

Albert, Max

**Working Paper**

## Bayesian learning and expectations formation: Anything goes

Diskussionsbeiträge - Serie I, No. 284

**Provided in Cooperation with:**

Department of Economics, University of Konstanz

*Suggested Citation:* Albert, Max (1996) : Bayesian learning and expectations formation: Anything goes, Diskussionsbeiträge - Serie I, No. 284, Universität Konstanz, Fakultät für Wirtschaftswissenschaften und Statistik, Konstanz

This Version is available at:

<https://hdl.handle.net/10419/68915>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

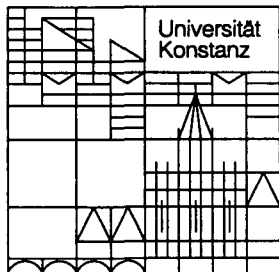
Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



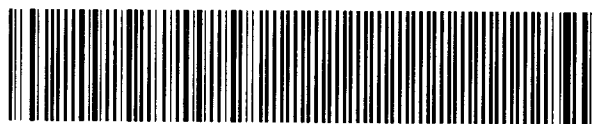
## Fakultät für Wirtschaftswissenschaften und Statistik

---

Max Albert

### Bayesian Learning and Expectations Formation: Anything Goes

W 284 (284)



12. DEZ. 1996 Weltwirtschaft  
Kiel

W 284 (284)

*mi. sig ja*

Diskussionsbeiträge

---

**Bayesian Learning and Expectations  
Formation: Anything Goes**

**Max Albert**

7195848

**Serie I - Nr. 284**

**November 1996**

Max Albert

# Bayesian Learning and Expectations Formation: Anything Goes

November 8, 1996\*

Correspondence:

Max Albert

Universität Konstanz

Fakultät für Wirtschaftswissenschaften und Statistik

P.O. Box 5560 D 140

D-78434 Konstanz, Germany

Phone: +49-7531-88-2212

Fax: +49-7531-88-4091

Internet: [max.albert@uni-konstanz.de](mailto:max.albert@uni-konstanz.de)

---

\* Previous versions of this paper have been presented at the 2<sup>nd</sup> *Florence-Konstanz Symposium* and at the Faculty of Economics of the University of Rostock. For useful hints and comments, I am grateful to Friedrich Breyer, Franco Gori, Ron A. Heiner, Karl-Josef Koch, Alberto Loayza-Grbic, Jürgen Meckl, Dieter Schmidtchen, Thusnelda Tivig and Heinrich Ursprung.

### **Abstract**

The paper proves that the Bayesian approach to learning and expectations formation implies no propositions that could conceivably be refuted by observation. For a (non-expanding) universe infinite in time but finite at any point of time, it is shown that by a suitable choice of priors, any betting strategy can be rationalized. Consequently, no sequence of future choices can be excluded on the basis of a decision maker's  $\psi$ . Neumann–Morgenstern utility function, her knowledge of past observations, and the assumption of Bayesian rationality. Since no betting strategy is irrational, Bayesianism is useless as a normative theory as well.

JEL Classification: B41, C11, D81, D83, D84

Keywords: Bayesianism, Bayesian learning, decision theory, expectations, learning, rational expectations, rationality, rationalizability, subjectively expected utility theory, subjective probability

# 1 Introduction

In the literature on expectations formation, the rational–expectations hypothesis (REH) is giving way to theories of learning.<sup>1</sup> In view of the history of the subject, this development is not surprising. However, to use the famous Knightian distinction, it again shifts the focus of interest from risk to uncertainty. The present paper argues that the accepted theory of decision making under uncertainty in economics, subjectively expected utility (SEU) theory or Bayesianism, is unable to deal with the problems of uncertainty and learning: Its only implication, as a normative (or prescriptive) and as a positive theory, is “anything goes”. This result does not imply that models employing Bayesian rationality are necessarily useless. However, it most certainly discredits the claim that *only such models* are respectable. Moreover, it sheds new light on the relation between Bayesian and bounded rationality.

## From Rational Expectations to Rational Expectations Formation

The REH is defined by two assumptions: (i) Agents’ expectations are subjective probability distributions over (usually: future) events. (ii) The subjective distributions conditional on the agents’ information coincide with the objective conditional distributions. Thus, the REH allows for risk but excludes uncertainty; there is nothing left to learn but the realizations of random variables with known probability distributions.<sup>2</sup> However, Muth’s (1961) basic idea, sometimes called the weak version of the REH, was that agents form expectations by rationally acquiring and processing information. Muth himself never developed this idea, which requires a theory of *rational learning*, but immediately jumped to a variant of the strong version of the REH. In a way, the recent literature walks where he jumped.

By assuming that not all the probability distributions for all the potentially influential variables are known, the weak version of REH introduces uncertainty and therefore implicitly invokes the SEU theory a. k. a. Bayesianism. The first name derives from the theory’s claim that in a situation of uncertainty, a rational decision

---

<sup>1</sup> Cf. Kirman & Salmon (1995) for a recent overview.

<sup>2</sup> On the REH, see, e. g., Frydman & Phelps (1983) and Pesaran (1987: ch. 1 and 2). Macroeconomic models often assume that objective and subjective distributions agree only w. r. t. the mean. In the present context, this variant of REH is of no importance.

maker maximizes his or her subjectively expected utility. The second name, which we will prefer, is explained by the fact that the theory implies a theory of rational learning, which proceeds via *Bayes' theorem* and therefore is referred to as *Bayesian learning*. According to Bayesianism, learning is a byproduct of intertemporal decision making. A rational decision maker derives expectations, which again take the form of subjective probability distributions, from an initially chosen subjective *prior distribution* by conditionalization on past observations. The derived *posterior distribution* is viewed as the result of learning.<sup>3</sup>

So while the REH assumes that subjective probability distributions just coincide with the objective distributions, Bayesian learning focuses on the dynamics of changes in subjective distributions. This immediately leads to the question of stability or convergence: Does Bayesian learning converge to rational expectations, i. e., does it lead to a convergence of the subjective towards the objective distributions? There are two aspects to this question. The minimal requirement for rational expectations is the convergence of subjective and objective distributions (expectations convergence). However, it is natural to assume, as it is often done, that rational expectations are based on knowledge of the true model of the economy, supplemented by knowledge of the objective distributions of the variables exogenous to this model. Accordingly, there arises the additional question of the convergence of beliefs concerning the model (model convergence).

In both of these senses, convergence is not guaranteed; moreover, no type of convergence implies the other.<sup>4</sup> As the literature on self-fulfilling prophecies shows, expectations convergence is possible without model convergence. Model convergence without expectations convergence is also possible, as an example by Blume & Easley (1995) shows, which will be discussed below.<sup>5</sup> Bayesianism as an account of ideal

---

<sup>3</sup> The classical exposition of Bayesianism is Savage (1954); see also Kiefer & Nyarko (1995) for a summary and discussion with special reference to learning and expectations formation.

<sup>4</sup> Of course, model convergence together with convergence of beliefs concerning the distribution of exogenous variables implies expectations convergence. However, note that our definition of model convergence does not say anything about the beliefs concerning exogenous variables.

<sup>5</sup> See Earman (1992: sections 6.3-6.5) for a short overview on the positive convergence results. On self-fulfilling prophecies see Pesaran (1987: 23) and, extensively, Farmer (1993). There is a further aspect to the convergence problem: not convergence to the truth but intersubjective convergence, which seems to be enough for some Bayesians, especially those who believe that

learning is therefore not without its problems: Even the beliefs of an ideal Bayesian learner who does *not* dismiss the truth from the outset and who faces *no* costs of gathering further information are not inevitably bound to converge to the truth.<sup>6</sup>

It is not clear, however, what lesson, if any, should be drawn from the possible failure of convergence. From an economist's point of view, the real issue is decision making, not learning, and a failure of beliefs to converge to the truth is not necessarily an indication that the decisions based on these beliefs are mistaken in some sense (cf. Kiefer & Nyarko 1995). On the contrary, Bayesians are bound to claim that they do not understand in which sense the maximization of subjectively expected utility could ever lead to mistakes. Is this not the very *definition* of rationality? Then which criterion could be behind a criticism of such decisions? Of course, *with hindsight* there are usually many decisions that would have been better than a decision maximizing subjectively expected utility. But there seems to be no criterion that can be employed to criticise the decisions of an ideally rational Bayesian *before* uncertainty is resolved.

So the possible failure of convergence, while certainly disappointing from a Bayesian perspective, provides no decisive argument against the approach. However, there remain other possible lines of criticism. Bayesian rationality may be unachievable in practice, inconsistent, or empty. The first claim is often heard, the second never (and rightly so, I believe). The present paper advances the third.

---

"objective" probabilities are just subjective probabilities held by several people in common. The latter position will be discussed only in passing; see the critical remarks on "rational priors" in section 4.

<sup>6</sup> The quintessential learning problems are posed by the sciences, social sciences included. Not surprisingly, therefore, Bayesianism provides a scientific methodology. See Howson & Urbach (1989) for an exposition and defense and Earman (1992) for a sympathetic but nonetheless critical overview. This methodology ties in neatly with the Bayesian perspective in statistical inference, eloquently defended by Leamer (1978) and others. The methodological literature is especially relevant in our context. For example, to those acquainted with the literature on the problem of induction it is of course quite obvious that there can never be a guarantee that a learning procedure will converge to the truth. Cf. Humphreys (1990) for a short introduction to the problem.



## Staking out the Claims

At first sight, the claim that Bayesianism is empty may not be very convincing. After all, the modern version of Bayesianism, based on Savage (1954), emerged as an extension of v. Neumann–Morgenstern (NM) utility theory (cf. Hacking 1990: 176), which in turn encompasses utility theory as used in the theory of demand. From the Bayesian perspective, NM utility theory is the special case resulting from the assumption that the objective probabilities of all events are known with certainty. And utility theory results if these objective probabilities are all either 0 or 1. In the latter case, there exist “operationally meaningful theorems” (OMTs), i. e., theorems that potentially could be refuted by observations.<sup>7</sup> These theorems are usually called the *axioms* of revealed preference (RP).<sup>8</sup> How can Bayesianism be empty if it implies the RP axioms for certain situations where the decision maker believes, among other things, that all relevant probabilities are 0 or 1? The answer is obvious: Bayesianism implies OMTs for situations that are characterized by certain beliefs on the part of the agent. If it is impossible to check *independently from the success or failure of the theorems* whether the agent holds these beliefs or not, the theorems have no empirical content.

The critical point is that beliefs are endogenous in the wider framework of Bayesianism. It is trivial that Bayesianism is *not empty in some sense* because it restricts actions once (some of) the beliefs of the decision maker are given. However, beliefs cannot be observed any better than preferences; in fact, it is one of the tenets of Bayesianism that beliefs have the same status as preferences in every respect (cf. Kiefer & Nyarko 1995). The unobservability of preferences is the rationale of the RP approach to the theory of demand. The RP argument showed that the

---

<sup>7</sup> Samuelson’s concept of “operational meaning” coincides with Popper’s (1959) definition of empirical content and is therefore not tied to the obsolete methodology of operationalism. We stick to this phrase since it is well-known in economics, esp. in connection with the revealed-preference approach.

<sup>8</sup> Modern developments can be summarized by Afriat’s theorem, which states that the GARP (generalized axiom of RP) is equivalent to the assumption of a locally nonsatiated, continuous, concave and monotonic utility function. GARP is inconsistent with many finite sequences of price vectors and corresponding consumption bundles. Thus GARP summarizes the empirical content of standard demand theory. See Varian (1992: 133) for an exposition.

theory of demand actually implies restrictions on observed behavior, setting limits to the rationalizability of choices.<sup>9</sup> The present paper considers a similar question. The focus is not on preferences but on beliefs. We assume that the NM utility function of an agent is known; his or her system of beliefs, however, is assumed to be unknown.<sup>10</sup> We furthermore assume that the agent is perfectly rational in the Bayesian sense of the term. The question is: Can we conclude on the basis of the NM utility function and some observed sequence of past choices that at least some sequences of future choices cannot occur?

The answer to this question cannot be found in the RP literature. The original RP approach does not apply to decision making under risk and uncertainty anyway. While some relevant extensions have been found, there are no OMTs covering Bayesian learning and expectations formation.<sup>11</sup> Such theorems would be

---

<sup>9</sup> Before Samuelson (1938, 1947) developed this approach, it was not clear whether anything about future choices could be learned from the observations of past choices of an individual. There was one simple and not very interesting restriction on individual demand functions implied by the theory, namely, zero homogeneity. Beyond that, it seemed that by postulating arbitrary (but still stable) preferences, almost any sequence of choices could be rationalized. It was unclear whether the Slutsky decomposition implied restrictions on *observable* behavior, i. e., reactions to *non-marginal* changes in prices and income (cf. Samuelson 1947: 107 fn. 13, Blaug 1992: 142).

<sup>10</sup> One obvious weakness of Bayesianism as a positive theory (which will, however, not be pursued here) is that it leaves open the choice of the utility function. Even in a certainty context, a sequence of choices does not reveal preferences if there is no intertemporal separability. However, in the present context we are prepared to grant knowledge of the NM utility function. This has to be done anyway whenever we discuss Bayesianism as a normative theory since it is natural to assume that at least decision makers themselves know their own NM utility functions.

<sup>11</sup> Border (1992) develops an RP approach to choice among lotteries with monetary rewards. The observer knows only that more money is preferred to less. If observer and observed agree on all (objective) probabilities, any choice behavior that is not statistically dominated can be rationalized by postulating a suitable utility function. The RP approach of Green & Osband (1991) is based on assumptions that deviate from Savage's (1954) framework in several ways. A direct comparison of results is therefore not possible. Kim (1992) considers choice under uncertainty but excludes learning, i. e., conditionalization on past observations. This can more reasonably be interpreted as subjective certainty concerning an objective probability distribution that is unknown to the observer. In Kim's set-up, the observed agent chooses from varying sets of uncertain prospects, where observed choices are assumed to reveal strict preference. Observed choices cannot be rationalized by a suitable probability distribution and a suitable NM utility function *iff* the following condition holds: There exists a probability distribution  $F$  which can be used to construct

relevant not only for Bayesianism as a positive theory of human behavior but also for Bayesianism as a theory or definition of rational behavior. Assume that Mike the Manager asks Betty the Bayesian for advice. Betty cannot take Mike's current beliefs for granted because it is an open question which beliefs are rational given Mike's knowledge of the past.<sup>12</sup> So it is natural for Mike to ask: Given what I know about the past, and given my NM utility function, is there anything I should *not* do? Is there a sequence of future choices, including reactions to new information, that can be classified as *irrational*? If, as I will argue, Betty can give no advice of this kind, Bayesianism is empty not only as a positive but also as a normative theory.

Whether we consider the normative or the positive theory of decision making, then, we are interested in theories that restrict the set of actions that can be rationalized. One such restriction is of course well-known. Let us assume that an NM utility function is given. Bayesians have always argued that their definition of rationality clashes with certain decision principles like the maximin principle.<sup>13</sup> Generalizing the respective argument, we can use the NM utility function to define a set of *dominated actions*: An action  $A$  is *dominated* if and only if for every assignment of probabilities to events, there always exists at least one action preferred to  $A$ . Bayesianism implies one restriction on behavior for a given NM utility function, namely, that no dominated action is chosen. Let us call this restriction the dominance principle. By definition, the dominance principle has nothing to do with learning: Whatever the agent learns about the true state of the world and the

---

a lottery  $A$  and a lottery  $B$  over the preferred and the rejected prospects, respectively, such that the individual is indifferent between  $A$  and  $B$ . Such violations of Bayesian rationality can only be observed in an experiment where the subject is confronted with lotteries  $A, B$  chosen by the experimenter and where both agree on the distribution  $F$ . The lotteries  $A, B$  cannot be inferred from observations. A violation is represented by a state of indifference; it will not be observed if the subject chooses  $A$ .

<sup>12</sup> Agents' beliefs concerning past observations are assumed to be true throughout the paper. Of course one may doubt whether all the observation statements one happens to believe in are actually true. This is the much-discussed problem of the "empirical basis", cf. Popper (1959: ch. 1). Assuming that the problem is absent amounts to a concession to Bayesianism.

<sup>13</sup> This point has been most often stressed in connection with statistical decision theory, cf. Lindley (1972: 13-15). Insofar as the literature uses loss instead of utility functions, it refers to the principle as the minimax principle.

probabilities this state implies for events, the set of dominated actions as defined above remains fixed. The present paper claims that nothing beyond the dominance principle follows from Bayesianism. This means that Bayesianism is empty *as a theory of learning and expectations formation*.

However, once this claim is accepted, it becomes difficult to defend the idea that the dominance principle actually yields some content. Identifying dominated actions already requires at least some knowledge about one's surroundings. For instance, burning money will often be a dominated action. But once I believe that I impress my business partners or the gods by lighting my cigar with a burning thousand-dollar note, things may look different. Of course, some such beliefs will be superstitious. But the superstitions of today are the wisdoms of yesterday. It takes learning to identify superstitious beliefs, which brings us back to square one. If Bayesianism is empty as a theory of learning, the dominance principle yields neither predictions nor advice.

Claims that the notion of rationality employed in economics is "almost empty" are not difficult to find. Hahn (1996: 186) states that in an intertemporal context, rationality "buys only a small bit of an answer" since it has to be supplemented by a theory on agents' beliefs. Concerning the content of Bayesian rationality in game theory, Blume & Easley (1995: 26) conclude that content mostly derives from restrictions on the set of beliefs. They make the assumption that rationality is common knowledge and restrict considerations to rationalizable strategies in the sense of Bernheim (1984). If even this restriction does not set limits to rational behavior, a similar conclusion for a truly general setting is very plausible. Bicchieri (1993: 14, esp. fn. 9) restricts the predictive usefulness of Bayesian rationality to stable environment and choice situations familiar to the agent; in a footnote, she mentions that there are convergence problems in the case of complicated priors. Arrow (1990: 29) writes that the rationality hypothesis by itself is "weak" and that its force derives from the addition of supplementary hypotheses. Ledyard (1986) demonstrates that Bayesianism is empty for a quite general game-theoretic setting. However, his results are derived mainly by variations of utility functions, assuming beliefs to be given. Maybe for this reason, he is still convinced of Bayesianism's value as a normative theory (Ledyard 1986: 60, 80f). Lucas (1976), in defending the strong version of the REH, claims that Bayesianism "in many applications" has

“little empirical content”.

The claim that Bayesianism is empty may be dubbed the folk theorem of decision theory. It shares the anything-goes character of the folk theorem in game theory. Moreover, it seems to be in the air; there is neither a completely general proof nor an original source.<sup>14</sup> Of course, there is a difference between the claim that Bayesian rationality is “weak” and that it is empty; the folk theorem as stated here is stronger than the statements just quoted. The difference stands behind Arrow’s (1990: 31) conviction, shared by many critics of Bayesian rationality, that rationality is difficult to achieve. This implies that rationality restricts behavior because otherwise there can be no difficulty. Those who are dissatisfied with the Bayesian approach seem to have mixed feelings: Rationality is almost empty but nevertheless somehow difficult to achieve—which implies that it cannot be completely empty.<sup>15</sup>

The paper undertakes to refute this idea. It is easy after all to be rational in the Bayesian sense, or so I will argue. Since any actions are known to be rationalizable, it would be sheer pedantry, and clearly irrational, to actually construct rationalizations. Rationalizations are interesting only if one is able to tell good ones from bad ones. Bayesianism gives no hint of how to do this. Even worse: Bayesianism cannot be supplemented by rules that allow to distinguish between good and bad rationalizations. Any such set of rules would be not a supplement but a substitute for Bayesianism: Since such rules would have to judge rationalizations on the basis of experience, they must provide a theory of learning different from Bayesianism. The latter would then be parasitic on another theory of learning. It is difficult to see why this other theory should be invoked only once as a first step in a learning process.

---

<sup>14</sup> Cf. Binmore (1992: 369ff, esp. fn. 17) on the folk theorem in game theory.

<sup>15</sup> Of course, one may hold the view that, while every behavior can be rationalized, such rationalizations are difficult to achieve. If one defines rational action as action accompanied by an appropriate rationalization, rationality is completely empty and nevertheless difficult to achieve. However, even this statement is true only for one meaning of “difficult”: According to the results of the present paper, rationalizations are *easy* to come by, although it would be *tiresome* to actually construct them.

## Chaos and Rationalizability

In order to demonstrate the emptiness of Bayesianism, we have to show that any sequence of actions can be rationalized on the basis of any kind of experience. More precisely: Consider the set of all contingent plans or *strategies* conceivable for an agent.<sup>16</sup> Bayesianism requires the agent to start from a subjective prior distribution over a set of hypotheses covering all the contingencies. The NM utility function together with this distribution determine the optimal strategy. We show that the agent can invert this procedure by choosing a strategy instead of a prior distribution. A rationalization of this strategy is provided by a set of hypotheses and a prior probability distribution over this set such that, *come what may*, the strategy maximizes the subjectively expected utility on the basis of the posterior distribution. That is: Every contingent action specified by the strategy is optimal in case the contingency arises.<sup>17</sup> Our proof is based on a set of quite simple hypotheses that can always be used to construct such a rationalization.

We start from a classical decision problem. A decision maker needs to predict the outcomes from an infinite sequence of experiments, where each single experiment results in either 0 or 1. Later, we will generalize the problem and consider the development of a finite and non-expanding universe, i. e., a universe with a finite description of given maximum length at every point in time. Such a description can be “digitalized”. Technically, we will therefore be concerned with predicting an infinite sequence of “experiments” where each single experiment results in a string of 0s or 1s of length  $n$ . The generalization will be trivial; so the simple problem is worth our attention.

---

<sup>16</sup> “Contingent” means: conditional on events. A contingent plan picks an action for every contingency. Note that many actions may have to be formulated in the “try to ...” form. If some experience leads you to believe that you are a ghost you may try to walk through a wall; however, at least according to my current system of beliefs, you will not succeed. If we are pedantic, even the idea that you may be able to try to walk through a wall presupposes some knowledge, namely, that it is possible to try to do this in certain situations. But I am not interested in blocking Bayesian arguments from the outset; so I will not pursue any questions as to the tacit presuppositions of Bayesian analysis. Instead, I am willing to make all concessions necessary to get the analysis going.

<sup>17</sup> Of course, a strategy may imply the same actions for all contingencies. In this case, it degenerates to a fixed course of actions.

As potential explanations of the sequence, we consider a set of *deterministic* instead of stochastic hypotheses. The basic assumption is that of a specific deterministic law describing a simple chaotic dynamic process known as baker-map dynamics. Learning takes place in a set of alternatives generated by combining the baker-map dynamics with different conjectures about an initial value or starting point for the dynamic process.

A similar example—the tent-map dynamics—has been discussed recently by Blume & Easley (1995). They consider the problem of asymptotic learning, showing that although Bayesian learning converges towards the true initial value (which is all there is to learn), predictions will not improve due to the chaotic nature of the deterministic process. The same result holds for the baker-map dynamics. It complements the no-convergence results mentioned earlier by demonstrating that model convergence does not imply expectations convergence.

However, we can get more mileage out of the machinery of chaotic dynamics. In Blume & Easley's analysis, the *true* law of the process generating the 0s and 1s is chaotic, and this law is *known* to the agent. Blume & Easley's result is based on the fact that information about the past is known to imply no restrictions for future developments. Chaotic dynamics, then, allows for cases where Bayesian learning does not improve the quality of predictions. This is an interesting point, but of course no decisive argument against Bayesianism. No procedure can improve on Bayesian learning in the case of the tent-map or the baker-map dynamics.<sup>18</sup>

The present paper is not at all concerned with the question of how Bayesian learning fails if the agent is confronted with a chaotic process. In our analysis, the question of the *true* law governing the sequence of 0s and 1s is not very interesting. The reality the Bayesian tries to come to grips with may be very simple. The chaos is in the Bayesian's head rather than outside. The problem is not the actual complexity of reality but the complexity of learning when the set of hypotheses becomes very large.

Our argument shows that Bayesian rationality becomes empty if the decision maker considers a sufficiently large set of hypotheses. The set of hypotheses described with the help of the baker map is such a large set. Whatever the real

---

<sup>18</sup> I am not implying that Blume & Easley do not agree with this evaluation.

process generating the sequence of 0s and 1s is like: Considering the baker–map dynamics as an explanation already implies that any experience can be accommodated without implications for future actions. Thus our set of hypotheses can be used to rationalize any strategy. This moreover shows that it is a very trivial task to consider such large sets. The Bayesian predicament is relevant even at a comparatively low level of sophistication.

Let me shortly explain the intuition behind the result. Note first that, as will be demonstrated, the result is trivial for finite sequences of observations. The baker–map dynamics comes into play only in the infinite case, where it can be used to formulate a dynamic system on the unit interval that produces an infinite sequence of 0s and 1s. This sequence just reveals the initial value or starting point of the system with ever increasing precision: The information provided by  $n$  observations is equivalent to the determination of a subinterval of the unit interval by  $n$  bisections. The sequence of bisections will eventually converge to the initial value. However, after any finite number of 0s and 1s, the remaining interval known to include the initial value still contains initial values that can produce any kind of extension to the sequence. In other words: the set of hypotheses represented by the set of initial values is so rich that, for any possible future, there is a hypothesis consistent with the past and predicting this future.

This is exactly the embarrassment of theoretical riches confronting an ideally rational decision maker in any context. Logically, one can never infer the laws governing the world from a finite number of past observations. While many theories may be eliminated over time, it is quite trivial that there always remain enough theories consistent with any kind of future. This is, in a nutshell, the problem of induction. If the situation were different, there would be something about the future that by logic alone could be deduced from past observations. The idea that this is impossible might be called Hume’s Impossibility Theorem (HIT). HIT implies that rational learning, if restricted to the application of deductive logic alone, yields no restrictions for decision making.<sup>19</sup> The present paper proves that Bayesian learning,

---

<sup>19</sup> The consequences of HIT for decision making are often cast into the form of another problem of induction. If the fact that logic and experience weed out too few theories is swept under the carpet, this creates a bulge—called the *pragmatic* (in contrast to the epistemic) problem of induction—one stumbles over when considering the question of which theories to use in practical



which employs more than just deductive logic, also yields no such restrictions.

This result is no surprise once it is made clear how many degrees of freedom Bayesianism leaves to the decision maker in setting up the initial beliefs. But it is surprising that a sufficiently rich set of hypotheses can be introduced in such a simple and compact way. If not much sophistication is needed to experience the problem generated by too many logical possibilities, maximum sophistication will necessarily lead to this problem. Perfect rationality involves logical omniscience (cf. Earman 1992: 121f), i. e., awareness of all logical possibilities and of all logical connections between all conceivable assumptions. Perfectly rational Bayesians have the option of restricting the set of hypotheses by choosing a prior distribution that assigns probability 0 to large subsets of hypotheses. But this requires consideration of the full set of logical possibilities, at least as a first step. Taking into account such a rich set of hypotheses, however, leads to a problem that is equivalent to the problem of predicting chaotic motion.

### **Further Proceedings**

Let us shortly summarize our line of argument. We will show that any strategy can be rationalized. Therefore, Bayesianism is empty from a normative point of view. Moreover, it is also empty from a positive point of view since without restrictions to rationalizability, past choices reveal nothing about future choices. Even if we make assumptions that allow for the existence of a set of dominated actions, nothing will ever be revealed but this set; there is no restriction on behavior that originates from rational learning. Even if we know the NM utility function of a Bayesian agent, we cannot predict behavior beyond the fact that the agent will choose no action that is dominated. So whatever we assume about the utility function, we get nothing out of the assumption that the agent processes information according to the Bayesian principles of rationality. Muth's (1961) conjecture that the theory of rational information processing can be used to solve the problem of expectations formation is thus refuted. Moreover, there is no reason to assume that Bayesian learning will ever converge to rational expectations in any sense (model or expectation convergence)

---

decision making. Cf. Musgrave (1989: section 4) and Miller (1994: 20-23, 38-45), whose solution rests on the assumption that it is possible to reduce the number of acceptable theories drastically.

even in a case where (i) the true model is exceedingly simple, (ii) there are no costs of gathering information and (iii) the true model is included among all the models considered by the Bayesian agent.

The paper proceeds as follows. Section 2 introduces a simple decision problem, which is then used to formulate precise questions. Preliminary answers to these questions are given for the finite case. Section 3 introduces the baker–map dynamics to create a set of hypotheses for the decision problem in the infinite case. The analysis of the finite case given in section 2 is extended to the infinite case. It is shown that no restriction on the choice of strategies emerges from Bayesian learning. Section 4 considers generalizations beyond the simple decision problem considered in sections 2 and 3. The last section offers some conclusions. The appendix contains proofs referred to in the text.

## 2 A Simple Decision Problem under Uncertainty

The present section introduces a simple decision problem under uncertainty. The problem is composed of an infinite sequence of subproblems, providing ample time for learning. This setup will be used to give a sharp formulation of our general problem in the form of three questions. A preliminary answer for the case of a finite sequence sketches all the main points of the general argument.

### Posing Questions

In this and the next section, we will be concerned solely with the following somewhat artificial situation.

*Adam & Eve and the Money-Spinner.* Adam the Agent owns a mysterious black box connected with a screen and a keyboard. The screen displays either 0 or 1; in fixed intervals the screen goes black and then again shows one of the digits.<sup>20</sup> The  $t$ -th observation is denoted by  $x_t$ ,  $t = 1, \dots, \infty$ . At every point in time  $t = 0, 1, \dots, \infty$ ,

---

<sup>20</sup> The set-up is equivalent to observations of consecutive tosses of a coin where there are two outcomes, heads and tails, corresponding to the outcomes of 0 and 1 in the present example. We do not use the more familiar coin-tossing example since the set of alternative hypotheses can be described more naturally in connection with the neutral black-box set-up.

Adam places his bets on the next digit by typing in a number  $y_t \in \{0, 1, \dots, Y\}$ . Doing nothing implies  $y_t = 0$ . So the sequence of events is as follows: At  $t = 0$ , Adam chooses  $y_0$ . At  $t = 1$ ,  $x_1$  appears and Adam chooses  $y_1$ . And so on. At time  $t$ , Adam knows  $\mathbf{x}^t \stackrel{\text{def}}{=} (x_1, \dots, x_t)$  and chooses  $y_t$  in the light of this information. This applies also to  $t = 0$ , since we let  $\mathbf{x}^0$  denote “no information”. If  $x_{t+1} = 0$ , the black box produces  $y_t$  perfect one-dollar notes. If  $x_{t+1} = 1$ , it produces  $Y - y_t$  equally perfect one-dollar notes. Adam is risk averse and strictly prefers more money to less at every point in time. His NM utility function  $\nu: \mathbb{R}_+ \mapsto \mathbb{R}$  at each point in time is increasing and strictly concave in money. For the purposes of numerical examples, we will choose  $\nu(y) = \ln(1 + y)$  and  $Y = 10$ . Adam’s information at time  $t$  encompasses all the facts just explained, all past observations  $\mathbf{x}^t$ , and all his past choices  $\mathbf{y}^t \stackrel{\text{def}}{=} (y_0, \dots, y_{t-1})$ , where  $\mathbf{y}^0$  again denotes “no information”.

Eve the Economist observes Adam. Her information at time  $t$  coincides with Adam’s; specifically, she knows his NM utility function.

In order to simplify the problem, we assume that both Adam and Eve know (i. e., rightly believe) that Adam’s choices have no influence on the sequence of digits appearing on the screen.  $\square$

The following three questions define the problem we are interested in by recourse to the situation of Adam and Eve.

**Question 1** *Can Eve exclude some logically possible sequences  $(\mathbf{x}^t, \mathbf{y}^t)$  as incompatible with the assumption that Adam is an ideally rational Bayesian agent?*

**Question 2** *Can Eve, after observing a sequence  $\mathbf{x}^t$ , give good advice to Adam from a Bayesian point of view?*

**Question 3** *Is there any restriction on the system of Eve’s conditional probabilities  $\Pr(y_{t+1} = i | (\mathbf{x}^{t+1}, \mathbf{y}^t))$ ,  $i = 0, 1, \dots, Y$  resulting from the hypothesis that Adam is an ideally rational Bayesian agent?*

Question 1 concerns Bayesianism as a positive theory that should yield predictions of Adam’s behavior. Question 2 concerns Bayesianism as a normative theory that could be used by Eve to advise Adam. Question 3 concerns Bayesianism as a methodology used by Eve to analyze the behavior of Adam.

Of course, all three questions are strongly interrelated. We will show that Adam can rationalize any choice of strategy. It follows that question 1 must be answered in the negative. The same goes for question 2. If any strategy can be rationalized, there is nothing a Bayesian advisor can say except “Do what you want”. The answer to question 3 is slightly more involved. There is a difference between questions 1 and 3. A negative answer to question 1 implies that no OMTs exist. But a Bayesian could still claim that Bayesianism as a methodology allows to conclude that certain sequences  $(\mathbf{x}^t, \mathbf{y}^t)$  become very improbable if Adam is rational. If that were possible, it would provide a Bayesian argument against Samuelson’s definition of “operationally meaningful”. A negative answer to question 3 presupposes that we demonstrate that Eve’s system of conditional probabilities for Adam’s actions is arbitrary.<sup>21</sup>

For all three questions, it is presupposed that Adam and Eve, the observed and the observer, argue from certain knowledge of Adam’s utility function. In this respect, we grant more knowledge than usually is available to observers, but this is necessary in case of question 2 and harmless otherwise. In another respect, we are not so liberal: Eve starts with no knowledge concerning Adam’s beliefs. One may argue that this is an unrealistic assumption. A real-world economist may actually know something about the beliefs of real-world agents. I do not want to deny this. However, knowledge about the beliefs of one’s fellow humans certainly is the product of learning. One is not born with the knowledge that other people do not (or do) believe in witches. The question is whether such knowledge will result from Bayesian learning.

## Nuts and Bolts

Let us shortly summarize the Bayesian analysis of Adam’s problem. The state space of observables for Adam’s problem is the set of all infinite sequences of 0s and 1s:  $\mathcal{S}_O \stackrel{\text{def}}{=} \{0, 1\}^\infty$ . Bayesianism requires that Adam chooses a probability measure  $\lambda$  on

---

<sup>21</sup> Actually, we could extend the problem by letting Eve bet on Adam’s behavior. This would enable us to consider Eve’s actions instead of her system of beliefs. However, this yields no further insights.

(a  $\sigma$ -algebra of subsets of)  $\mathcal{S}_O$ .<sup>22</sup> Let  $\mathbf{x}^t$  be a potential sequence of  $t$  observations. Since  $\lambda$  must be chosen such that Adam can solve his decision problem, we know that it has to imply conditional probabilities for the events  $x_{t+1} = 0$  and  $x_{t+1} = 1$ , denoted by  $\Pr(x_{t+1} = i | \mathbf{x}^t)$ ,  $i = 0, 1$ . The unconditional probabilities  $\Pr(x_1 = i | \mathbf{x}^0)$  are included in this notation. The probabilities  $\Pr(x_{t+1} = i | \mathbf{x}^t)$  yield Adam's expectations concerning the next observation if he actually observes  $\mathbf{x}^t$ . Technically, Adam revises his prior probability measure  $\lambda$ , or learns in the Bayesian sense of the term, by conditioning on the subset on  $\mathcal{S}_O$  identified by past observations. The result of learning at time  $t$  are the conditional probabilities for all potential future observations  $\Pr(x_{t+n} = i | \mathbf{x}^t)$ ,  $n = 1, \dots, \infty$ , where  $\mathbf{x}^t$  are the actual past observations.<sup>23</sup>

A problem results if Adam observes a sequence  $\mathbf{x}^t$  that has subjective probability 0 under his prior. In this case, he is free to choose a new prior distribution. We will

---

<sup>22</sup> According to the Bayesian approach of Savage (1954), subjective probabilities and the NM utility function are determined jointly. However, as Binmore (1993: 207) and Aumann (1987: 13 fn. 13) point out, it is one of the fundamental aspects of Bayesianism that preferences and beliefs can be separated (although they share a comparable status, see the remarks by Kiefer & Nyarko referred to above). This implies that an agent can be sure as to his NM utility function without being clear about what his beliefs should be. See also the remarks of Binmore (1993: 207) on Savage's position and on "massaging the priors".

<sup>23</sup> Note that Adam does not know enough in order to identify a set of dominated actions or strategies. As already argued, the identification of dominated sets depends on knowledge. For example, if Adam additionally knew that there never will be more than three 1s in a row, there are clearly many dominated strategies. On the other hand, we could modify the decision problem such that Adam's knowledge as specified so far is sufficient to identify dominated strategies. Consider the following modification: Adam can type in two numbers,  $y_{t0}$  and  $y_{t1}$ , where  $(y_{t0}, y_{t1}) \in \{0, 1, \dots, 10\}^2$  and  $y_{t0} + y_{t1} \leq 10$ . The number of one-dollar notes he gets is  $y_{ti}$  if  $x_{t+1} = i$ . Moreover, let Adam have the strictly convex utility function of a risk lover. Then the only non-dominated choices are  $(y_{t0}, y_{t1}) = (0, 10)$  and  $(y_{t0}, y_{t1}) = (10, 0)$ . However, this results from the assumption that Adam knows the sequence of digits  $x_t$  to be independent from his choices. There is always a degree of uncertainty that does not allow for the identification of dominated strategies. In the modified example, the best action becomes  $(y_{t0}, y_{t1}) = (9, 1)$  if Adam believes that  $x_{t+1} = 0$  if and only if  $y_{t1} = 10$ . So once we do not grant the knowledge of independence between the digits on the screen and Adam's choices, there will again be no dominated set. The independence assumption allows for a simpler presentation because the observable events  $x_t$  already form a state space, while a potential influence of choices on the sequence of 0s and 1s requires a state space of hidden variables explaining such potential influences.

be able to avoid this problem (see p. 19 below), which in any case only increases the difficulties of predicting Adam's choices.

Bayesian rationality, as traditionally received, requires that Adam chooses the probability measure on  $\mathcal{S}_O$  in an indirect way. He should choose a set  $\mathcal{H}$  of mutually exclusive hypotheses, each of which implies objective probabilities for all potential future observations. Let  $I$  be a set indexing the hypotheses, i. e.,  $\mathcal{H} = \{H_\theta\}_{\theta \in I}$ . Then the probabilities specified by hypothesis  $H_\theta$  will be denoted by  $\Pr(x_t = i | H_\theta)$  and  $\Pr(\mathbf{x}^t | H_\theta)$ .<sup>24</sup> Furthermore, he has to choose a subjective probability measure  $\mu$  on (a  $\sigma$ -algebra of subsets of)  $\mathcal{H}$ . The set of hypotheses  $\mathcal{H}$  together with the probability measure  $\mu$  implies a prior probability distribution on the state space  $\mathcal{S}_O$ . However, we consider another, extended state space  $\mathcal{S}_U \stackrel{\text{def}}{=} \mathcal{S}_O \times I$ , where each element is a pair consisting of an infinite sequence of observations and a number denoting a hypothesis from  $\mathcal{H}$ . The subjective probability measure  $\mu$  on  $\mathcal{H}$ , together with the objective probabilities specified by the hypotheses in  $\mathcal{H}$ , implies a probability measure on the extended state space  $\mathcal{S}_U$ . Given the subjective probability measure on the set of hypotheses, we can compute subjective probabilities for events, notably the unconditional probabilities<sup>25</sup>

$$(1) \quad \begin{aligned} (a) \quad \Pr_\mu(x_t = i | \mathcal{H}) &\stackrel{\text{def}}{=} \int_I \Pr(x_t = i | H_\theta) d\mu(\theta) \\ (b) \quad \Pr_\mu(\mathbf{x}^t | \mathcal{H}) &\stackrel{\text{def}}{=} \int_I \Pr(\mathbf{x}^t | H_\theta) d\mu(\theta). \end{aligned}$$

$\Pr_\mu(x_t = i | \mathcal{H})$  is the subjective probability for  $x_t = i$  before any observations have been made;  $\Pr_\mu(\mathbf{x}^t | \mathcal{H})$  is the corresponding subjective probability for a sequence of  $t$  observations.

---

<sup>24</sup> Some Bayesians reject any notion of objective probability; cf. Hacking (1990: 173f) on different views. Such strict subjectivists will admit only deterministic hypotheses, and they have to translate the joint assumptions of absence of uncertainty and presence of risk (typical of the REH) into the Common Prior Assumption, as it is done by Aumann (1987: 12ff). In order to simplify the exposition, we admit only objective probabilities on the level of hypotheses. Nothing will depend on this, however. Strict subjectivists can easily restate all the arguments in their language. Only when we assume the true hypothesis to be stochastic (in section 4), strict subjectivists can no longer follow; they are not in the target group of the respective argument.

<sup>25</sup>  $\int_A f(\theta) d\mu(\theta)$  denotes the integral of  $f(\theta)$  w. r. t. the probability measure  $\mu$  taken over  $\theta \in A \subseteq I$ . Note that, despite this notation,  $\mu$  is a set function.

Again, each observation identifies a subset of the extended state space  $\mathcal{S}_U$ , and again learning proceeds by conditionalizing on this subset. The difference to the case without hypotheses is that changes in the probabilities of the observable states can be traced back to changes in the probabilities of the hypotheses. The conditional probability for subsets of hypotheses indexed by  $A \subseteq I$  can be defined as

$$(2) \quad \mu(A | \mathbf{x}^t) \stackrel{\text{def}}{=} \frac{\int_A \Pr(\mathbf{x}^t | H_\theta) d\mu(\theta)}{\Pr_\mu(\mathbf{x}^t | \mathcal{H})},$$

which allows to define conditional probabilities for the next observation:<sup>26</sup>

$$(3) \quad \Pr_\mu(x_{t+1} = 0 | \mathcal{H} \wedge \mathbf{x}^t) \stackrel{\text{def}}{=} \int_I \Pr(x_{t+1} = 0 | H_\theta) d\mu(\theta | \mathbf{x}^t)$$

Adam knows that there is no influence from choice at one point in time to consequences at other points in time. The only connection between choices is learning. Given the obtaining beliefs, each act of choice can be considered separately from the others. Therefore, Adam is maximizing his subjectively expected utility at each point in time  $t$ , solving the problem

$$(4) \quad \max_{y_t} \{p_t \cdot \nu(y_t) + (1 - p_t) \cdot \nu(Y - y_t) : y_t \in \mathbf{N}_0, Y - y_t \geq 0\},$$

where  $p_t = \Pr_\mu(x_{t+1} = 0 | \mathcal{H} \wedge \mathbf{x}^t)$ .

Basically, we want to show that Adam can rationalize in advance any betting strategy, i. e., any sequence of contingent choices. A rationalization can obviously be found if it possible to fix the posterior probabilities in advance, before any observations are actually made.<sup>27</sup> More precisely: A rationalization can always be found if for any predetermined infinite sequence of probabilities  $p_t$ ,  $t = 0, \dots, \infty$ , there exists a set of hypotheses  $\mathcal{H}$  and a subjective probability measure  $\mu$  on  $\mathcal{H}$  such that

$$(5) \quad \Pr_\mu(x_{t+1} = 0 | \mathcal{H} \wedge \mathbf{x}^t) = p_t \quad \text{for all } \mathbf{x}^t \in \{0, 1\}^t \text{ and } t = 0, \dots, \infty$$

<sup>26</sup> “ $\wedge$ ” denotes logical conjunction (“and”). Propositions or sets of propositions are either given a name like  $\mathcal{H}$ , or some obvious notation is used as in the case “ $\mathbf{x}^t$ ”, which in a conjunction like “ $\mathcal{H} \wedge \mathbf{x}^t$ ” stands for the  $t$  propositions “At time  $s$ ,  $\mathbf{x}_s$  is observed”,  $s = 1, \dots, t$ . “ $\mathcal{H} \wedge \mathbf{x}^0$ ” is of course equivalent to “ $\mathcal{H}$ ”.

<sup>27</sup> This is of course a sufficient, not a necessary condition.

Such a set of hypotheses and attached prior distribution will be called a rationalization of the betting strategy. The analysis will be simplified by the assumption that a sequence of bets, if it can be rationalized at all, can be rationalized without assuming any probabilities  $\Pr_\mu(x_{t+1} = i | \mathcal{H} \wedge \mathbf{x}^t)$  to be 0. This is the case because  $y_t$  is a discrete variable.<sup>28</sup> Actually, all variables figuring in realistic decision problems should be discrete, even time. You cannot plan to meet a friend at  $\pi$  past eleven because there is no clock that is infinitely precise. In general, there are no measurement processes that could ever determine a real number with infinite precision. Given discreteness, any decision that is optimal if some event has zero probability will also be optimal if the probability of the respective event is small enough. Therefore it is never necessary to rationalize decisions by setting certain probabilities to zero.

### Approaching Answers

Let us turn to the task of finding a set of hypotheses that could be used to rationalize every betting strategy. It is instructive to begin with a set that will not do, for instance, the set of hypotheses described by the condition that 0 and 1 always alternate ( $x_{t+1} \neq x_t$  for all  $t$ ). It contains two hypotheses, one implying  $x_1 = 0$ , the other  $x_1 = 1$ . This set is obviously not rich enough to yield a rationalization for every betting strategy since already the first observation fixes the probabilities for all further events at either 0 or 1. Given the NM utility function, this uniquely determines the rational betting strategy after one observation. What is required for our purposes is a set rich enough for Adam to adjust his beliefs to any past

---

<sup>28</sup> It is moreover necessary to assume that the utility is real-valued. If the utility functions can take on values in the extended reals  $[-\infty, +\infty]$ , discreteness is not sufficient for this result. We then have to admit conditional probabilities of 0 or 1. This makes it necessary to consider the problem of how Adam proceeds if a zero-probability event occurs. Kiefer & Nyarko (1995: 49) provide the arch-Bayesian answer: Adam should choose in advance new prior distributions that will be used after such an event. This gives enough degrees of freedom to generate all our results. However, since this case yields no additional insights, we restrict the analysis to real-valued utility functions. Note that Kiefer & Nyarko consider restrictions on the new priors (“proper conditional probabilities”). On the general question of constraining the choice of priors, see section 4 on “rational priors”.



without thereby restricting his probability assignments to future events, whatever these assignments are. The set described by  $x_{t+1} \neq x_t$  is much too small to fulfill this requirement.

Sets of hypotheses that are rich enough are provided by deterministic processes with chaotic behavior. Most of the paper will be concerned with such a potential explanation of the inner workings of the box. In this subsection, however, we prepare the ground by analyzing a very trivial case. We assume that Adam, instead of having to choose infinitely often, has to choose only twice. This allows for a simple description of a sufficiently rich set of hypotheses.

We start with Adam's state space  $\mathcal{S}_O \stackrel{\text{def}}{=} \{0, 1\}^2$ , conveniently described by a cross table:

		$X_2$		
		0	1	$\Sigma$
$X_1$	0	$p_{00}$	$p_{01}$	$\Pr(X_1 = 0)$
	1	$p_{10}$	$p_{11}$	$\Pr(X_1 = 1)$
	$\Sigma$	$\Pr(X_2 = 0)$	$\Pr(X_2 = 1)$	1

$p_{ij} = \Pr(X_1 = i \wedge X_2 = j)$  factors into  $\Pr(X_2 = j | X_1 = i) \cdot \Pr(X_1 = i)$ ,  $i, j = 0, 1$ . The distribution for  $X_1$  and the two conditional distributions for  $X_2$  can be chosen independently from each other, without any restrictions. If Bayesianism does not require more than that some such distribution is chosen, it is empty, since then Adam can rationalize any contingent plan. For example, he may decide to choose  $y_0 = Y$  and, independently from the first observation,  $y_1 = 0$ . This can be achieved by setting  $p(X_1 = 0)$ ,  $p(X_2 = 1 | X_1 = 1)$  and  $p(X_2 = 1 | X_1 = 1)$  all smaller than  $\frac{12}{100}$ . The following values do the trick:

		$X_2$		
		0	1	$\Sigma$
$X_1$	0	0.855	0.095	0.95
	1	0.0445	0.0055	0.05
	$\Sigma$	0.8995	0.1005	1

The question is whether the Bayesian requirement of blowing up the state space

$\mathcal{S}_O$  by formulating some hypotheses yields restrictions. Quite obviously, this is not the case for our simple example. It is useful, however, to go through the argument since this delivers the clue to the extension to infinite sequences of observations.

Let us describe the four possible outcomes of the two observations by a random variable  $\chi(X_1, X_2) \stackrel{\text{def}}{=} \frac{X_1}{2} + \frac{X_2}{4}$  taking on values in  $\{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$ . Let us assume that Adam considers a set  $\mathcal{H}$  of four hypotheses  $H_\theta$ , where for reasons that will become apparent we index the hypotheses by  $\theta \in \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$ . Assuming the probabilities given above, we have the following state space  $\mathcal{S}_U$ , where the state space  $\mathcal{S}_O$  now occurs as the last column:

		$\theta$				
		0	1/4	1/2	3/4	$\Sigma$
$\chi$	0	$q_{00}$	$q_{01}$	$q_{02}$	$q_{03}$	0.855
	1/4	$q_{10}$	$q_{11}$	$q_{12}$	$q_{13}$	0.095
	1/2	$q_{20}$	$q_{21}$	$q_{22}$	$q_{23}$	0.0445
	3/4	$q_{30}$	$q_{31}$	$q_{32}$	$q_{33}$	0.0055
$\Sigma$		$\Pr(H_0)$	$\Pr(H_{\frac{1}{4}})$	$\Pr(H_{\frac{1}{2}})$	$\Pr(H_{\frac{3}{4}})$	1

The probabilities  $q_{ji}$  factor into  $\Pr(\chi = j | H_\theta) \cdot \Pr(H_\theta)$ . The requirement to generate the probabilities of the state space  $\mathcal{S}_O$  by choosing a distribution over the hypotheses in  $\mathcal{H}$  can be fulfilled in many ways; the most trivial is to consider the following deterministic hypotheses:  $H_0$  and  $H_{\frac{3}{4}}$  state that the screen will always show 0 or 1, respectively.  $H_{\frac{1}{4}}$  and  $H_{\frac{1}{2}}$  state that 0 and 1 always alternate, i.e.,  $x_{t+1} \neq x_t$ , where  $H_{\frac{1}{4}}$  assumes  $x_1 = 0$  and  $H_{\frac{1}{2}}$  assumes  $x_1 = 1$ . Translating these hypotheses concerning  $x_t$  into hypotheses concerning  $\chi$ , we find that  $H_\theta$  implies  $\chi = \theta$ . Thus  $q_{ij}$  is 0 for  $i \neq j$ . The probabilities  $\Pr(H_0) = 0.855$ ,  $\Pr(H_{\frac{1}{4}}) = 0.095$ ,  $\Pr(H_{\frac{1}{2}}) = 0.0445$  and  $\Pr(H_{\frac{3}{4}}) = 0.0055$  form the unique prior over  $\mathcal{H}$  that generates the specified marginal distribution over  $\mathcal{S}_O$ .

Obviously, any betting strategy for finitely many bets has such a trivial rationalization. For a sequence of  $T$  observations, there are  $2^T$  points in  $\mathcal{S}_O$  corresponding to  $2^T$  possible values of  $\chi$ . Let  $I$  be the set of values for  $\chi$ . Specify a hypothesis  $H_\theta$  for every point in  $\mathcal{S}_O$  that predicts that  $\chi = \theta$  will be realized. Only the probabilities on the main diagonal of  $\mathcal{S}_U$ , then, can be different from 0. A probability distribution

over  $\mathcal{S}_O$  fixes the probabilities  $\Pr(\chi = \theta)$ . Any such distribution can be rationalized by setting  $\Pr(H_\theta) = \Pr(\chi = \theta)$ . The deterministic hypotheses form the corners of a  $2^T$ -dimensional simplex. The probability distribution over  $\mathcal{S}_O$  is a point on the simplex. Each such point results from a unique convex combination of the corners. The deterministic hypotheses form a basis for all these distributions.

Of course, it is possible to additionally consider stochastic hypotheses. These hypotheses correspond to additional points in the above-mentioned simplex. Since the deterministic hypotheses already form a basis, the inclusion of stochastic hypotheses will make the distribution over the set of hypotheses non-unique in many cases. It is then possible to get rid of some of the deterministic hypotheses. In fact, since we can rationalize any choices without recourse to degenerate probabilities, we could restrict considerations to distributions on  $\mathcal{S}_O$  in the interior of the simplex. For such distributions, we can always find any number of stochastic hypotheses that rationalize the distribution. In the extreme case, we take just a single hypothesis rationalizing expectations in terms of objective probabilities. However, for any set of hypotheses that does not include all the deterministic hypotheses, there are always prior distributions over the state space  $\mathcal{S}_O$  that cannot be rationalized. So the set of all the deterministic hypotheses is the minimal set that is always rich enough to generate all priors over  $\mathcal{S}_O$ .

In the finite case, a rationalization is a trivial matter. The question is whether this is different once we consider an infinite sequence of observations and actions. This is where chaos comes in. We are going to show in the next section that the baker-map dynamics provides a set of hypotheses that is rich enough to generate all possible distributions on the state space. There will be a strong analogy to the finite case as analyzed in this section. One could of course argue that it is enough to concentrate on the finite case since every real agent can only take a finite number of decisions. However, much of the literature and many models in economics are concerned with asymptotic properties. Therefore, we will extend the analysis to infinite sequences in order to make clear that Bayesian rationality is empty even asymptotically.

### 3 Chaotic Beliefs

#### A Chaotic Clock

The following set of hypotheses describes the evolution of the inner states of Adam's black box as a deterministic process depending on an initial value. The law of the deterministic process is the baker-map dynamics, which can be graphically illustrated as the output of a chaotic clock (fig. 1). The clock has only one pointer; the dial shows the real numbers in the interval  $I = [0, 1)$ . Initially, the pointer deviates by an angle  $\omega = 2\theta\pi$  from the 0 position, thus pointing at  $\theta$ . At  $t = 1, 2, \dots, \infty$ , the pointer moves by doubling the angle  $\omega$ . Adam's black box shows 0 if and only if the pointer is in the first half of the dial.

Insert fig. 1 from the end of the paper!

According to the chaotic-clock hypothesis, the inner states of the black box at  $t = 0, 1, \dots, \infty$  are described by a real variable, the pointer position  $z_t \in I$ . These inner states evolve deterministically, but it can only be observed whether the pointer position is  $z_t \in [0, \frac{1}{2})$  or  $z_t \in [\frac{1}{2}, 1)$ . These states are indicated by  $x_t = 0$  and  $x_t = 1$ , respectively. The deterministic law by itself does not allow for a prediction of future observations; an assumption concerning the starting point or initial value  $z_1 = \theta$  is also necessary. Thus there is a set  $\mathcal{H}^*$  of chaotic-clock hypotheses  $H_\theta$ , one for each initial value  $\theta \in I$ . The dynamical system corresponding to  $\mathcal{H}^*$  is<sup>29</sup>

$$(6) \quad \begin{aligned} (a) \quad x_t &= g(z_t) \stackrel{\text{def}}{=} 2 \cdot z_t \text{ div } 1 \\ (b) \quad z_{t+1} &= h(z_t) \stackrel{\text{def}}{=} 2 \cdot z_t \text{ mod } 1 \\ (c) \quad z_1 &= \theta. \end{aligned}$$

Note that we will use the chaotic clock again to generalize our results to the case of a finite universe: *The chaotic clock provides a universal theory that is capable*

---

<sup>29</sup> "div" denotes integer division; "mod" denotes the indivisible rest of the integer division, i. e.,  $x \text{ mod } n \stackrel{\text{def}}{=} x - (x \text{ div } n)$ . On the baker-map dynamics, see Ford (1983), Devaney (1989: 18 example 3.4, 39, 52) and Schuster (1988: 107f). The graphical illustration is due to Davies (1987: ch. 4). On coin tossing and dyadic development see also Brémaud (1988: 28-31).

of explaining everything. Of course, this kind of explanation would be completely uninteresting from a scientific perspective. However, this is beside the point. The significance of the chaotic clock lies in the fact that it leads to the typical problems connected with a very rich set of hypotheses.

In order to analyze the consequences of uncertainty concerning  $\theta$ , Adam has to know how  $x_t$  develops for a given  $\theta$ . This is very simple in principle. Every  $\theta \in I$  can be expressed in binary form as

$$\theta = \sum_{n=1}^{\infty} \frac{\theta_n}{2^n} = 0.\theta_1\theta_2\theta_3 \dots$$

where  $\theta_n$  is 0 or 1 (dyadic development). We adopt the rule that the representation must have infinitely many 0s. Thus  $\frac{1}{2}$  should be represented as  $0.1\bar{0}$  and not as  $0.0\bar{1}$ , where a bar over a digit denotes infinite repetition. We use  $\bar{0}$  as a reminder, since this sequence of 0s has a meaning in our context. Moreover, in order to avoid cumbersome notation, we use  $\bar{1}$  in expressions like  $[0.a\bar{0}, 0.a\bar{1})$  to denote the upper boundary of an interval.

Every real number has a unique dyadic development with infinitely many 0s. The sequence generated by the chaotic clock is just the dyadic development of the starting point:

$$(7) \quad \begin{aligned} \theta &= 0.\theta_1\theta_2\theta_3 \dots \theta_t \dots \\ x_t &= \theta_t \end{aligned}$$

This follows from two mathematical facts: (i) We have  $\theta \geq \frac{1}{2}$  if and only if  $\theta_1 = 1$ . (ii) Doubling the angle  $\omega = 2\theta\pi$  shifts all  $\theta_i$  to the left by one position except for  $\theta_1$ , which is eliminated since an angle of  $2\pi$  is interpreted as an angle of 0.

Every sequence of 0s and 1s with an infinite number of 0s represents a starting point. For finite sequences, the deterministic law can generate any sequence whatsoever. Any restrictions on sequences and, correspondingly, on beliefs could only derive from considering infinite sequences.

Why should Adam ever consider a set of hypotheses like  $\mathcal{H}^*$ ? The point is that all these hypotheses represent logical possibilities and that Adam, being perfectly rational, is logically omniscient. He therefore is aware of all these (and many more) hypotheses and their implications. Of course, he may throw out many hypotheses in the process of selecting priors. But in doing so, he is aware of the fact—which

will be demonstrated in the next subsection—that he can rationalize any betting strategy by choosing suitable priors. The selection of priors on such a large set of hypotheses will be seen to be equivalent to the problem of predicting chaotic motion.

It is a further important point that  $\mathcal{H}^*$  fulfills all the formal requirements of a scientific theory: It assumes a simple mechanism governed by a law of motion that produces different results according to the initial position of the mechanism. While not even the Swiss could actually produce the chaotic clock (*exactly* doubling an angle requires infinite precision), physical processes that lead to chaotic dynamics are not rare, and imperfect observability can produce the kind of irregular behavior characteristic of the chaotic clock.

We go on to show that the system of conditional probabilities  $\Pr(x_{t+1} = 0 | \mathbf{x}^t)$  can be chosen arbitrarily and still be derived from a prior probability measure on  $\mathcal{H}^*$ . This implies that any sequence of actions can be rationalized *in advance*, i. e., before any observations are made, since one of the possible assignments always is  $\Pr(x_{t+1} = 0 | \mathbf{x}^t) = p_t$  for all  $\mathbf{x}^t$ , which fixes the posterior probabilities. For such an assignment, Bayesianism is just a fig leaf, allowing Adam to pass predetermined values off as the result of learning. Of course, Adam can also allow for any influence of observations on his behavior instead. He can choose freely between any sequence of actions, conditional on the past or completely predetermined or any mixture of both. There is, of course, one restriction on his system of beliefs: At no point in time, Adam can assign a positive probability to the event  $F_0$  that the number of 0s will be finite. Or in other words: Adam has to believe that it is impossible that the sequence of observations will end in an unbroken infinite sequence of 1s. This restriction just results from the fact that the baker-map dynamics is unable to produce an unbroken infinite sequence of 1s: For  $x_n = 1$  for all  $n \geq t$ , the pointer of the chaotic clock has to point to 1 at time  $t$ , which is impossible since 1 is not on the dial. We will denote this restriction by  $\Pr(F_0 | \mathcal{H}^*) = 0$ , since it holds independently from the subjective probability measure on  $\mathcal{H}^*$ . At the end of this section, we will show that even this restriction on beliefs can easily be removed.

The following subsection is slightly more technical than previous explanations. It is possible to skip it without losing the thread of the argument.

## The Dynamics of Belief

We turn to the dynamics of beliefs and predictions generated by the chaotic-clock hypotheses  $\mathcal{H}^*$ . Since these hypotheses are deterministic, the probabilities  $\Pr(x_t = 0 | H_\theta)$  are either 0 or 1, i. e.,  $H_\theta$  yields certain or point predictions. Uncertainty enters via the uncertainty concerning  $\theta \in I$ . As a Bayesian, Adam chooses a subjective probability measure on (a  $\sigma$ -algebra of subsets of)  $I$ . The posterior probabilities  $\Pr(H_\theta | \mathbf{x}^t)$  of the hypothesis  $H_\theta$  after  $t$  observations is 0 if the first digits in the dyadic development of  $\theta$  do *not* agree with the observations. It is important to understand how subsets of  $\{H_\theta\}$  are eliminated or falsified by successive observations.

We use the following trivial connection between dyadic expansions and bisections. For each  $\theta \in I$ , there exists a unique sequence of intervals (henceforth called basic intervals) of the form

$$(8) \quad I_t(m) \stackrel{\text{def}}{=} \left[ \frac{m}{2^t}, \frac{m+1}{2^t} \right), \quad m \in \{0, \dots, 2^t - 1\}$$

with the following properties: (a) The sequence starts with  $I$ . (b) Any other interval in the sequence is generated by bisecting its precursor and selecting either the upper or the lower interval resulting from the bisection as member of the sequence. (c) The intersection of the intervals converges to  $\theta$ .<sup>30</sup> There is a unique sequence of such bisections converging to a given  $\theta$  since the basic intervals are open to the right. In each step of the bisection, write down the digit 0 if the lower and the digit 1 if the upper of the two intervals generated by the bisection is part of the sequence. The sequence of 0s and 1s so generated is just the dyadic development of  $\theta$ . The reverse also holds: The dyadic expansion of a number in  $I$  determines the sequence of bisections of  $I$  converging to this number.

Now compare the dyadic development of starting points with the observations. To any sequence  $\mathbf{x}^t$  of  $t$  observations corresponds a number  $\chi = 0.x_1x_2 \cdots x_t$  in dyadic expansion and therefore a sequence of bisections. These bisections agree with the bisections generated by the dyadic expansion of the starting point. The

---

<sup>30</sup> For definitions and some facts concerning the basic intervals and bisections see the respective section of the appendix. Fig. 3 in the appendix illustrates the definition of the basic intervals  $I_t(m)$ .

latter is unknown; however, the observations  $\mathbf{x}^t$  reveal the first  $t$  digits of its dyadic expansion and therefore the first  $t$  bisection steps. If  $\chi \in I_t(m)$ , Adam knows that  $\theta \in I_t(m)$ .

If Adam wants to have well-defined probabilities  $\Pr_\mu(x_{t+1} = i | \mathbf{x}^t)$  for any potential sequences of observation  $\mathbf{x}^t$ , he should include all the basic intervals into his  $\sigma$ -algebra of subsets of  $I$ . We are going to show that choosing such a  $\sigma$ -algebra yields no restrictions on his decisions. Therefore, we can safely assume that  $\mu(I_t(m))$  is well-defined for all  $m$  and  $t$ .

The foregoing analysis established a one-to-one relation between a potential sequence of observations  $\mathbf{x}^t$  and basic intervals  $I_t(m) = [0.x_1x_2 \cdots x_t\bar{0}, 0.x_1x_2 \cdots x_t\bar{1})$ . The prior probability of the sequence  $\mathbf{x}^t$  is the probability of the interval  $I_t(m)$ . Thus instead of  $\Pr_\mu(\mathbf{x}^t | \mathcal{H})$ , we use

$$(9) \quad p_t(m) \stackrel{\text{def}}{=} \mu(I_t(m)).$$

As already mentioned, the posterior probabilities of the hypotheses  $H_\theta$  with  $\theta \notin I_t(m)$  are 0: They have been falsified by the observations. Given that Adam knows that  $\theta \in I_t(m)$ , he can easily compute the posterior probability of  $x_{t+1} = 0$ . The next observation will be 0 (1) if and only if the starting point  $\theta$  is in the lower (upper) of the two subintervals  $I_{t+1}(2m)$  and  $I_{t+1}(2m+1)$ . Thus we find

$$(10) \quad \begin{aligned} (a) \quad \Pr_\mu(x_{t+1} = 0 | \mathcal{H}^* \wedge \mathbf{x}^t) &= \frac{p_{t+1}(2m)}{p_t(m)} \\ (b) \quad \Pr_\mu(x_{t+1} = 1 | \mathcal{H}^* \wedge \mathbf{x}^t) &= \frac{p_{t+1}(2m+1)}{p_t(m)}. \end{aligned}$$

Let us denote these conditional probabilities by

$$(11) \quad q_{t+1}(s) \stackrel{\text{def}}{=} \begin{cases} \frac{p_{t+1}(s)}{p_t\left(\frac{s}{2}\right)} & \text{iff } s \text{ even or zero} \\ \frac{p_{t+1}(s)}{p_t\left(\frac{s-1}{2}\right)} & \text{otherwise} \end{cases}$$

for  $t \geq 0$ . Then we can express the prior probability  $p_T(m)$  as a product of conditional probabilities  $q_t(s(t))$  with  $t \leq T$ :

$$(12) \quad p_T(m) = \prod_{t=1}^T q_t(s(t)).$$



We illustrate these considerations by a diagram, where a tree structure is used to derive the conditional probabilities from the prior distribution and vice versa (fig. 2). Every time-point  $t = 0, \dots, T$  is assigned a set of  $2^t$  nodes  $N_t(m)$ ,  $m = 0, \dots, 2^{t-1}$  residing in the intervals  $I_t(m)$ . The nodes and the intervals are numbered from left to right. From each node there emerge two branches, the left (right) one representing the observation  $x_{t+1} = 0$  ( $x_{t+1} = 1$ ). Every branch emerging from a node leads to a further node.

Insert fig. 2 from the end of the paper!

Every starting point  $\theta \in I$  leads to a path through the tree determined by its dyadic development, a zero at the  $j$ -th dyadic place implying that at node  $N_{j-1}(m)$  the left one of the two branches is taken, which means that the left of the two intervals resulting from the next bisection step is taken. The information concerning past observations at time  $T$  can therefore be expressed by specifying the current node or interval. Moreover, the node  $N_T(m)$  will be reached if and only if  $\theta \in I_T(m)$ . Thus the probability of reaching node  $N_T(m)$  is  $p_t(m)$  as defined in (9). This probability is the product of probabilities  $q_t(s)$  for  $t \leq T$  as expressed by (12). The probabilities  $q_t(s)$  are the probabilities of going left or right from a given node; they are therefore, as already defined above in (11), the conditional probabilities of 0 and 1 given the past connected with the node. Of course, the two probabilities  $q_t(s)$  and  $q_t(s+1)$  relevant at node  $N_{t-1}(s)$  ( $s$  even or zero) add up to 1. We may assume that no  $q_t(s)$  ever becomes 0; so the subjective probability of reaching a node may never be 0. Given a probability distribution over the unit interval, all the  $q_t(s)$  can be determined since all the  $p_t(m)$  are determined. The latter is trivial since  $p_t(m)$  is just the probability assigned to the basic interval  $I_t(m)$ .

However, we are more interested in the inverse operation. If one goes through a certain path, one in effect carries out a sequence of bisections of the unit interval. These bisections determine with ever greater precision the starting point  $\theta$  that would give rise to the path, if, as explained before, an  $x_t = 0$  ( $x_t = 1$ ) is interpreted as an instruction to take the left (right) turn at the current node. Assigning probabilities  $q_t(s)$  to all the branchings of all the possible paths determines probabilities

$p_t(m)$  for all the intervals. For finite time, this obviously results in a probability distribution over the unit interval  $I$ . There are always infinitely many densities compatible with the initial choice of the conditional probabilities. The simplest of these densities is a unique piecewise continuous density that is constant on each  $I_T(m)$ .

In the appendix it is proved that this is also true for the case of an infinite number of observations. This is not obvious. In fact, as already mentioned, it is not literally true in general but only if the restriction  $\Pr(F_0 | \mathcal{H}^*) = 0$  is fulfilled, i. e., if the probability implicitly assigned by the conditional probabilities  $q_t(s)$  to the event  $F_0$  that there will be a finite number of 0s is 0. Under this condition, the probabilities  $q_t(m)$  attached to the basic intervals  $I_t(m)$  are shown to induce a probability measure on the unit interval.

Adam as an ideally rational person is always aware of the implications of all the assumptions he is considering. When choosing the priors over  $\mathcal{H}^*$ , he is aware of the implications for the conditional probabilities  $q_t(s)$ . Thus instead of choosing a prior over  $\mathcal{H}^*$ , Adam can as well choose the conditional probabilities  $q_t(s)$ . What he does not know in advance are the conditional probabilities that will turn out to be his posterior probabilities because they are picked out among the probabilities (10) by the *actual* sequence  $\mathbf{x}^t$ .

## Implications

The result of the last subsection is summarized in an informal way in the following proposition.<sup>31</sup>

**Proposition 1** *Let  $F_0$  be the event that there will be a finite number of 0s. Let  $\mathcal{H}^*$  be the set of chaotic-clock hypotheses described by (6). Consider an arbitrary assignment of numerical values to all the conditional probabilities  $\Pr(x_{t+1} = i | \mathbf{x}^t)$ ,  $i = 0, 1$ . If and only if  $\Pr(F_0) = 0$  under this assignment, there exists a probability measure  $\mu$  on  $\mathcal{H}^*$  such that the conditional probabilities  $\Pr_\mu(x_{t+1} = i | \mathcal{H}^* \wedge \mathbf{x}^t)$  coincide with the values assigned to  $\Pr(x_{t+1} = i | \mathbf{x}^t)$ . The probability measure  $\mu$  is unique.*

---

<sup>31</sup> A formal statement of the proposition is theorem 9 in the appendix. The appendix moreover contains a proof of the theorem.

The conditional probabilities  $\Pr_\mu(x_{t+1} = i | \mathbf{x}^t)$  are the posterior probabilities if the agent actually observes  $\mathbf{x}^t$ . Proposition 1 states that Adam is almost completely free in his choice of the conditional probabilities. For example, assume that the prior is a uniform distribution with density  $f(\theta) = 1$  over  $I$ . Then the subjective probability of  $x_{t+1} = 0$  is equal to  $\frac{1}{2}$ , independently from the past observations  $\mathbf{x}^t$ .<sup>32</sup> However, we have to take one final step in order to get rid of the restriction  $\Pr(F_0) = 0$ . We introduce a very slight modification of the chaotic-clock hypothesis. We assume that the angle between the pointer and the zero position is not doubled as before but quadrupled:

$$(13) \quad \begin{aligned} (a) \quad & x_t = g(z_t) \\ (b) \quad & z_{t+1} = h(2 \cdot z_t) \\ (c) \quad & z_1 = \theta. \end{aligned}$$

The functions  $g, h$  are defined as in (6). For any  $\theta = 0.\theta_1\theta_2\theta_3 \cdots \theta_t \cdots$ , the original system (6) yields the observations  $x_t = \theta_t$ . The modified system (13) yields  $x_t = \theta_{2t-1}$ , i. e., every second digit of the dyadic representation of the starting point is irrelevant. Therefore a starting point  $0.\overline{10}$  generates an unbroken infinite number of 1s. The same is true for  $0.\overline{1011}$ . Since both dyadic representations have an infinite number of 0s, they are legitimate. The sequences generated by the modified system result from deleting every second digit from the sequences generated by the original system. This removes the restriction that the sequence has to contain an infinite number of 0s. It implies, however, that an infinite number of starting points will lead to the same sequence of observations. This creates no problems; it just means that there will be many priors that rationalize the same set of conditional probabilities.

Let us call the set of modified chaotic-clock hypotheses  $\mathcal{H}^{**}$ . Our results are then summarized informally by the following proposition.<sup>33</sup>

---

<sup>32</sup> Cf. Brémaud (1988: 29) who, however, does not discuss *subjective* probabilities or other than uniform distributions on  $I$ . Blume & Easley (1995: 19f, 36f) demonstrate for the tent-map dynamics that in the case of a density continuous at the starting point, the posterior distribution will converge to the uniform distribution. It is easy to show that the same result holds for the baker-map dynamics.

<sup>33</sup> The appendix shows that deleting every second digit still yields a probability measure on the set of all sequences.

**Proposition 2** *Let  $\mathcal{H}^{**}$  be the set of modified chaotic-clock hypotheses described by (13). Consider an arbitrary assignment of numerical values to all the conditional probabilities  $\Pr(x_{t+1} = i | \mathbf{x}^t)$ ,  $i = 0, 1$ . There exist infinitely many probability measures  $\mu$  on  $\mathcal{H}^{**}$  such that the conditional probabilities  $\Pr_{\mu}(x_{t+1} = i | \mathcal{H}^{**} \wedge \mathbf{x}^t)$  coincide with the values assigned to  $\Pr(x_{t+1} = i | \mathbf{x}^t)$ .*

Proposition 2 implies that Adam can always find a rationalization for any betting strategy. This answers question 1. Bayesian rationality is empty as a positive theory; Eve cannot exclude any sequence of actions of Adam on the basis of the hypothesis that Adam is rational in the Bayesian sense. Nor can Eve give any advice to Adam, even if she knows his NM utility function, since no sequence of actions is irrational, whatever observations have been made. This answers question 2: Bayesianism is empty as a normative theory. The answer to question 3 seems to be quite obvious: How can there be a restriction on Eve's subjective probabilities concerning Adam's actions if no sequence of actions can be excluded? However, we will provide a formal answer to question 3 that will confirm this intuition. Eve may consider Adam and the money-spinner as a big black box. Knowing that no sequence of actions can be excluded on the basis of her knowledge, she confronts a similar, although slightly more general problem as Adam. To this problem we can apply the generalization considered in the next section.

## 4 Generalizations

This section generalizes the results of the last section and ties up some loose ends, namely, the inclusion of stochastic hypotheses and the question of rational priors.

### A Theory of Everything

We turn to the variant of Eve's problem connected with question 3, i.e., to the problem of an economist trying to predict the behavior of rational agents and using Bayesianism as a methodology. Since Eve is a Bayesian herself, she is not overly impressed by the fact that Bayesianism yields no OMTs concerning Adam's behavior since the Bayesian methodology allows a continuum of beliefs between "ruled out"

and “still possible”. However, this is not going to help her. Eve’s problem is not different from Adam’s except for the number of states at each point of time. She is perfectly rational and therefore logically omniscient. Thus she confronts the set of all logically possible hypotheses, a set that is rich enough to accommodate any past without restricting the scope of expectations concerning the future. She knows that Adam is rational, but she also knows the results of the previous section, namely, that this implies no restriction on Adam’s betting strategy. Therefore she is in no better position to predict Adam than to predict the digits generated by the black box. For a formal proof, we have to generalize the results of the last section to larger spaces of observables. This presents no difficulties.

Eve’s state space of observables,  $\mathcal{S}_O$ , consists of all sequences of 0s and 1s that could conceivably be generated by Adam’s black box and of all choices open to Adam. Moreover, she might observe other things like Adam’s facial expression or his pattern of consumption that are or are not related to Adam’s betting behavior. Sticking to our premise that all observable variables can only range over a finite set of values, we assume that Eve’s observable universe can be “digitalized”: Each state can be described by a finite string of 0s and 1s. Since the number of different states is not necessarily  $2^n$  for some  $n > 0$ , we have to allow for the possibility that several strings describe the same state. With this caveat, Eve’s space of observable states is  $\mathcal{S}_O \stackrel{\text{def}}{=} \{0, 1, \dots, 2^n - 1\}^\infty$ ,  $n > 1$ .

Again, we describe a sufficiently rich set of hypotheses with the help of a chaotic clock. Eve considers Adam and the money–spinner as a big black box that displays one of  $2^n$  different configurations, i. e., combination of observables, at each point in time. The configurations are determined by a chaotic clock that makes  $n$  angle-doubling ticks at each point in time. The dial is divided into  $2^n$  sections:

$$I_n(i) \stackrel{\text{def}}{=} \left[ \frac{i}{2^n}, \frac{i+1}{2^n} \right), \quad i \in \{0, 1, \dots, 2^n - 1\}$$

Configuration  $i$  comes up if the pointer is in  $I_n(i)$ . The behavior of the black box is described by the following dynamical system:

$$(14) \quad \begin{aligned} (a) \quad x_t &= g_n(z_t) \stackrel{\text{def}}{=} 2^n \cdot z_t \text{ div } 1 \\ (b) \quad z_{t+1} &= h_n(z_t) \stackrel{\text{def}}{=} 2^n \cdot z_t \text{ mod } 1 \\ (c) \quad z_1 &= \theta \end{aligned}$$

This dynamical system works very much like (6), with the difference that the shift from  $z_t$  to  $z_{t+1}$  now concerns a string of  $n$  digits instead of 1 digit, if we look at the dyadic expansion of the  $z_t$ . Accordingly,  $x_t$  is now a natural number between 0 and  $2^n - 1$ . If we express  $x_t$  as a dyadic number, it will be of maximum length  $n$ . We include leading 0s to bring each  $x_t$  up to this maximum length. The dynamical system thus generates a sequence of strings of length  $n$ , where each string consists of 0s and 1s. The sequence generated by the chaotic clock is again the dyadic development of the starting point:

$$(15) \quad \begin{aligned} \theta &= 0.\overbrace{\theta_1 \cdots \theta_n}^1 \overbrace{\theta_{n+1} \cdots \theta_{2n}}^2 \cdots \overbrace{\theta_{(t-1)n+1} \cdots \theta_{tn}}^t \cdots \\ \dot{x}_t &= \underbrace{\theta_{(t-1)n+1} \cdots \theta_{tn}}_{t\text{-th string}} \end{aligned}$$

The new chaotic clock is completely equivalent to the old one, except that it reveals the dyadic expansion of the starting point blockwise instead of digit by digit. For this reason, the same restriction holds: The chaotic clock is unable to produce an unbroken infinite sequence of strings consisting of 1s only. The remedy is the same as before: We ignore every second digit. This leads to the following system:

$$(16) \quad \begin{aligned} (a) \quad x_t &= g_n(z_t) \\ (b) \quad z_{t+1} &= h_n(2 \cdot z_t) \\ (c) \quad z_1 &= \theta \end{aligned}$$

The set of hypotheses corresponding to the starting points  $\theta \in I$  is denoted by  $\mathcal{H}_n^{**}$ . It provides a “theory of everything” for universes that can be digitalized at every point in time but are infinite in time. Since the possibilities of assigning probabilities to sequences are not affected by the fact that these sequences are now revealed in a blockwise fashion, the previous results still apply, and we can generalize proposition 2 to the case of arbitrarily large but finite universes.

**Proposition 3** *Let  $\mathcal{H}_n^{**}$  be the set of modified chaotic-clock hypotheses described by (16). Consider an arbitrary assignation of numerical values to all the conditional probabilities  $\Pr(x_{t+1} = i | \mathbf{x}^t)$ ,  $i = 0, 1, \dots, 2^n - 1$ ,  $n \geq 1$ . There exist in-*

finitely many probability measures  $\mu$  on  $\mathcal{H}_n^{**}$  such that the conditional probabilities  $\Pr_\mu(x_{t+1} = i | \mathcal{H}_n^{**} \wedge \mathbf{x}^t)$  coincide with the values assigned to  $\Pr(x_{t+1} = i | \mathbf{x}^t)$ .

This answers question 3. Bayesianism as a methodology is completely useless in predicting rational behavior because there are no OMTs covering this behavior. Eve's expectations concerning Adam's behavior will be completely arbitrary. Moreover, the analysis of this section shows that the restriction of Adam's problem to predicting single digits is immaterial. Everything works as before as long as he considers a universe the observable aspects of which are finite.

### Stochastic Hypotheses and Non-Convergence

In section 2, we have already seen that for a finite sequence of observations, nothing changes if we additionally include stochastic hypotheses among a sufficiently rich set of deterministic hypotheses. We now extend this result to infinite sequences. We avoid the notational complications raised by large universes and go back to Adam's problem. It will be obvious that the results generalize.

The observable state space is again  $\mathcal{S}_O = \{0, 1\}^\infty$ . The most general form of stochastic hypothesis Adam could consider specifies conditional *objective* probabilities  $\Pr(x_{t+1} = i | \mathbf{x}^t)$ ,  $i = 0, 1$  for all potential sequences  $\mathbf{x}^t$ . We already know that from a Bayesian perspective, each such hypothesis is equivalent to the set of hypotheses  $\mathcal{H}^{**}$  in conjunction with a subjective probability measure on  $\mathcal{H}^{**}$ . The set of all deterministic hypotheses again forms a basis generating (subjectivist equivalents of) all the stochastic hypotheses from subjective probability measures. This is another way of stating proposition 2. Moreover, we can generate any probability measure on  $\mathcal{H}^{**}$  as a probabilistic mixture of several such measures. From this it already follows that in general we can rationalize betting strategies partly or completely in terms of stochastic hypotheses. Nevertheless, it is instructive to consider an example where additionally to  $\mathcal{H}^{**}$  a stochastic hypothesis is introduced explicitly. This allows the construction of illustrative non-convergence examples.

Let us consider a very simple stochastic hypothesis, namely, the fair-coin hypothesis  $H$  implying  $\Pr(x_{t+1} = i | H) = \frac{1}{2}$ . Assume that Adam has already decided to explain the behavior of the money-spinner by the set of chaotic-clock hypotheses  $\mathcal{H}^{**}$ . He has already chosen a subjective probability measure on this set. Now it

occurs to him that, additionally, he should consider  $H$ . Is it necessary for him to adjust his prior to do this? It depends. We have already mentioned the fact that the rationalization of a betting strategy never requires the use of degenerate probabilities. More precisely, without ruling out any betting strategy, we can assume that Adam has chosen a probability measure  $\mu$  characterized by

$$\Pr_{\mu}(x_{t+1} = 0 | \mathcal{H}^{**} \wedge \mathbf{x}^t) \in \left[ \frac{l}{2}, 1 - \frac{l}{2} \right], \quad 0 < l < 2$$

for all  $\mathbf{x}^t$ . From this assumption it already follows that the prior  $\mu$  takes the fair-coin hypothesis  $H$  into account, if implicitly.

In order to see this, remember that from a Bayesian viewpoint,  $H$  is equivalent to the uniform prior  $u$  on  $\mathcal{H}^{**}$ , i. e., we have  $\Pr_u(x_{t+1} = 0 | \mathcal{H}^{**} \wedge \mathbf{x}^t) = \frac{1}{2}$ . Consider another prior  $\lambda$  on  $\mathcal{H}^{**}$  that is related to the old prior  $\mu$  in the following way: If  $\Pr_{\mu}(x_{t+1} = 0 | \mathcal{H}^{**} \wedge \mathbf{x}^t) = p$ , then  $\Pr_{\lambda}(x_{t+1} = 0 | \mathcal{H}^{**} \wedge \mathbf{x}^t) = f(p)$ , where

$$f(p) \stackrel{\text{def}}{=} \frac{p - l/2}{1 - l}.$$

Proposition 2 guarantees the existence of  $\lambda$ . Obviously, the old prior  $\mu$  is equivalent to a probabilistic mixing of  $\lambda$  and  $u$ : Just give a weight of  $l$  to  $u$  and a weight of  $1 - l$  to  $\lambda$ . If  $\Pr_{\mu}(x_{t+1} = 0 | \mathcal{H}^{**} \wedge \mathbf{x}^t) = p$ , then  $\Pr_{l \cdot u + (1-l) \cdot \lambda}(x_{t+1} = 0 | \mathcal{H} \wedge \mathbf{x}^t)$  is  $l \cdot \frac{1}{2} + (1 - l) \cdot f(p)$ , which just equals  $p$  by the definition of  $f$ . In other words: The old prior  $\mu$  on  $\mathcal{H}^{**}$  can be interpreted as implicitly assigning a subjective probability of  $l > 0$  to (the subjectivist equivalent of) the fair-coin hypothesis  $H$ . Along these line, we could easily show that  $\mu$  implicitly assigns a positive probability to (the subjectivist equivalent of) any stochastic hypothesis anybody would like to consider. Sets of stochastic hypotheses can be considered analogously.

This result can be used to construct very strong counterexamples to convergence. Just assume that the fair-coin hypothesis  $H$  is true. Even if Adam assigns a positive probability to the truth—which is a very strong bias in favor of the truth since there is a continuum of hypotheses—, he can rationalize any betting strategy. For example, Adam may stick in all eternity to the strategy  $y_{t+1} = x_t \cdot Y$ , i. e., putting all the money on  $x_{t+1} = i$  if  $x_t = i$ . This is perfectly rational, whatever evidence in favor of the fair-coin hypothesis somebody else, e. g., Eve, might discover in the observations. Of course, since Eve is a Bayesian herself, she might also start with a



---

bias in favor of the truth without ever converging either to the truth or to Adam's opinions. Any result according to which Bayesian learning leads to convergence to the truth (i. e., to rational expectations) or to intersubjective agreement is based on a suitable restriction of the priors. If one wants to argue that convergence is an implication of rationality, one has to find an argument in favor of rational priors.

### **Rational Priors**

There is a tradition in economics and game theory that tries to rule out the arbitrariness of priors. Some Bayesians, e. g., Aumann (1987: 7, 13f), defend the view that, even though there are no restrictions on priors, all rational agents should hold the same subjective probabilities if they have been exposed to the same experiences. This view is known as the Harsanyi Doctrine or the Common Prior Assumption (CPA). Aumann refers to Savage in this context (without giving a reference) and conjectures that Savage would have accepted the CPA. I disagree (cf. Savage 1962: 11, 13, 14). However, Savage was convinced that *in practice* experience often leads to convergence of opinion. But this is not a starting point for Bayesianism; it is a fact in need of explanation. For convergence, one needs priors that are not too different. The CPA just begs the question in assuming identical priors.

The CPA makes sense only if there exist canonical or rational priors before any experience. This leads to the classical problem of whether there is an acceptable "principle of insufficient reason" determining probabilities before experience. This idea, going back to Laplace, has been criticized by many authors (cf. Leamer 1978: 22-39, 61-63; Howson & Urbach 1989: 45-48, 285, 289; Earman 1992: 14-17, 138-141). It had been revived by Keynes and others in the form of a theory of "logical" probabilities, i. e., uniquely determined *a priori* probabilities of the logically possible hypotheses. One of the arguments in favor of Bayesianism has been the discovery that such probabilities do not exist.<sup>34</sup> It seems not to be a promising way of further

---

<sup>34</sup> This appears already to have been a conjecture of Ramsey, the earliest of the modern Bayesians, who made this argument against Keynes, one of the early proponents of logical or, as Hacking calls them, rational probabilities; cf. Hacking (1990: 165, 170). The Keynesian program was taken up later by Carnap; it was intended to provide one of the cornerstones of logical positivism. There is a widespread agreement today that this program foundered in just the way Ramsey conjectured: There are no logical probabilities; cf. Howson & Urbach (1989: 48-56).

development to revive this idea again. As the history of the subject presents itself, the burden of proof that there is an acceptable “principle of insufficient reason” rests with those in favor of the CPA.

However, it is possible to distill an argument in favor of rational priors from the results presented so far. It is worth the effort to spell out this potential argument and to refute it.

As already mentioned, Blume & Easley (1995: 19f, 36f) demonstrate for the tent-map dynamics that in the case of a density continuous at the starting point, the posterior distribution will converge to the uniform distribution. The latter implies a probability of  $\frac{1}{2}$  for  $x_t = 0$ . It is tempting to interpret this result, which also holds for the baker-map dynamics, as saying that in the long run, i. e., for sufficiently large  $t$ , everybody with a “reasonably simple” prior over  $I$  will assign a subjective probability of roughly  $\frac{1}{2}$  to  $x_t = 0$ . “Reasonably simple” priors are those that can be described by a continuous prior. In other words: Reasonable folks learn after a while that this chaotic process is more or less equivalent to tossing a fair coin. This provides an argument for starting with a uniform prior: If Adam and Eve foresee that any reasonable prior  $\mu$  for large  $t$  leads to  $\Pr_\mu(x_{t+1} = 0 | \mathcal{H}^{**} \wedge \mathbf{x}^t) \approx \frac{1}{2}$  independently from the actual observations  $\mathbf{x}^t$ , they should anticipate the inevitable result of the learning process and already start with the uniform prior  $u$  that from the beginning implies  $\Pr_u(x_{t+1} = 0 | \mathcal{H}^{**} \wedge \mathbf{x}^t) = \frac{1}{2}$ . If this is indeed a rational prior, then Adam and Eve should both adopt it, and Eve would be able to perfectly predict Adam on account of the assumption that he is rational (if he actually is rational).

However, this argument is fundamentally mistaken. The result of Blume & Easley shows that the identification of continuous densities with “reasonably simple” priors would be a big mistake. It ignores the fact that simplicity is not a property inherent to a proposition but, if there is a tenable definition of simplicity at all, a property of the way a proposition is expressed. Let  $\lambda$  be a prior with  $\Pr_\lambda(x_{t+1} = 0 | \mathcal{H}^{**} \wedge \mathbf{x}^t) = \frac{1}{3}$ .  $\lambda$  cannot be described with the help of a continuous density on  $\mathcal{H}^{**}$ . Viewed from this perspective, the prior is very complicated. Nevertheless, the description we have actually used is as simple as that for the uniform prior and arguably much simpler than any description of other priors based on continuous but complicated densities.

Moreover, it would be a mistake to give a special status to the long-run results

of the learning process. The decision to adopt a continuous prior is a decision to consider 0 and 1 to be equally probable in the long-run. These long-run results are not a result of learning; they are imposed on the learning process. Therefore, even if one restricts priors to those describable by a continuous density, it is not the case that one anticipates the results of learning by adopting the uniform prior from the outset. The actual (in contrast to the possible) results of learning cannot be anticipated *by definition*.

The attitude of most Bayesians concerning rational priors is probably best summarized by Edward Leamer. He states that rational priors, in order to reflect ignorance, would have to be dominated in their influence on posterior distributions by any set of data. Such priors simply do not exist (Leamer 1978: 62, 111, 114). Although defending the Bayesian standpoint, he views it as a "critical defect" of Bayesianism that priors are different to find for the individual and vary among different persons (Leamer 1978: 11). The importance of this fact for statistical analysis is not our concern. In the case of decision theory, at least, it has serious consequences.

## 5 Conclusion

With the help of our simple example we have shown that a perfectly rational Bayesian decision maker can rationalize any betting strategy in advance. Bayesianism is empty as a positive and as a normative theory of decision making. Moreover, Bayesianism as a methodology for predicting the behavior of Bayesian agents is as helpless as any other methodology.

This does not imply that every model using Bayesian learning is useless. If such a model generates a restricted set of solutions, this results from restrictions on prior beliefs of the agents. The combination of Bayesianism with restrictions on priors could generate solutions that coincide with the results from a more reasonable theory of rational behavior, if there is one. The solutions could even, as the proponents of as-if reasoning never tire to point out, agree with actual behavior. However, accidental success of this kind is cold comfort to those who are interested in a theory of rational behavior. What we have shown in effect is that any application of Bayesian-

ism just provides a special example of bounded rationality, probably without the psychological plausibility often associated with bounded-rationality theorizing.

Let us explain this point more carefully. The real-world counterpart to the perfectly rational (PR) Bayesian is the boundedly rational (BR) Bayesian. Let us call these hypothetical individuals Priscilla and Brian. Brian is not logically omniscient; he does not—although, as we have shown, he easily could—consider a set of hypotheses sufficiently rich to rationalize any behavior. However, Brian knows the results we have derived. He is aware of the fact that clever Priscilla is able to rationalize any behavior and that, even if he was unable to do the same, she could do it for him if there were a chance to ask her. Brian starts with a restricted set of hypotheses. Moreover, there are only some prior distributions that appeal to him, for whatever reasons. All these priors yield restrictions for his actions. Should Brian accept these restrictions as requirements of rational behavior?

The main problem is that BR Bayesianism is a substitute for PR Bayesianism. Brian believes that ideally he should ask Priscilla for advice. If he knows that Priscilla would offer to rationalize anything he would like to do, why should he accept the limits set by BR Bayesianism? It is not as if the rules of BR Bayesianism protect against *mistakes* that could be identified as such by Priscilla. To Priscilla, there are no mistakes; therefore there cannot be any protection. Why, then, should anyone be particular about “learning” procedures? And what could be the difference between perfect rationality and bounded rationality? If Brian applies Bayesian procedures, he is doing nothing essentially different from applying a rule of thumb. From Priscilla’s point of view, the results can be rationalized as those of any other rule of thumb.

Of course, Brian could reject Priscilla’s point of view. If we find another theory of perfect rationality, different rules of thumb may no longer appear to be equivalent. And even if Bayesian procedures under certain circumstances get better marks than other procedures, it will not be for Bayesian reasons. Actually, I believe that there are bits and pieces of another theory of perfect rationality, to be found in the methodology of science (or rather the non-Bayesian accounts thereof). But this is another story.

## References

- Aliprantis, Charalambos D./ **Border**, Kim C. (1994), *Infinite Dimensional Analysis*, Berlin: Springer
- Arrow**, Kenneth (1990), "Economic theory and the hypothesis of rationality", in: Eatwell et al. (1990), 25-37
- Aumann**, Robert J. (1987) "Correlated equilibrium as an expression of Bayesian rationality", *Econometrica* 55, 1-18
- Bauer**, Heinz (1978), *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie*, 3<sup>rd</sup> ed., Berlin: de Gruyter
- Bernheim**, David (1984), "Rationalizable strategic behavior", *Econometrica* 52, 1007-1028
- Bicchieri**, Cristina (1993), *Rationality and Coordination*, Cambridge: Cambridge University Press
- Binmore**, Ken (1992), *Fun and Games*, Lexington/MA: Heath
- Binmore**, Ken (1993), "De-Bayesing game theory", in: Ken **Binmore**/ Alan **Kirman**/ Piero **Tani** (eds), *Frontiers of game theory*, Cambridge/Mass.: MIT Press 1993, 321-339
- Blaug**, Mark (1992), *The methodology of economics*, 2<sup>nd</sup> rev. ed., Cambridge: Cambridge University Press
- Blume**, Lawrence E./ **Easley**, David (1995), "What has the rational learning literature taught us?", in: **Kirman** & **Salmon** (1995), 12-39
- Border**, Kim C. (1992), "Revealed preference, stochastic dominance and choice of lotteries", *Journal of Economic Theory* 56, 20-42
- Brémaud**, Pierre (1988), *An Introduction to Probabilistic Modelling*, New York: Springer
- Davies**, Paul (1987), *Cosmic Blueprint*, London: Heinemann
- Devaney**, Robert L. (1989), *An Introduction to Chaotic Dynamical Systems*, 2<sup>nd</sup> ed., Redwood City/Cal.: Addison-Wesley
- Earman**, John (1992), *Bayes or Bust?*, Cambridge/Mass.: MIT Press
- Eatwell**, John/ **Milgate**, Murray/ **Newman**, Peter (eds) (1990), *The New Palgrave: Utility and Probability*, New York: Norton
- Farmer**, Roger E. A. (1993), *The Macroeconomics of Self-Fulfilling Prophecies*,

Cambridge/Mass.: MIT Press

- Ford**, Joseph (1983), "How random is a coin toss?" *Physics Today*, April 1983
- Frydman**, Roman/ **Phelps**, Edmund S. (1983), "Introduction", in: Roman **Frydman**/ Edmund S. **Phelps** (eds), *Individual Forecasting and Aggregate Outcomes*, Cambridge: Cambridge University Press 1983, 1-30
- Green**, Edward J./ **Osband**, Kent (1991), "A revealed preference theory for expected utility", *Review of Economic Studies* 58, 677-696
- Hacking**, Ian (1990), "Probability", in: **Eatwell** et al. (1990), 163-177
- Hahn**, Frank (1996), "Rerum cognoscere causas", *Economics and Philosophy* 12, 183-195
- Howson**, Colin/ **Urbach**, Peter (1989), *Scientific Reasoning: The Bayesian Approach*, La Salle/Ill.: Open Court
- Humphreys**, Paul W. (1990), "The problem of induction", in: **Eatwell** et al. (1990), 116-120
- Kiefer**, Nicholas M./ **Nyarko**, Yaw (1995), "Savage-Bayesian models of economics", in: **Kirman** & **Salmon** (1995), 40-62
- Kim**, T. (1992), "The subjective expected utility hypothesis and revealed preference", *Economic Theory* 1: 251-263
- Kirman**, Alan/ **Salmon**, Mark (eds) (1995), *Learning and Rationality in Economics*, Oxford: Blackwell
- Leamer**, Edward E. (1978), *Specification searches*, New York: Wiley
- Ledyard**, John O. (1986), "The scope of the hypothesis of Bayesian equilibrium", *Journal of Economic Theory* 39, 59-82
- Lindley**, Dennis V. (1972), *Bayesian Statistics. A Review*, Philadelphia SIAM
- Lucas**, Robert E. (1976), "Econometric policy evaluation: a critique", in: Karl **Brunner**/ Allan H. **Meltzer** (eds), *Phillips Curve and Labor Markets*, Amsterdam: North-Holland, 19-46
- Miller**, David (1994), *Critical Rationalism. A Restatement and Defense*, Chicago and La Salle/Ill.: Open Court
- Musgrave**, Alan (1989), "Saving science from scepticism", in: Fred **D'Agostino**/ Ian C. **Jarvie** (eds), *Freedom and rationality*, Dordrecht: Kluwer 1989, 297-323
- Muth**, John F. (1961), "Rational expectations and the theory of price movements", *Econometrica* 29, 315-335

- Pesaran, M. Hashem** (1987), *The Limits to Rational Expectations*, Oxford: Blackwell
- Popper, Karl R.** (1959), *The Logic of Scientific Discovery*, London: Hutchinson
- Samuelson, Paul A.** (1938), "A note on the pure theory of consumer's behavior", *Economica* 5, 61-71
- Samuelson, Paul A.** (1947), *Foundations of Economic Analysis*, Cambridge/Mass.: Harvard University Press
- Savage, Leonard J. et al.** (1962), *The Foundations of Statistical Inference*, London: Methuen/New York: Wiley
- Savage, Leonard J.** (1954), *The Foundations of Statistics*, New York: Wiley
- Schuster, Heinz Georg** (1988), *Deterministic Chaos*, 2<sup>nd</sup> rev. ed., Weinheim: VCH
- Varian, Hal R.** (1992); *Microeconomic Analysis*, 3<sup>rd</sup> rev. ed., New York: Norton

## Appendix

This appendix contains a proof of the central result of the paper. This result is used but not proved in the text. In order to make the paper self-contained, the following subsection gives (some of the) necessary basic definitions and results presupposed later on.

### Some Important Definitions and Theorems

**Def. 1 (Partition)** *A partition of a set  $A$  is a finite system of disjoint subsets  $A_i \subseteq A$ ,  $i = 1, \dots, n$  with  $\bigcup_{i=1}^n A_i = A$ .*

**Def. 2 (Semiring)**  $\mathcal{R} \subseteq \mathcal{P}(\Omega)$  *is a semiring in  $\Omega$  if*

1.  $\mathcal{R} \neq \emptyset$ ,
2.  $A, B \in \mathcal{R} \rightarrow A \cap B \in \mathcal{R}$ , and
3. *there exists a partition of  $A \setminus B$  in  $\mathcal{R}$  for all  $A, B \in \mathcal{R}$ .*

**Def. 3 (Ring)**  $\mathcal{R} \subseteq \mathcal{P}(\Omega)$  *is a ring in  $\Omega$  if*

1.  $\mathcal{R} \neq \emptyset$ ,
2.  $A, B \in \mathcal{R} \rightarrow A \cup B \in \mathcal{R}$ , and
3.  $A, B \in \mathcal{R} \rightarrow A \setminus B \in \mathcal{R}$ .

**Def. 4 (Algebra)** *A ring  $\mathcal{A} \subseteq \mathcal{P}(\Omega)$  with  $\Omega \in \mathcal{A}$  is called algebra.*

**Def. 5 ( $\sigma$ -Algebra)**  $\mathcal{A} \subseteq \mathcal{P}(\Omega)$  *is called  $\sigma$ -algebra in  $\Omega$  if*

1.  $\Omega \in \mathcal{A}$ ,
2.  $A \in \mathcal{A} \rightarrow \Omega \setminus A \in \mathcal{A}$ , and
3.  $\forall n \in \mathbf{N} [A_n \in \mathcal{A}] \rightarrow \bigcup_{i=1}^{\infty} A_n \in \mathcal{A}$ .



**Lemma 1** For every  $\mathcal{A} \in \mathcal{P}(\Omega)$  there exists a smallest  $\sigma$ -algebra containing  $\mathcal{A}$ , called the  $\sigma$ -algebra generated by  $\mathcal{A}$  and denoted by  $\sigma(\mathcal{A})$ .

**Def. 6** Given a sequence  $\{A_n\}$  of sets,  $A_n \downarrow A$  means  $A_{n+1} \subseteq A_n$  and  $A = \bigcap_{n=1}^{\infty} A_n$ , and  $A_n \uparrow A$  means  $A_n \subseteq A_{n+1}$  and  $A = \bigcup_{n=1}^{\infty} A_n$ .

**Def. 7 (Set Functions)** A function  $\mu: \mathcal{R} \mapsto \bar{\mathbb{R}}$ , where  $\bar{\mathbb{R}}$  denotes the extended reals and  $\mathcal{R} \subseteq \mathcal{P}(\Omega)$ , is called a set function. For the following definitions, let  $\mathcal{R}$  be a semiring. Moreover, let  $A \in \mathcal{R}$  and  $C_i \in \mathcal{R}$  for all  $i$ , where the  $C_i$  are pairwise disjoint. Then  $\mu$  is

- non-negative, if  $\mu(A) \geq 0$  for all  $A$ ;
- finite, if  $|\mu(A)| < \infty$  for all  $A$ ;
- additive, if  $\bigcup_{i=1}^n C_i \in \mathcal{R} \rightarrow \mu(\bigcup_{i=1}^n C_i) = \sum_{i=1}^n \mu(C_i)$ ;
- $\sigma$ -additive, if  $\bigcup_{i=1}^{\infty} C_i \in \mathcal{R} \rightarrow \mu(\bigcup_{i=1}^{\infty} C_i) = \sum_{i=1}^{\infty} \mu(C_i)$ .

**Def. 8 (Charge)** A charge is a set function  $\mu$  on a semiring that is additive and non-negative with  $\mu(\emptyset) = 0$ .

**Def. 9 (Content)** A content is a charge on a ring.

**Def. 10 (Premeasure)** A premeasure is a  $\sigma$ -additive content.

**Theorem 1** A finite content  $\mu$  on a ring  $\mathcal{R}$  is a premeasure if and only if for every sequence  $\{A_n\}$  in  $\mathcal{R}$  with  $A_n \downarrow \emptyset$   $\lim_{n \rightarrow \infty} \mu(A_n) = 0$ .

*Proof.* Cf. Bauer (1978: 23, Satz 3.2). □

**Def. 11 (Measure)** A measure is a premeasure on a  $\sigma$ -algebra.<sup>35</sup>

**Theorem 2** Every finite premeasure on an algebra  $\mathcal{A} \subseteq \mathcal{P}(\Omega)$  can be extended to a unique measure on  $\sigma(\mathcal{A})$ .

---

<sup>35</sup> Aliprantis & Border (1994: 282) use the term measure more generally for a non-negative, finite and  $\sigma$ -additive charge on a semiring. They have no special term for “measure on a  $\sigma$ -algebra”.

*Proof.* Cf. Bauer (1978: 35, Satz 5.7). Note that an algebra  $\mathcal{A} \subseteq \mathcal{P}(\Omega)$  is a ring containing  $\Omega$ , and that a finite content (and therefore a finite premeasure) on an algebra is also  $\sigma$ -finite (cf. Bauer 1978: 34, Def. 5.6, 35, Lemma 5.8).  $\square$

**Def. 12 (Probability Measure)** *A probability measure is a measure  $\mu$  on a  $\sigma$ -algebra  $\Sigma$  with  $\mu(\Sigma) = 1$ .*

**Def. 13 (Borel Sets)** *The  $\sigma$ -algebra generated by the system of half-open intervals  $[a, b)$  (where  $[a, b) = \emptyset$  if  $b \leq a$ ) is called the  $\sigma$ -algebra of the Borel sets of  $\mathbb{R}$ .*

**Lemma 2** *The restriction of the  $\sigma$ -algebra of the Borel sets to the interval  $[0, 1)$  is a  $\sigma$ -algebra generated by the half-open intervals  $[a, b) \subseteq [0, 1)$ .*

*Proof.* Obvious, since the intersection of intervals of the form  $[a, b)$  is always an interval of the same form.  $\square$

**Def. 14 (Measurable Space)** *A set  $X$  equipped with a  $\sigma$ -algebra of subsets  $\mathcal{A}$  is called a measurable space and denoted by  $(X, \mathcal{A})$ .*

**Def. 15 (Measurable Function)** *A function  $f: (X, \mathcal{A}) \mapsto (Y, \mathcal{B})$ , where  $\mathcal{A}$  and  $\mathcal{B}$  are algebras of subsets of  $X$  and  $Y$  respectively, is measurable if  $f^{-1}(B) \in \mathcal{A}$  for each  $B \in \mathcal{B}$ .*

**Lemma 3** *A function  $f: (X, \mathcal{A}) \mapsto (Y, \mathcal{B})$  between measurable spaces, where  $\mathcal{C}$  is a family of subsets of  $Y$  and  $\mathcal{B} = \sigma(\mathcal{C})$ , is measurable if and only if  $f^{-1}(C) \in \mathcal{A}$  for every  $C \in \mathcal{C}$ .*

*Proof.* See Aliprantis & Border (1994: 279).  $\square$

## Basic Intervals

Definitions, lemmata and theorems in this and the next section are mostly specific to this paper.

We consider the set of intervals  $[a, b) \in [0, 1)$  that result from iterative bisections of the unit interval. We call these intervals *basic intervals*. The basic intervals are bounded by certain fractions.

**Def. 16 (Basic Intervals)** *The elements of the set*

$$\mathcal{I} \stackrel{\text{def}}{=} \emptyset \cup \{I_n(m) : 0 \leq m < 2^n, m, n \in \mathbb{N}_0\}$$

*are called basic intervals. The interval  $\left[\frac{m}{2^n}, \frac{m+1}{2^n}\right) \in \mathcal{I}$ , where  $0 \leq m < 2^n$  and  $n \in \mathbb{N}_0$ , is called the  $m$ -th basic interval on level  $n$  and denoted by  $I_n(m)$ . The unit interval  $I_0(0)$  is also denoted by  $I$ .*

The basic interval  $I_n(m)$  is one of the intervals we can find by  $n$  bisections of the unit interval. Fig. 3 illustrates the numbering of the basic intervals introduced by the definition of  $I_n(m)$ .

Insert fig. 3 from the end of the paper!

**Def. 17 (Interval Bounds)** *The set*

$$\mathbb{P}_0 \stackrel{\text{def}}{=} \left\{ \frac{m}{2^n} : n, m \in \mathbb{N}_0, 0 \leq m \leq 2^n \right\}$$

*is called the set of bounds for the basic intervals. The set  $\mathbb{P} \stackrel{\text{def}}{=} \mathbb{P}_0 \setminus \{0\}$  is called the set of the upper bounds.*

Note that, with the exception of 1, the upper bound of one interval is the lower bound of another.

The following properties of the basic intervals can easily be verified by a glance at fig. 3.

**Lemma 4**

- a) *The  $2^n$  basic intervals  $\{I_n(m)\}_{m=0}^{2^n-1}$  form a partition of the unit interval.*
- b) *Let  $I_n(m)$  be a basic interval and  $k \geq n \geq 0$ . The  $2^{k-n}$  basic intervals defined by  $\{I_k(l)\}_{l=a}^b$ , where  $a = m \cdot 2^{k-n}$  and  $b = (m+1) \cdot 2^{k-n} - 1$ , form a partition of  $I_n(m)$ .*

c) Let  $I_n(m), I_k(l)$  be basic intervals with  $k \geq n \geq 0$ . Then

$$I_n(m) \cap I_k(l) = \begin{cases} I_k(l) & \text{iff } l \in [a, b] \\ \emptyset & \text{otherwise,} \end{cases}$$

where  $a = m \cdot 2^{k-n}$  and  $b = (m + 1) \cdot 2^{k-n} - 1$ .

We will use some simple properties of infinite sequences of basic intervals. Note that we can always transform a finite into an infinite sequence by endlessly repeating the last element. With this convention, we subsequently assume all sequences of sets to be infinite. Moreover, we assume that the first element of a sequence has index 0 if nothing else is said. We therefore write  $\{A_n\}$  for  $\{A_n\}_{n=0}^\infty$ .

**Lemma 5** For any sequence  $\{A_n\}$  in  $\mathcal{I}$  with  $A_n \downarrow \emptyset$  and  $A_n \neq \emptyset$  for all  $n$ , there exists a  $p \in \mathbb{P}$  and an  $\bar{n}$  such that

$$A_n = \left[ p - \frac{1}{2^{k(n)}}, p \right)$$

for all  $n \geq \bar{n}$ , where  $k: \mathbb{N}_0 \rightarrow \mathbb{N}_0$  is non-decreasing with  $\lim_{n \rightarrow \infty} k(n) = \infty$ .

*Proof.* Consider the sequence  $\{B_n\}$  of closures of the  $A_n$ . This sequence forms a standard bisection converging to a real number  $p$ , i. e.,  $B_n \downarrow \{p\}$ . In order for  $A_n \downarrow \emptyset$ , there has to be an  $\bar{n}$  such that  $p \notin A_n$  for all  $n \geq \bar{n}$ . Moreover,  $p$  must be excluded from the  $A_n$  with  $n \geq \bar{n}$  but included in their closure. Thus  $p$  must be in the set  $\mathbb{P}$  of upper bounds of the basic intervals. Since  $B_n \downarrow \{p\}$ ,  $p$  must be the upper bound for all  $A_n$  with  $n \geq \bar{n}$ . Thus the  $A_n$  form a subsequence of  $\left[ p - \frac{1}{2^n}, p \right)$  for  $n \geq \bar{n}$ .  $\square$

Subsequently, we will mostly be concerned with finite unions of basic intervals.

**Def. 18 (Configurations)** A finite union of basic intervals is called a configuration. The set of all configurations is denoted by  $\mathcal{F}$ .

**Def. 19 (Uniform Partitions)** A partition of a configuration into basic intervals of level  $n$  is called a uniform partition on level  $n$ .

The following result on uniform partitions is obvious in view of lemma 4.

**Lemma 6**

- a) Every configuration has a uniform partition on some level.
- b) If a configuration has a uniform partition on level  $\bar{n}$ , it has a uniform partition on every level  $n > \bar{n}$ .
- c) Any two configurations have a uniform partition on the same level.

**Lemma 7** *The set  $\mathcal{I}$  is a semiring.*

*Proof.*  $\mathcal{I}$  is (i) not empty and (ii) contains  $\emptyset$ . Lemma 4 shows that (iii)  $\mathcal{I}$  contains the intersection of any two elements. Lemma 6 ensures that  $A, B \in \mathcal{I}$  have uniform partitions on a common level since basic intervals are also configurations. Therefore, (iv)  $A \setminus B$  is a union of basic intervals.  $\square$

**Theorem 3** *The set  $\mathcal{F}$  of configurations is an algebra in  $\mathcal{P}(I)$ .*

*Proof.*  $\mathcal{F}$  is (i) not empty, (ii) contains  $\emptyset$ , and (iii) contains  $I$ . Let  $A, B$  be two configurations. The definition of a configuration ensures that  $A \cup B$  is a configuration. Thus we have (iv) closure under finite unions. Lemma 6 ensures that  $A$  and  $B$  have uniform partitions on some common level  $n$ . Therefore,  $A \cap B$  and  $A \setminus B$  are unions of basic intervals of level  $n$  and thus configurations themselves. Thus, we have (v) closure under finite intersections and (vi) closure under relative complementation. Therefore,  $\mathcal{F}$  is an algebra.  $\square$

The following result goes one step further than theorem 3.

**Theorem 4** *The  $\sigma$ -algebra generated by  $\mathcal{I}$ , denoted by  $\sigma(\mathcal{I})$ , is the  $\sigma$ -algebra of the Borel sets of the unit interval  $I$ .*

*Proof.*  $\sigma(\mathcal{I})$  contains every interval  $[a, b)$  with  $a < b \in I$ . This can be seen as follows. Any  $a$  and any  $b$  can be generated as limes of a countable sequence of bisections, i. e., as intersection of a sequence of basic intervals of higher and higher levels:

$$a = \bigcap_{n=0}^{\infty} I_n[m(a, n)]$$

$$b = \bigcap_{n=0}^{\infty} I_n[m(b, n)]$$

We get the interval  $[a, b)$  as countable intersection of sets  $M_n$ , i. e., as

$$[a, b) = \bigcap_{n=0}^{\infty} M_n,$$

where

$$M_n \stackrel{\text{def}}{=} \bigcup_{k=m(a,n)}^{m(b,n)} I_n(k).$$

$\sigma(\mathcal{I})$  therefore contains *all* the intervals  $[a, b)$ . By lemma 2, the set of these intervals generates the  $\sigma$ -algebra of the Borel sets.  $\square$

**Def. 20** For every sequence  $\{A_n\}$  in  $\mathcal{F}$  with  $A_{n+1} \subseteq A_n$  for all  $n$ , let the strictly increasing function  $k: \mathbb{N}_0 \mapsto \mathbb{N}_0$  select a level on which a uniform partition of  $A_n$  exists. Let the function  $l: \mathbb{N}_0 \mapsto \mathbb{N}_0$  select an element  $I_{k(n)}(l(n))$  of this uniform partition such that  $I_{k(n+1)}(l(n+1)) \subset I_{k(n)}(l(n))$ . If such an element does not exist, then use  $\emptyset$  instead of some  $I_{k(n+1)}(m)$ . Then the sequence  $\{I_{k(n)}(l(n))\}_{n=0}^{\infty}$  in  $\mathcal{I}$  is called a basic subsequence of the sequence  $\{A_n\}$ .

**Lemma 8** Every sequence  $\{A_n\}$  in  $\mathcal{F}$  with  $A_{n+1} \subseteq A_n$  for all  $n$  has a basic subsequence.

*Proof.* Follows from lemma 6.  $\square$

**Lemma 9** Let there be a sequence  $\{A_n\}$  in  $\mathcal{F}$  with  $A_n \downarrow \emptyset$  and  $A_n \neq \emptyset$  for all  $n$ . Then the sequence has at least one basic subsequence  $\{U_n\}$  with  $U_n \neq \emptyset$  for all  $n$ . Moreover, for all basic subsequences  $\{U_n\}$  of  $\{A_n\}$  we have  $U_n \downarrow \emptyset$ .

*Proof.* Assume that all basic subsequences of  $\{A_n\}$  were identical with  $\emptyset$  for all  $n > \bar{n}$ . Then  $A_n = \emptyset$  for all  $n > \bar{n}$  by the definition of a basic subsequence, which contradicts the assumptions of the lemma.

Assume that  $U_n \downarrow U$  where  $U \neq \emptyset$ . Then  $A_n \downarrow A$  and  $U \subseteq A$  since it follows from the definition of a basic subsequence that  $A_n \cap U_n = U_n$  for all  $n$ , which contradicts the assumptions of the lemma.  $\square$

**Def. 21 (Limit Point)** Given a sequence  $\{A_n\}$  in  $\mathcal{F}$  with  $A_n \downarrow A$  and  $A \in \mathcal{F}$ , a point  $x \in \{0, 1\}$  with  $A_n \cap [x - \frac{1}{2^n}, x) \neq \emptyset$  for all  $n$  is called a limit point of  $\{A_n\}$ .

**Lemma 10** *Let there be a sequence  $\{A_n\}$  in  $\mathcal{F}$  with  $A_n \downarrow \emptyset$  and  $A_n \neq \emptyset$  for all  $n$ .*

- a)  $\{A_n\}$  has at least one limit point.
- b) All limit points of  $\{A_n\}$  are in  $\mathbb{P}$ .

*Proof.* According lemma 9, there exists a basic subsequence  $\{U_n\}$  of  $\{A_n\}$  with  $U_n \downarrow \emptyset$  and  $U_n \neq \emptyset$  for all  $n$ . According to lemma 5, there exists a  $p \in \mathbb{P}$  and an  $\bar{n}$  such that

$$U_n = \left[ p - \frac{1}{2^{k(n)}}, p \right)$$

for  $n \geq \bar{n}$  and some non-decreasing function  $k: \mathbb{N}_0 \mapsto \mathbb{N}_0$  with  $\lim_{n \rightarrow \infty} k(n) = \infty$ . In fact, by the definition of subsequence,  $U_{n+1} \subset U_n$ , which requires that  $k(n)$  is strictly increasing in the relevant region, i. e., for  $n \geq \bar{n}$ . Because  $U_n$  is a subsequence, we have  $U_n \cap A_n \neq \emptyset$  for all  $n$ . Therefore  $p$  is a limit point (see def. 21).

Let us assume that there was a limit point  $x$  of  $\{A_n\}$  with  $x \notin \mathbb{P}$ . Then  $A_n \cap \left[ x - \frac{1}{2^n}, x \right) \neq \emptyset$  for all  $n$ . However, since  $x \notin \mathbb{P}$ , there exists no basic interval such that  $x$  is its upper bound. Thus there is always some  $U_n \in \mathcal{I}$  such that  $x \in U_n \subseteq A_n$ . This implies  $A_n \downarrow A$  with  $x \in A$ , which contradicts the assumptions of the lemma.  $\square$

**Lemma 11** *Let  $\mathcal{L}$  be the set of all limit points of a sequence  $\{A_n\}$  in  $\mathcal{F}$  with  $A_n \downarrow \emptyset$  and  $A_n \neq \emptyset$  for all  $n$ . Then between any two elements of  $\mathcal{L}$  there must be some basic interval that contains no elements of  $\mathcal{L}$ . Obviously, this implies  $\mathcal{L} \neq \mathbb{P} \cap I_n(m)$ , i. e.,  $\mathcal{L}$  never contains all points of  $\mathbb{P}$  in some basic interval.*

*Proof.* We derive a contradiction from the assumptions of the theorem and the assumption that there are two points  $a < b \in \mathcal{L}$  such that every non-empty basic interval  $[x, y)$  with  $x \geq a$  and  $y \leq b$  contains further elements of  $\mathcal{L}$ . Consider a uniform partition of  $A_n$  at level  $k$ . By the definition of limit points, every point in  $[a, b)$  must be in some basic interval of the partition. Therefore, every  $A_n$  contains  $[a, b)$ . Thus we have proved that  $A_n \downarrow A$  with  $[a, b) \in A$ , which contradicts  $A_n \downarrow \emptyset$ .  $\square$

**Lemma 12** *The set  $\mathcal{L}$  of all limit points of a sequence  $\{A_n\}$  in  $\mathcal{F}$  with  $A_n \downarrow \emptyset$  and  $A_n \neq \emptyset$  for all  $n$  is compact.*

*Proof.* Since  $\mathcal{L} \in \mathbb{P}$ ,  $\mathcal{L}$  is bounded. It remains to be shown that a sequence of limit points converges to a limit point if it converges at all. Let  $\{q_i\}_{i=0}^{\infty}$  be a sequence in  $\mathcal{L}$  converging to some point  $\bar{q}$ . This means that for every open set  $U_n \subseteq \mathbb{R}$  with  $\bar{q} \in U$  there exists an  $k$  such that  $q_i \in U_n$  for  $i > k$ . Let  $U_n = \left(\bar{q} - \frac{1}{2^n}, \bar{q}\right) \cup \left[\bar{q}, \bar{q} + \frac{1}{2^n}\right)$ . Since with every limit point some open set containing the limit point is in  $U_n$ , and since every open set containing a limit point has a non-empty intersection with every  $A_n$ , we have  $A_n \cap U_n \neq \emptyset$  for all  $n$ . Now assume that the upper part  $\left[\bar{q}, \bar{q} + \frac{1}{2^n}\right)$  of each  $U_n$  always contained a point of the sequence. From this it would follow that  $A \downarrow A$  with  $\bar{q} \in A$ :  $A_n \cap \left[\bar{q}, \bar{q} + \frac{1}{2^n}\right)$  is never empty, which implies that each  $A_n$  has to contain some basic interval  $\left[\bar{q}, \bar{q} + \frac{1}{2^n}\right)$ . Since  $A \downarrow A \neq \emptyset$  contradicts the assumptions of the lemma, there has to be an  $l$  such that  $q_i \in \left(\bar{q} - \frac{1}{2^n}, \bar{q}\right)$  for  $i > l$ . This implies that  $A_n \cap \left(\bar{q} - \frac{1}{2^n}, \bar{q}\right) \neq \emptyset$  for all  $n$ . Applying definition 21, we find that  $\bar{q} \in \mathcal{L}$ . Therefore  $\mathcal{L}$  is compact.  $\square$

Our considerations are based on intervals  $[a, b)$ . There exist analogous definitions and results for intervals  $(a, b]$ . All proofs can be applied mutatis mutandis. We therefore omit all proofs for the respective lemmata and theorems.

**Def. 22 (Inverse Basic Intervals)** *The elements of the set*

$$\bar{\mathcal{I}} \stackrel{\text{def}}{=} \{(a, b) : a, b \in \mathbb{P}, a \leq b\}$$

*are called inverse basic intervals.*

**Def. 23 (Inverse Configurations)** *A finite union of inverse basic intervals is called an inverse configuration. The set of all inverse configurations is denoted by  $\bar{\mathcal{F}}$ .*

**Def. 24 (Inverse Counterparts)** *The inverse counterpart of a basic interval  $I_n(m)$  is the inverse basic interval  $\left(\frac{m}{2^n}, \frac{m+1}{2^n}\right]$  denoted by  $\bar{I}_n(m)$ . The inverse counterpart of a configuration  $A$  is the inverse configuration  $\bar{A}$  that results from replacing the basic intervals in any uniform partition by their inverse counterparts. The inverse counterparts of the elements of  $\bar{\mathcal{I}}$  and  $\bar{\mathcal{F}}$  are the elements in  $\mathcal{I}$  and  $\mathcal{F}$  of which they are inverse counterparts.*



**Theorem 5** *The set  $\bar{\mathcal{I}}$  of inverse basic intervals is a semiring. The set  $\bar{\mathcal{F}}$  of inverse configurations is an algebra in  $\mathcal{P}(I)$ .*

**Theorem 6** *For every sequence  $\{\bar{A}_n\}$  of inverse configurations with  $\bar{A}_n \downarrow \emptyset$ , there exists either an  $\bar{n}$  such that  $\bar{A}_n = \emptyset$  for all  $n \geq \bar{n}$ , or a compact set  $\mathcal{L}$  of limit points with  $\mathcal{L} \subset \mathbb{P}_0 \setminus 1$ .*

The following theorem connects limit sets of sequences with limits of sequences of inverse counterparts.

**Theorem 7** *For every sequence  $\{A_n\}$  of configurations with  $A_n \downarrow \emptyset$ , there exists either an  $\bar{n}$  such that  $A_n = \emptyset$  for all  $n \geq \bar{n}$ , or a compact set  $\mathcal{L}$  of limit points with  $\mathcal{L} \subset \mathbb{P}$ . In the latter case we moreover have  $\bar{A}_n \downarrow \mathcal{L}$ , where  $\bar{A}_n$  denotes the inverse counterpart of  $A_n$ .*

*Proof.* The first sentence is just a summary of previous results on limit points and therefore requires no additional proof. Thus we concentrate on the second sentence.

From the definition of a limit point, we know that for all  $p \in \mathcal{L}$  the intersection  $A_n \cap \left[p - \frac{1}{2^n}, p\right)$  is never empty. Thus there exists some non-decreasing function  $k: \mathbb{N}_0 \mapsto \mathbb{N}_0$  with  $\lim_{n \rightarrow \infty} k(n) = \infty$  such that

$$U_n = \left[ p - \frac{1}{2^{k(n)}}, p \right)$$

is a basic subsequence of  $\{A_n\}$ . Then

$$\bar{U}_n = \left( p - \frac{1}{2^{k(n)}}, p \right]$$

is a subsequence of  $\{\bar{A}_n\}$ , and  $\bar{A}_n \downarrow A$  with  $p \in A$ . Since the same argument holds for every  $p \in \mathcal{L}$ , we have shown that  $\mathcal{L} \subseteq A$ . It remains to be shown that  $A \subseteq \mathcal{L}$ . So let  $x \in A$ . This implies  $x \in \bar{A}_n$  for all  $n$  and therefore  $A_n \cap \left[x - \frac{1}{2^n}, x\right) \neq \emptyset$ , which implies  $x \in \mathcal{L}$  by the definition of a limit point.  $\square$

## Basic Charges

We define a finite charge on  $\mathcal{I}$  (called basic charge) by attaching numbers between one and zero to each basic interval  $I_n(m)$ . The extension of a finite charge on  $\mathcal{I}$  to a

finite content on the algebra  $\mathcal{F}$  is trivial. We then use the concept of limit points of a sequence  $A_n \downarrow \emptyset$  in  $\mathcal{F}$  to find a necessary and sufficient condition for the extension of the content on  $\mathcal{F}$  to a probability distribution on the  $\sigma$ -algebra of the Borel sets of the unit interval  $I$ . This extension follows from standard theorems once we have shown that the content on  $\mathcal{F}$  is  $\sigma$ -additive and hence a premeasure.

**Def. 25 (Basic Charge)** *Let there be a real number  $q_n(m) \in [0, 1]$  attached to each basic interval  $I_n(m)$ . Let these numbers have the property  $q_n(m-1) + q_n(m) = 1$  if  $m > 0$  is odd. The set function  $\mu: \mathcal{I} \rightarrow \mathbb{R}_+$  recursively defined by*

$$\mu(A) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if } A = \emptyset \\ 1 & \text{if } A = I \\ \mu\left[I_{n-1}\left(\frac{m-1}{2}\right)\right] \cdot q_n(m) & \text{if } A = I_n(m) \text{ with } m > 0 \text{ odd} \\ \mu\left[I_{n-1}\left(\frac{m}{2}\right)\right] \cdot q_n(m) & \text{if } A = I_n(m) \text{ with } m > 0 \text{ even} \end{cases}$$

*is called a basic charge.*

The recursion works as follows: Consider a sequence  $\{A_n\}_{n=0}^k$  of  $k+1$  basic intervals with the property  $A_n \downarrow A_k$  where no level from 0 to  $k$  is left out. So  $A_0$  is  $I$  and  $A_k = I_k(m)$  for some  $0 \leq m < 2^k$ . Then  $q_k(m)$  is the product of all the  $q$ -values connected with the precursors of  $A_k$  in the sequence. In constructing the sequence, there are always two possibilities for the next step given the first  $n$  steps: taking the higher or the lower of the two subintervals on the next level. The  $q$ -values for these two possibilities add to 1. This construction principle obviously leads to the following properties of the function  $\mu$ .

**Lemma 13** *The basic charge has the following properties, where the second is a generalization of the first:*

- a)  $\sum_{m=0}^{2^n-1} \mu(I_n(m)) = 1$
- b)  $\sum_{m \in \mathcal{M}} \mu(I_n(m)) = \mu(I_k(l))$  if the basic intervals  $I_n(m)$  with  $m \in \mathcal{M}$  form a uniform partition of  $I_k(l)$

**Lemma 14** *The basic charge is a finite charge on  $\mathcal{I}$ .*

*Proof.* (i)  $\mathcal{I}$  is a semiring. (ii) Lemma 13 in effect states that the basic charge is an additive set function. (iii)  $\mu(\emptyset) = 0$  as well as (iv) non-negativity are already part of the definition. (v) Finiteness again follows from lemma 13 and the recursive definition of the basic charge.  $\square$

An extension to the algebra  $\mathcal{F}$  of configurations is trivial.

**Lemma 15** *Let  $\mu$  be a basic charge on  $\mathcal{I}$ . Let the function  $\mu^*: \mathcal{F} \mapsto [0, 1]$  be defined by*

$$\mu^*(A) \stackrel{\text{def}}{=} \sum_{n=1}^k \mu(C_n),$$

where the  $C_n$  form some uniform partition of  $A$ .  $\mu^*$  then is a finite content on  $\mathcal{F}$ .

Again, all the definitions and results apply mutatis mutandis to inverse basic intervals and configurations. We summarize these trivial facts in the following theorem.

**Theorem 8** *Let  $\mu$  be a basic charge on  $\mathcal{I}$  and  $\mu^*$  its extension to  $\mathcal{F}$ . Let  $\bar{\mu}: \bar{\mathcal{I}} \mapsto \mathbb{R}_+$  and  $\bar{\mu}^*: \bar{\mathcal{F}} \mapsto \mathbb{R}_+$  be defined by the condition that inverse counterparts are assigned the same value by  $\mu$  and  $\bar{\mu}$  and  $\mu^*$  and  $\bar{\mu}^*$ , respectively. Then  $\bar{\mu}$  is a finite charge (called inverse basic charge) on  $\bar{\mathcal{I}}$  and  $\bar{\mu}^*$  is a finite content on  $\bar{\mathcal{F}}$ .*

We now give a necessary and sufficient condition for the basic charge to induce a probability measure on  $\sigma(\mathcal{I})$ .

**Theorem 9** *Let  $\mu$  be a basic charge on  $\mathcal{I}$  and  $\mu^*$  its extension to  $\mathcal{F}$ . Let  $\bar{\mu}$  be the inverse basic charge and  $\bar{\mu}^*$  its extension to  $\bar{\mathcal{F}}$ .  $\mu^*$  can be extended to a probability measure on  $\sigma(\mathcal{I})$  if and only if the following condition holds: For any  $\mathcal{L} \subseteq \mathbb{P}$  and for all sequences  $\{\bar{A}_n\}$  in  $\bar{\mathcal{F}}$  with  $\bar{A}_n \downarrow \mathcal{L}$ ,  $\lim_{n \rightarrow \infty} \bar{\mu}^*(\bar{A}_n) = 0$ . The extension is unique.*

*Proof.* We first show that  $\mu^*$  is a premeasure on  $\mathcal{F}$ , which is a necessary condition for its extension to a probability measure.  $\mu^*$  is a finite content on an algebra. According to theorem 1,  $\mu^*$  then is a premeasure if and only if  $\lim_{n \rightarrow \infty} \mu^*(A_n) = 0$  for all  $A_n \downarrow \emptyset$ . According to theorem 7, either (a)  $A_{\bar{n}} = \emptyset$  for some  $\bar{n}$  or (b) there exists a compact set of limit points  $\mathcal{L} \in \mathbb{P}$ .  $\lim_{n \rightarrow \infty} \mu^*(A_n) = 0$  is trivially

fulfilled in case (a), as the definition of the basic charge shows. We therefore consider only sequences  $A_n \downarrow \emptyset$  with  $A_n \neq \emptyset$  for all  $n$ . Then theorem 7 shows that  $\bar{A}_n \downarrow \mathcal{L}$ , where  $\bar{A}_n$  is the inverse counterpart of  $A_n$ . However, by the definition of the inverse basic charge and its extension,  $\bar{\mu}^*(\bar{A}_n) = \mu^*(A_n)$  for all  $A_n$  and thus  $\lim_{n \rightarrow \infty} \bar{\mu}^*(\bar{A}_n) = \lim_{n \rightarrow \infty} \mu^*(A_n)$ . Therefore  $\lim_{n \rightarrow \infty} \bar{\mu}^*(\bar{A}_n) = 0$  is necessary and sufficient for  $\lim_{n \rightarrow \infty} \mu^*(A_n) = 0$ . Since we obviously can construct a sequence  $A_n$  for any predetermined set of limit points (by considering an infinite union of sequences of basic intervals generating these limit points), the condition of the theorem is necessary and sufficient for  $\lim_{n \rightarrow \infty} \mu^*(A_n) = 0$  to hold for all sequences  $\{A_n\}$  with  $A_n \downarrow \emptyset$  and  $A_n \neq \emptyset$  for all  $n$ .

The rest of the proof is simple. We have already shown that the  $\sigma$ -algebra generated by  $\mathcal{F}$  is the  $\sigma$ -algebra of the Borel sets of  $I$ . Theorem 2 allows us to extend the premeasure  $\mu^*$  in a unique way to the  $\sigma$ -algebra of the Borel sets. Since  $\mu^*(I) = 1$ , we have shown that the basic charge under the conditions of the theorem induces a unique probability measure on  $I$ .  $\square$

### A Probability Measure on the Cantor set

We now consider the Cantor set  $\mathcal{C} \stackrel{\text{def}}{=} \{0, 1\}^\infty$  of all infinite sequences of 0s and 1s. In the text this set is denoted by  $\mathcal{S}_O$ , the state space of observables. The following considerations are less formal than those of the last two subsections.

Consider the metric on  $\mathcal{C}$  generated by

$$d(\mathbf{x}, \mathbf{y}) \stackrel{\text{def}}{=} \sum_{n=1}^{\infty} \frac{|x_n - y_n|}{3^n},$$

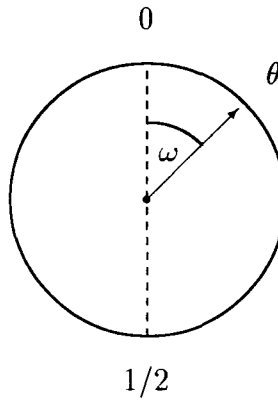
where  $x_n$  and  $y_n$  are the  $n$ -th element in the sequences  $\mathbf{x}$  and  $\mathbf{y}$ , respectively. The topology  $\mathcal{T}$  induced by this metric is a system of open sets determined by

$$\{\mathbf{y} \in \mathcal{C} : d(\mathbf{x}, \mathbf{y}) < r \in \mathbb{R}_+\}.$$

Each sequence in this set agrees on the first  $t$  elements  $\mathbf{x}^t = (x_0, \dots, x_t)$ , where  $t$  falls if  $r$  rises (cf. Aliprantis & Border 1994: 93). We consider the  $\sigma$ -algebra  $\sigma(\mathcal{T})$  generated by the topology  $\mathcal{T}$ . This is the  $\sigma$ -algebra on the state space of observables used in the text.

As has been demonstrated in the text, each  $\mathbf{x}^t$  corresponds to a unique basic interval and vice versa. There is a one-to-one function  $f: \sigma(\mathcal{I}) \mapsto \sigma(\mathcal{T})$ ; this function is described in the text with the help of the baker-map dynamics. The function is measurable since the inverse image of an element of  $\mathcal{T}$  is always in the  $\sigma$ -algebra  $\sigma(\mathcal{I})$ . Therefore, a probability measure on  $\sigma(\mathcal{I})$  induces a probability measure on  $\sigma(\mathcal{T})$ . However, by theorem 9, the measure on the Cantor set will always have the property that measurable sets of sequences with a finite number of 0s have probability 0.

In the text, this restriction is overcome by using another mapping from  $\sigma(\mathcal{I})$  to  $\sigma(\mathcal{T})$ . This mapping can be defined in the following way. Let  $g: \sigma(\mathcal{C}) \mapsto \sigma(\mathcal{C})$  be a function that takes the open set corresponding to  $\mathbf{x}^t$  to the open set corresponding to  $\mathbf{y}^{t'}$ , where  $t' = t \operatorname{div} 2$  and  $\mathbf{y}^{t'}$  is defined by deleting every odd-numbered element in  $\mathbf{x}^t$ . Let  $h \stackrel{\text{def}}{=} g \circ f$ . Then  $h$  is measurable since both  $f$  and  $g$  are measurable. The inverse image under  $h$  of an element of  $\mathcal{T}$  is always a finite union of disjoint intervals from  $\mathcal{I}$ . Again, the measurability of  $h$  ensures that a probability measure on  $\sigma(\mathcal{I})$  induces a probability measure on  $\sigma(\mathcal{T})$ .



*Fig. 1: A chaotic clock.* The figure illustrates the baker-map dynamics as behavior of a chaotic clock. There is one pointer that can point to all real numbers in the interval  $I = [0, 1)$ , where the vertically upward position is zero and the vertically downward position is  $\frac{1}{2}$ . Initially, the pointer deviates by an angle  $\omega = 2\theta\pi$  from the vertically upward position, thus pointing at the real number  $\theta$ . At  $t = 1, 2, \dots, \infty$ , the pointer moves by doubling the angle  $\omega$ . If the pointer comes to rest in the first half of the dial, the screen of Adam's black box shows 0; otherwise it shows 1.

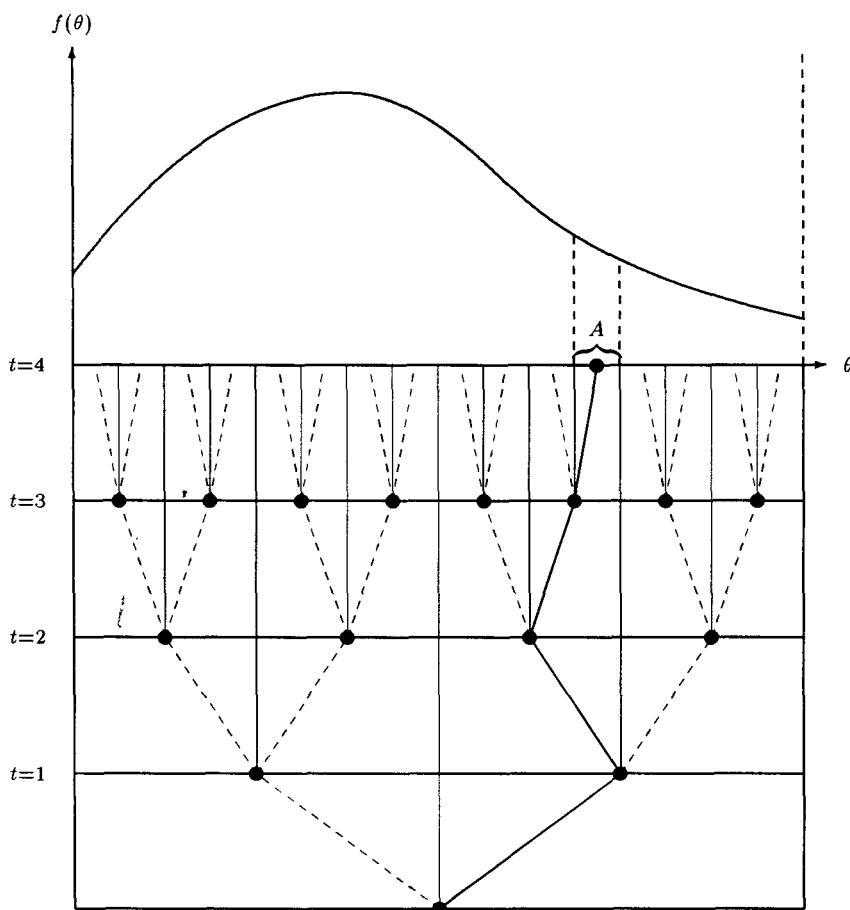
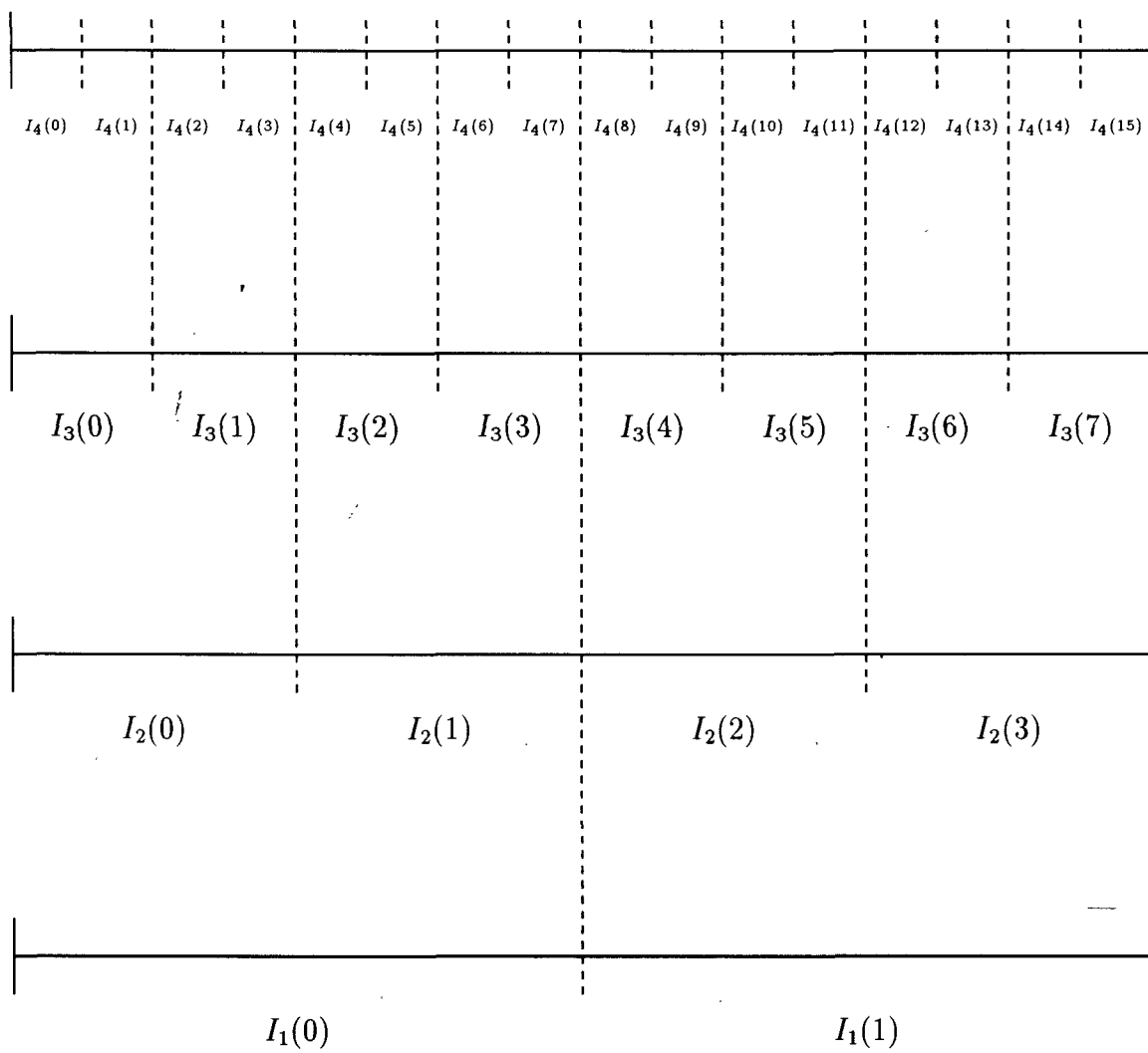


Fig. 2: Illustration of the example by a tree diagram. The abscissa shows the set  $I$  of all possible starting points  $\theta$  for the process described by (6). The ordinate shows a subjective density  $f(\theta)$  on  $I$ . Below the coordinate system is the tree of possible sequences of 0s and 1s for  $T = 4$  periods. The left branch of each node corresponds to the occurrence of  $x_t = 0$ . The sequence of 0s and 1s determined by  $\theta \in A$  (where  $A$  is the basic interval  $I_4(11)$ ) is given by the unique path through the tree leading to  $A$ . For  $T = 4$ , all  $\theta$  in  $A$  lead to the same path (indicated by solid lines) since their dyadic development is identical for the first four places. Thus the subjective probability for the solid-line path is  $\int_A f(\theta) d\theta$ .



*Fig. 3: Basic intervals.* The figure demonstrates the numbering of the basic intervals.