Plans of Action:

Supplement to Exploration and Correlation

Evan Piermont[*]and Roee Teper[†]

April 24, 2019

Abstract

In this supplemental material to Piermont and Teper (2019) 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.

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 2 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

[*]Royal Holloway, University of London, Department of Economics, evan.piermont@rhul.ac.uk
[†]University of Pittsburgh, Department of Economics, rteper@pitt.edu

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?

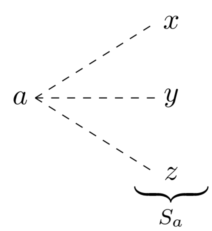

Figure 1: An action, a, and its support, S_a.

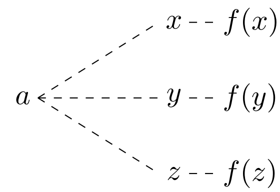

Figure 2: A degenerate PoA, (a, f).

Theorem 3 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 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.

1.1 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.[1] 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.[2] To begin, let $Q_0 = R_0 = \Delta^{\mathcal{B}}(\mathcal{A})$ and, for define recursively for each $n \geqslant 1$

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

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

[1] The requirement that for all x there exists an action that yields x with certainty (i.e., $S_a = \{x\}$) facilitates the identification of utilities over outcomes. Although the DM has preferences over such actions, this does not mean they are available in an arbitrary exploration problem, just as degenerate lotteries are part of the primitive choice set in von-Neumann-Morgenstern, but are not feasible in every decision problem.

[2] 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.

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

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

for all n. Let Q denote the restriction of $Q*$ that satisfies (1) and $R = Q \cap R^*$.

Proposition 1. *There exists a homeomorphism, $\lambda : R \to \Delta^{\mathcal{B}}(\mathcal{A} \otimes R))$ such that*

$$\mathrm{marg}_{\mathcal{A} \times \mathcal{K}(X \times R_{n-1})}(\lambda(r)) = \mathrm{proj}_n r. \tag{2}$$

Proof. In Section 1.6 ■

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

Proof. In Section 1.6 ■

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 $p_{-n}q \equiv \bigcup_{h \in \mathcal{H}(p,n)} p_{-h}q$.

For $h = (p^1, a^1, f^1, x^1 \ldots p^n, a^n, f^n, x^n)$ 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.

1.2 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.[3] 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 $\sigma, \sigma' \in \Sigma$,*

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

[3]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.

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.

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

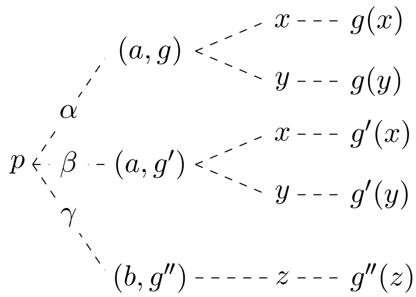

Figure 3: A PoA, p, defined by $p(a, g) = \alpha$, $p(a, g') = \beta$ and $p(b, g'') = \gamma = 1 - \alpha - \beta$.

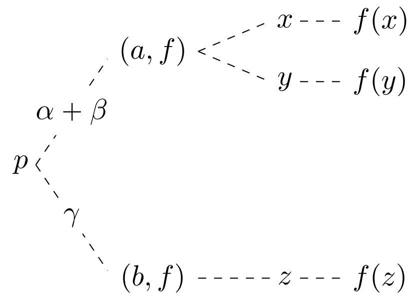

Figure 4: The PoA, $p.f$ where $f : X \to P$, and p is defined in figure 3. Notice, $p.f(a, f) = p(a, g) + p(a, g') = \alpha + \beta$.

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.[4]

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

[4]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.

replace the continuation problems with objectively equivalent plans, the DM should be indifferent between p and 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 1.4 when we discuss the behavioral restrictions of exchangeability in the current setup.

1.3 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 3 (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*

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

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 1.7. ∎

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.

1.4 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 $h \in \mathcal{H}(p, n)$, $h' \in \mathcal{H}(q, n)$, $\mathrm{proj}_{\mathcal{A}^n} h = \pi\big(\mathrm{proj}_{\mathcal{A}^n} h'\big)$.*

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 4 (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 1.7. Conditions 2, 3, and 4 are equivalent due to Theorem 2 in Piermont and Teper (2019). ■

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.

1.5 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 (2019). 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, **?** 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. (2017) 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').

1.6 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*

$$\text{marg}_{\mathcal{A}\times\mathcal{K}(X\times Q_{n-1})}(\lambda(q)) = q_n. \tag{3}$$

Proof. [Step 1: Extension Theorem.]

Let $C_n = \{(q_0, \dots q_n) \in \prod_{k=0}^n Q_k | q_k = F_{k+1}(q_{k+1}), \forall k = 1 \dots n-1\}$, and $T_n = \mathcal{K}(X \times C_n)$ for $n \geqslant 0$. Let $T^* = \prod_{n=0}^\infty T_n$ and $T = \{t \in T^* | (\text{proj}_{T_n} t_{n+1} = t_n\}$. Let $Y_0 = \Delta^{\mathcal{B}}(\mathcal{A})$ and for $n \geqslant 1$ let $Y_n = \Delta^{\mathcal{B}}(\mathcal{A} \times T_0 \times \dots \times T_n))$. We say the the sequence of probability measures $\{\nu_n \in Y_n\}_{n\geqslant 0}$ is consistent if $\text{marg}_{A\dots T_{n-1}} \nu_{n+1} = \nu_n$ 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 $\text{marg}_{\mathcal{A}} \nu = \nu_0$ and $\text{marg}_{A\dots T_n} \nu = \nu_n$. 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 $D_n = \{(t_0, \dots t_n) \in \times_{n=0}^n T_n | t_k = \text{proj}_{T_n}(t_{k+1}), \forall k = 1 \dots n-1\}$. Let $Y^d = \{\{\nu_n\} \in Y^c | \nu_n(\mathcal{A} \times D_n) = 1, \forall n \geqslant 0\}$. 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 $\text{marg}_{\mathcal{A}\times\mathcal{K}(X\times Q_{n-1})}(\text{marg}_{\mathcal{A}\times T_{n-1}}(\nu_n)) = q_n$ 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:

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

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 $k \in 1 \ldots n$) q_n uniquely determines $q_0 \ldots q_{n-1}$, which, (given the projection mappings) uniquely determines $T_0 \ldots T_{n-1}$— m_i is also injective. As for, m_n^{-1}, continuity follows from the continuity of F_k for $k \in 1 \ldots n$ 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, $\psi(Y^d) = \{\nu \in \Delta^{\mathcal{B}}(\mathcal{A} \times T^*) | \nu(\mathcal{A} \times T) = 1\}$. □

[Step 3: Extending Forwards.]

Let τ denote the map from $\mathcal{A} \times \mathcal{K}(X \times Q) \to \mathcal{A} \times T$ as

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

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 $\tau^{-1}(E) = \bigcap_{n\geqslant 0} \{\{x, q_0, \ldots q_n \times_{k=n_1^\infty} Q_k\} \in K(X \times Q^*)\}$, 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, $\mathrm{marg}_{\mathcal{A}}(\kappa(\nu)) = \mathrm{marg}_{\mathcal{A}}(\nu)$ and

$$\mathrm{marg}_{\mathcal{A}\times\mathcal{K}(X\times Q_{n-1})}(\kappa(\nu)) = \mathrm{marg}_{\mathcal{A}\times\mathcal{K}(X\times Q_{n-1})}(\mathrm{marg}_{\mathcal{A}\times T_{n-1}}(\nu))$$

for all $n \geqslant 1$.

$\square$

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

Proof of Proposition 1. 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 $\Gamma^r_n = \{(a, \{x, q\}) \in \mathcal{A} \otimes Q | q_k \in R_k,\ k = 0 \ldots n\}$. Then $\lambda(r)(\Gamma^r_n) = \mathrm{marg}_{A\times K(X\times Q_n)}(\lambda(r))(\mathcal{A} \otimes R_n) = r_{n+1}(\mathcal{A} \otimes R_n) = 1$ for all $n \geqslant 1$. So $\lambda(r)(\mathcal{A}\otimes R) = \lambda(r)(\cap_{n\geqslant 0}\Gamma^r_n) = \lim \lambda(r)(\Gamma^r_n) = 1$. Now, fix $q \in Q$ with $\lambda(q)(\mathcal{A}\otimes R) = 1$, then $q_n(\mathcal{A}\otimes R_{n-1}) = \mathrm{marg}_{A\times K(X\times Q_{n-1})}(\lambda(q))(\mathcal{A}\otimes R_{n-1}) = \lambda(r)(\Gamma^r_n) \geqslant \lambda(r)(\mathcal{A}\otimes R) = 1$ for all $n \geqslant 0$ and so $q \in R$. $\blacksquare$

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

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

Where W^n *is* $W(W(\ldots W(E)\ldots))$ *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 2. 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 1 that $r \in \prod_{n=0}^{\infty} P_n$. Finally, note that $W^\star(r) \subseteq W^\star(W(r))$. $\blacksquare$

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

$$U_h(p) = \mathbb{E}_p \left[v_h(a,f) \right] \tag{4}$$

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:

$$\hat{U}(\sigma) = \mathbb{E}_\sigma \left[\hat{u}(x,\rho) \right] \tag{5}$$

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,

$$\hat{U}(\sigma) = \mathbb{E}_\sigma \left[u(x) + W(\rho) \right] \tag{6}$$

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 $\beta = 0$:

$$\hat{U}(\sigma) = \mathbb{E}_\sigma \left[u(x) + \delta \hat{U}(\rho) \right] \tag{7}$$

Both representing $\hat{\succcurlyeq}$ and being unique up to affine translations, we can normalize each

U_h to coincide with $\hat{U}$ over Σ. $\square$

[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

$$\hat{V}_{h,a}(f) = \mathbb{E}_{\mu_{h,a}}\left[\hat{U}(f(x))\right] \tag{8}$$

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

[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 (8), we know

$$(a, g) \succcurlyeq_h (a, g') \iff \mathbb{E}_{\mu_{h,a}}\left[\hat{U}(g(x))\right] \geqslant \mathbb{E}_{\mu_{h,a}}\left[\hat{U}(g'(x))\right] \tag{9}$$

From (7) we have

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

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

$$\rho.g \succcurlyeq_h \rho.g' \iff \mathbb{E}_{\mu_{h,a}}\left[\hat{U}(g(x))\right] \geqslant \mathbb{E}_{\mu_{h,a}}\left[\hat{U}(g'(x))\right] \tag{10}$$

Combining the implications of (9) and (10), 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 $\mathrm{marg}_X\tau^{1,(a,f)} = \mu_{h,a}$. 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,

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

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,

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

and so, for (a, f'):

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

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

$$v_h(a, f) = \mathbb{E}_{\mu_{h,a}}\left[u(x) + \delta U_{h(a,x)}(f(x))\right]. \tag{11}$$

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 (11) into (4) provides

$$U_h(p) = \mathbb{E}_p\left[\mathbb{E}_{\mu_{h,a}}\left[u(x) + \delta U_{h'(a,x)}(f(x))\right]\right] \tag{12}$$

as desired. □

■

Proof of Theorem 4. 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),

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

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

References

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

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

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

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. Subjective Information Choice Processes. 2017. *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.

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. The Dynamics of Preferences, Predictive Probabilities, and Learning. 2019. *Working paper.*

Evan Piermont and Roee Teper. Exploration and correlation. *Games and Economic Behavior*, 2019.

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