t)}) < \delta^{t/2}$ for all $x_{(t)}$, we also have $\lambda(B_t) < \delta^{t/2}$.

Next observe that the bounds on the probabilities of the various B_t events converge. This implies that for the given $\varepsilon > 0$ there is a large enough T_0 such that

$$\sum_{t \geqslant T_0} \delta^{t/2} < \varepsilon$$

and thus, for this T_0,

$$\lambda\left(\bigcup_{t\geqslant T_0} B_T\right) < \varepsilon$$

and

$$\lambda\left(\left\{\omega \mid \phi\left(\mathcal{R}\left(h_t(\omega)\right)\right) \leqslant \delta^{t/2}\ \forall t \geqslant T_0\right\}\right) > 1-\varepsilon.$$

A.3.2. Completion of the proof of Proposition 3

Consider a given $\varepsilon > 0$ and let T_0 be the period provided by Lemma 1. Then, on the corresponding event (whose probability is at least $1-\varepsilon$)

$$\phi\left(\mathcal{R}\left(h_t(\omega)\right)\right) \leqslant \delta^{t/2} \quad \forall t \geqslant T_0$$

and this, together with the assumption that $\phi \in \Phi_+^p$, that is, $\sum_{i<t} \phi(A_{i,t,x,z}) \geqslant ct^\gamma$ for $c > 0$ and $\gamma < -1$, implies that

$$\frac{\phi(\mathcal{R}(h_t(\omega)))}{\phi(\mathcal{CB}(h_t(\omega)))} < \frac{\delta^{t/2}}{ct^\gamma}$$

where the right hand side converges to 0 as t tends to ∞.

Considering a sequence $\varepsilon_n \searrow 0$, one concludes that the convergence to 0 occurs everywhere apart from a set whose λ-measure is zero.

A.4. Proof of Proposition 4

Let there be given $\omega \in \Omega$. For simplicity of notation, we define ϕ without guaranteeing the normalization $\phi(\mathcal{A}) = 1$. It will be obvious from the construction, however, that $0 < \phi(\mathcal{A}) < \infty$ so that ϕ can be normalized.

We first define ϕ on the case-based conjectures. For every $t \geqslant 1$, let

$$\phi(A_{i,t,x,z}) = \frac{1}{(t+5)^3}.$$

(We use a "lag" of 5 periods to make sure that the rate of decay between any two consecutive periods is not too fast. Specifically, we wish to guarantee that each element in the sequence is at least half of its predecessor.)

Clearly,

$$\phi(\mathcal{CB}) \leqslant |X|^2 \sum_{t\geqslant 1}\left[t\frac{1}{(t+5)^3}\right] < \infty.$$

We now turn to define ϕ on $\mathcal{R}$. In the proof we wish to assign weights to subsets of conjectures in $\mathcal{R}$. Note that for every subset $\mathcal{R}' \subset \mathcal{R}$ and every $a > 0$ one may assign a positive weight $\phi(f) > 0$ to each $f \in \mathcal{R}$ such that $\phi(\mathcal{R}') = a$, say by considering an enumeration of $\mathcal{R}'$, $f_1, f_2, \ldots$ and setting $\phi(f_j) = a/2^j$. In the rest of this proof, we will simply say "assign a weight $a > 0$ to the subset $\mathcal{R}'$", referring to such an assignment.

If $\omega \in \mathbb{S}$, there exists a theory $f \in \mathcal{R}$ such that $\omega \in [f]$. In this case, assign $\phi(f) = 1$ and assign the weight $a = 1/4$ to the set of all the other theories, $\mathcal{R}\backslash\{f\}$. It is easily observed that, at each $t \geqslant 0$, $\omega_Y(t) \in \arg\max_{y\in Y} \phi(\mathcal{A}(h_t, \{y\}))$ and thus $\omega \in \Omega_\phi$ is established, while $\phi \in \Phi_+^p$ holds.

Next assume that $\omega \notin \mathbb{S}$. Denote, for $t \geqslant 0$,

$$\mathcal{R}_t = \mathcal{R}\left(h_t(\omega)\right).$$

$\mathcal{R}_t$ denotes the set of theories that are unrefuted by history $h_t(\omega)$. Observe that they are all relevant for prediction at period t. Clearly, $\mathcal{R}_0 = \mathcal{R}$, as $h_0(\omega)$ contains only the value of x_0 and no theory makes any prediction about the x's. Moreover, $\mathcal{R}_{t+1} \subset \mathcal{R}_t$, because any theory that agrees with ω for the first $(t+1)$ observations also agrees with it for the first t observations. Finally,

$$\bigcap_t \mathcal{R}_t = \varnothing$$

because $\omega \notin \mathbb{S}$. We can thus define, for $t > 1$, the set of theories that are proven wrong at period t to be

$$\mathcal{W}_t = \mathcal{R}_{t-1}\backslash\mathcal{R}_t.$$

Observe that

$$\mathcal{R} = \bigcup_t \mathcal{W}_t$$

and

$$\mathcal{W}_t \cap \mathcal{W}_{t'} = \varnothing$$

whenever $t \neq t'$.

Thus, at period t $\mathcal{R}_t$ consists of all theories that were unrefuted by $h_t(\omega)$, and it is the disjoint union of $\mathcal{R}_{t+1}$, namely the theories that correctly predict $y_t = \omega_Y(t)$ and $\mathcal{W}_{t+1}$, namely the theories that predict different values for y_t, and that will be proven wrong.

If we ignore the case-based conjectures, the prediction made by the theories in $\mathcal{R}_t$ is guaranteed to be the "correct" prediction $\omega_Y(t)$ if

$$\phi(\mathcal{R}_{t+1}) > \phi(\mathcal{W}_{t+1}).$$

(Observe that, as compared to $h_t(\omega)$, $h_{t+1}(\omega)$ specifies two additional pieces of information: the realization of y_t, $\omega_Y(t)$, and the realization of x_{t+1}, $\omega_X(t+1)$. However, theories do not predict the x values, and thus the theories in $\mathcal{R}_{t+1}$ are all those that were in $\mathcal{R}_t$ and that predicted $y_t = \omega_Y(t)$; the observation of x_{t+1} does not refute any additional theories.)

A simple way to construct $\phi \in \Phi_+^p$ is to make sure that the prediction at each period is dominated by the rule-based conjectures, despite the existence of the case-based conjectures. To guarantee that this is the case, we set

$$\phi(\mathcal{R}_t) = \frac{3}{(t+5)^2}$$

at each $t \geqslant 0$.

Observe that, for $t \geqslant 0$,

$$\begin{aligned}\phi(\mathcal{R}_{t+1}) &= \frac{3}{(t+6)^2},\\ \phi(\mathcal{W}_{t+1}) &= \phi(\mathcal{R}_t) - \phi(\mathcal{R}_{t+1})\\ &= \frac{3}{(t+5)^2} - \frac{3}{(t+6)^2}.\end{aligned}$$

This dictates the definition of ϕ on $\mathcal{R}$: we start with $\phi(\mathcal{R}) = \phi(\mathcal{R}_0) = \frac{3}{5^2}$, and assign the weight $3[(t+5)^{-2} - (t+6)^{-2}]$ to the subset of theories $\mathcal{W}_{t+1}$. Since $\bigcup_t \mathcal{W}_t = \mathcal{R}$, this defines ϕ on all of $\mathcal{R}$. Clearly, $\phi(\mathcal{R})$ is finite.

Next, observe that at each $t \geqslant 0$, $\omega_Y(t) \in \arg\max_{y \in Y} \phi(\mathcal{A}(h_t, \{y\}))$. Specifically, at $t = 0$ we only have to compare the rule-based hypotheses. We have

$$\begin{aligned}\phi(\mathcal{R}_1) &= \frac{3}{6^2},\\ \phi(\mathcal{W}_1) &= \frac{3}{5^2} - \frac{3}{6^2}\end{aligned}$$

so that

$$\phi(\mathcal{R}_1) - \phi(\mathcal{W}_1) = 2\frac{3}{6^2} - \frac{3}{5^2} > 0.$$

For each $t \geqslant 1$, the total weight of the case-based conjectures is

$$t\frac{1}{(t+5)^3}.$$

We wish to show that the weight of the theories that predict the "correct" continuation $\omega_Y(t)$, $\mathcal{R}_{t+1}$, is larger than that of the theories that predict other continuations, even when the latter is combined with all case-based conjectures. Indeed,

$$\phi(\mathcal{R}_{t+1}) - \phi(\mathcal{W}_{t+1}) = 2\frac{3}{(t+6)^2} - \frac{3}{(t+5)^2} > t\frac{1}{(t+5)^3}.$$

This completes the proof that $\omega_Y(t) \in \arg\max_{y \in Y} \phi(\mathcal{A}(h_t, \{y\}))$ for all t, and it is easily verified that after normalization we obtain $\phi \in \Phi_+^p$ such that $\omega \in \Omega_\phi$. □

A.5. Proof of Proposition 5

Assume that $\omega \in \mathbb{S}$. Then there exists a theory $f \in \mathcal{R}$ such that $\omega \in [f]$. Since $\phi \in \Phi_+$, $\phi(f) > 0$ and this implies that $\phi(\mathcal{R}(h_t(\omega))) > \phi(f) > 0$ for all t. By contrast, $\phi(\mathcal{CB}(h_t(\omega))) \searrow 0$. Similarly, $\phi(\mathcal{R}(h_t(\omega)) \backslash \mathcal{R}(h_{t+1}(\omega))) \searrow 0$ because the sets $\{\mathcal{R}(h_t(\omega)) \backslash \mathcal{R}(h_{t+1}(\omega))\}_t$ are pairwise disjoint (and the sum of their weights is bounded). Hence, from some T onwards, theory f dominates prediction and $\omega \in \Omega_{RB\phi}$. □

A.6. Proof of Observation 1

Let there be given a theoretically closed $\phi \in \Phi_+^p$ and assume that $\omega \in \Omega_{CB\phi}$. This implies that, from some T' onwards, $\omega_Y(t) \in \arg\max_{y\in Y} \phi(\mathcal{A}(h_t, \{y\}))$. Assume, without loss of generality, that $T' \geqslant T$, so that for $t \geqslant T'$ $\arg\max_{y\in Y} \phi(\mathcal{A}(h_t, \{y\}))$ is a singleton. By theoretical closedness of ϕ, for $h_{T'}(\omega)$, there exists $f \in \mathcal{R}$ such that $[h_{T'}] \subset [f]$ and, for every $t \geqslant T'$, and every continuation h_t of $h_{T'}$, $f(h_t) \in \arg\max_{y\in Y} \phi(\mathcal{CB}(h_t, \{y\}))$, hence $f(h_t) = \omega_Y(t)$. It follows that $\omega \in \mathbb{S}$ and $\omega \in \Omega_{RB\phi}$. Clearly, $\Omega_{RB\phi} \cap \Omega_{CB\phi} = \varnothing$, hence this is a contradiction to the assumption that $\omega \in \Omega_{CB\phi}$. It thus follows that $\Omega_{CB\phi} = \varnothing$. □

References

Akaike, H., 1954. An approximation to the density function. Ann. Inst. Stat. Math. 6, 127–132.
Blackwell, D., Dubins, L., 1962. Merging of opinions with increasing information. Ann. Math. Stat. 33, 882–886.
Dempster, A.P., 1967. Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Stat. 38, 325–339.
Domingosu, P., 1996. Unifying instance-based and rule-based induction. Mach. Learn. 24, 141–168.
Gilboa, I., Schmeidler, D., 2001. A Theory of Case-Based Decisions. Cambridge University Press, Cambridge.
Gilboa, I., Schmeidler, D., 2003. Inductive inference: an axiomatic approach. Econometrica 71, 1–26.
Gilboa, I., Samuelson, L., Schmeidler, D., 2013. Dynamics of inductive inference in a unified model. J. Econ. Theory 148, 1399–1432.
Goodman, N., 1955. Fact, Fiction, and Forecast. Harvard University Press, Cambridge, MA.
Hume, D., 1748. Enquiry Into the Human Understanding. Clarendon Press, Oxford.
Kalai, E., Lehrer, E., 1993. Rational learning leads to Nash equilibrium. Econometrica 61, 1019–1045.
Kolodner, J., 1992. An introduction to case-based reasoning. Artif. Intell. Rev. 6 (1), 3–34.
Parzen, E., 1962. On the estimation of a probability density function and the mode. Ann. Math. Stat. 33, 1065–1076.
Riesbeck, C.K., Schank, R.C., 1989. Inside Case-Based Reasoning. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ.
Rissland, E.L., Skalak, D.B., 1989. Combining case-based and rule-based reasoning: A heuristic approach. In: Proceedings IJCAI-89, pp. 524–530.
Schank, R.C., 1986. Explanation Patterns: Understanding Mechanically and Creatively. Lawrence Erlbaum Associates, Hillsdale, NJ.
Shafer, G., 1976. A Mathematical Theory of Evidence. Princeton University Press, Princeton.
Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall, London and New York.
Slade, S., 1991. Case-based reasoning: A research paradigm. AI Mag., 42–55.
Solomonoff, R., 1964. A formal theory of inductive inference I. Inf. Control 7, 1–22. II. Inf. Control 7 (1964) 224–254.
Wittgenstein, L., 1922. Tractatus Logico Philosophicus. Routledge and Kegan Paul, London.