Exploration and Correlation[*]

Evan Piermont[†]and Roee Teper[‡]

June 19, 2017

Abstract

We study the extent to which contemporaneous correlations across actions affect an agent's preferences over the different strategies in exploration problems. We show that such correlations carry no economic content and do not affect the agent's preferences and, in particular, her optimal strategy. We argue that for similar reasons there is an inherent partial identification of the beliefs in exploration problems. Nevertheless, even under the partial identification, we show there are explicit conditions allowing the modeler to test whether the agent is acting according to some Bayesian model.

Key words: Bandit problems; correlated arms; strong exchangeability.

---

[*] The authors wish to thank Eddie Dekel, Shaowei Ke, Ehud Lehrer, Kyoungwon Seo, Teddy Seidenfeld and Eran Shmaya for their comments and remarks. Special acknowledgment to Kfir Eliaz for discussions which motivated us to think about this project.

[†] University of Pittsburgh, Department of Economics, `ehp5@pitt.edu`

[‡] University of Pittsburgh, Department of Economics, `rteper@pitt.edu`

1 INTRODUCTION

Exploration models capture a common trade off between an immediate payoff and arrival of new information, which can potentially impact future decisions and payoffs.[1] In such models an agent has to choose, every period, one project out of several in which to invest. By observing the outcome of an investment, the agent learns both about the chosen project and, in case the outcomes across different projects are correlated, about other projects as well. Each decision is predicated on the tradeoff between the immediate value of the investment and the future value of the information obtained by observing the outcome. Therefore, the agent's optimal investment strategy is a function of the history of observed outcomes, the projects that will be feasible in the future, and her beliefs regarding the true process generating the outcomes of each project.

Can the agent's beliefs about the generation process be identified from her preferences over different investment strategies? While the true generating process jointly determines all projects' outcomes each period, when considering a fixed investment strategy, the agent cares only about the outcome of the chosen project. As such, the agent's choices can reveal only the margins of her beliefs—her beliefs about each individual project conditional on the observed history. To substantiate this conclusion, Appendix A provides a decision theoretic model, wherein we introduce a novel framework capturing the exploration-exploitation tradeoffs, and axiomatize the representation typically applied in bandit models. The representation uniquely pins down the agent's marginal beliefs regarding the underlying process. In light of this observation, we answer the following three questions: (i) What restrictions on the marginal beliefs ensure they arise from an exchangeable process jointly determining the projects' outcomes? (ii) When the marginals do arise from an exchangeable process, is it unique? and (iii) Can we draw insights from our identification on the theory of bandit problems?

Recall, an exchangeable process is one in which the belief does not depend on the order of information arrival. Exchangeability has long been the cornerstone of the subjectivist, Bayesian paradigm in the context of repeated experimentation,[2] and our interest in exchangeability is tantamount to an assumption that the agent places no special importance to the period in which an outcome was observed. Note, however, that de Finetti's exchangeability condition can not be directly tested in our framework, since the agent does not observe the outcomes of different projects simultaneously. Thus, to answer (i), we provide a condition termed *Across-Arm Symmetry*, which dictates that the marginal beliefs are invariant to jointly permuting both the order in which projects are chosen and the corresponding outcomes. Across-Arm Symmetry holds if and only if the agent's beliefs (i.e., the marginal beliefs assessable to the modeler) coincide with the marginals of an exchangeable process.

The answer to (ii) is more subtle. In the subsequent examples, we show that in the finite horizon case, where the agent chooses investment strategies over n-periods, her beliefs are sometimes (but not always) uniquely identified. Curiously, when considering the infinite horizon, the identification problem is more severe despite the fact that the modeler has access to more data. When the agent's beliefs arise from her preferences over infinite investment strategies, the consistent exchangeable model

---

[1] Exploration models were introduced by Robbins (1952) and have been extensively studied in the statistics literature (as bandit problems), and widely incorporated in economic models (as search problems, stopping problems, research and development, experimentation, portfolio design, etc). See Berry and Fristedt (1985) for an overview of classic results within the statistics literature. For a survey of economic applications see Bergemann and Välimäki (2008).

[2] See for example de Finetti (1972); Diaconis (1977); Schervish (2012).

is *never* completely identified. In particular, *contemporaneous correlations* (i.e., the likelihood of an outcome of project a in a period given the outcome of project b in the *same* period) carry no economic content in such exploration problems. And so, to answer (iii), the optimal strategy in infinite horizon bandit problems do not depend on contemporaneous correlations if the process generating the data is exchangeable. While this is a negative result from the modeler's vantage—the general stochastic process governing beliefs can only be partially identified—it is a boon to the agent: when solving an exploration problem, contemporaneous correlations can be ignored without changing the set of optimal strategies, simplifying her decision problem.

Exploration Models and Elicited Beliefs. Because in exploration environments the agent can chose only one project in each period, her preferences over the different strategies depend only on the *margins* of her beliefs. And vice versa, the agent can only reveal—through choice or preference over investment strategies—her history dependent beliefs over each project separately. Therefore, we take as our observable data the marginals of a stochastic process.[3] Specifically, there is a set of actions, $\mathcal{A}$, each element of which, a, is associated with the outcome space S_a. We are considering the family of processes over the outcomes of the different actions, where each period one and only one action is observed. Let $\mathbf{T} = (T_1, T_2, ...)$, where $T_i \in \{S_a\}_{a\in\mathcal{A}}$ for every i. Let $\mathcal{T}$ denote the set of all such sequences. For any $\mathbf{T}$ in $\mathcal{T}$, let $\zeta_{\mathbf{T}}$ be a distribution over $\mathbf{T} = (T_1, T_2, ...)$. We refer to these distributions as our observables; and, denoting $S = \prod_{a\in\mathcal{A}} S_a$, we assume a distribution, ζ, over $S^{\mathbb{N}}$ is not observable. When considering a finite horizon problem, $\mathcal{T}$ refers to all sequences of length N.

In this framework, we introduce a condition referred to as *Across-Action Symmetry (AA-SYM)* and Theorem 1 shows that it is necessary and sufficient for the observables to be consistent with an infinite horizon exchangeable process over the joint realizations (that is, $S^{\mathbb{N}}$). AA-SYM states that (i) if an event, E, which is in the intersection of $\mathbf{T}$ and $\mathbf{T}'$, then $\zeta_{\mathbf{T}}(E) = \zeta_{\mathbf{T}'}(E)$ and (ii) if $\mathbf{T}$ and $\mathbf{T}'$ are permutations of one another, and $E \in \mathbf{T}$ is the corresponding permutation of $E' \in \mathbf{T}'$ then $\zeta_{\mathbf{T}}(E) = \zeta_{\mathbf{T}'}(E')$. Informally, the probability of obtaining outcome x when taking action a followed by outcome y when taking action b, is the same as the probability of obtaining outcome y when taking action b followed by outcome x when taking action a. This is reminiscent of the symmetry (exchangeability) property, except, in each period the outcome space may change as different actions can be taken.

Example 1.A. *Consider the following example of a two-period problem where in each of the periods the agent has to chose between two projects, a and b, each of which can either succeed or fail:* $S_a = \{s_a, f_a\}$ *and* $S_b = \{s_b, f_b\}$. *The agent believes that each project will have exactly one success, equally likely to be in either period, and, moreover, believes the two projects will succeed and fail jointly.*[4] *The corresponding process* $\zeta \in \Delta(S^2)$ *is*

---

[3] This is precisely the set of parameters identified by the decision theoretic exercise in Appendix A.

[4] Examples 1.A and 1.B would have similar implications if we considered a somewhat less extreme point of view in terms of the probabilities. For instance, the same conclusions would have been reached had we considered projects yielding s or f in the first period with the same probability, while in the second period, a project that was a success in the first period also yields s in the second period with probability $\frac{4}{9}$, and given a failure in the first period, a project yields s in the second with probability $\frac{5}{9}$. Such negative auto-correlated processes are the typical representative of exchangeability in finite horizon models, and have been used extensively in the finance literature (see for example Poterba and Summers (1988); Berk and Green (2004) and references therein.)

		$n = 1$			
		s_a, s_b	s_a, f_b	f_a, s_b	f_a, f_b
$n = 0$	s_a, s_b	0	0	0	$\frac{1}{2}$
	s_a, f_b	0	0	0	0
	f_a, s_b	0	0	0	0
	f_a, f_b	$\frac{1}{2}$	0	0	0

(1A:J)

We assume the modeler cannot observe ζ *itself but instead the marginals,* $\zeta_{x,y}$ *for* $(x, y) \in \{a, b\}^2$*. It is easy to see that* (1A:J) *imparts the following restrictions.*

$$\begin{aligned} \zeta_{x,y}(s_x, s_y) &= \zeta_{x,y}(f_x, f_y) = 0 \\ \zeta_{x,y}(s_x, f_y) &= \zeta_{x,y}(f_x, s_y) = \frac{1}{2}. \end{aligned} \tag{1A:M}$$

Assume further that the per-period utility associated with each outcome is $u(s_a) = 9$, $u(f_a) = -9$, $u(s_b) = 18$, *and* $u(f_b) = -18$*. The agent is an expected utility maximizer, and her total utility is the sum across the two periods. Given these restrictions on preferences,* (1A:M) *determines the agent's valuation of investment plans. Indeed, for* $x, y, z \in \{a, b\}$*, let* $(x, (y, z))$ *denote the investment plan in which the agent takes action* x *in the first period, and* y *conditional on* x*'s success and* z *on* x*'s failure. The agent's valuations for investment plans are given as follows:* $V(x, (y, z)) = 0$ *if* $y = z$*, and*

$$\begin{aligned} V(a, (a, b)) = V(b, (a, b)) &= \frac{9}{2} \\ V(a, (b, a)) = V(b, (b, a)) &= -\frac{9}{2}. \end{aligned} \tag{1A:P}$$

First, notice that the marginal beliefs, (1A:M), satisfy AA-SYM: $\zeta_{a,b}(s_a, f_b) = \frac{1}{2} = \zeta_{b,a}(f_b, s_a)$, etc. Second, while tedious, it is routine to verify that we could have conversely begun with (1A:P) to uniquely recover the marginal beliefs, (1A:M). More generally in this vein, the marginal beliefs $\{\zeta_{\boldsymbol{T}}\}_{\boldsymbol{T} \in \mathcal{T}}$ can be identified from the agent's preferences over investment plans (see Appendix A for a complete analysis).

Lastly, the joint exchangeable belief, ζ as defined by (1A:J), is in fact uniquely determined by the marginal beliefs. Although we do not assume that ζ is observed, it is identified by the agent's preferences over investment plans. In particular, we can identify that the agent believes the two projects are perfectly correlated.

Example 1.B. *If, instead, the agent believed the projects were independent of each other (but otherwise the same as in the above example), she would entertain the following distribution:*

		$n = 1$			
		s_a, s_b	s_a, f_b	f_a, s_b	f_a, f_b
$n = 0$	s_a, s_b	0	0	0	$\frac{1}{4}$
	s_a, f_b	0	0	$\frac{1}{4}$	0
	f_a, s_b	0	$\frac{1}{4}$	0	0
	f_a, f_b	$\frac{1}{4}$	0	0	0

(1B:J)

which would impart a different set of marginals

$$\begin{aligned}\zeta_{x,x}(s_x, s_x) &= \zeta_{x,x}(f_x, f_x) = 0 \\ \zeta_{x,y}(s_x, f_y) &= \zeta_{x,y}(f_x, s_y) = \frac{1}{2} \qquad \textit{if } x = y \\ \zeta_{x,y}(s_x, f_y) &= \zeta_{x,y}(f_x, s_y) = \frac{1}{4} \qquad \textit{if } x \neq y.\end{aligned} \tag{1B:M}$$

This of course has a corresponding change in the agent's valuations: $V(x, (y, z)) = 0$ *if* $y = z$*, and*

$$\begin{aligned} V(a, (a, b)) &= -\frac{9}{2} \quad & V(b, (a, b)) &= \quad 9 \\ V(a, (b, a)) &= \quad \frac{9}{2} \quad & V(b, (b, a)) &= -9.\end{aligned} \tag{1B:P}$$

While the above example shows that the correlation between projects can potentially affect (or, be recovered from) the agent's preferences—or equivalently her marginal beliefs—it is not typical. The inherent observability constraint in the standard framework of experimentation generally bears a cost; the exchangeable process with which our observables are consistent is typically highly non-unique.

Example 2. *Let the actions and outcome be the same as Example 1. Consider the following joint distributions:*

		$n = 1$			
		s_a, s_b	s_a, f_b	f_a, s_b	f_a, f_b
$n = 0$	s_a, s_b	$\frac{1}{4}$	0	0	$\frac{1}{4}$
	s_a, f_b	0	0	0	$\frac{1}{4}$
	f_a, s_b	0	0	0	0
	f_a, f_b	$\frac{1}{4}$	0	0	$\frac{1}{4}$

		$n = 1$			
		s_a, s_b	s_a, f_b	f_a, s_b	f_a, f_b
$n = 0$	s_a, s_b	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
	s_a, f_b	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
	f_a, s_b	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
	f_a, f_b	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$

Each project's (unconditional) success is i.i.d. with probability $\frac{1}{2}$*; On the left panel, the two projects per-period successes and failures are perfectly correlated, whereas in the right they are independent. However, both joint distributions impart the exact same restrictions on marginal beliefs:* $\zeta_{x,y}(s_x, s_y) = \zeta_{x,y}(f_x, f_y) = \zeta_{x,y}(s_x, f_y) = \zeta_{x,y}(f_x, s_y) = \frac{1}{4}$ *for* $x, y \in \{a, b\}$*. The agent values* every *investment plan at 0. Therefore, the agent's valuation of all investment plans, and in particular her optimal strategy, are identical regardless of the correlation between the two actions.*

In Example 1 the agent's preferences over investment strategies perfectly revealed her perceived *contemporaneous* correlation between the two projects. In Example 2, we can infer nothing about how the agent perceives the contemporaneous correlation. The latter proves to be the rule. In the sequel, we show that in the infinite horizon problem, beliefs can *never* be fully identified. Fortunately, the obstruction can be precisely delineated; contemporaneous correlations stand as the only obstacle thwarting the identification of the agent's preferences.

To characterize the agent's beliefs in the absence of perfect identification, we introduce a subclass of the widely studied exchangeable processes termed *strongly exchangeable* processes. We elaborate. Assume there is an underlying distribution governing the joint realization of actions that is *intertemporally* i.i.d. This distribution is not known exactly, but there exists a prior probability over what it might be. The prior is updated every period upon the observation of the realization of actions. Due to de Finetti (1931); Hewitt and Savage (1955), these classical Bayesian updating processes are referred to as exchangeable. In a strongly exchangeable process, where the periodic state-space takes a product structure, the set of possible underlying distributions are such that outcomes across actions are independent. Thus, a strongly exchangeable process is one in which, conditional on the distributional parameter, outcomes are both inter-temporally and *contemporaneously* independent.

Despite strong exchangeability having more structure than classic exchangeability, it imparts no additional restrictions in our model. Theorem 3 shows that a family of observables satisfying AA-SYM is consistent with a strongly exchangeable process, and this process is unique. We conclude, strong exchangeability is the full characterization of exchangeability in our statistical framework and the lack of contemporaneous correlations carry no constraints beyond AA-SYM. Strong exchangeability does not introduce any restriction on the exploration problem the agent is facing—the agent's preferences over the different investment strategies, and in particular, her optimal strategy, are not affected by contemporaneous correlations.

Organization. The subsequent Section 2 briefly introduces the general exploration problem and discusses the extent to which the agent's beliefs of the underlying uncertainty play a role in her contemplation of the optimal strategy. Section 3 introduces the observable processes that represent the agent's beliefs. Here, we provide a statistical condition on observable processes, AA-SYM, so that the agent's beliefs are consistent with an exchangeable process. Section 4 introduces the notion of strong exchangeability and presents our identification (non) uniqueness result. Section 5 discusses the point of disagreement among Bayesians in environments of experimentation. Proofs are Section 6. A technical appendix contains a decision theoretic exercise in which we axiomatize an agent's preferences over investment strategies and provide behavioral conditions for the identification of beliefs as assumed in main text.

2 Beliefs and the Value of Investment Strategies

Consider a standard exploration problem. There is a set of consumption outcomes X, over which a utility function $u : X \to \mathbb{R}$ is defined, and a set of actions $\mathcal{A}$, where each action $a \in \mathcal{A}$ can yield any of the outcomes in $S_a \subseteq X$. Each period the agent has to chose one (and only one) action, the outcomes of which she observes and derives utility from. A history of length $n \in \mathbb{N}$ is a sequence of n action-outcome pairs, $(a_1, x_1; a_2, x_2; \cdots ; a_n, x_n)$, where $x_i \in S_{a_i}$ for every $i \in n$. The set of finite histories is denoted by $\mathcal{H}$. Similarly, an infinite history is an infinite sequence of action-outcome pairs, $(a_1, x_1; a_2, x_2; \cdots ; a_i, x_i; \cdots)$, where $x_i \in S_{a_i}$ for every $i \in \mathbb{N}$. Future payoffs are discounted by $\delta \in (0, 1)$, thus an infinite history $h = (a_1, x_1; a_2, x_2; \cdots)$ is valued according to discounted utility,

$$U(h) = \sum_{i \in \mathbb{N}} \delta^{i-1} u(x_i).$$

A strategy in this environment is a function, $\sigma : \mathcal{H} \to \mathcal{A}$, determining which action to take following every possible finite history. Since the outcome of a given action is uncertainty, the agent's beliefs determine which action she prefers to take following every history, and in sum, her optimal strategy. Towards formalizing this, let $\mathcal{S}_{\mathcal{A}} \equiv \prod_{a \in \mathcal{A}} S_a$, and $\mathcal{S} \equiv \prod_{n \geqslant 1} \mathcal{S}_{\mathcal{A}}$. The set $\mathcal{S}$ represents the grand state-space; a state determines the realization of each action in each period. The uncertainty over the state space, that is the agent's beliefs over what is the state generating the actions' outcomes, is captured through a probability (or, a process) ζ.

Given such a belief, ζ, every strategy, σ, induces a unique countably additive probability $\mathbf{P}_\sigma$ over the set of infinite histories.[5] The agent values a strategy according to its expected utility with respect to the probability it induces over infinite histories. That is,

$$V(\sigma) = \mathbf{E}_\sigma \left(U(h) \right), \tag{1}$$

where $\mathbf{E}_\sigma$ denotes the expectation operator with respect to $\mathbf{P}_\sigma$. The agent's optimal strategy is the one maximizing $V(\cdot)$.

Remark 1. *Denote by $\mu_{h,a}(x)$ the ζ-probability, that conditional on an n-period history h, action a yields outcome x in period $n+1$. Given two agents (U, ζ) and (U, ζ') such that $\{\mu_{h,a}\}_{h \in \mathcal{H}, a \in \mathcal{A}} = \{\mu'_{h,a}\}_{h \in \mathcal{H}, a \in \mathcal{A}}$ then $V = V'$: the agents rank strategies identically.*

To obtain Remark 1, notice that for a history of length n, $h \in \mathcal{H}$, and an action, $a \in \mathcal{A}$, $\mathbf{P}_\sigma(h; a, x)$ is defined by

$$\mathbf{P}_\sigma(h; a, x) = \mathbf{P}_\sigma(h) \mu_{h,a}(x) \tag{2}$$

if $\sigma(h) = a$ and $x \in S_a$. Otherwise, the probability is 0. By standard arguments $\mathbf{P}_\sigma$ is uniquely determined by its measure of finite histories. By examining (2) it is clear that $\boldsymbol{P}_\sigma$, and therefore $\boldsymbol{E}_\sigma$, depends only on $\{\mu_{h,a}\}_{h \in \mathcal{H}, a \in \mathcal{A}}$.

While we completely specify the underlying uncertainty over the joint realization of all actions following every history, the agent herself only takes into account the margins of this process: $\{\mu_{h,a}\}_{h \in \mathcal{H}, a \in \mathcal{A}}$. Equivalently, had we not specified ζ nor U and instead asked the agent her preferences over the different exploration strategies—assuming she is a discounted expected utility maximizer—we would only be able to uniquely identify the marginals, $\{\mu_{h,a}\}_{h \in \mathcal{H}, a \in \mathcal{A}}$ (and the utility parameters, u and δ). This latter point is made in a formal fashion in Appendix A—we conduct a decision theoretic exercise, construct the set of investment strategies, and provide the axioms allowing us to determine whether the agent is indeed a discounted subjective expected utility maximizer, as in Eq. (1).

3 The Statistical Framework

In order for a modeler to understand the DM's updating process, and whether it follows Bayes rule, we need to construct her beliefs regarding not only each action individually but also her beliefs regarding the correlation between actions. As we will see, in the generic case we have insufficient data to uniquely identify a (subjective) joint distribution. We will still, however, be able to identify a representative with unique properties.

Observable Processes. Consider the family $\mathcal{T}$ of all sequences of individual experiments (i.e., individual actions), where different experiments can be taken in the different periods. Let $\mathbf{T} = (T_1, T_2, ...)$ where $T_i \in \{S_a : a \in \mathcal{A}\}$ for every $i \geqslant 1$; so, each T_i corresponds to taking an action, say a, and expecting one of its outcomes, S_a. (Like before S_a corresponds to the set of possible outcomes.) $\mathcal{T}$ is then the collection of all such $\mathbf{T}$'s. For each $\mathbf{T} = (T_1, T_2, ...)$ let $\zeta_{\mathbf{T}} \in \Delta^{\mathcal{B}}(\mathbf{T})$ be a process over $\mathbf{T}$; a distribution over all possible outcomes when taking action T_1, followed by T_2, followed by T_3, etc.[6] For a given history of outcomes $h \in (T_1, T_2, ..., T_n)$, we say $h \in \mathbf{T}$ whenever $\mathbf{T} = (T_1, T_2, ..., T_n, T_{n+1}, ...)$.

---

[5] Endowed with the Borel sigma-algebra generated by all finite histories. We identify each finite history with the set of infinite histories that extend it.

[6] For any metric space M, denote $\Delta^{\mathcal{B}}(M)$ as the set of Borel probability distributions over M, endowed with the weak*-topology, and $\Delta(M)$ the subset of distributions with denumerable support.

Lastly, for a sequence of experiments $\mathbf{T} = (T_1, ..., T_n, T_{n+1}, ...)$ and a permutation $\pi : n \to n$, denote $\pi\mathbf{T} = (T_{\pi(1)}, ..., T_{\pi(n)}, T_{n+1}, ...)$.

A *Subjective Expected Experimentation (SEE) belief structure* is a family of processes $\{\zeta_{\mathbf{T}}\}_{\mathbf{T}\in\mathcal{T}}$, such that for any two sequences of experiments, $\mathbf{T}, \mathbf{T}' \in \mathcal{T}$, if there is some history, $h \in \mathbf{T} \cap \mathbf{T}'$, then $\zeta_{\mathbf{T}}(h) = \zeta_{\mathbf{T}'}(h)$. This latter condition imposes that the probability of outcomes today do not depend on which experiments are to be conducted in the future.

Remark 1 in the previous section and Theorem 6 in Appendix A state that the DM's belief over $\mathcal{S}$, ζ, is identified only up to the marginal beliefs, $\{\mu_{h,a}\}_{h\in\mathcal{H},a\in\mathcal{A}}$. Each such family of marginals uniquely determines an SEE belief structure in the obvious manner. Given $\{\mu_{h,a}\}_{h\in\mathcal{H},a\in\mathcal{A}}$ and a sequence $\mathbf{T} = (S_{a_1}, S_{a_2}, ...)$, $\zeta_{\mathbf{T}}$ is the unique (countably additive) process satisfying

$$\zeta_{\mathbf{T}}(h) = \mu_{\varnothing,a_1}(x_1) \cdot \mu_{(a_1,x_1),a_2}(x_2) \cdots \mu_{(a_1,x_1,...,a_{n-1},x_{n-1}),a_n}(x_n)$$

for every finite history $h \in \mathcal{H}$. In fact, SEE belief structures are exactly the set of processes that can be constructed from a family of marginal beliefs, $\{\mu_{h,a}\}_{h\in\mathcal{H},a\in\mathcal{A}}$. Indeed the inverse is as follows: fix $\{\zeta_{\mathbf{T}}\}_{\mathbf{T}\in\mathcal{T}}$ and some history $h = (a_1, x_1; \ldots; a_n, x_n)$. Let $\mathbf{T}$ be any sequence of experiments containing $(h; a, x)$. Define $\mu_{h,a}(x) = \zeta_{\mathbf{T}}(T_{n+1} = x|h)$. This is well defined since for any $\mathbf{T}$ containing $(h; a, x)$, the probability of $(h; a, x)$ is the same by the restriction on SEE belief structures.[7]

Exchangeable Processes and Consistency. Recall, $\mathcal{S}_{\mathcal{A}} \equiv \prod_{a\in\mathcal{A}} S_a$, and $\mathcal{S} \equiv \prod_{n\geqslant 0} \mathcal{S}_{\mathcal{A}}$. $\mathcal{S}$ represents the grand state-space; a state, s, determines the realization of each action in each period –an entity which is unobservable to the modeler.

Definition. *An SEE belief structure* $\{\zeta_{\mathbf{T}}\}_{\mathbf{T}\in\mathcal{T}}$ *is* ***consistent*** *with* $\zeta \in \Delta^{\mathcal{B}}(\mathcal{S})$ *if* $\zeta|_{\mathbf{T}} = \zeta_{\mathbf{T}}$ *for every* $\mathbf{T} \in \mathcal{T}$.

That is, $\{\zeta_{\mathbf{T}}\}_{\mathbf{T}\in\mathcal{T}}$ is consistent with some process ζ over $\mathcal{S}$ if for every sequence of experiments $\mathbf{T}$, the marginal of ζ to $\mathbf{T}$ (denoted by $\zeta|_{\mathbf{T}}$) coincides with $\zeta_{\mathbf{T}}$. In such a case the processes ζ, which we cannot observe, explains all our data.

Because it forms the basis subjective Bayesianism and for the statistical literature on bandit problems, we will pay particular attention to the class of *exchangeable* processes.

Definition. *Let* Ω *be a probability space and* $\hat{\Omega} = \prod_{n\geqslant 1} \Omega$. *The process* $\zeta \in \Delta^{\mathcal{B}}(\hat{\Omega})$ *is* ***exchangeable****, if there exists probability measure* θ *over* $\Delta^{\mathcal{B}}(\mathcal{S}_{\mathcal{A}})$*, such that*

$$\zeta(E) = \int_{\Delta^{\mathcal{B}}(\Omega)} \hat{D}(E) d\theta(D), \tag{3}$$

where for any $D \in \Delta^{\mathcal{B}}(\Omega)$*,* $\hat{D}$ *is the corresponding product measure over* $\hat{\Omega}$*.*

Remark 2. *If* ζ *is exchangeable, then* θ *is unique.*

Exchangeable processes were first characterized by de Finetti (1931, 1937) and later extended by Hewitt and Savage (1955). Their fundamental result states that a process $\zeta \in \Delta^{\mathcal{B}}(\hat{\Omega})$ is exchangeable if and only if for any finite permutation $\pi : \mathbb{N} \to \mathbb{N}$ and event $E = \prod_{n\in\mathbb{N}} E_n$, we have

$$\zeta(E) = \zeta(\prod_{n\in\mathbb{N}} E_{\pi(n)}). \tag{4}$$

---

[7] It is routine to check the above maps are continuous so that the bijection is in fact a homeomorphism.

From the economic vantage, a DM who understands there to be an exchangeable process governing the outcome of actions would be considered Bayesian. This is because, given the representation in Eq. 3, the DM (acts as if she) entertains a second order distribution, which she updates following every observation.

We would like to understand under what circumstances an SEE belief structure is a result of Bayesian updating. If we could infer from preferences the beliefs over the joint realizations of all actions, that is $\prod_{a\in\mathcal{A}} S_a$, then our questions would boil down to verifying whether this process satisfies exchangeability. However, we can only infer the beliefs over each action separately, and thus, our task remains. We need to find a condition on the family of $\zeta_{\mathbf{T}}$'s that determines whether it follows Bayes rule.

Definition. *An SEE belief structure* $\{\zeta_{\mathbf{T}}\}_{\mathbf{T}\in\mathcal{T}}$ *is* **Across-Arm Symmetric (AA-SYM)** *if*

$$\zeta_{\mathbf{T}}(h) = \zeta_{\pi\mathbf{T}}(\pi h)$$

for every $\mathbf{T} \in \mathcal{T}$*,* $h \in \mathbf{T}$ *and a permutation* $\pi : n \to n$*.*

Intuitively, AA-SYM requires that if we consider a different order of experiments, then the probability of outcomes (in the appropriate order) does not change. The next theorem states that across-arm symmetry is a necessary and sufficient condition for an SEE belief structure to be consistent with Bayesian updating of some belief over the *joint* realizations of all actions.

Theorem 1. *An SEE belief structure* $\{\zeta_{\mathbf{T}}\}_{\mathbf{T}\in\mathcal{T}}$ *satisfies AA-SYM if and only if it is consistent with an exchangeable process* $\zeta \in \Delta^{\mathcal{B}}(\mathcal{S})$*.*

Theorem 1 is stated without proof. Necessity is trivial and sufficiency will be a straightforward application of Theorem 3. Although the theorem as stated concerns only infinite-horizon process, AA-SYM is also a necessary and sufficient condition for an finite horizon processes to be consistent with some exchangeable process, provided there exists some consistent joint distribution.[8]

4 Strong Exchangeability and Contemporaneous Correlations

Unfortunately, AA-SYM is not sufficient to obtain a unique exchangeable process consistent with an SEE belief structure. This lack of identification stems directly from the inability to observe the DM's beliefs regarding *contemporaneous* correlations. Consider an infinite horizon version of Example 2: two coins are flipped each period. There are two possible governing processes, which are i.i.d. across time periods. (1) the coins are perfectly correlated (with equal probability on HH and TT), or (2) the coins are identical and independent (and both have equal probability on H and T). Notice, the two cases induce the same marginal distributions over each coin *individually.* Thus, if the modeler has access only to the DM's marginal beliefs, the two processes are indistinguishable.

In this section we introduce a strengthening of exchangeability, which we aptly call *strongly exchangeable*, under which stochastic independence is preserved both inter-temporally (as in vanilla

---

[8] The proof in the finite horizon case is quite intuitive. Let η be a consistent joint distribution. For each event E let $E^\star$ denote the union of πE for all $n!$ permutations $\pi : n \to n$, where n is the number of periods. Construct ζ as follows: $\zeta(E) = \frac{\eta(E^\star)}{n!}$. The process ζ is well defined and it is clearly exchangeable. Moreover, $\{\zeta_{\mathbf{T}}\}_{\mathbf{T}\in\mathcal{T}} = \{\eta_{\mathbf{T}}\}_{\mathbf{T}\in\mathcal{T}}$, since for all $E \in \boldsymbol{T}$, $\eta(E)$ is equal to $\eta(\pi E)$ and therefore also to $\zeta(E)$.

exchangeability) and *contemporaneously.*[9]

Definition. *A process $\zeta \in \Delta^{\mathcal{B}}(\mathcal{S})$ is* ***strongly exchangeable*** *if there exists a probability measure θ over $\Delta^{IN} \equiv \prod_{a \in \mathcal{A}} \Delta(S_a)$, such that*

$$\zeta(E) = \int_{\Delta^{IN}} \hat{D}(E) d\theta(D),$$

where for any $D \in \Delta^{IN}$, $\hat{D}$ is the corresponding product measure over $\mathcal{S}$.

Under a strongly exchangeable process the outcomes of actions that occur at the same time are independently resolved. Of course, this does not impose that there is no informational cross contamination between actions. Information regarding the distribution of action a is informative about the underlying parameter governing the exchangeable process, and therefore, also about the distribution of action b. Since exchangeable processes were first characterized as being invariant to permutations, for the sake of completeness we provide a similar characterization of strongly exchangeable processes.

Theorem 2. *The process $\zeta \in \Delta^{\mathcal{B}}(\mathcal{S})$ is strongly exchangeable if and only if for any set of finite permutations $\{\pi_a : \mathbb{N} \to \mathbb{N}\}_{a \in \mathcal{A}}$ and event $E = \prod_{n \in \mathbb{N}} \prod_{a \in \mathcal{A}} E_{n,a}$, we have*

$$\zeta(E) = \zeta(\prod_{n \in \mathbb{N}} \prod_{a \in \mathcal{A}} E_{\pi_a(n),a}). \tag{5}$$

Proof. In Appendix A.8. ■

Following the intuition above, it should come as no surprise that under AA-SYM strong exchangeability can never be ruled out. In other words, there is no SEE belief structure, therefore no preferences over PoAs, that distinguishes exchangeability from strong exchangeability. Strongly exchangeable processes are ones where each dimension can be permuted independently. If $\pi_a = \pi_b$ for all $a, b \in \mathcal{A}$, the condition is exactly exchangeability. Strongly exchangeable process are especially relevant with respect to the current focus because they act as representative members to the equivalence classes of exchangeable processes consistent with the same SEE belief structure.

Theorem 3. *An SEE belief structure $\{\zeta_{\mathbf{T}}\}_{\mathbf{T} \in \mathcal{T}}$ satisfies AA-SYM if and only if it is consistent with a strongly exchangeable process. Furthermore, such a strongly exchangeable process is unique.*

Proof. In Appendix A.8. ■

5 A COMMENT ON BAYESIANISM IN ENVIRONMENTS OF EXPERIMENTATION

The results in Section 4 have two related implications to Bayesianism in general models of experimentation. First, it is well known that the beliefs of two Bayesians observing the same sequence of signals will converge in the limit. Our results imply that in a setup of experimentation, different Bayesians obtaining the same information, might still hold different views of the world in the limit. Their beliefs over the uncertainty underlying each action will be identical, but they can hold different beliefs over the joint distribution.

The second point has to do with the possible equivalence with non-Bayesian DMs. Theorem 3 states that AA-SYM is necessary and sufficient for an SEE belief system to be consistent with some

---

[9] We feel reasonably certain that strong exchangeability must have been studied previously in the statistics literature. However, we have found no references.

exchangeable process. As discussed in the Introduction, AA-SYM projected to stochastic processes is weaker than the standard symmetry axiom applied in the literature, because the standard assumption requires that histories fully specify the evolution of the state, while in our setup, histories can only specify cylinders. Because AA-SYM is a weaker assumption, de Finetti's theorem implies that processes satisfying such an assumption need not be exchangeable and have a Bayesian representation as in Eq. (3).

Consider the following example of a stochastic process. In every period two coins are flipped. In odd periods the coins are perfectly correlated (with equal probability on HH and TT), and in even periods the coins are identical and independent (and both have equal probability on H and T). The associated observable processes satisfy AA-SYM, but the process itself is clearly not exchangeable. Nevertheless, Theorem 3 guarantees that there is a (unique) strongly exchangeable process that is consistent with the SEE belief structure. In this case it is easy to see that that process would be the one in which every period we toss two coins that are identical and independent (and both have equal probability on H and T).

6 Proofs

Proof of Theorem 2. First we show, if a strongly exchangeable process ζ over $\mathcal{S}$ is induced by an i.i.d distribution D over $\mathcal{S}_{\mathcal{A}}$, then it must be that the marginals of D (on $\{S_a\}_{a\in\mathcal{A}}$) are independent, that is $D \in \Delta^{IN}$. Indeed, consider two non-empty, disjoint collection of actions, $\hat{\mathcal{A}}, \hat{\mathcal{A}}' \subset \mathcal{A}$. Let $E, F \in S_{\hat{\mathcal{A}}}$, $E', F' \in S_{\hat{\mathcal{A}}'}$, be measurable events. Identify E^n with the cylinder it E generates in $\mathcal{S}$ when in the n^{th} coordinate: $E^n = \{s \in \mathcal{S} | s_{n,\mathcal{B}} \in E\}$. Since ζ is strongly exchangeable we have that

$$\zeta\left(E^n \cap E'^n \cap F^{n+1} \cap F'^{n+1}\right) = \zeta\left(E^n \cap F'^n \cap F^{n+1} \cap E'^{n+1}\right). \tag{2Sym}$$

We will refer to the latter weaker property as *two symmetry*. Now, since ζ is i.i.d generated by D, we have that (abusing notation by identifying E with the cylinder it generates in $S_{\mathcal{A}}$)

$$D(E \cap E') \cdot D(F \cap F') = D(E \cap F') \cdot D(F \cap E').$$

Substituting via the rule of conditional probability:

$$D(E|E') \cdot D(E') \cdot D(F|F') \cdot D(F') = D(E|F') \cdot D(F') \cdot D(F|E') \cdot D(E').$$

This implies that

$$\frac{D(E|E')}{D(E|F')} = \frac{D(F|E')}{D(F|F')}.$$

Since this is true for all events, we have that $D(E|E') = D(E|F')$ for every $E \in S_{\hat{\mathcal{A}}}$ and $E', F' \in S_{\hat{\mathcal{A}}'}$, implying $\hat{\mathcal{A}}$ and $\hat{\mathcal{A}}'$ are independent.

We now move to show that strong exchangeability is sufficient for the representation specified in the statement of the result. Since strong exchangeability implies exchangeability, we can apply de Finetti's theorem and represent the process ζ by

$$\zeta(\cdot) = \int_{\Delta(S_{\mathcal{A}})} \hat{D}(\cdot) d\psi(D).$$

We need to show that ψ's support lies in Δ^{IN}.

For $s \in \mathcal{S}$ and $t \in \mathbb{N}$ let s_t be the projection of s into the first t periods. Now, let $\zeta(\cdot|s_t) : S_{\mathcal{A}} \to [0, 1]$

be the *one period ahead predictive probability*, given that the history of realizations in the first t periods is s_t. Since ζ is exchangeable, $\zeta(\cdot|s_t)$ converges (as $t \to \infty$) with ζ probability 1. Moreover, the set of all limits is the support of ψ. Denote the limit for a particular s by D_s. Of course, the exchangeability of ζ also guarantees that $\zeta(\cdot,\cdot|s_t) : S_{\mathcal{A}} \times S_{\mathcal{A}} \to [0,1]$, that is the *two period ahead predictive probability*, converges to $D_s \times D_s$. Furthermore, ζ is strongly exchangeable; the limit itself satisfies (2Sym), and the arguments above imply that $D_s \in \Delta^{IN}$ with ζ probability 1. ∎

Proof of Theorem 3. Fix an SEE belief structure $\{\zeta_{\mathbf{T}}\}_{\mathbf{T}\in\mathcal{T}}$. We first construct a pre-measure $\hat{\zeta}$ on the semi-algebra of cylinder sets. Fix any ordering over $\mathcal{A}$. Set $\hat{\zeta}(\emptyset) = 0$ and $\hat{\zeta}(\mathcal{S}) = 1$. Let $E \neq \mathcal{S}$ be an arbitrary cylinder, i.e., $E = \prod_{n\in\mathbb{N}}\prod_{a\in\mathcal{A}} E_{n,a}$, such that for only finitely many (n,a), is $E_{n,a} \neq S_a$. Clearly, there are a finite number of $a \in \mathcal{A}$ such that $E_{k,a} \neq S_a$ for any k. By the ordering on $\mathcal{A}$ denote these $a_1 \dots a_n$. For each a_i let m_i denote the number of components such that $E_{k,a_i} \neq S_{s_i}$, and for $j = 1 \dots m_i$, let $k_{i,j}$ denote the j^{th} such component. Finally, for each a_i, let π_{a_i} denote any permutation such that $\pi_{a_i}(k_{i,j}) = j + \sum_{i'<i} m_{i'}$. Consider $\hat{E} = \prod_{n\in\mathbb{N}}\prod_{a\in\mathcal{A}} E_{\pi_a(n),a}$, where $\pi_a = \pi_{a_i}$ if $a \in a_1 \dots a_n$ and the identity otherwise. That is, for $n \in 1 \dots m_1$, $\hat{E}_{n,a} = S_a$ for all a except a_1, for $n \in m_1 \dots m_1 + m_2$, $\hat{E}_{n,a} = S_a$ for all a except a_2, etc. Let $\mathbf{T}(E)$ denote the sequence such that $T_n = S_i$ for $\sum_{i'<i} m_{i'} < n \leqslant \sum_{i'\leqslant i} m_{i'}$. Again that is, for $n \in 1 \dots m_1$, $T_n = S_{a_1}$, for $n \in m_1 \dots m_1 + m_2$, $T_2 = S_{a_2}$, etc.

For the remainder of this proof, for any cylinder E, $\hat{E}$ denotes the corresponding cylinder generated by the above process, in which at most a single action is restricted in each period. Let $\mathbf{T}(E)$ denote any observable process which observes the sequence of restricted actions. Finally, for any cylinder, E, which is restricted in most one action each period, and any $\mathbf{T}$ which observes each restricted set, identify E the relevant event in $\mathbf{T}$. So, Set $\hat{\zeta}(E) = \zeta_{\mathbf{T}(E)}(\hat{E})$.

To apply the Carathéodory extension theorem for semi-algebras, we need to show that for any sequence of disjoint cylinders $\{E^k\}_{k\in\mathbb{N}}$ such that $E = \bigcup_{k\in\mathbb{N}} E^k$ is a cylinder, $\hat{\zeta}(E) = \sum_{k\in\mathbb{N}} \hat{\zeta}(E^k)$. Towards this, assume that E, E' are disjoint cylinders such that $E \cup E'$ is a cylinder. Then it must be that there exists a unique (n,a) such that $E_{n,a} \cap E'_{n,a} = \emptyset$ and for all other (m,b), $E_{m,b} = E'_{m,b}$. Indeed, if this was not the case, then there exists some (m,b) and some x such that (WLOG) $x \in E_{m,b} \backslash E'_{m,b}$. But then, for all $s \in E \cup E'$, $s_{m,b} = x \implies s_{n,a} \in E_{n,a} \neq (E \cup E')_{n,a}$ a contradiction to $E \cup E'$ being a cylinder. But this implies $\hat{E}$ and $\hat{E}'$ induce the same sequence of restricted coordinates, differing on the restriction of single coordinate, and therefore, $\mathbf{T}(E) = \mathbf{T}(E')$. This implies that $\hat{E} \cup \hat{E}' \subseteq \mathbf{T}(E)$. Since $\zeta_{\mathbf{T}(E)}$ is finitely additive, so therefore $\hat{\zeta}(E \cup E') = \zeta_{\mathbf{T}(E)}(\hat{E} \cup \hat{E}') = \zeta_{\mathbf{T}(E)}(\hat{E}) + \zeta_{\mathbf{T}(E)}(\hat{E}') = \hat{\zeta}(E) + \hat{\zeta}(E')$.

Since $\hat{\zeta}$ is finitely additive over cylinder sets, countable additivity follows if we show that for all decreasing sequences of cylinders $\{E^k\}_{k\in\mathbb{N}}$, such that $\inf_k \hat{\zeta}(E^k) = \epsilon > 0$, we have $\bigcap_{k\in\mathbb{N}} E^k \neq \emptyset$. But this follows immediately from the finiteness of S_a. Since $E^{k+1} \subseteq E^k$, it must be that $E^k_{n,a} \subseteq E^k_{n,a}$. But each $E^k_{n,a}$ is finite, hence compact, and nonempty, because $\zeta(E^k) \geqslant \epsilon$. Therefore $\bigcap_{k\in\mathbb{N}} E^k_{n,a} \neq \emptyset$. The result follows by noting that the intersection of cylinder sets is the cylinder generated by the intersection of the respective generating sets. Let ζ denote the unique extension of $\hat{\zeta}$ to the σ-algebra on $\mathcal{S}$.

That ζ is consistent with $\{\zeta_{\mathbf{T}}\}_{\mathbf{T}\in\mathcal{T}}$ is immediate. We need to show that ζ is strongly exchangeable.

Let E be a cylinder. Let $\bar{\pi}_a$ denote a finite permutation for each $a \in \mathcal{A}$. Let $F = \prod_{n \in \mathbb{N}} \prod_{a \in \mathcal{A}} E_{\bar{\pi}_a(n),a}$. Let π_{a_i} denote the permutation given by the construction of $\hat{F}$. Then $\hat{F} = \prod_{n \in \mathbb{N}} \prod_{a \in \mathcal{A}} E_{\pi_a(\bar{\pi}_a(n)),a}$. This implies there exists some permutation π^* such that $\hat{F} = \prod_{n \in \mathbb{N}} \prod_{a \in \mathcal{A}} E_{\pi^*(n),a}$. By AA-SYM, $\zeta_{\mathbf{T}(\hat{E})}(\hat{E}) = \zeta_{\pi^*\mathbf{T}(\hat{E})}(\pi^*\hat{E}) = \zeta_{\mathbf{T}(\hat{F})}(\hat{F})$. Therefore, $\zeta(E) = \zeta(F)$ and so, by Theorem 2, ζ is strongly exchangeable.

Finally, the similar logic show that ζ is unique. Towards a contradiction, assume there was some distinct, strongly exchangeable ζ', also consistent with $\{\zeta_{\mathbf{T}}\}_{\mathbf{T} \in \mathcal{T}}$. Then, since the cylinder sets form a π-system, there must be some cylinder such that $\zeta(E) \neq \zeta'(E)$. But, by strong exchangeability, $\zeta(\hat{E}) = \zeta(E)$ and $\zeta'(\hat{E}) = \zeta'(E)$, so $\zeta(\hat{E}) \neq \zeta'(\hat{E})$ –a contradiction to their joint consistency with $\{\zeta_{\mathbf{T}}\}_{\mathbf{T} \in \mathcal{T}}$. ■

References

Francis J Anscombe and Robert J Aumann. A definition of subjective probability. *The annals of mathematical statistics*, 34(1):199–205, 1963.

Dirk Bergemann and Juuso Välimäki. Bandit Problems. In Steven N. Durlauf and Lawrence E. Blume, editors, *The New Palgrave Dictionary of Economics*. Second edition, 2008.

Jonathan B Berk and Richard C Green. Mutual fund flows and performance in rational markets. *Journal of political economy*, 112(6):1269–1295, 2004.

Donald A Berry and Bert Fristedt. *Bandit problems: sequential allocation of experiments (Monographs on statistics and applied probability)*. Springer, 1985.

Adam Brandenburger and Eddie Dekel. Hierarchies of Beliefs and Common Knowledge. *Journal of Economic Theory*, 59(1):189–198, 1993.

Kevin Cooke. Preference Discovery and Experimentation. 2016. *Working paper.*

Bruno de Finetti. Funzione caratteristica di un fenomeno aleatorio. 1931.

Bruno de Finetti. La prévision: ses lois logiques, ses sources subjectives. In *Annales de l'institut Henri Poincaré*, volume 7, pages 1–68, 1937.

Bruno de Finetti. *Probability, Induction, and Statistics.* New York: John Wiley, 1972.

Eddie Dekel, Barton L. Lipman, and Aldo Rustichini. Representing Preferences with a Unique Subjective State Space. *Econometrica*, 69(4):891–934, 2001.

Persi Diaconis. Finite forms of de finetti's theorem on exchangeability. *Synthese*, 36(2):271–281, 1977.

David Dillenberger, Juan Sebastián Lleras, Philipp Sadowski, and Norio Takeoka. A theory of subjective learning. *Journal of Economic Theory*, pages 1–31, 2014.

David Dillenberger, R. Vijay Krishna, and Philipp Sadowski. Dynamic Rational Inattention. 2015. *Working paper.*

Larry G Epstein and Kyoungwon Seo. Symmetry of evidence without evidence of symmetry. *Theoretical Economics*, 5(3):313–368, 2010.

L.G. Epstein and M. Le Breton. Dynamically consistent beliefs must be Bayesian. *Journal of Economic Theory*, 61:1–22, 1993.

Faruk Gul and Wolfgang Pesendorfer. Self-Control and the Theory of Consumption. *Econometrica*, 72(1):119–158, 2004.

Edwin Hewitt and Leonard J. Savage. Symmetric measures on Cartesian products. *Transactions of the American Mathematical Society*, 80(2):470, 1955.

Kazuya Hyogo. A subjective model of experimentation. *Journal of Economic Theory*, 133(1):316–330, 2007.

Alexander Kechris. *Classical descriptive set theory*, volume 156. Springer Science & Business Media, 2012.

Peter Klibanoff, Sujoy Mukerji, and Kyoungwon Seo. Perceived Ambiguity and Relevant Measures. *Econometrica*, 82(5):1–46, 2013.

David M Kreps. A Representation Theorem for "Preference for Flexibility". *Econometrica*, 47(3): 565–577, 1979.

R. Vijay Krishna and Philipp Sadowski. Dynamic Preference for Flexibility. *Econometrica*, 82(133): 655–703, 2014.

Ehud Lehrer and Roee Teper. Who is a Bayesian? 2015. *Working paper*.

Evan Piermont, Norio Takeoka, and Roee Teper. Learning the krepsian state: Exploration through consumption. *Games and Economic Behavior*, 100:69–94, 2016.

James M Poterba and Lawrence H Summers. Mean reversion in stock prices: Evidence and implications. *Journal of financial economics*, 22(1):27–59, 1988.

Herbert Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–536, 1952.

Mark J Schervish. *Theory of statistics*. Springer Science & Business Media, 2012.

A Plans of Action: Supplementary Material

In this appendix we provide a decision theoretic framework and axiomatization for an agent who is facing the classic exploration and exploitation tradeoff. We show that only the margin of the decision maker's belief can be identified from her ranking of the different strategies available in a bandit problem. The setup is one of infinite horizon, but note that even if we look at the restriction of the domain to finite problems (or a subset of the actions), then identification might be incomplete. Nevertheless, the axioms must still hold, and therefore, any invariance to contemporaneous correlations persists.

A.1 The Decision Theoretic Framework

A DM is tasked with ranking sequential and contingent choice objects: the action taken by the agent at any stage depends on the outcomes of previous actions. Formally, our primitive is a preference over *plans of action* (*PoAs*). Each action, a, is associated with a set of consumption prizes the action might yield, S_a. Then, a PoA is recursively defined as a *lottery over pairs* (a, f), where a is an action and f is a mapping that specifies the continuation PoA for each possible outcome in S_a. Theorem 5 shows that the construction of PoAs is well defined. So, a PoA specifies an action to be taken each period that can depend on the outcome of all previously taken actions. See Figures 1 and 2, where $f(x), f(y), f(z)$ are themselves PoAs. Each node in a PoA can be identified by a *history* of action-outcome realizations preceding it.

The actions in our model is in direct analogy to the arms of bandit problem (or actions in a repeated game). PoAs correspond to the set of all (possibly mixed) strategies in these environments. Note, however, the DM's perception of which outcome in S_a will result form taking action a is not specified. This is subjective and should be identified from the DM's preferences over PoAs. As discussed above, the main question is to what extent these beliefs can be identified and what are the economic implications of belief identification in this framework?

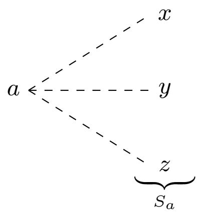

Figure 1: An action, a, and its support, S_a.

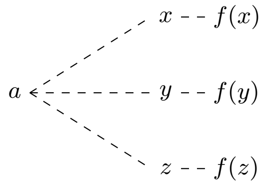

Figure 2: A degenerate PoA, (a, f).

Theorem 6 axiomatizes preferences over PoAs of a DM who at each history entertains a belief regarding the outcome of future actions. That is, at each history h and for every action a, the DM entertains a belief $\mu_{h,a}$ over the possible outcomes S_a; $\mu_{h,a}(x)$ is the DM's subjective probability that action a will yield outcome x, contingent on having observed the history h. Given this family of beliefs, the DM acts as a subjective discounted expected utility maximizer, valuing a PoA p, after observing h, according to a *Subjective Expected Experimentation* (*SEE*) representation:

$$U_h(p) = \mathbb{E}_p \left[\mathbb{E}_{\mu_{h,a}} \left[u(x) + \delta U_{h'}(f(x)) \right] \right], \tag{SEE}$$

where h' is the updated history (following h) when action a is taken and x is realized. All the parameters of the model—the consumption utility over outcomes, u, the discount factor, δ, and the history dependent subjective beliefs, $\{\mu_{h,a}\}_{h\in\mathcal{H},a\in\mathcal{A}}$—are identified uniquely.

The identification result accompanying the representation concerns the marginal beliefs, $\{\mu_{h,a}\}_{h\in\mathcal{H},a\in\mathcal{A}}$, and not a joint stochastic process over all actions, as is the starting point in the standard approach to bandit problems. In the main text, we explore the statistical information encoded in these marginal beliefs, and the extent to which a joint distribution can be identified. It is immediate that each behavioral strategy available to the agent in a bandit problem defines a unique plan of action, and vice versa . Moreover, simple algebra shows that the classical (time-separable discounted expected utility) valuation of behavioral strategies (see Eq. (1)) is the restriction of (SEE) to such plans.

A.2 Constructing Plans of Action.

Let X be a finite set of outcomes, endowed with a metric d_X. Outcomes are consumption prizes. For any metric space, M, let $\mathcal{K}(M)$ denote the set non-empty compact subsets of M, endowed with the Hausdorff metric. Likewise, for any metric space M, denote $\Delta^{\mathcal{B}}(M)$ as the set of Borel probability distributions over M, endowed with the weak*-topology, and $\Delta(M)$ the subset of distributions with denumerable support.

Let $\mathcal{A}$ be a compact and metrizable set of actions. Each action, a, is associated with a set of outcomes, $S_a \in \mathcal{K}(X)$, which is called the support of the action. We assume the map $a \mapsto S_a$ is continuous and surjective. For any metric space M, let $\mathcal{A} \otimes M = \{(a, f) | a \in \mathcal{A}, f\colon S_a \to M\} = \{(a, \{(x_i, m_i)\}_{i \in I}) \in \mathcal{A} \times \mathcal{K}(X \times M) | \bigcup_{i \in I}\{x_i\} = S_a \text{ and } x_i \neq x_j, \forall i \neq j \in I)\}$, endowed with the subspace topology inherited from the product topology. By the continuity of $a \mapsto S_a$ we know that the relevant subspace is closed and hence the topology on $\mathcal{A} \otimes M$ is compact whenever M is. We can think of f as the assignment into M for each outcome in the support of action a. For any $f : X \to M$ we will abuse notation and write (a, f) rather than $(a, f|_{S_a})$.

We will begin by constructing a more general notion of plans.[10] To begin, let $Q_0 = R_0 = \Delta^{\mathcal{B}}(\mathcal{A})$ and, for define recursively for each $n \geqslant 1$

$$\begin{aligned} Q_n &= \Delta^{\mathcal{B}}(\mathcal{A} \times \mathcal{K}(X \times Q_{n-1})) \text{ and,} \\ R_n &= \{r_n \in Q_n | r_n(\mathcal{A} \otimes R_{n-1}) = 1\} \end{aligned}$$

Define $Q^* = \prod_{n \geqslant 0} Q_n$ and $R^* = \prod_{n \geqslant 0} R_n$.

We restrict ourselves to the set of *consistent* elements of R^*: those elements such that, the $(n-1)$-period plan implied by the n-period plan is the same as the $(n-1)$-period plan. Let $G_1 : \mathcal{A} \times \mathcal{K}(X \times Q_0) \to \mathcal{A}$ as the mapping $(a, \{x, q_0\}) \mapsto a$. Let $F_1 : Q_1 \to Q_0$ as the mapping $F_1 : q_1 \mapsto \big(E \mapsto q_1(G_1^{-1}(E))\big)$, for any $E \in \mathcal{B}(\mathcal{A})$. Therefore, for any $E \in \mathcal{B}(\mathcal{A})$, $F_1(p_1)(E)$ is the probability of event E in period 0 as implied by p_1; $F_1(p_1)$ is the distribution over period 0 actions implied by p_1. From here we can recursively define $G_n : \mathcal{A} \times \mathcal{K}(X \times Q_n) \to \mathcal{A} \times \mathcal{K}(X \times Q_{n-1})$ as:

$$G_n : (a, \{x, q_{n-1}\}) \mapsto (a, \{x, F_{n-1}(q_0)\}))$$

and $F_n : Q_n \to Q_{n-1}$ as:

$$F_n : q_n \mapsto \big(E \mapsto q_n(G_n^{-1}(E))\big)$$

for any E in $\Delta^{\mathcal{B}}(\mathcal{A} \times \mathcal{K}(X \times Q_{n-1}))$. A consistent generalized plan is one such that

$$F_n(q_n) = q_{n-1}, \tag{6}$$

for all n. Let Q denote the restriction of $Q*$ that satisfies (6) and $R = Q \cap R^*$.

Proposition 4. *There exists a homeomorphism, $\lambda : R \to \Delta^{\mathcal{B}}(\mathcal{A} \otimes R))$ such that*

$$\text{marg}_{\mathcal{A} \times \mathcal{K}(X \times R_{n-1})}(\lambda(r)) = \text{proj}_n r. \tag{7}$$

Proof. In Section A.8 ■

---

[10] This methodology serves two purposes. First, the more general approach allows us to use standard techniques for the construction of infinite horizon choice objects. Second, generalized plans may be of direct interest in future work, when, for example, denumerable support is not desirable.

Finally, we want to consider plans whose support is denumerable. It is easy enough to set $P_0 = \Delta(\mathcal{A}) \subset R_0$, and define recursively $P_n = \Delta(\mathcal{A} \otimes P_{n-1}) \subset R_n$. Of course, there is a potential pitfall still lurking: for a given $\prod_{n \geqslant 0} P_n$, although each p_n is a denumerable lottery, the associated element, $\lambda(p)$ might live in $\Delta^B(\mathcal{A} \otimes P)$ rather than $\Delta(\mathcal{A} \otimes P)$. Indeed, we need also to restrict our attention to the set of plans that have countable support not just for each finite level, but also "in the limit," and whose implied continuation plans are also well behaved in such a manner. Fortunately, this can be done.

Theorem 5. *There exists maximal set $P \subset R$ such that for each $p \in P$, $\text{proj}_n p \in P_n$, and λ is a homeomorphism between P and $\Delta(\mathcal{A} \otimes P)$.*

Proof. In Section A.8 ■

The set P is our primitive. As a final notational comment, we would like to consider a further specification of *objective* plans, denoted by $\Sigma \subset P$. Σ denotes the set of plans which contain no subjective uncertainty; in every period, every possible action yields some outcome with certainty. Recall, for each $x \in X$ there is an associated action, a_x such that $S_{a_x} = \{x\}$. Associate this set of actions with X. Then $\Sigma_0 = \Delta(X)$ and, recursively, $\Sigma_n = \Delta(X \times \Sigma_{n-1})$. Finally $\Sigma = P \bigcap \prod_{n \geqslant 0} \Sigma_n$. That is, these plans specify only actions with deterministic outcomes at every stage. It is straightforward to show λ takes Σ to $\Delta(X \times \Sigma)$.

Histories. PoAs are infinite trees; each node, therefore, is itself the root of a new PoA—a distribution over action-continuation pairs. Each action-continuation, (a, f), in the support of a node contains branches to new nodes (PoAs). The branches emanating from an action coincide with the outcomes in the support of that action, $x \in S_a$. The node that follows x is the PoA specified by $f(x)$. Each node, therefore, is reached after a unique history: the history specifies the realization of the distribution of each pervious node, and outcome of the action realized. Thus, for a given PoA, p, each history of length n is an element of $\prod_{t=1}^{n} P \times [\mathcal{A} \otimes P] \times X$ such that $p^1 = p$ and

$$\begin{aligned} (a^t, f^t) &\in supp(p^t) \\ x^t &\in S_{a^t} \\ p^{t+1} &= f^t(x^t) \end{aligned}$$

Define the set of all histories of length n for p as $\mathcal{H}(p, n)$ and the set of all finite histories as $\mathcal{H}(p)$. Let $\mathcal{H}(n) = \bigcup_{p \in P} \mathcal{H}(p, n)$ and, $\mathcal{H} = \bigcup_{n \in \mathbb{N}} \mathcal{H}(n)$. For each $h \in \mathcal{H}(p, n)$, h corresponds to the node (PoA) defined by $f^n(x^n)$. Lastly, for any $p, q \in P$ and $h \in \mathcal{H}(p)$ define $p_{-h}q$ as the (unique!) element of P that coincides with p everywhere except after h in which case $f^n(x^n)$ is replaced by q. Note that the n period plan implied p and $p_{-h}q$ are the same. For any $p, q \in P$ and $n \in \mathbb{N}$, let $p_{-n}q \equiv \bigcup_{h \in \mathcal{H}(p,n)} p_{-h}q$.

Finally, for any $h = (p^1, a^1, f^1, x^1 \ldots p^n, a^n, f^n, x^n)$ and $\hat{h} = (\hat{p}^1, \hat{a}^1, \hat{f}^1, \hat{x}^1 \ldots \hat{p}^n, \hat{a}^n, \hat{f}^n, \hat{x}^n)$ both in $\mathcal{H}(n)$, we say that h and h' are *$\mathcal{A}$-equivalent*, denoted by $h \overset{\mathcal{A}}{\sim} h'$ if $a^i = \hat{a}^i$ and $x^i = \hat{x}^i$ for $i \leqslant n$. That is, two histories of length n are $\mathcal{A}$-equivalent, whenever they correspond to the same sequence of action-realization pairs, ignoring the objective randomization stage of each period and the continuation assignment to outcomes that did not occur. It will turn out, we are only interested in the $\mathcal{A}$-equivalence classes of histories. Technically, this is the consequence of the linearity of preference and indifference to the resolution of uncertainty (as shown in Lemma 3); conceptually, this is because all uncertainty in the model regards the realization of actions, and so, observing objective lotteries has no informational benefit.

A.3 THE AXIOMS

The primitive in our model is a preference relation $\succcurlyeq \subseteq P \times P$ over all PoAs. When specific PoA and history are fixed, the preferences induce history dependent preferences as follows: for any $p \in P$, and

$h \in \mathcal{H}(p)$ define $\succcurlyeq_h \subseteq P \times P$ by

$$q \succcurlyeq_h r \iff p_{-h}q \succcurlyeq p_{-h}r.$$

The following axioms will be employed over all history induced preferences.[11] A history is *null* if $\succcurlyeq_h$ is a trivial relation. This first four axioms are variants on the standard fare for discounted expected utility. They guarantee the expected utility structure, non-triviality, stationarity and separability (regarding objects over which learning cannot take place), respectively.

A1. (vNM). *The binary relation, $\succcurlyeq_h$ satisfies the expected utility axioms. That is: weak order, continuity, and independence.*

We require a stronger non-triviality condition that is standard, because of the subjective nature of the dynamic problem. We need to ensure the DM believes *some* outcome will obtain. Therefore, not all histories following a given action can be null.

A2. (NT). *For any non-null h, and any (a, f), not all $h' \in h \times \mathcal{H}((a, f), n)$ are null.*

Of course, the nature of the problem at hand precludes stationarity and separability in full generality. Since the objective is to let the DM's beliefs depend on prior outcomes explicitly, her preferences will as well. However, the DM's beliefs do not influence her assessment of objective plans (i.e., elements of Σ), and so it is over this domain that stationarity and separability are retained. This means, the DM's preferences *in utility terms* are stationary and separable, but we still allow the conversion between actions and utils to depend on her beliefs which change responsively.

A3. (SST). *For all non-null $h \in \mathcal{H}$, and $\sigma, \sigma' \in \Sigma$,*

$$\sigma \succcurlyeq \sigma' \iff \sigma \succcurlyeq_h \sigma'.$$

A4. (SEP). *For all $x, x' \in X$, $\rho, \rho' \in \Sigma$ and $h \in \mathcal{H}$,*

$$\big(\frac{1}{2}(x, \rho) + \frac{1}{2}(x', \rho')\big) \sim_h \big(\frac{1}{2}(x, \rho') + \frac{1}{2}(x', \rho)\big).$$

Because of the two-stage nature of the resolution of uncertainty each period (first, the resolution of lottery over $\mathcal{A} \otimes P$, and then the resolution of the action over X), we need an additional separability constraint. From the point of view of period n, and when considering the continuation problem beginning in period $n + 1$, the DM should not care if uncertainty is resolved in period n (when the action-continuation pair is realized), or in period $n + 1$. That is, we also assume the DM is indifferent to the timing of objective lotteries *given a fixed action.*

A5. (IT). *For all $a \in \mathcal{A}$, $h \in \mathcal{H}$, $\alpha \in (0, 1)$, and $(a, f), (a, g) \in \hat{P}$,*

$$\alpha(a, f) + (1 - \alpha)(a, g) \sim_h (a, \alpha f + (1 - \alpha)g),$$

where mixtures of f and g are taken point-wise.

Thus far the axioms introduced are somewhat standard. However, in our particular framework these assumptions do not guarantee that the value of the action is in any way related with its realization of consumption alternatives. This is because, unlike other environments, the set of outcomes, X, plays a dual role in exploration models: representing both the space of outcomes and the state space regarding future actions.

---

[11] It is via the use of this construction that our appeal to denumerably supported lotteries provides tractability. If we were to employ lotteries with uncountable support, then histories would, in general, be zero probability events; under the expected utility hypothesis, $\succcurlyeq_h$ would be null for all $h \in \mathcal{H}$. This could be remedied by appealing to histories as *events* in $\mathcal{H}$, measurable with respect to the filtration induced by previous resolutions of lottery-action-outcome tuples. We believe that this imposes a unnecessary notational burden.

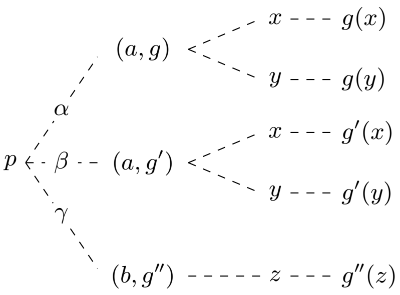

Figure 3: A PoA, p, defined by $p(a,g) = \alpha$, $p(a,g') = \beta$ and $p(b,g'') = \gamma = 1-\alpha-\beta$.

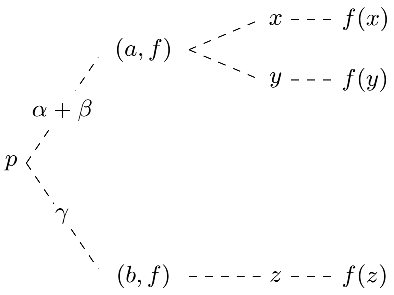

Figure 4: The PoA, $p.f$ where $f : X \to P$, and p is defined in figure 3. Notice, $p.f(a,f) = p(a,g)+p(a,g') = \alpha+\beta$.

The realization of an outcome x delivers utility according to both of these roles, and, to ensure consistency between them requires two steps. First, construct a subjective distribution over each action by treating X as a state space. This will be done by looking at the ranking of continuation mappings for each action (i.e., (a,f) compared to (a,g)). Interpreting X as the periodic state space, these continuation mappings are analogous to "acts" in the standard subjective expected utility paradigm– and so, standard techniques allow for the identification of such a subjective belief. Second, we need to ensure that the value assigned to arbitrary PoAs is the expectation according to these beliefs. Towards this, the following notation is introduced.

Definition. *For any function* $f : X \to \mathcal{P}$*, define* $p.f \in P$ *as* $p.f[(a,g)] = p[\{(b,h)|b = a\}]$ *if* $g = f$*, and* $p.f[(a,g)] = 0$ *if* $g \neq f$*.*

Take note, because we are dealing with distributions of denumerable support, we have no measurability concerns. The plan of action $p.f$ has the same distribution over actions in the first period, but the continuation plan is unambiguously assigned by f, as shown in Figures 3 and 4. If the original plan is in $\mathcal{A}\otimes P$, then the dot operation is simply a switch of the continuation mapping: $(a,g).f = (a,f)$. This operation is introduced because it allows us to isolate the subjective distribution of the first period's action.

Definition. $p,q \in P$ *are* ***h-proportional*** *if for all* $f,g : X \to \Sigma$*.*

$$p.f \succcurlyeq_h p.g \iff q.f \succcurlyeq_h q.g$$

Since the images of f and g are in Σ, there is no informational effect from observing the outcome of p. Hence, f and g can be thought of as objective assignments into continuation utilities. The ranking '$p.f \succcurlyeq p.g$' is really a ranking over f and g as functions from $X \to \mathbb{R}$. Thus, h-proportionality states that the DM's subjective uncertainty regarding X is the same when faced with p or with q.[12]

A6. (PRP). *For all* $p,q \in P$*, and* $f : X \to \Sigma$ *if* p *and* q *are* h*-proportional then* $p.f \sim_h q.f$*.*

The outcomes of an action represent not only the uncertainty regarding continuation, but also the utility outcome for the current period. So, when p and q are h-proportional, and thus induce the same uncertainty regarding X, the DM's uncertainty about her current period utility is the same across the plans. Therefore, if we replace the continuation problems with objectively equivalent plans, the DM should be indifferent between p and q.

---

[12] To see this, note that the relation R on $\mathbb{R}^X \times \mathbb{R}^X$ defined by fRg if and only if $p.f \succcurlyeq p.g$ is a preference relation over acts that satisfies the Anscombe and Aumann (1963) axioms, and therefore encodes the DM's subjective likelihood of each $E \subset X$. From a functional standpoint, h-proportionality states the subjective distribution over X induced by p is the same as that induced by q.

PRP states that, when future discounted expected utilities have been identified, the entire exploration/exploitation tradeoff collapses to a simple 2-stage intertemporal tradeoff. Of course, this requires the identification of continuation values, and therefore a full understanding of future utilities via the beliefs. In the most general model, there need not be *any* connection between today's beliefs and tomorrow's, hence the only behavior associated with exploration models is that which can be derived from the recursive structure. This need not be viewed as a negative result. Instead, we have shown that sharp behavioral markers of exploration behavior must arise from conditions on the evolution of beliefs. An example for that is provided in Section A.5 when we discuss the behavioral restrictions of exchangeability in the current setup.

A.4 A Representation Result and Belief Elicitation

The following is our general axiomatization result. It states that the properties above characterize a DM who, when facing a PoA, calculates the subjective expected utility according to a collection of history dependent beliefs over action-outcome pairs, and among different PoAs contemplates the benefits of consumption versus learning.

Theorem 6 (Subjective Expected Experimentation Representation)**.** *$\succcurlyeq_h$ satisfies vNM, NT, SST, SEP, IT and PRP if and only if there exists a utility index $u : X \to \mathbb{R}$, a discount factor $\delta \in (0,1)$, and a family of beliefs $\{\mu_{h,a} \in \Delta(S_A)\}_{h \in \mathcal{H}, a \in \mathcal{A}}$ such that*

$$U_h(p) = \mathbb{E}_p\left[\mathbb{E}_{\mu_{h,a}}\left[u(x) + \delta U_{h'(a,x)}(f(x))\right]\right], \tag{SEE}$$

jointly represents $\{\succcurlyeq_h\}_{h \in \mathcal{H}}$, where $h'(a,x) = (h, p, (a,f), x)$. Moreover, u is cardinally unique, δ is unique, the family of beliefs is unique, and $\mu_{h,a} = \mu_{h',a}$ whenever $h \overset{\mathcal{A}}{\sim} h'$.

Proof. In Section A.8. ■

The theorem states that we can (uniquely) elicit the beliefs, following every history, over the outcomes of each action separately. We will henceforth refer to such beliefs as an *SEE belief structure*. The axioms do not impose any restrictions on the dynamics of such beliefs. More importantly, the theorem shows that, when ranking the different strategies in a bandit problem, the decision maker does not reveal her beliefs over the *joint* realizations of the different actions.

A.5 AA-SYM as a Behavioral Restriction

In the main text we introduced the notion of *across-arm symmetry* or AA-SYM, which stated that the DM's beliefs were invariant under joint permutations of the order of actions and observations. AA-SYM is a necessary and sufficient condition for consistency with an exchangeable process. In this section we introduce the axiomatic counterpart of AA-SYM, and so we can identify Bayesianism in exploration environments directly from preferences over the strategies.

Definition. *Let π be an n-permutation and $p, q \in P$. We say that q is π-permutation of p if for all $h \in \mathcal{H}(p,n)$, $h' \in \mathcal{H}(q,n)$, $\text{proj}_{\mathcal{A}^n} h = \pi\big(\text{proj}_{\mathcal{A}^n} h'\big)$.*

If p admits any π-permutations it must be that the first n actions are assigned unambiguously (i.e., it does not depend on the realization of prior actions nor the objective randomization).

A7. (AA-SYM). *Let π be an n-permutation and $p, p' \in P$ with p' a π-permutation of p. Then, for all $a \in \mathcal{A}$, $\tau, \sigma, \sigma' \in \Sigma$, and $h \in \mathcal{H}(p,n)$, $h' \in \mathcal{H}(p',n)$, if h is a permutation of h' then*

$$p_{-n}\tau \succcurlyeq (p_{-n}\sigma)_{-h}\sigma' \iff p_{-n}\tau \succcurlyeq (p_{-n}\sigma)_{-h'}\sigma'.$$

After n periods the plan $p_{-n}\tau$ provides τ with certainty, while the plan $(p_{-n}\sigma)_{-h}\sigma'$ provides σ *unless* the history h occurs. Hence, the DM's preference between the plans depends on their ex-ante subjective assessment of how likely h is to occur. Similarly to the logic behind h-proportionality, AA-SYM states that the DM's assesses h to be exactly as probable as h'. In other words, the DM's likelihood of outcome realizations is invariant to the order in which the actions are taken. The intuition behind the next result is correspondingly straightforward.

Proposition 7 (Correlated Arms, Exchangeable Process)**.** *Let $\succcurlyeq$ admit an SEE representation with the associated observable processes $\{\zeta_{\mathbf{T}}\}_{\mathbf{T}\in\mathcal{T}}$. Then, the following are equivalent:*

1. *$\succcurlyeq_h$ satisfies AA-SYM;*
2. *$\{\zeta_{\mathbf{T}}\}_{\mathbf{T}\in\mathcal{T}}$ satisfies AA-SYM;*
3. *$\{\zeta_{\mathbf{T}}\}_{\mathbf{T}\in\mathcal{T}}$ is consistent with an exchangeable process; and*
4. *$\{\zeta_{\mathbf{T}}\}_{\mathbf{T}\in\mathcal{T}}$ is consistent with a (unique) strongly exchangeable process.*

Proof. The proof that condition 1 is equivalent to condition 2 is provided Section A.8. Conditions 2, 3, and 4 are equivalent due to Theorem 3. ■

The proposition implies that strong-exchangeability carries no additional restrictions, beyond those of exchangeability, on agents' preferences over the different strategies in bandit problems, and in particular on their optimal strategies.

A.6 FURTHER DISCUSSION

Related Literature. Within decision theory, the literature on learning broadly considers how a DM incorporates new information, generally via notions of Bayesianism and Exchangeability, and often in the domain of uncertainty: see Epstein and Le Breton (1993); Epstein and Seo (2010); Klibanoff et al. (2013); Lehrer and Teper (2015). Recently, there has been an interest in subjective learning, or, the identification of the set of possible "signals" that the DM believes she might observe. At it's most simple, this is the elicitation of the set of potential tastes (often referred to as subjective states) the decision maker anticipates, accomplished by examining the DM's preference over *menus* of choice objects: see Kreps (1979); Dekel et al. (2001). By also incorporating consumption goods that contract on an objective state space, the modeler can interpret the DM's preference for flexibility as directly stemming from her anticipation of acquiring information regarding the likelihood of states, as in Dillenberger et al. (2014); Krishna and Sadowski (2014).

There is also a small but highly relevant literature working on the identification of responsive learning. Hyogo (2007) considers a two-period model, with an objective state space, in which the DM ranks action-menu pairs. The action is taken in the first period and provides information regarding the likelihood of states, after the revelation of which, the DM choose a state-contingent act from the menu. The identification of interest is the DM's subjective interpretation of actions as signals. Similarly, Cooke (2016) entertains a similar model without the need for an objective state-space, and in which the consumption of a single object in the first period plays the role of a fully informative action. Cooke, therefore, identifies both the state-space and the corresponding signal structure. Piermont et al. (2016) consider a recursive and infinite horizon version of Kreps' model, where the DM deterministically learns about her preference regarding objects she has previously consumed. Dillenberger et al. (2015) consider a different infinite horizon model where the DM makes separate choices in each period regarding her information structure and current period consumption. It is worth pointing out, all of these models, unlike the this paper, capitalize on the "preference for flexibility" paradigm to characterize learning. We are able to identify subjective learning without appealing to the menu structure because of the purely responsive aspect of our model. In other words, flexibility is "built in" to our setup, as a different action can be taken after every possible realization of the signal (action).

Subjective Learning with Endogenous and Exogenous Information. As witnessed the literature covered above, there seems to be a divide in the literature regarding subjective learning. In one camp, are models that elicit the DM's perception of exogenous flows of information (as a canonical example, take Dillenberger et al. (2014)), and in the other are models that assume information is acquired only via actions taken by the DM (where this paper lies). Realistically, neither of these information structures capture the full gamut of information transmission in economic environments.

Consider the following example within the setup of the current paper. A firm is choosing between two projects (actions), a and b. Assume that each project has a high-type and a low type. The firm believes (after observing h) the probability that each project is the high-type is $\mu_{h,a}$ and $\mu_{h,b}$, respectively. By experimenting between a and b the firm's beliefs and preferences will evolve.

But, what happens if the firm anticipates the release of a comprehensive report regarding project a just before period 1? This report will declare project a high quality with probability $\alpha^h > \frac{1}{2}$ if the projects true type is high and with probability $\alpha^l < \frac{1}{2}$ if it is low. Hence, the report is an informative signal. Now, if the firms belief after observing h in period 0 is given by $[\mu_{h,a}, \mu_{h,b}]$ then, according to Bayes rule, the firms belief regarding project a being the high-type, at the beginning of period 1 will be $\mu^+_{h,a} = \frac{\alpha^h \cdot \mu_{h,a}}{\alpha^h \cdot \mu_{h,a} + \alpha^l (1 - \mu_{h,a})}$, if the report is positive, and $\mu^-_{h,a} = \frac{(1-\alpha^h) \cdot \mu_{h,a}}{(1-\alpha^h) \cdot \mu_{h,a} + (1-\alpha^l) \cdot (1-\mu_{h,a})}$ if the report is negative.

Unfortunately, however, the ex-ante elicitation of preferences in our domain cannot capture the anticipation of information. The firm is ranking PoAs according to its aggregated belief from the ex-ante perspective, and thus, so as to maximize its expected belief:

$$\big(\alpha^h \mu_{h,a} + \alpha^l (1 - \mu_{h,a})\big)\mu^+_{h,a} + \big((1 - \alpha^h)\mu_{h,a} + (1 - \alpha^l)(1 - \mu_{h,a})\big)\mu^-_{h,a} = \mu_{h,a}.$$

Because of the Bayesian structure, the DM's beliefs must form a martingale, so her expectation of her anticipated beliefs are exactly her ex-ante beliefs. This fact, coupled with the linearity of expected utility, imply that the DM's ex-ante preference over PoAs is unaffected by her anticipation of exogenous information arrival.

All hope is not lost, however, of fully characterizing the DM's subjective information structure. The approach of Dillenberger et al. is orthogonal to our's, leading us to conjecture that the two models can co-exist and impart a clean separation between exogenous and endogenous information flows. Going back to the example, imagine there are two PoAs, p and q such that p is preferred to q under beliefs μ^+_h, and q to p under μ^-_h. The DM would therefore strictly desire flexibility after period 0, even after she is able to condition her decision on h. Of course, because the report is released after period 0, irrespective of the action taken by the DM, for any 0-period history h', there must exist some other PoAs, p' and q', for which flexibility is strictly beneficial (after h').

A.7 PROOFS REGARDING THE CONSTRUCTION OF PLANS OF ACTION.

Lemma 1. *There exists a homeomorphism, $\lambda : Q \to \Delta^{\mathcal{B}}(\mathcal{A} \times \mathcal{K}(X \times Q))$ such that*

$$\mathrm{marg}_{\mathcal{A} \times \mathcal{K}(X \times Q_{n-1})}(\lambda(q)) = q_n. \tag{8}$$

Proof. [STEP 1: EXTENSION THEOREM.] Let $C_n = \big\{(q_0, \ldots q_n) \in \prod_{k=0}^n Q_k | q_k = F_{k+1}(q_{k+1}), \forall k = 1 \ldots n-1\big\}$, and $T_n = \mathcal{K}(X \times C_n)$ for $n \geqslant 0$. Let $T^* = \prod_{n=0}^\infty T_n$ and $T = \big\{t \in T^* | (\mathrm{proj}_{T_n} t_{n+1} = t_n\big\}$. Let $Y_0 = \Delta^{\mathcal{B}}(\mathcal{A})$ and for $n \geqslant 1$ let $Y_n = \Delta^{\mathcal{B}}(\mathcal{A} \times T_0 \times \ldots \times T_n))$. We say the the sequence of probability measures $\{\nu_n \in Y_n\}_{n \geqslant 0}$ is consistent if $\mathrm{marg}_{A \ldots T_{n-1}} \nu_{n+1} = \nu_n$ for all $n \geqslant 0$. Let Y^c denote the set of all consistent sequences. Then we know by Brandenburger and Dekel (1993), for every $\{\nu_n\} \in Y^c$ there exists a unique $\nu \in \Delta^{\mathcal{B}}(\mathcal{A} \times T^*)$ such that $\mathrm{marg}_{\mathcal{A}} \nu = \nu_0$ and $\mathrm{marg}_{A \ldots T_n} \nu = \nu_n$. Moreover, the map $\psi : Y^c \to \Delta^{\mathcal{B}}(\mathcal{A} \times T^*)$:

$$\psi : \{\nu_n\} \mapsto \nu$$

is a homeomorphism. □

[STEP 2: EXTENDING BACKWARDS.] Let $D_n = \big\{(t_0, \ldots t_n) \in \times_{n=0}^n T_n | t_k = \mathrm{proj}_{T_n}(t_{k+1}), \forall k = 1 \ldots n-1\big\}$. Let $Y^d = \big\{\{\nu_n\} \in Y^c | \nu_n(\mathcal{A} \times D_n) = 1, \forall n \geqslant 0\big\}$. We will now show, for each $q \in Q$, there exists a unique $\{\nu_n\} \in Y^d$, such that $\nu_0 = q_0$ and $\mathrm{marg}_{\mathcal{A} \times \mathcal{K}(X \times Q_{n-1})}(\mathrm{marg}_{\mathcal{A} \times T_{n-1}}(\nu_n)) = q_n$ for all

$n \geqslant 1$. Indeed, let m_0, m_1 be the identify function on $\mathcal{A}$ and $\mathcal{A} \times \mathcal{K}(X \times Q_0)$, respectively. Then for each $n \geqslant 2$ let $m_n : \mathcal{A} \times D_{n-1} \to \mathcal{A} \times \mathcal{K}(X \times Q_{n-1})$ as follows:

$$m_{n+1} : (a, \{x^0, q_0^0\}, \{x^1, q_0^1, q_1^1\} \dots \{x^n, q_0^n \dots q_n^n\}^n) \mapsto (a, \{x^n, q_n^n\}).$$

Note: for $n \geqslant 0$, each m_n is a Borel isomorphism. Indeed, continuity of m_n is obvious, and measurability follows immediately from the fact that canonical projections are measurable in the product σ-algebra. It is clear that m_n is surjective, and —since (given F_k for $k \in 1 \dots n$) q_n uniquely determines $q_0 \dots q_{n-1}$, which, (given the projection mappings) uniquely determines $T_0 \dots T_{n-1}$— m_i is also injective. As for, m_n^{-1}, continuity follows from the continuity of F_k for $k \in 1 \dots n$ and the projection mappings. Lastly, measurability of m_n^{-1} comes from the fact that a continuous injective Borel function is a Borel isomorphism (see Kechris (2012) corollary 15.2).

So, let $\psi : Q \to Y^d$ as the map

$$\phi : q \mapsto \{E_n \mapsto q_n(m_n(E_n))\}_{n \geqslant 0},$$

for any $E_n \in \mathcal{B}(A \times T_0 \times \ldots \times T_n)$. The continuity of ϕ and ϕ^{-1} follow from the fact that they are constructed from the pushforward measures of m_n^{-1} and m_n, respectively, which are themselves continuous (or, explicitly, see GP lemma 4).

Finally, let $\Gamma_n = \mathcal{A} \times D_n \times_{k=n+1}^{\infty} T_k$. Let $\nu = \psi(\{\nu_n\})$ for some $\{\nu_n\}$ in Y^d. Then $\nu(\Gamma_n) = \nu(\mathcal{A} \times D_n) = 1$. So, $\nu(\mathcal{A} \times T) = \nu(\cap_{n \geqslant 0} \Gamma_n) = \lim \nu(\Gamma_n) = 1$. Also, note, if $\nu(\mathcal{A} \times T) = 1$, then $\nu(\Gamma_n) = 1$ for all $n \geqslant 0$. So, $\nu \in Y^d$ if and only if $\nu(\mathcal{A} \times T) = 1$, i.e., if, $\psi(Y^d) = \big\{\nu \in \Delta^{\mathcal{B}}(\mathcal{A} \times T^*) | \nu(\mathcal{A} \times T) = 1\big\}$. □

[Step 3: Extending Forwards.] Let τ denote the map from $\mathcal{A} \times \mathcal{K}(X \times Q) \to \mathcal{A} \times T$ as

$$\tau : \big(a, \{x, q\}\big) \mapsto \big(a, (\{x, q_0\}, \{x, q_0, q_1\}, \ldots)\big)$$

That τ it is a bijection follows from the consistency conditions on Q, T, and C_n for $n \geqslant 1$. Now takes some measurable set $E \subseteq T$. Then $\tau^{-1}(E) = \bigcap_{n \geqslant 0} \big\{\{x, q_0, \dots q_n \times_{k=n_1^{\infty}} Q_k\} \in K(X \times Q^*)\big\}$, the countable intersection of measurable sets, and hence measurable. That τ and τ^{-1} are continuous is immediate. Therefore, by the same argument as in [Step 2], τ is a Borel isomorphism and $\kappa : \Delta^{\mathcal{B}}\big(\mathcal{A} \times T\big) \to \Delta^{B}\big(\mathcal{A} \times \mathcal{K}(X \times Q)\big)$,

$$\kappa : \nu \mapsto \big(E \mapsto \nu(\tau(E))\big)$$

for all E in $\Delta^{\mathcal{B}}\big(\mathcal{A} \times \mathcal{K}(X \times Q)\big)$. Clearly, $\text{marg}_{\mathcal{A}}(\kappa(\nu)) = \text{marg}_{\mathcal{A}}(\nu)$ and $\text{marg}_{\mathcal{A} \times \mathcal{K}(X \times Q_{n-1})}(\kappa(\nu)) = \text{marg}_{\mathcal{A} \times \mathcal{K}(X \times Q_{n-1})}(\text{marg}_{\mathcal{A} \times T_{n-1}}(\nu))$ for all $n \geqslant 1$.

□

Behold, $\lambda = \kappa \circ \psi \circ \phi$ is the desired homeomorphism. ■

Proof of Proposition 4. We show that λ is a homeomorphism between R and $\Delta^{\mathcal{B}}(\mathcal{A} \otimes R)$. Identify $\Delta^{\mathcal{B}}(\mathcal{A} \otimes R)$ with $\big\{\nu \in \Delta^{\mathcal{B}}(\mathcal{A} \times K(X \times Q)) | \nu(\mathcal{A} \otimes R) = 1\big\}$. Let $r \in R$. For each $n \geqslant 0$ let $\Gamma_n^r = \{(a, \{x, q\}) \in \mathcal{A} \otimes Q | q_k \in R_k,\ k = 0 \dots n\}$. Then $\lambda(r)(\Gamma_n^r) = \text{marg}_{A \times K(X \times Q_n)}(\lambda(r))(\mathcal{A} \otimes R_n) = r_{n+1}(\mathcal{A} \otimes R_n) = 1$ for all $n \geqslant 1$. So $\lambda(r)(\mathcal{A} \otimes R) = \lambda(r)(\cap_{n \geqslant 0} \Gamma_n^r) = \lim \lambda(r)(\Gamma_n^r) = 1$. Now, fix $q \in Q$ with $\lambda(q)(\mathcal{A} \otimes R) = 1$, then $q_n(\mathcal{A} \otimes R_{n-1}) = \text{marg}_{A \times K(X \times Q_{n-1})}(\lambda(q))(\mathcal{A} \otimes R_{n-1}) = \lambda(r)(\Gamma_n^r) \geqslant \lambda(r)(\mathcal{A} \otimes R) = 1$ for all $n \geqslant 0$ and so $q \in R$. ■

Definition. *Set $W, W^\star : \mathcal{P}(R) \to \mathcal{P}(R)$ as the functions:*

$$\begin{aligned} W &: E \mapsto \{r' \in R | r' \in Im(f) \textit{ for some } (a, f) \in \text{supp}(\lambda(r)), r \in E\} \textit{ and,} \\ W^\star &: E \mapsto \bigcup_{n \geqslant 0} W^n(E) \end{aligned}$$

Where W^n is $W(W(\dots W(E)\dots))$ with n applications of W.

Definition. *Let $P_0 = \Delta(\mathcal{A})$ and recursively, $P_n = \{p_n \in R_n | p_n \in \Delta(\mathcal{A} \otimes P_{n-1})\}$. Set $P = \big\{p \in \prod_{n=0}^{\infty} P_n | \lambda(W^\star(r)) \subseteq \Delta(\mathcal{A} \otimes R)\big\}$.*

Proof of Theorem 5. We show that λ is a homeomorphism between P and $\Delta(\mathcal{A} \otimes P)$. First note, by construction, for all $r \in R$, $\lambda(r) \in \Delta^{\mathcal{B}}(A \otimes W(r))$. Let $p \in P$; by the conditions on P, $\lambda(p) \in \Delta(\mathcal{A} \otimes R)$. Therefore, it suffices to show that for any $p \in P$, and $r \in W(p)$, $r \in P$. So fix some $r \in W(p)$. It follows from an analogous argument to Corollary 4 that $r \in \prod_{n=0}^{\infty} P_n$. Finally, note that $W^\star(r) \subseteq W^\star(W(r))$. ■

A.8 Proofs Regarding the SEE Representation.

Lemma 2. *If $\succcurlyeq_h$ satisfies* vNM *and* IT*, then $\succcurlyeq_h$ satisfies the* sure thing principal*:*

A8. (STP). *For all $a \in \mathcal{A}$ and $f, f', g, g' : X \to P$, such that, for all $x \in X$, either (i) $f(x) = f'(x)$ and $g(x) = g'(x)$ or (ii) $f(x) = g(x)$ and $f'(x) = g'(x)$. Then,*

$$(a, f) \succcurlyeq_h (a, g) \iff (a, f') \succcurlyeq_h (a, g').$$

Proof. Assume this was not true and, without loss of generality, that $(a, f) \succcurlyeq_h (a, g)$ but $(a, g') \succ_h (a, f')$. Now notice, when mixtures are taken point-wise, $\frac{1}{2}f + \frac{1}{2}g' = \frac{1}{2}g + \frac{1}{2}f'$. Therefore,

$$\begin{aligned}\big(\frac{1}{2}(a, f) + \frac{1}{2}(a, g')\big) &\succ_h \big(\frac{1}{2}(a, g) + \frac{1}{2}(a, f')\big)\\ &\sim_h (a, \frac{1}{2}g + \frac{1}{2}f') = (a, \frac{1}{2}f + \frac{1}{2}g')\\ &\sim_h \big(\frac{1}{2}(a, f) + \frac{1}{2}(a, g')\big),\end{aligned}$$

where the first line follows from vNM, and the indifference conditions from IT. This is a contradiction. ■

Lemma 3. *If $\succcurlyeq_h$ satisfies* vNM *and* IT *for all $h \in \mathcal{H}$, then, if $h \overset{\mathcal{A}}{\sim} h'$ then $\succcurlyeq_h = \succcurlyeq_{h'}$.*

Proof. We will show the claim on induction by the length of the history. So let $h, h' \in \mathcal{H}(1)$ such that $h \overset{\mathcal{A}}{\sim} h'$. Therefore, $h = (p, (a, f), x)$ and $h' = (p', (a, g), x)$. Notice, by definition we have, $p = \alpha(a, f) + (1 - \alpha)r$ and $p' = \alpha'(a, g) + (1 - \alpha')r'$, for some $\alpha, \alpha' \in (0, 1]$ and $r, r' \in P$.

Let $q, q' \in P$; we want to show that $q \succcurlyeq_h q' \iff q \succcurlyeq_{h'} q'$. So let $q \succcurlyeq_h q'$, or by definition, $p_{-h}q \succcurlyeq p_{-h}q'$, which by the above observation is equivalent to

$$\alpha(a, f)_{-((a,f),(a,f),x)}q + (1 - \alpha)r \succcurlyeq \alpha(a, f)_{-((a,f),(a,f),x)}q + (1 - \alpha)r.$$

By independence (i.e., vNM) this is if and only if $(a, f)_{-((a,f),(a,f),x)}q \succcurlyeq (a, f)_{-((a,f),(a,f),x)}q'$, which by STP is if and only if $(a, g)_{-((a,g),(a,g),x)}q \succcurlyeq (a, g)_{-((a,g),(a,g),x)}q'$. Using independence again, this is if and only if $p'_{-h'}q \succcurlyeq p'_{-h'}q'$. This completes the base case.

So assume the claim holds for all histories of length n. So let $h, h' \in \mathcal{H}(n + 1)$ such that $h \overset{\mathcal{A}}{\sim} h'$. Therefore, $h = (h_n, p, (a, f), x)$ and $h' = (h'_n, p', (a, g), x)$, for some $h_n, h'_n \in H(n)$ such that $h_n \overset{\mathcal{A}}{\sim} h'_n$. By the inductive hypothesis $\succcurlyeq_{h_n} = \succcurlyeq_{h'_n}$.

Let $q, q' \in P$, and $q \succcurlyeq_h q'$, or by definition, $p_{-(p,(a,f),x)}q \succcurlyeq_{h_n} p_{-(p,(a,f),x)}q'$. By independence and the sure thing principle this is if and only if $(a, g)_{-((a,g),(a,g),x)}q \succcurlyeq_{h_n} (a, g)_{-((a,g),(a,g),x)}q'$, which by independence again (and the equivalence of $\succcurlyeq_{h_n}$ and $\succcurlyeq_{h'_n}$), is if and only if $p'_{-(p',(a,g),x)}q \succcurlyeq_{h'_n} p'_{-(p',(a,g),x)}q'$. ■

Proof of Theorem 6. [Step 0: Value Function.] Since $\succcurlyeq_h$ satisfies vNM, there exists a $v_h : \mathcal{A} \otimes P \to \mathbb{R}$ such that

$$U_h(p) = \mathbb{E}_p\left[v_h(a,f)\right] \tag{9}$$

represents $\succcurlyeq_h$, with v_h unique un to affine translations. □

[Step 1: Recursive structure.] To obtain the skeleton of the representation, lets consider $\hat{\succcurlyeq}$, the restriction of $\succcurlyeq$ to Σ (i.e., using the natural association between streams of lotteries and degenerate trees). The relation $\hat{\succcurlyeq}$ satisfies vNM (it is continuous by the closure of Σ in P). Hence there is a linear and continuous representation: i.e., an index $\hat{u} : X \times \Sigma \to \mathbb{R}$ such that:

$$\hat{U}(\sigma) = \mathbb{E}_\sigma\left[\hat{u}(x,\rho)\right] \tag{10}$$

unique upto affine translations.

Following Gul and Pesendorfer (2004), (henceforth GP), fix some $(x',\rho') \in \Sigma$. From SEP we have $\hat{U}(\frac{1}{2}(x,\rho) + \frac{1}{2}(x',\rho')) = \hat{U}(\frac{1}{2}(x,\rho') + \frac{1}{2}(x',\rho))$, and hence, $\hat{u}(x,\rho) = \hat{u}(x,\rho') + \hat{u}(x',\rho) - \hat{u}(x',\rho')$. Then setting $u(x) = \hat{u}(x,\rho') - \hat{u}(x',\rho')$ and $W(\rho) = \hat{u}(x',\rho)$, we have,

$$\hat{U}(\sigma) = \mathbb{E}_\sigma\left[u(x) + W(\rho)\right] \tag{11}$$

Now, consider $p' = (x',\rho)$. Notice that p' has unique 1-period history: $h = (p',p',x')$. By NT, h cannot be null. So, by SST, $\hat{\succcurlyeq}_h = \hat{\succcurlyeq}$. This implies, of course that $W = \delta\hat{U} + \beta$ for some $\delta > 0$ and $\beta \in \mathbb{R}$. Following Step 3 of Lemma 9 in GP exactly, we see that $\delta < 1$ and without loss of generality we can set $\beta = 0$:

$$\hat{U}(\sigma) = \mathbb{E}_\sigma\left[u(x) + \delta\hat{U}(\rho)\right] \tag{12}$$

Both representing $\hat{\succcurlyeq}$ and being unique up to affine translations, we can normalize each U_h to coincide with $\hat{U}$ over Σ. □

[Step 2: The existence of subjective probabilities.] For each $a \in \mathcal{A}$ consider

$$\mathcal{F}(a) = a \otimes \Sigma$$

i.e., the elements of $\hat{P}$ that begin with action a and from period 2 onwards are in Σ. Associate $\mathcal{F}(a)$ with the set of "acts": $f : S_a \to \Sigma$, in the natural way. For any acts f, g let $f_{-x}g$ denote the act that coincides with f for all $x' \in S_a$, $x' \neq x$, and coincides with g after x. For each $h \in \mathcal{H}$, and acts $f, g \in \mathcal{F}(a)$, say $f \mathrel{\dot{\succcurlyeq}_{h,a}} g$ if and only if $(a,f) \succcurlyeq_h (a,g)$.

It is immediate that $\dot{\succcurlyeq}_{h,a}$ is a continuous weak order (where, as before, continuity follows from the closure of $\mathcal{F}$ in P). Further, $\dot{\succcurlyeq}_{h,a}$ satisfies independence. Indeed: fix $f, g, h \in \mathcal{F}(a)$ with $f \mathrel{\dot{\succcurlyeq}_{h,a}} g$. Then

$$\begin{aligned} f \mathrel{\dot{\succcurlyeq}_{h,a}} g &\implies (a,f) \succcurlyeq_h (a,g) \\ &\implies \alpha(a,f) + (1-\alpha)(a,h) \succcurlyeq_h \alpha(a,g) + (1-\alpha)(a,h) \\ &\implies (a, \alpha f + (1-\alpha)h) \succcurlyeq_h (a, \alpha g + (1-\alpha)h) \\ &\implies \alpha f + (1-\alpha)h \mathrel{\dot{\succcurlyeq}_{h,a}} \alpha g + (1-\alpha)h, \end{aligned}$$

where the third line uses IT. Lastly, $\dot{\succcurlyeq}_{h,a}$ satisfies monotonicity, a direct consequence of SST and STP. Hence, we have state-independence which gives us the full set of Anscombe and Aumann (1963) axioms for an SEU representation of $\dot{\succcurlyeq}_{h,a}$ with state space S_a. That is, a belief $\mu_{h,a} \in \Delta(S_a)$ and a utility index from $\Sigma \to \mathbb{R}$ (which is of course, $\hat{U}$, and so will be denoted as such), such that

$$\hat{V}_{h,a}(f) = \mathbb{E}_{\mu_{h,a}}\left[\hat{U}(f(x))\right] \tag{13}$$

represents $\dot{\succcurlyeq}_{h,a}$. ▫

[STEP 3: PROPORTIONAL ACTIONS.] Now, fix some $h \in \mathcal{H}$ and consider an arbitrary $(a, f) \in \mathcal{A} \otimes P$. Let $\rho \in \Sigma$ be such that $\text{marg}_X \rho = \mu_{h,a}$. We claim, (a, f) and ρ are h-proportional. Fix some $g, g' : X \to \Sigma$. From (13), we know

$$(a, g) \succcurlyeq_h (a, g') \iff \mathbb{E}_{\mu_{h,a}}\left[\hat{U}(g(x))\right] \geqslant \mathbb{E}_{\mu_{h,a}}\left[\hat{U}(g'(x))\right] \tag{14}$$

From (12) we have

$$\begin{aligned}\hat{U}(\rho.g) &= \mathbb{E}_{\rho}\left[u(x) + \delta\hat{U}(g(x))\right]\\ &= \mathbb{E}_{\text{marg}_X \rho}\left[u(x) + \delta\hat{U}(g(x))\right]\\ &= \mathbb{E}_{\mu_{h,a}}\left[u(x)\right] + \delta\,\mathbb{E}_{\mu_{h,a}}\left[\hat{U}(g(x))\right]\end{aligned}$$

In corresponding fashion we obtain the analogous representation for $\hat{U}(\rho.g')$, and hence

$$\rho.g \succcurlyeq_h \rho.g' \iff \mathbb{E}_{\mu_{h,a}}\left[\hat{U}(g(x))\right] \geqslant \mathbb{E}_{\mu_{h,a}}\left[\hat{U}(g'(x))\right] \tag{15}$$

Combining the implications of (14) and (15), we see that (a, f) and ρ are h-proportional. ▫

[STEP 4: PROPORTIONAL PLANS.] We now claim that for any $h \in \mathcal{H}$ and $p \in P$ there exists some $\sigma \in \Sigma$ such that $p \sim_h \sigma$. Fix some $p \in P$, and for each $n \in \mathbb{N}$ define p^n to be any PoA that agrees with p on the first n periods, then provides elements of Σ unambiguously. Note that $p_n \to p$ point-wise and hence converges in the product topology. Therefore, the claim reduces to finding a convergent sequence $\{\sigma_n\}_{n \in \mathbb{N}} \subset \Sigma$ such that $\sigma^n \sim_h p^n$, as continuity ensures the limits are indifferent.

We will prove the subsidiary claim by induction. Consider p^1, for each $(a, f) \in$ *supp*$[p^1]$, note, by assumption, $f : X \to \Sigma$. Let $\tau^{1,(a,f)} \in \Sigma$ be such that $\text{marg}_X \tau^{1,(a,f)} = \mu_{h,a}$. By [STEP 3], (a, f) and $\tau^{1,(a,f)}$ are h-proportional. And thus, $\tau^{1,(a,f)}.f \sim_h (a, f).f = (a, f)$, by PRP. Let $\sigma^1 \in \Sigma$ be such that $\sigma^1[E] = p^1[\{(a, f) | \tau^{1,(a,f)}.f \in E\}]$. Therefore,

$$\begin{aligned}U_h(p^1) &= \mathbb{E}_{p^1}\left[v_h(a, f)\right]\\ &= \mathbb{E}_{p^1}\left[\hat{U}(\tau^{1,(a,f)}.f)\right]\\ &= \mathbb{E}_{\sigma^1}\left[\hat{U}(\rho)\right]\\ &= \hat{U}(\sigma^1)\end{aligned}$$

where the third line comes from the change of variables formula for pushforward measures. This completes the base case.

Now, assume the claim hold for all h and $m \leqslant n-1$ for some $n \in \mathbb{N}$. Consider p^n. Note that for all h' of the form $h(x) = (h, p^n, (a, f), x)$, the implied continuation problem $p^n(h')$ satisfies the inductive hypothesis. Therefore, there exists a $\sigma^{n-1,h'} \sim_{h'} p(h')$ for all such h'.

Let $\star$ denote the mapping: $(a, f) \mapsto (a, f)^\star = (a, x \mapsto \sigma^{n-1,h(a,x)})$, where $h(a, x) = (h, p^n, (a, f), x)$. By construction, for each (a, f) in *supp*(p^n), and $x \in S_a$ we have $(a, f) \sim_h (a, f_{-x}\sigma^{n-1,h(a,x)})$ (using the notation from [STEP 2]). Employing STP we have $(a, f) \sim_h (a, f)^\star$ (i.e., enumerating the outcomes in S_a and changing f one entry at a time, where STP ensures that each iteration is indifferent to the last).

Let $\hat{p}^n \in P$ be such that $\hat{p}^n[E] = p^n[\{(a,f)|(a,f)^\star \in E\}]$. So,

$$\begin{aligned} U_h(p^n) &= \mathbb{E}_{p^n}\big[v_h(a,f)\big] \\ &= \mathbb{E}_{p^n}\big[v_h((a,f)^\star)\big] \\ &= \mathbb{E}_{\hat{p}^n}\big[v_h(b,g)\big] \\ &= U_h(\hat{p}^n) \end{aligned}$$

Applying the base case to $\hat{p}^n$ concludes the inductive step. Notice also, the convergence of $\{\sigma^n\}_{n\in\mathbb{N}}$ is easily verified, following the fact that the marginals on p_n are fixed for any σ^m with $m \geqslant n$. □

[Step 5: Representation.] Consider any $(a,f) \in \mathcal{A}\otimes P$. We claim that there exists an $(a,f') \in \mathcal{F}(a)$ such that $(a,f) \sim_h (a,f')$. Indeed, by [Step 4], for any $x \in S_a$, there exists some $\rho(a,x)$ such that $\rho(a,x) \sim_{h(a,x)} f(x)$, where $h(a,x) = (h,(a,f),(a,f),x)$. Define $f' \in \mathcal{F}(a)$ as $x \mapsto \rho(a,x)$. It follows from STP that $(a,f) \sim_h (a,f')$.

We know by [Step 3] that there exists a $\rho \in \Sigma$, h-proportional to (a,f), with $\text{marg}_X\rho = \mu_{h,a}$. Hence $(a,g) = (a,f).g \sim_h \rho.g$ for all $g : X \to \Sigma$. We have,

$$\begin{aligned} v_h(a,g) &= \hat{U}(\rho.g) \\ &= \mathbb{E}_{\mu_{h,a}}\big[u(x) + \delta\hat{U}(g(x))\big], \end{aligned}$$

and so, for (a,f'):

$$v_h(a,f') = \mathbb{E}_{\mu_{h,a}}\big[u(x) + \delta\hat{U}(\rho(a,x))\big].$$

By the indifference condition $\rho(a,x) \sim_{h(a,x)} f(x)$,

$$v_h(a,f) = \mathbb{E}_{\mu_{h,a}}\big[u(x) + \delta U_{h(a,x)}(f(x))\big]. \tag{16}$$

Notice, $h(a,x) \overset{\mathcal{A}}{\sim} h'(a,x) = (h,p,(a,f),x)$, so by Lemma 3, $\succcurlyeq_{h(a,x)} = \succcurlyeq_{h'(a,x)}$. Applying this fact, and plugging (16) into (9) provides

$$U_h(p) = \mathbb{E}_p\left[\mathbb{E}_{\mu_{h,a}}\big[u(x) + \delta U_{h'(a,x)}(f(x))\big]\right] \tag{17}$$

as desired. □

■

Proof of Theorem 7. Let $\{\mu_{h,a}\}_{h\in\mathcal{H},a\in A}$ be an SEE structure for $\succcurlyeq$ that satisfies AA-SYM. Let $\{\zeta_\mathbf{T}\}_{\mathbf{T}\in\mathcal{T}}$ be the associated family of observable processes. Fix $\mathbf{T}$ and some n period history $h \in \mathbf{T}$. Let, $(a_1,x_1)\dots(a_n,x_n)$, where for each $i \leqslant n$ let a_i is such that $T_i = S_{a_i}$ and x_i is the i^{th} component of h. This represents an $\mathcal{A}$-equivalence class of decision theoretic histories. In out standard abuse of notation, let h also denote this class of histories. Following this abuse, when it is not confusing to do so, let πh denote both the permuted statistical history and the $\mathcal{A}$-equivalence class represented by $(a_{\pi(1)},x_{\pi(1)})\dots(a_{\pi(n)},x_{\pi(n)})$.

Fix some n-permutation π. Let p denote the PoA that assigns a_i in the i^{th} period with certainty. Let p' be the π-permutation of p. We have

$$\alpha = \zeta_\mathbf{T}(h) = \mu_{\varnothing,a_1}(x_1)\cdot\mu_{(a_1,x_1),a_2}(x_2)\cdots\mu_{(a_1,x_1,\dots,a_{n-1},x_{n-1}),a_n}(x_n.$$

Let $\sigma,\sigma' \in \Sigma$ be such that $U_h(\sigma) = 1$ and $U_h(\sigma') = 0$. Then, by (SEE) we have

$$p_{-n}(\alpha\sigma + (1-\alpha)\sigma') \sim (p_{-n}\sigma')_{-h}\sigma$$

so, by AA-SYM, we have,

$$p'_{-n}(\alpha\sigma + (1-\alpha)\sigma') \sim (p'_{-n}\sigma')_{-h'}\sigma$$

which implies, again by (SEE),

$$\alpha = \mu_{\varnothing,a_{\pi(1)}}(x_{\pi(1)}) \cdot \mu_{(a_{\pi(1)},x_{\pi(1)}),a_{\pi(2)}}(x_{\pi(2)}) \cdots \mu_{(a_{\pi(1)},x_{\pi(1)},\ldots,a_{\pi(n-1)},x_{\pi(n-1)}),a_{\pi(n)}}(x_{\pi(n)}) = \zeta_{\pi\mathbf{T}}(\pi h).$$

Hence, $\zeta_{\mathbf{T}}(h) = \zeta_{\pi\mathbf{T}}(\pi h)$ as desired. ■