INDUCTIVE LEARNING IN SMALL AND LARGE WORLDS

SIMON M. HUTTEGGER

1. Introduction

Bayesian treatments of inductive inference and decision making presuppose that the structure of the situation under consideration is fully known. We are, however, often faced with having only very fragmentary information about an epistemic situation. This tension was discussed in decision theory by Savage (1954) in terms of 'small worlds' and 'large worlds' ('grand worlds' in Savage's terminology). Large worlds allow a more fine grained description of a situation than small worlds. The additional distinctions may not be readily accessible, though. Savage realized that planning ahead is difficult, if not impossible, whenever we don't have sufficient knowledge of the the basic structure of an epistemic situation. Since planning is at the heart of Bayesian inductive inference and decision making, it is not easy to see how a Bayesian—or any learner, for that matter—can deal with incomplete information.

The aim of this paper is to outline how the mathematical and philosophical foundations of inductive learning in large worlds may be developed. First I wish to show that there is an important sense in which Bayesian solutions for inductive learning in large worlds exist. The basic idea is the following: Even if one's knowledge of a situation is incomplete and restricted, Bayesian methods can be applied based on the information that is available. This idea is more fully developed in §§4, 5 and 6 for two concrete inductive learning rules that I introduce in §2. Importantly, however, this does not always lead to *fully rational* inductive learning: the analysis of a learning rule within the confines of the available information is by itself not sufficient to establish its rationality in larger contexts. In §3 I couch this problem in terms of Savage's discussion of small worlds and large worlds. I take this thread up again in §7, where inductive learning in small and large worlds is examined in the light of bounded rationality and Richard Jeffrey's epistemology of 'radical probabilism'.

2. Two Learning Rules

In order to fix ideas we introduce two abstract models of learning. Both models provide rules for learning in a decision context. The decision context is convenient for developing a theory of learning in large and small worlds, but it is by no means necessary, as I hope to make clear at the end of this section.

Suppose that nature has K states $\{1, \dots, K\}$ which are chosen repeatedly, resulting in a sequence of states. We don't make any assumptions about the process that

Date: July 2015.
I would like to thank Brian Skyrms, Elliott Wagner and Kevin Zollman for comments on earlier drafts of this paper. This paper was presented at the University of Groningen, where the audience provided helpful .

generates this sequence. For example, in a stationary environment nature might choose each state i with a fixed probability p_i, the choices being independent. Alternatively, the sequence of states of nature could be a stationary Markov chain. But mother nature need not be stationary: the states might also be the strategic choices of players in a game theoretic environment.

We assume also that there is an agent who can choose among M acts $\{1, \ldots, M\}$. The outcome of a choice by nature and a choice by the agent is a state-act pair. For each such pair there is assumed to be a von Neumann-Morgenstern utility which represents the agent's preferences over outcomes.[1] Suppose that the agent can observe states of nature. She may then have predictive probabilities about the state that will be observed on the next trial. Let X_n denote the state of nature at the nth trial. In the language of probability theory, X_n is a random variable that takes on values in $\{1, \ldots, K\}$. For a sequence of states $X_1, \ldots, X_n$, let n_i be the number of times state i has been observed. Then the following is a natural way to calculate the predictive probability of observing state i on trial $n + 1$:

$$\mathbb{P}[X_{n+1} = i | X_1, \ldots, X_n] = \frac{n_i + \alpha_i}{n + \sum_j \alpha_j} \tag{1}$$

for $i = 1, \ldots, K$ and some non-negative constants α_j. This rule is a generalization of Laplace's rule of succession. If all α_i are equal to zero, (1) is Reichenbach's straight rule (Reichenbach, 1949). Provided that some α_j are positive, then prior to making observations ($n = 0$) the predictive probability for state i is equal to $\alpha_i / \sum_j \alpha_j$. By calculating the expected utility of each act relative to the predictive probabilities (1), the agent can choose an act that maximizes this expected utility. Such an act is called a *(myopic) best response*. The resulting rule is called *fictitious play*. Fictitious play was originally conceived as a device for calculating Nash equilibria in games. Since then it has become one of the most important rules studied in the theory of learning in games, and many of its properties are well understood.[2]

Fictitious play is a simple form of the dominant Bayesian paradigm; it combines Bayesian conditioning with maximizing expected payoffs. As such, it assumes that the agent has the conceptual resources for capturing what states of nature there are. Note, however, that fictitious play is a fairly simple type of Bayesian learning—Bayesian learners can be vastly more sophisticated. A Bayesian might not just choose a myopic best response, but might also contemplate the effects of her choice on future payoffs. Moreover, a Bayesian agent is not required to have predictive probabilities that are given by a rule of succession; in general, posterior probabilities might have a much more complex structure.

Some modes of learning need less fine-grained information. The rule we consider here does not need information about states but only about acts and payoffs. Suppose that there are L possible payoffs $\pi_1 < \pi_2 < \cdots < \pi_L$ (which can again be thought of as von Neumann-Morgenstern utilities of consequences, ordered from the least to the most preferred). Now suppose that the agent has chosen her ith action n_i times, n_{ij} of which resulted in payoff π_j. Then our second learning rule

[1] Von Neumann-Morgenstern utilities are given by a cardinal utility function. The utility scale determined by such a function is unique up to a choice of a zero point and the choice of a unit. See von Neumann and Morgenstern (1944).

[2] Fictitious play was introduced by Brown (1951). For more information see Fudenberg and Levine (1998) and Young (2004). Strictly speaking, one also needs to specify a rule for breaking payoff ties. In the present context it doesn't matter which one is adopted.

recommends to choose an act i that maximizes

$$\frac{\sum_j \pi_j n_{ij} + \sum_j \pi_j \alpha_{ij}}{n_i + \sum_j \alpha_{ij}}. \tag{2}$$

Up to the parameters α_{ij}, this is the average payoff that was obtained by act i. If the α_{ij} are positive, then $\sum_j \pi_j \alpha_{ij} / \sum_j \alpha_{ij}$ can be viewed as the agent's initial estimate of act i's future payoff.

The rule (2) is a type of *reinforcement learning scheme*, where $\sum_j \pi_j \alpha_{ij}$ is the agent's initial propensity for choosing act i and $\sum_j \pi_j n_{ij}$ is the cumulative payoff associated with act i. After having chosen an act i, the payoff is added to the cumulative payoff for i. The total reinforcement for an act is the sum of the initial propensity and the cumulative payoff for that act. The act with maximum average payoff is then chosen in the next trial. For this reason we call this rule *average reinforcement learning*.[3] Averaging mitigates the effect of choosing an act very often. Such an act may accrue a large cumulative reinforcement even if the payoffs at each trial are very small, and so may look attractive despite not promising a significant payoff on the next trial.

Fictitious play can also be regarded as a reinforcement learning scheme; but it is of a quite different kind, known as 'hypothetical reinforcement learning'. In hypothetical reinforcement learning cumulative payoffs are not just gotten by adding actually obtained payoffs; the agent also adds the 'counterfactual' payoffs she would have gotten if she had chosen other acts (Camerer and Ho, 1999). Fictitious play can use counterfactual payoff information because knowing the state of nature allows her to infer what the alternative payoffs would have been. Average reinforcement learning does not have the conceptual resources to make these inferences.

The difference between average reinforcement learning and fictitious play can also be expressed in another way by looking again at their inputs. One fundamental classification of learning rules for decision situations is whether or not they are *payoff based*.[4] Fictitious play is not payoff based because it does not just keep track of the agent's own payoffs. Average reinforcement learning, on the other hand, is payoff based since its only input is information about payoffs.

Both fictitious play and average reinforcement learning choose an act that seems best from their point of view. And both rules are *inductive* in the sense that they take into account information about the history before choosing an act. There are other inductive rules, to be sure. In repeated decision problems of the kind described above, any inductive learning rule has

(i) a 'conceptual system' capturing the information that must be available in order to use that rule and
(ii) a specification of how the rule maps information about the past to future choices.

The conceptual system of fictitious play consists of a set of states, a set of acts and a set of outcomes. The conceptual system of average reinforcement learning consists of a set of acts and a set of outcomes. Fictitious play maps finite sequences of states to acts (up to a rule that applies when two acts have the same expected utility). Average reinforcement learning maps sequences of pairs of acts and outcomes to

[3]Rules such as (2) are well known in the literature on bandit problems as 'greedy' learning rules, because they always choose what currently seems best (Berry and Fristedt, 1985).
[4]See Fudenberg and Levine (1998) and Young (2004).

acts (again up to a rule that breaks ties). Alternatively, both rules may be thought of as mapping histories to vectors whose components represent the values of acts (expected values and average reinforcements, respectively), or as ordering acts. Other learning rules use different kinds of inputs and map them to choices, either deterministically or probabilistically.

Inductive learning is of course not restricted to repeated decision problems. In general, we may take an inductive learning rule as a map from finite sequences of inputs to some types of outputs so that the rule is always defined for increasing sequences of inputs. More formally, let us say that the *conceptual system* of a learning rule is an ordered set $\mathcal{C}$ of sets $S_1, \ldots, S_k$, where the elements of each S_i are objects that are in some way epistemically accessible to an agent using that rule. The inputs are given by $S_1, \ldots, S_m$ and the outputs are given by $S_{m+1}, \ldots, S_k$. A finite history of inputs is a sequence of elements in $S_1 \times \cdots \times S_m$. An inductive learning rule maps each finite history to an output (an element in, or a subset of $S_{m+1} \times \cdots \times S_k$). Roughly speaking, sequences of inputs describe what the agent has learned, while outputs specify what is adjusted in response to learning.[5]

In the following section we refer to fictitious play, average reinforcement learning, and a decision context. But keep in mind that the approach developed here applies to inductive learning more generally.

3. Small and Large Worlds

Our two learning rules are interesting not just because of their relevance to learning in repeated decision situations, but also because the conceptual system of average reinforcement learning is a 'coarsening' of the conceptual system of fictitious play. This can be seen as follows. We may think of states, acts, and outcomes as propositions, as Richard Jeffrey does in his logic of decision (Jeffrey, 1965). Then the sets of states, acts, and outcomes together define a partition. This partition captures the knowledge structure underlying fictitious play. The conceptual system of average reinforcement learning, on the other hand, only has the set of acts and the set of outcomes as elements. The partition determined by these two sets is a coarsening of the partition underlying fictitious play. Thus the knowledge structure of fictitious play is more refined than the knowledge structure of average reinforcement learning.

As an example, consider the decision problem given in the following table:

	S_1	S_2	S_3
A	\$1	\$2	\$2
B	\$2	\$2	\$1

There are three states of the world (S_1, S_2 and S_3) and two acts (A and B). If A is chosen and the state of the world is S_1, then the agent gets \$1; likewise for the other entries. The conceptual resources of a fictitious player allow her to capture events such as 'I choose B, the state of the world is S_2, and I get \$2', which is one element

[5]A conceptual system might also contain other objects, such as relations between sets, which are not going to be relevant for us. Furthermore, this is not the most general characterization of an inductive learning rule. For instance, inductive learning does not require that the agent maps any finite sequence of inputs to outputs. An agent's memory could be limited. In this case only sequences of some fixed length are relevant. Learning might also not go on forever, so that there is some upper bound for sequences of inputs. Incorporating these changes would not make any difference for the arguments in this paper.

in the partition given by states, acts and outcomes. If she is told that the true state is neither S_1 nor S_3, she knows that she cannot win \$1. If she is told that she wins \$1, then she knows that she cannot be in state S_2. The information available to average reinforcement learning is less fine grained. Here we can only express events such as 'I choose B and I get \$2'. Each such event is a proper superset of a fictitious play event. If in the new partition you are told that you don't get \$1, you only know that you get \$2 without knowing which state of the world is the true one. Thus, in a very precise sense, average reinforcement learning does in general not require as much information about a decision situation as fictitious play. This is true for all payoff based learning rules in any decision problem where states are not uniquely identifiable in terms of acts and payoffs, which will be the case whenever the payoff of an act is the same in more than one state.

Using a term introduced by Savage (1954), the conceptual system of a payoff based learning rule may be thought of as a *small world*. Savage distinguishes small worlds from *large worlds*. The large world that pertains to a decision situation is a fully considered description of the decision problem at hand. This means that *every* potentially relevant act, state of the world or outcome has entered the description of the decision problem.[6] If we again view a decision problem in terms of its associated partition, the large world is the finest partition that is relevant for a given decision situation. Any small world, on the other hand, is a coarsening of that finest partition that ignores some large world distinctions. In between the large world and a particular small world we may have worlds that are smaller or larger relative to one another. In particular, the world underlying a payoff based learning rule such as average reinforcement learning is a smaller world than the world of fictitious play or other non-payoff based learning methods.

The distinction between small and large worlds is related to another important topic in decision theory: bounded rationality.[7] Bounded rationality goes back mainly to the work of Herbert Simon, who maintained that real world reasoning and decision making is not captured adequately by the standards of high rationality models (e.g. Simon, 1955, 1957). To illustrate this point, consider the extreme case of a large world, namely, a person's whole life (Savage, 1954, p. 83). In this large world one is choosing how to live, once and for all. This choice is made after having considered the decision situation in full detail. This is evidently unrealistic even for agents with fairly sophisticated reasoning powers. For most kinds of agents it is possible to find less extreme large worlds that are beyond the bounds given by plausible epistemic constraints for that type of agent.

What we should take from this discussion is that decision making and inductive learning nearly always takes place in a small world, that is, a coarsening of an underlying large world. This allows us to discuss the rationality of learning rules such as fictitious play or average reinforcement learning at the small world level or the large world level. At the small work level we may ask whether a learning rule is rational within its conceptual system. The kind of rationality I am referring to here aims at identifying the principles underlying a learning rule. Consider fictitious

[6]In other words, for a large world decision problem we require that there is no proposition that, when added to the description of the decision situation, would disrupt the agent's preferences. See Joyce (1999, p. 73) for a more precise definition.

[7]This relationship was recently discussed by, e.g., Binmore (2009), Gigerenzer and Gaissmaier (2011) or Brighton and Gigerenzer (2012).

play, and take its take its conceptual systems as given. We may wonder whether the way it calculates predictive probability is just arbitrary or can be based on reasonable principles. The same can be asked with regard to average reinforcement learning.

This part of the project is to some extent a purely mathematical venture. Its methodology is the axiomatic method, which was used very successfully in many fields of modern mathematics, starting with Hilbert's foundations of geometry and used by Kolmogorov in his theory of probability. The probabilistic theories of inductive inference due to Bruno de Finetti and Rudolf Carnap are especially salient for our project. Both de Finetti and Carnap derive inductive learning rules from a set of basic postulates. In the next section (§4) I explain how these postulates can be used to apply the axiomatic method to fictitious play. At this level, the resulting theory of inductive learning could be treated as a mathematical theory without interpretation. This would not be satisfying from a philosophical perspective, and both Carnap and de Finetti thought of their projects as normative epistemological theories. I'll explain their positions briefly in §5, arguing for a position very close to de Finetti's view. On this view, the postulates from which a learning rule can be derived are *inductive assumptions* about the basic structure of the learning situation. If an agent's beliefs conform to those inductive assumptions, then—since fictitious play follows deductively from the postulates—it is the only admissible learning rule. The agent should adopt fictitious play on pain of inconsistency.

What is new about my approach is that the same methodology can be applied to average reinforcement learning. Given the differences between average reinforcement learning and fictitious play, it might not be immediately obvious that this is possible. In §6 I show two things: (i) Based on the conceptual system of average reinforcement learning there is a set of plausible postulates from which that learning rule can be derived, and (ii) these postulates can be thought of as inductive assumptions. Given that an agent believes a particular set of inductive assumptions, she has to be an average reinforcement learner on pain of inconsistency. This yields the same kind of theory of rational learning as for fictitious play, but for a learning rule with bounded resources.

The normative status of different modes of inductive learning (fictitious play, average reinforcement learning) is based on the consistency between inductive assumptions and the learning procedure. Rationality here is correct reasoning within an abstract small world model of a learning situation. But, as in the case of decision theory, the question of whether learning procedures are rational goes beyond the small world context. Are the inferences that are judged to be rational in the small world also rational in larger worlds? And how is this related to the conceptual abilities of agents? These questions are especially relevant for payoff based learning rules because of their coarse conceptual basis.

The relationship between small and large worlds is an important but also very complex question[8], and I don't claim to have a fully satisfying answer to all these issues. For now let me set aside this discussion. I'll pick it up again after having laid out the small world foundations of learning rules which was outlined in this section. This will put us in a better position to gain some qualitative insights into the more complex issues.

[8]Cf. the discussion in Savage (1954, p. 82-91) and Joyce (1999, p. 70-77).

4. De Finetti's Theorem and Inductive Logic

The well-known key for providing a foundation for fictitious play is the notion of *exchangeability*. Exchangeability originated with W. E. Johnson's permutation postulate (Johnson, 1924). But not before Bruno de Finetti's work on exchangeable sequences of random events was it used with full force.

Suppose that $X_1, X_2, \ldots$ is an infinite sequence of random variables (which may as in §2 be thought of as recording the occurrences of the states of nature $\{1, \ldots, K\}$). The sequence $X_1, X_2, \ldots$ is said to be *exchangeable* if the probability $\mathbb{P}$ of any finite initial sequence of states only depends on the number of times each state occurs and not on their order. More precisely, let $(j_1, \ldots, j_n)$ be a vector whose components j_i are elements of $\{1, \ldots, K\}$. Then the sequence is exchangeable if

$$\mathbb{P}[X_1 = j_1, X_2 = j_2, \ldots, X_n = j_n] = \mathbb{P}[X_1 = j_{\sigma(1)}, X_2 = j_{\sigma(2)}, \ldots, X_n = j_{\sigma(n)}]$$

for any permutation σ of $\{1, \ldots, n\}$ and any $n = 1, 2, \ldots$. If, for example, $K = 3$, then

$$\mathbb{P}[1,2,3] = \mathbb{P}[3,1,2] = \mathbb{P}[2,3,1] = \mathbb{P}[1,3,2] = \mathbb{P}[3,2,1] = \mathbb{P}[2,1,3]$$

because all these sequences have one state 1, one state 2, and one state 3. But it need not be the case that, e.g., $\mathbb{P}[1,1,2] = \mathbb{P}[1,2,3]$.

If the sequence $X_1, X_2, \ldots$ is exchangeable, then de Finetti's celebrated representation theorem states that the probability measure $\mathbb{P}$ is a mixture of independent multinomial probabilities.[9] For $i = 1, \ldots, K$, let n_i be the number of times i is found among $(j_1, \ldots, j_n)$. Also, let Δ^K be the set of all probability distributions $(p_1, \ldots, p_K)$ on the set $\{1, \ldots, K\}$, where p_i is the probability of i. If $X_1, X_2, \ldots$ is exchangeable, then there exists a unique prior measure $d\mu$ on Δ^K such that for every n and every $(j_1, \ldots, j_n)$

$$\mathbb{P}[X_1 = j_1, X_2 = j_2, \ldots, X_n = j_n] = \int_{\Delta^K} p_1^{n_1} \cdots p_K^{n_K} \, d\mu(\mathbf{p}) \tag{3}$$

with $n_1 + \cdots n_K = n$ and $\mathbf{p} = (p_1, \ldots, p_K) \in \Delta^K$. This theorem has several remarkable consequences.[10] One concerns the metaphysics of chance. A radical subjectivist such as de Finetti thinks of objective chances as illusions. Because of the representation theorem, a subjectivist is nonetheless licensed to use chances if her subjective probabilities are exchangeable. The elements in Δ^K can be viewed as chance parameters. Given the chance parameters, the agent believes that observations are like trials which are independently and identically distributed according to those chance parameters. The mixing measure $d\mu$ can be viewed as the agent's prior beliefs over chances. Yet again, a strict subjectivist would insist that the measure $d\mu$ is nothing but a useful fiction that one is allowed to entertain *because* of the representation theorem.[11] An agent does not need to believe in true chances. But if her degrees of beliefs are exchangeable, the agent behaves as if she did.

Another consequence of de Finetti's theorem is relevant for inductive reasoning. The formula (3) can be used to calculate the conditional probability of observing a certain state given past observations. The past observations determine a posterior

[9] See de Finetti (1937). There are much more general versions of the theorem. For an excellent survey see Aldous (1985).

[10] See Zabell (1989).

[11] This is the main difference between classical Bayesians, such as Laplace and Bayes himself, and modern subjective Bayesians.

from the prior $d\mu$. The conditional probability of a state given past observations is then just the chance expectation relative to the posterior.

One concrete instance of this conditional probability turns out to be especially salient for our purposes. Suppose the prior measure $d\mu$ has a Dirichlet distribution on Δ^K.[12] A straightforward calculation shows that the conditional probability of observing state i at time $n+1$ given all past observations is equal to

$$\frac{n_i + \alpha_i}{n + \sum_j \alpha_j}.$$

Thus, the predictive probabilities in (1) result from the representation theorem whenever one's degrees of belief are exchangeable and one has a Dirchlet prior over chances.

De Finetti himself emphasizes the qualitative implications of his theorem for the problem of induction, and the artificiality of particular choices of prior distributions over chances.[13] This is in line with the view that chance priors are only useful fictions. Now, Dirichlet priors are certainly mathematically convenient—but is there a deeper reason for using them as one's mixing measure?

Inductive logic provides an answer to this question. The type of inductive logic I'm referring to goes back to W. E. Johnson, and was later independently developed by Rudolf Carnap. Inductive logic also starts with exchangeability. But instead of using Dirichlet priors, Johnson and Carnap assume what is often referred to as 'Johnson's sufficientness postulate.' The sufficientness postulate states that the probability of observing category i on the next trial only depends on i, the number n_i of i's observed to date, and the total number of trials n:

$$\mathbb{P}[X_{n+1} = i | X_1, \ldots, X_n] = f_i(n_i, n) \tag{4}$$

In order for the conditional probabilities in (4) to be well defined, we must have

$$\mathbb{P}[X_1 = j_1, \ldots, X_n = j_n] > 0 \quad \text{for all } j_1, \ldots, j_n \in \{1, \ldots, K\}. \tag{5}$$

The regularity condition (5) says that each finite initial sequence of observations has positive prior probability.

It can be proved that if $K \geq 3$ the predictive probabilities given in (1) follow from exchangeability together with (4) and (5).[14] Johnson's sufficientness postulate can thus be viewed as a characterization of Dirichlet priors. If an agent thinks that knowledge of i, n_i and n is sufficient for determining the probability of i on the

[12] That is, $d\mu$ is

$$\frac{\Gamma\left(\sum_{j=1}^K \alpha_j\right)}{\prod_{j=1}^K \Gamma(\alpha_j)} p_1^{\alpha_1 - 1} \cdots p_K^{\alpha_K - 1} dp_1 \cdots dp_{K-1},$$

where Γ is the gamma function.

[13] E.g. de Finetti (1938). On p. 203 of the translation (where Φ denotes the prior) he writes: "It must be pointed out that precise applications, in which Φ would have a determinate analytic expression, do not appear to be of much interest: as in the case of exchangeability, the principal interest of the present methods resides in the fact that the conclusions depend on a gross, qualitative knowledge of Φ, the only sort we can reasonably suppose to be given (except in artificial examples)."

[14] For details, see Zabell (1982), who reconstructs the proof sketch from Johnson's posthumously published article (Johnson, 1932). It has to be assumed that there are at least three states, for otherwise the sufficientness postulate is empty. If there are only two states, one can either assume that the predictive probabilities are linear in n_i, or consider relevance quotients as in Costantini (1979).

next trial given the outcomes so far, then she acts *as if* she has a Dirchlet prior over the unknown chances of the K types of observations.

5. Rational Foundations

We see that there are two main approaches to deriving the probabilities used by fictitious play. One rests on subjective Bayesian probability theory. The other one is based on inductive logic.[15] Both attempts succeed in providing an axiomatic basis for fictitious play (together with von Neumann-Morgenstern utility theory). However, going beyond the purely mathematical aspect of the theories, we may ask in what sense we get a rational foundation from the axiomatic approach.

One view is that assumptions like exchangeability, the sufficientness postulate, or regularity are requirements of rationality—something rational beliefs have to conform to. Let's consider exchangeability as an example. At least in his early work on inductive logic, Carnap seems to have thought of exchangeability as a logical principle. Carnap was working within the context of probability functions on sentences of a first-order language. In this context exchangeability is the invariance of probability under permutations of the individual constants of the language. Based on the idea of Alfred Lindenbaum and Alfred Tarski that logical properties have to be invariant under certain transformations of the elements of the language, Carnap viewed exchangeability as a basic a priori principle of inductive logic. Accordingly, one is rationally required to have exchangeable probabilities (Carnap, 1950, §§90, 91).

The problem with this idea is that in applications to problems of inductive inference exchangeability has usually a temporal component, or otherwise refers to known and unknown observations. The information conditioned on involves observations that have already been made, and what is predicted is unknown as of yet. Within the context of inductive logic on a first-order language this means that the individual constants are temporally ordered. This results in what Carnap calls a 'coordinate language.' But in such a language the principle of exchangeability loses its innocence and makes a substantial claim about what the agent expects to observe.[16] The claim is that the observations one expects to make are homogenous. It is not easy to see how one can uphold the logicality of exchangeability or other symmetry principles in the face of the very concrete claims they make about random processes.[17]

De Finetti and other subjective Bayesians offer a different vision of the rationality of learning rules. On this view, exchangeability or Johnson's sufficientness postulate are not taken as having the status of necessary principles. Instead, they are an agent's personal judgements, or *inductive assumptions*, regarding the basic structure of her observations.[18] Inductive assumptions are basic facts about an agent's prior probability. They express the agent's beliefs about the whole learning situation rather than just some aspects of it. Exchangeability is a prime example of an inductive assumption. It says that *all* trials are completely analogous; it doesn't make a difference whether you observe an outcome on the first trial or on a later

[15]The two are closely related; see Skyrms (1996).

[16]Carnap was well aware of this point; see Carnap (1971, 118-120)

[17]For a thorough discussion of these issues see Zabell (1989).

[18]I borrow the term 'inductive assumptions' from Howson (2000).

trial. Exchangeability is thus a *symmetry* or *invariance* principle. Inductive assumptions are often expressed as symmetry principles. Such principles tell us what an agent thinks is of relevance about the whole learning situation.

What distinguishes de Finetti's subjectivistic view from the logical view of Carnap is that an agent is not required to have exchangeable degrees of belief but is allowed to use other symmetries (like the one discussed in the next section). But then, what makes inductive learning rational on the subjectivistic view if inductive assumptions (exchangeability, sufficientness postulate) are not themselves requirements of rationality? The rationality of an inductive learning rule is due to its being *consistent* with the agent's inductive assumptions. For instance, the Carnapian predictive probabilities (1) follow from exchangeability, Johnson's sufficientness postulate, and regularity. If the agent accepts these assumptions, she is bound to learn from experience according to a generalized rule of succession. In fact, it is easy to see that all three conditions are necessary for Carnapian predictive probabilities. It would be inconsistent to claim that one's conditional probabilities are Carnpian predictive probabilities while to deny at the same time that, say, one's unconditional probabilities are exchangeable.

The modest subjectivist understanding of rational learning is thus not as sweeping as some might wish. This is not the place to explain why it is the correct position; de Finetti and others have forcefully argued for it in many places.[19] I'm only going to make a few remarks concerning the epistemic status of inductive assumptions. If they are not rationality postulates, how are they justified? Let's again consider exchangeability. It is a property of an agent's degrees of belief. How does an agent come to hold exchangeable degrees of belief? In some cases, she might have good reasons for it. For example, our agent 's judgment that a sequence of coin flips is exchangeable might be based on past experiences with similar coins. But de Finetti and other subjective Bayesians do not insist upon inductive assumptions having this kind of epistemic warrant. The reason is that subjective Baysianism is not a foundationalist epistemology (like Carnap's early inductive logic) where beliefs must be grounded on some bedrock of fundamental beliefs and principles.[20] Bayesians take a much more permissive attitude towards inductive assumptions. They may be just what you currently happen to believe. On this picture inductive assumptions don't need to have an ultimate foundation. They are taken as a starting point for learning from experience. But of course, they are themselves revisable as well. If you detect a pattern in a sequence of coin flips—the sequence you observe is, say, $10101010\ldots$—you may give up the assumption of exchangeability at some point. Bayesian epistemology thus shares features with what Harman (2002) calls a 'general foundations theory'.[21] In a general foundations epistemology at least some of an agent's beliefs at a time are taken to be justified by default or until proven incorrect. Belief revision changes initial beliefs only in the face of sufficiently strong evidence. The rational justification a subjective Bayesian gets for a learning rule (such as Carnapian predictive probabilities) is therefore a relative one: it is relative to her inductive assumptions, which are beliefs that she currently holds.

---

[19]E.g. de Finetti (1930, 1937, 1959), Savage (1967), Jeffrey (1992), or Howson (2000).

[20]For a location of Carnap's program as a foundational theory see the opening essay in Jeffrey (1992).

[21]Harman traces it back to Goodman (1955) and Rawls (1971). See also Harman (2007).

Even if you think that the subjectivistic position is too weak, and that there is more to rational learning, we can take that position as a minimally plausible one for our next goal—developing the foundations for average reinforcement learning. My claim is that average reinforcement learning can be approached with the same axiomatic methodology as fictitious play. Now, if there is a defensible stronger sense of rational induction than the one underlying the subjectivistic point of view—something along the logical position of Carnap—that sense could then also be used for average reinforcement learning or other learning rules. Thus, while the following axiomatic theory is based on the subjective Bayesian understanding of rational learning, it could be made to fit a stricter approach.

6. Foundations of Average Reinforcement Learning

We now show that average reinforcement learning involves predictive probabilities for payoff events that are based on a generalization of exchangeability and some additional inductive assumptions.

Suppose that the agent has chosen her ith action n_i times, n_{ij} of which resulted in payoff π_j. Given that she chooses i on trial $n+1$, the conditional probability of obtaining payoff π_j on this trial may be equal to

$$\frac{n_{ij}+\alpha_{ij}}{n_i+\sum_j \alpha_{ij}} \quad \text{for } i=1,\ldots,M \text{ and } j=1,\ldots,L. \tag{6}$$

The α_{ij} are non-negative parameters. With respect to the predictive probabilities, the agent's expected payoff when choosing act i is

$$\sum_j \pi_j \frac{n_{ij}+\alpha_{ij}}{n_i+\sum_j \alpha_{ij}} = \frac{\sum_j \pi_j n_{ij}+\sum_j \pi_j \alpha_{ij}}{n_i+\sum_j \alpha_{ij}}. \tag{7}$$

The expression on the right side is the central quantity of average reinforcement learning (2). Hence, an agent using (2) chooses acts that maximize expected payoffs relative to the predictive probabilities (6).

In order to examine the predictive probabilities in (6) more thoroughly we consider a generalization of exchangeability called 'partial exchangeability'. This notion goes back to de Finetti, too (de Finetti, 1938, 1959). Partial exchangeability is important in various developments of Bayesian inductive logic.[22] The basic idea of partial exchangeability is that exchangeability obtains only within particular types of outcomes and not across types. De Finetti gives the example of tossing two coins which might not look altogether the same. The tosses of each coin are exchangeable, while the tosses of both coins together need not be. Persi Diaconis and David Freedman discuss an example of observations of patients that fall into four categories (given by two pairs of distinctions, male-female and treated-untreated). Again, observations within each category may be judged exchangeable without assuming exchangeability across categories. Exchangeability is a limiting case where categories are judged to be completely analogous, whereas partial exchangeability allows us to express weaker kinds of analogies.

To be more precise, suppose that there are M types $\{1,\ldots,M\}$. For average reinfrocement learning a type is an act i. For each type i there is an infinite sequence of random variables $X_{i1}, X_{i2}, \ldots$ taking values in a set of outcomes $\{1,\ldots,L\}$. The intended interpretation of outcome j is that a payoff π_j is obtained. Types and

[22] See for instance Freedman (1962) or Diaconis and Freedman (1980).

outcomes might in general capture other aspects of a random process. Notice that we assume that each type is observed infinitely often. The observations can be thought of as an infinite array:

$$\begin{array}{l} X_{11}, X_{12}, \dots \\ X_{21}, X_{22}, \dots \\ \vdots \\ X_{M1}, X_{M2}, \dots \end{array} \tag{8}$$

Let n_i be the number of times type i was observed in the first n trials, and let n_{ij} be the number of times outcome j was observed together with type i. Naturally, $\sum_i n_i = n$ and $\sum_j n_{ij} = n_i$. Then the array (8) is partially exchangeable if, for any n, all finite arrays $X_{11}, \dots, X_{1,n_1}; X_{21}, \dots, X_{2,n_2}; \dots; X_{M1}, \dots, X_{M,n_M}$ that have the same counts n_{ij} $(i = 1, \dots M, j = 1, \dots L)$ have the same probability. The numbers n_{ij} are a *sufficient statistic* for a partially exchangeable probability distribution. Thus the array (8) is partially exchangeable if the probability $\mathbb{P}$ of any finite initial sequence only depends on the number of times each outcome occurs within each type and not on their order.

de Finetti showed that his representation theorem for exchangeable random variables extends to partial exchangeability (de Finetti, 1938, 1959). Let $(\Delta^L)^M$ be the M-fold product of the set of probability distributions Δ^L on $\{1, \dots, L\}$. Let $X_{11}, X_{12}, \dots; X_{21}, X_{22}, \dots; \dots; X_{M1}, X_{M2}, \dots$ be a partially exchangeable array where each type i occurs infinitely often. Let the components of

$$(j_{11}, \dots, j_{1,n_1}; j_{21}, \dots, j_{2,n_2}; \dots; j_{M1}, \dots, j_{M,n_M})$$

be elements of $\{1, \dots, L\}$, and denote by n_{ij} the number of events j of type i. Then there exists a unique probability measure μ on $(\Delta^L)^M$ such that

$$\begin{aligned} \mathbb{P}[X_{11} = j_{11}, \dots, X_{1,n_1} = j_{1,n_1}; \dots; X_{M,1} = j_{M,1}, \dots, X_{M,n_M} = j_{M,n_m}] \\ = \int_{(\Delta^L)^M} \prod_{i=1}^{M} p_{i1}^{n_{i1}} \cdots p_{iL}^{n_{iL}} d\mu(\mathbf{p_1}, \dots, \mathbf{p_M}), \end{aligned} \tag{9}$$

where $\mathbf{p_i} = (p_{i1}, \dots, p_{iL}) \in \Delta^L$ for $i = 1, \dots, M$. This means that partially exchangeable sequences of random variables also are mixtures of independent trials. But, unlike the exchangeable case, the probabilities of the independent trials can be unequal. Equiprobability only obtains within each type i. Different types may have different probabilities.

We want to predict the probability of getting the payoff π_j for the next time act i is chosen. Under the assumption of partial exchangeability, the conditional probability of obtaining π_j is

$$\frac{\int_{(\Delta^L)^M} p_{ij} \prod_{k=1}^{M} p_{k1}^{n_{k1}} \cdots p_{kL}^{n_{kL}} d\mu(\mathbf{p_1}, \dots, \mathbf{p_M})}{\int_{(\Delta^L)^M} \prod_{k=1}^{M} p_{k1}^{n_{k1}} \cdots p_{kL}^{n_{kL}} d\mu(\mathbf{p_1}, \dots, \mathbf{p_M})}. \tag{10}$$

We now show how the probabilities (6) can be derived from this expression. If the mixing measure $d\mu$ is a product measure $d\mu_1 \times \cdots \times d\mu_M$, then

$$\int_{(\Delta^L)^M} \prod_{k=1}^{M} p_{k1}^{n_{k1}} \cdots p_{kL}^{n_{kL}} d\mu(\mathbf{p_1}, \dots, \mathbf{p_M}) = \prod_{k=1}^{M} \int_{\Delta^L} p_{k1}^{n_{k1}} \cdots p_{kL}^{n_{kL}} d\mu_k(\mathbf{p_k}),$$

and the conditional probability (10) reduces to

$$\frac{\int_{\Delta^L} p_{ij} p_{i1}^{n_{i1}} \cdots p_{iL}^{n_{iL}} d\mu_i(\mathbf{p_i})}{\int_{\Delta^L} p_{i1}^{n_{i1}} \cdots p_{iL}^{n_{iL}} d\mu_i(\mathbf{p_i})}. \tag{11}$$

Let's also assume that each measure $d\mu_i$ has a Dirichlet distribution on Δ^L. That is, the measure $d\mu$ is given by the product

$$\prod_{i=1}^{M} \frac{\Gamma\left(\sum \alpha_{ij}\right)}{\prod_j \Gamma\left(\alpha_{ij}\right)} p_{i1}^{\alpha_{i1}-1} \cdots p_{iL}^{\alpha_{iL}-1} dp_{i1} \cdots dp_{i,L-1}.$$

Then, just as in §4, the predictive probabilities (6) of average reinforcement learning follow from (11). Notice the key assumption that $d\mu$ is a product measure. This means that choosing act i does not influence opinions about payoffs resulting from choosing some other act k. This assumption essentially brings us back to the case of exchangeability (as noted in de Finetti, 1938).

One might wonder if inductive logic could again be used to provide a deeper foundation for a product of Dirichlet priors. To see that this is possible, we again assume that the infinite array (8) is partially exchangeable. Then a version of Johnson's sufficientness postulate expresses the idea that only payoffs obtained whenever act i is chosen are relevant for predicting payoffs when i is chosen the next time:

$$P[X_{i,n_{i+1}} = j | X_{11}, \ldots, X_{1,n_1}; \ldots; X_{M1}, \ldots, X_{M,n_M}] = f_{ij}(n_{ij}, n_i) \tag{12}$$

In words, the probability of observing j the next time n_{i+1} that type i obtains only depends on i, j, n_{ij}, and n_i. For these conditional probabilities to be well defined we assume that

$$\mathbb{P}[X_{11} = j_{11}, \ldots, X_{1,n_1} = j_{1,n_1}; \ldots; X_{M,1} = j_{M,1}, \ldots, X_{M,n_M} = j_{M,n_M}] > 0 \tag{13}$$

for all possible combinations of elements $j_{11}, \ldots, j_{1,n_1}; \ldots; j_{M1}, \ldots, j_{M,n_M}$ of outcomes in $\{1, \ldots, L\}$.

In the appendix I show that the predictive probabilities in (6) can be derived from these assumptions. The leading idea is this: Observe that the subsequences of payoffs for each type are infinitely exchangeable sequences. Applying the sufficientness postulate (12) to these subsequences then reduces the setup to the case described in §4. The sufficientness postulate (12) can therefore be viewed as a subjective characterization of a prior product measure of Dirichlet distributions.

The arguments presented here make a number of significant assumptions. First, we assume a finite number of possible payoffs. To deal with other types of situations we would need to use a more general kind of inductive logic, such as the one suggested by Skyrms (1993). A second requirement is that the mixing measure $d\mu$ be a product measure. It would be interesting to investigate more general cases that allow analogical influences between different types of outcomes. Both assumptions could be relaxed by using a more general approach to predictive probabilities.[23]

With an axiomatic foundation of the predictive probabilities (7) in place, the average reinforcement learner can, just like the fictitious player, be viewed as choosing acts that maximize expected payoffs with respect to these predictive probabilities. This choice rule can be built on the von Neumann-Morgenstern theory of expected utility where the agent considers her acts as lotteries. If the agent's preferences

[23]On analogical predictive rules, see Romeijn (2006) and references therein.

meet the von Neumann-Morgenstern axioms, then there is a cardinal utility function for each outcome and she will choose an act that maximizes expected utility relative to this function and her predictive probabilities. This can be modified in various ways; we need not suppose that our agents are expected utility maximizers. Predictive probabilities can be used in other ways for making choices.[24]

The assumptions underlying average reinforcement learning—partial exchangeability, the modified sufficientness postulate and regularity—can also be thought of as inductive assumptions as explained in §5. On that view, any agent whose beliefs conform to these inductive assumptions is bound to learn according to average reinforcement learning.

7. Radical Probabilism and Bounded Rationality

What we have seen is that our inductive learning rules can be given the same kind of foundation in terms of inductive assumptions, even though one of them is more bounded than the other. One might object that we have only shown this for two special learning rules. Let me indicate how the methodology outlined for fictitious play and average reinforcement learning can also be applied to tother learning rules. In particular, I have in mind payoff based learning rules such as other types of reinforcement learning or trial and error learning rules[25]

A learning rule maps inputs to outputs (see §2). Thus, prior to developing an axiomatic foundation we need to specify the conceptual resources it needs for inputs and outputs. The resulting conceptual system captures the basic structure of the learning process. Once the conceptual system is in place, the next and crucial step is to derive the learning rule from a set of postulates. The postulates should be assumptions that (partially) describe an agent's degrees of belief about the learning process. Like exchangeability or partial exchangeability, they can be thought of as inductive assumptions. As such, they can be interpreted as those conditions on an agent's beliefs that mandate a belief dynamics which follows the learning rule.

This is a very general outline, and I do not claim that its strategy will always be successful. There may be learning rules that are too complex to find an underlying set of plausible and simple inductive assumptions. Even if it is possible in principle to do this, it could be very hard to actually prove the desired result. Nonetheless, it seems clear that our methodology can be implemented for many learning rules. Substantiating this claim is a task for future research, but there are some related issues that I'd like to say a bit more about. One set of issues was discussed above and is about the relationship between small and large worlds and rationality. The other one (not unrelated) aims at identifying a general epistemological framework that can be used as an umbrella for our learning rules. Let's start with the second issue and work our way to the first.

Many payoff based learning rules are quite non-Bayesian, much more so than average reinforcement learning. Can they be fit within the kind of Bayesian framework we have appealed to so far? An answer surely depends on what exactly we mean by 'Bayesian framework'. Both average reinforcmement learning and fictitious

[24]One could use, for example, a randomized form of expected utility maximization where an agent chooses suboptimally with some probability.

[25]The probably most important kind of reinforcement learning, Herrnstein-Erev-Roth learning, is discussed in Young (2004). An example for a trial and error rule is probe and adjust, which is introduced in Skyrms (2010).

play maximize expected payoffs relative to predictive probabilities. Many learning rules will not have that feature, and so will not be optimal in a decision theoretic sense. But since they are learning procedures it might be possible to relate them to *Bayesian learning* and establish them as rational learning procedures.

Indeed, the variant of Bayesianism dubbed 'radical probabilism' by Richard Jeffrey allows us to subsume many inductive learning rules under a coherent epistemological program. One central aspect of radical probabilism is its rejection of Bayesian conditioning as the only legitimate form of learning from experience.[26] Bayesian conditioning presupposes that there is an observational proposition that is learned for certain. If this is the case, then conditioning requires one to adopt as one's new probability the prior probability conditional on the observational proposition. As an example think of the formation of predictive probabilities in the fictitious play process where the agent conditions posterior probabilities on observations of states of the world.

Now Jeffrey points out that learning an observational proposition is not the only mode of learning from experience. His generalization of Bayesian conditioning, known as 'probability kinematics' or 'Jeffrey conditioning', provides an alternative rule for uncertain evidence (Jeffrey, 1965). Uncertain evidence does not come in terms of factual propositions. Rather, what is learned is expressed by how probabilities change. Jeffrey did not think that radical probabilism ends with probability kinematics. He suggested that there might be many other forms of learning from experience that are legitimate under various kinds of assumptions (Jeffrey, 1992).

Radical probabilism allows us to give a more definite interpretation of average reinforcement learning. Strictly speaking, payoffs or utilities are not facts that can be observed like a state of the world. They are the result of choosing an act and an unknown state of the world. And while they do not express the factual content of an agent's learning experience, payoffs register the effects of learning on the average reinforcement learner. Hence, it can be thought of as a probabilistic learning rule even though there is no factual proposition that expresses its dynamics. As such average reinforcement learning is a mode of learning in the radical probabilist's sense, just like Jeffrey conditioning. We can say more, in fact: The axiomatic foundation specifies those epistemic situations where it is adequate as a learning rule.

Radical probabilism also encompasses other payoff based learning rules or other probabilistic learning methods. Each such method gives rise to a probability space that captures an agent's beliefs about the dynamics of learning with increasing information. For some of them it is possible to provide a deeper foundation in terms of inductive assumptions.

Everything considered so far takes place within a given conceptual system. If that conceptual system fully exhausts an agent's conceptual abilities, then we have a fairly convincing account of rational learning. Even if from the outside one might be able to say that the agent's conceptual system is not ideal for capturing a learning situation, this seems to be irrelevant for judging the agent's internal rationality. Her mode of learning is rational given her non-ideal conceptual abilities. Let's call this 'learning at full capacity'.

But what if our agent is not learning at full capacity? By this I mean the following. An agent's conceptual abilities often allow for a wide variety of conceptual

[26]On this and other aspects see various essays in Jeffrey (1992) and also Bradley (2005).

systems, some of them being more fine grained and richer then others. Now, for a given learning situation she might adopt a learning rule whose conceptual system is not as fine grained as could be—her conceptual abilities would allow her to learn based on a richer conceptual system. Both in this case and the full capacity setting the agent may be called boundedly rational, but there is an important difference: A full capacity agent exhausts her conceptual apparatus, and the boundedness is due to inherent conceptual limits. In the language of small and large worlds, the learning process is set in what is from the agent's perspective the large world. In the second case, however, the boundedness is something the agent could overcome, at least in principle. The learning process takes place in what is from the agent's perspective a small world.

To illustrate the difference between the two cases, consider two average reinforcement learners. One of them learns at full capacity. That is, the conceptual system of average reinforcement learning is her large world. The other agent has the conceptual abilities to make more distinctions. Maybe she can also observe states of the world. But for whatever reasons she also learns according to average reinforcement learning. In the tradition of Herbert Simon both agents may be viewed as boundedly rational since they don't process all the information that could possibly be exploited. But in the full capacity case that is a judgement that we make from an external perspective—the full capacity learner cannot do better—while in the other case the judgement can be made from the agent's perspective.

Let's look at this a bit more closely. For the agent that learns at less than full capacity it is not clear whether she learns rationally even if her learning rule is consistent with her inductive assumptions in the small world. There are larger worlds in which her small world assessments could change. Consider again an average reinforcement learner who is not learning at full capacity. This learner probably could refine the partition given by acts and consequences. A refinement may also yield a more refined learning dynamics with more acts or more consequences. An agent might be capable to refine more radically, for example by throwing in states of the world. In this case we could have an average reinforcement learner who might in principle be a fictitious player, provided that she refines her conceptual system accordingly.

This shows that an agent who does not learn at full capacity has the option of learning quite differently in larger worlds. It is plausible that the agent would be better off learning in the larger world, all other things being equal, since it is a more considered version of the learning situation. From the large world perspective the small world estimates of probabilities and choices seem less than ideal because these estimates don't take into account all the information available in the larger world.

We arguably find ourselves often in a situations that. We don't take into account all the information that could potentially be relevant because it does not seem worth it or because the situation does not lend itself to obtain that information easily. In particular, an exhaustive list of relevant states of the world might be difficult to come by except in the most simple decision problems—hence the significance of payoff based learning rules. On the other hand, we often feel that the small world is largely sufficient for our purposes. This is the point where Herbert Simon's bounded rationality becomes crucial. It allows us to put the small world rationality of learning methods (consistency with inductive assumptions) in perspective. The

inductive inferences drawn might be correct within the small world of the learning rule without claiming that they would continue to be correct in a larger world. If an agent learns at less than full capacity, then her inductive inferences are boundedly rational since they need not hold up to the standards of larger worlds. This kind of "non-optimality" is a hallmark of bounded rationality (Selten, 2001). What we get in our new approach is a minimal condition on boundedly rational learning procedures: that they be rational in their small worlds.

Sometimes it might be possible to claim more. This point is stressed by Joyce (1999, p. 74–77) in the closely related context of decision theory. Joyce argues that for a small world decision to be rational it should meet two requirements. First, it should be rational in the small world; and second, the decision maker must be committed to the view that she would make the same decision in the large world. We can think about inductive learning in much the same way. An agent could be committed to the view that her small world inductive inferences, or some of their relevant qualitative features, would be the same if she was learning at full capacity (i.e. in the large world). For an agent who is so committed, small world inductive inferences are a kind of *best estimate* of their large world counterparts. If the agent thought that in the large world her inferences would be different, she couldn't be committed to her small world inferences in that way.

I'd like to emphasize that the full commitment is not necessary for a boundedly rational agent. We can be rational in a small world without thinking that our views are best estimates in the large world. This might be the most plausible view if we don't have a sense of about what the largest world that would potentially be accessible to us might be. In a situation like that we could still be able to think some refinements ahead to larger worlds which are not *the* large world. Concerning these larger worlds our small world opinions may be our best estimates for larger worlds. Taking this into consideration, there is not a simple dichotomy between bounded rationality learning (no commitment to the large world) and full rationality learning (full commitment to the large world). We instead can be committed to some more refined scenarios than the one we actually adopt without having that commitment for all larger ones.

This gives us a more nuanced understanding of rationality and bounded rationality. Bounded rationality is a graded concept ranging from small world rationality to large world rationality. That we have gradations does not mean that we can always compare different kinds of learners concerning the extent of their bounded rationality. The reason is that even if we have two learners with the same small world, their more fine grained partitions need not be comparable—neither partition is a refinement or a coarsening of the other.[27] Even so, what we have is an explication of bounded rationality in terms of small world rationality and commitments to larger worlds.

8. Conclusion

Bounded rationality has two sources: access to information and computational capacities (Simon, 1955). By considering learning rules where one has strictly less information than the other, we have focused on access to information. We have shown that there is an important sense in which both learning rules can be thought of as rational. Each learning rule can be consistent with an agent's

[27]A set of refinements of a given small world partitions is in general only partially ordered.

inductive assumptions. We have also explained how our learning procedures can be understood as learning from experience in the radical probabilist's sense. Finally, we have analyzed the concepts of bounded rationality and rationality with respect to our treatment of the two learning rules in terms of small and large worlds.

There are a number of open questions. Although we have indicated how our methodology carries over to other learning rules, much work remains to be done to fill out the details. On the formal side many learning rules await an axiomatic treatment. On the conceptual side it is open whether they can be fit within radical probabilism and what bounded rationality means in these cases.

Appendix: Inductive Logic

We assume that the following statements are true:

(1) The infinite array (8) is partially exchangeable.
(2) The sufficientness postulate (12) holds for all $n_1, \dots, n_M$.
(3) The regularity axiom (13) holds for all $n_1, \dots, n_M$ and for all possible combinations of elements $j_{11}, \dots, j_{1,n_1}; \dots; j_{M1}, \dots, j_{M,n_M}$ of outcomes in $\{1, \dots, L\}$.

In addition, we suppose that there are at least three outcomes ($L \geq 3$). Otherwise, like in the original Johnson-Carnap continuum of inductive methods, the sufficientness postulate is empty.

With these assumptions in place, we wish to prove that for all for all $n_1, \dots, n_M$:

$$\mathbb{P}[X_{i,n_i+1} = j | X_{11}, \dots, X_{1,n_1}; \dots; X_{M1}, \dots, X_{M,n_M}] = \frac{n_{ij} + \alpha_{ij}}{n_i + \sum_j \alpha_{ij}} \tag{14}$$

The proof is a slight variant of an argument by Sandy Zabell for Markov exchangeable sequences.[28] By (1) there is an infinite number of occurrences of type i. Let $Y_n = X_{i,n}$. Then the resulting embedded sequence is exchangeable.

Lemma. *The sequence $Y_1, Y_2, \dots$ is exchangeable.*

Proof. Suppose that the array (8) is independently distributed with parameters p_{ij}, where p_{ij} is the probability that j occurs when the trial belongs to group i (so outcomes are identically distributed within a group). Then the probability of the cylinder set $\{Y_1 = j_1, \dots, Y_n = j_n\}$ is

$$\mathbb{P}[Y_1 = j_1, \dots, Y_n = j_n] = p_{ij_1} \cdots p_{ij_n}$$

since at each time type i occurs the probability of outcome j_m is p_{ij_m} for $m = 1, \dots, n$. In this case the embedded sequence is i.i.d. Now suppose that the array (8) is partially exchangeable. By de Finetti's theorem for partial exchangeability the array is a mixture of independently distributed sequences. Hence the distribution of the embedded sequence is a mixture of i.i.d. trials by the above argument and thus exchangeable. □

The goal now is to show that (i) the embedded sequence $Y_1, Y_2, \dots$ satisfies Johnson's sufficientness postulate, and (ii) to apply the Johnson-Zabell theorem to it (Johnson, 1932; Zabell, 1982).

Let $(j_1, \dots, j_n)$ be a sequence of outcomes in $\{1, \dots, L\}$. Consider the cylinder set $\{Y_1 = j_1, \dots, Y_n = j_n\}$. This event corresponds to observing the outcomes $j_1, \dots, j_n$ in the first n trials of a fixed type i. Now consider all finite arrays that

[28] See Zabell (1995).

have exactly n outcomes of type i which result in the sequence $j_1, \ldots, j_n$. To each such finite array there corresponds a cylinder set, and the resulting countable family of cylinder sets $\{E_m\}$ is a countable partition of the event $\{Y_1 = j_1, \ldots, Y_n = j_n\}$. We now need the following lemma (a proof can be found in Zabell (1995)).

Lemma. *If $\{E_m\}$ is a countable partition of B such that $\mathbb{P}[B \cap E_m] > 0$ for all m and $\mathbb{P}[A|B \cap E_m]$ has the same value for all m, then $\mathbb{P}[A|B] = \mathbb{P}[A|B \cap E_m]$.*

The sufficientness postulate (2) implies that

$$\mathbb{P}[Y_{n+1} = j | \{Y_1 = j_1, \ldots, Y_n = j_n\} \cap E_m]$$

is independent of m. By the regularity axiom, $\mathbb{P}[\{Y_1 = j_1, \ldots, Y_n = j_n\} \cap E_m] > 0$ for all m. It follows from the lemma that

$$\mathbb{P}[Y_{n+1} = j | Y_1 = j_1, \ldots, Y_n = j_n] = \mathbb{P}[X_{i,n+1} = j | X_{11}, \ldots, X_{1,n_1}; \ldots; X_{M1}, \ldots, X_{M,n_M}].$$

The expression on the left side depends only on $n_i = n$ and on n_{ij}. In addition, by the first lemma, the sequence $Y_1, Y_2, \ldots$ is an infinite exchangeable sequence. Thus the predictive probabilities (14) follow from the Johnson-Zabell theorem applied to the infinite sequence $Y_1, Y_2, \ldots$. The parameters α_{ij} must be nonnegative.

References

Aldous, D. J. (1985). Exchangeability and related topics. *École d'Été de Probabailités de Saint-Fleur XIII – 1983. Lecture Notes in Mathematics*, 1117:1–198.

Berry, D. A. and Fristedt, B. (1985). *Bandit Problems: Sequential Allocation of Experiments*. Chapman & Hall, London.

Binmore, K. (2009). *Rational Decisions*. Princeton University Press, Princeton.

Bradley, R. (2005). Radical probabilism and Bayesian conditioning. *Philosophy of Science*, 72:342–364.

Brighton, H. and Gigerenzer, G. (2012). Are rational actor models "rational" outside small worlds. In Okasha, S. and Binmore, K., editors, *Evolution and Rationality*, pages 84–109. Cambridge, Cambridge University Press.

Brown, G. W. (1951). Iterative solutions of games by fictitious play. In Koopmans, T. C., editor, *Activity Analysis of Production and Allocation*, pages 374–376. Wiley, New York.

Camerer, C. and Ho, T. (1999). Experience-weighted attraction learning in normal form games. *Econometrica*, 67:827–874.

Carnap, R. (1950). *Logical Foundations of Probability*. University of Chicago Press, Chicago.

Carnap, R. (1971). A basic system of inductive logic, part 1. In Rudolf Carnap and Richard C. Jeffrey, editors, *Studies in Inductive Logic and Probability I*, pages 33–165. University of California Press, Los Angeles.

Costantini, D. (1979). The relevance quotient. *Erkenntnis*, 14:149–157.

de Finetti, B. (1930). Probabilismo. Saggio critico sulla teoria delle probabilit'a e sul valore della scienza. *Logos (Biblioteca di Filosofia, diretta da A. Alivera)*, pages 163–219. Translated as 'Probabilism. A Critical Essay on the Theory of Probability and on the Value of Science' in *Erkenntnis*, 31:169–223, 1989.

de Finetti, B. (1937). La prevision: ses lois logiques ses sources subjectives. *Annales d l'institut Henri Poincaré*, 7:1–68. Translated in Kyburg, H. E. and Smokler, H. E., editors, *Studies in Subjective Probability*, pages 93–158, Wiley, New York, 1964.

de Finetti, B. (1938). Sur la condition d'equivalence partielle. In *Actualités Scientifiques et Industrielles No. 739: Colloques consacré à la théorie des probabilités, VIième partie*, pages 5–18. Paris. Translated in Jeffrey, R. C., editor, *Studies in Inductive Logic and Probability II*, pages 193–205, University of California Press, Los Angeles, 1980.

de Finetti, B. (1959). La probabilita e la statistica nei raporti con l'induzione, secondo i dwersi punti di vista. In *Corso C.I.M.E su Induzione e Statistica*. Cremones, Rome. Translated in de Finetti, B, *Probability, Induction and Statistics*, chapter 9, Wiley, New York, 1974.

Diaconis, P. and Freedman, D. (1980). De Finetti's generalizations of exchangeability. In R. C. Jeffrey, editor, *Studies in Inductive Logic and Probability II*, pages 233–249. University of California Press, Los Angeles.

Freedman, D. (1962). Mixtures of Markov processes. *Annals of Mathematical Statistics*, 33:114–118.

Fudenberg, D. and Levine, D. K. (1998). *The Theory of Learning in Games*. MIT Press, Cambridge, Mass.

Gigerenzer, G. and Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, 62:451–482.

Goodman, N. (2000). *Fact, Fiction, and Forecast*. Cambridge University Press, Cambridge.

Harman, G.. (2002). Reflections on Knowledge and its Limits. *The Philosophical Review*, 111:417–428.

Harman, G. and S. Kulkarni (2007). *Reliable Reasoning: Induction and Statistical Learning Theory*. MIT Press, Cambridge.

Howson, C. (2000). *Hume's Problem. Induction and the Justification of Belief*. Clarendon Press, Oxford.

Jeffrey, R. C. (1965). *The Logic of Decision*. McGraw-Hill, New York. 3rd revised edition Chicago: University of Chicago Press, 1983.

Jeffrey, R. C. (1992). *Probability and the Art of Judgement*. Harvard University Press, Cambridge MA.

Johnson, W. E. (1924). *Logic, Part III: The Logical Foundations of Science*. Cambridge University Press, Cambridge, UK.

Johnson, W. E. (1932). Probability: The deductive and inductive problems. *Mind*, 41:409–423.

Joyce, J. M. (1999). *The Foundations of Causal Decision Theory*. Cambridge University Press, Cambridge.

Rawls, J. (1971). *A Theory of Justice*. Harvard University Press, Cambridge MA.

Reichenbach, H. (1949). *Theory of Probability*. University of California Press, Los Angeles.

Romeijn, J.-W. (2006). Analogical predictions for explicit similarity. *Erkenntnis*, 2006:253–280.

Savage, L. J. (1954). *The Foundations of Statistics*. Dover Publications, New York.

Savage, L. J. (1967). Implications of personal probability for induction. *The Journal of Philosophy*, 64:593–607.

Selten, R. (2001). What is bounded rationality? In G. Gigerenzer and R. Selten, editors, *Bounded rationality: The adaptive toolbox*, pages 13–36, MIT Press, Cambridge MA.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69:99–118.

Simon, H. A. (1957). *Models of Man*. Wiley, New York.

Skyrms, B. (1993). Carnapian inductive logic for a value continuum. *Midwest Studies in Philosophy*, 18:78–89.

Skyrms, B. (1996). Carnapian inductive logic and Bayesian statistics. In T. S. Ferguson, L. S. Shapley, J. B. M., editor, *Statistics, Probability, And Game Theory: Papers in Honor of David Blackwell*, pages 321–336. Institute of Mathematical Statistics, Hayward, CA.

Skyrms, B. (2010). *Signals: Evolution, Learning, and Information*. Oxford University Press, Oxford.

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

Young, H. P. (2004). *Strategic Learning and its Limits*. Oxford University Press, Oxford.

Zabell, S. L. (1982). W. E. Johnson's "sufficientness" postulate. *The Annals of Statistics*, 10:1091–1099.

Zabell, S. L. (1989). The Rule of Succession. *Erkenntnis*, 31:283–321.

Zabell, S. L. (1995). Characterizing Markov exchangeable sequences. *Journal of Theoretical Probability*, 8:175–178.

Department of Logic & Philosophy of Science, UC Irvine