Equivalent Conditional Probability Systems

Byung Soo Lee*

November 13, 2024

Abstract

By defining conditional probabilities as primitives, Rényi (1955) provides a framework for understanding scenarios that challenge classical probability theory, such as conditioning on events with zero prior probability. The paper proposes a notion of equivalence among such *conditional probability systems* (CPSs) and shows its appealing properties (e.g., existence of a canonical form). Additionally, we demonstrate an application of the equivalence concept by continuing the work started in Brandenburger et al. (2023) to show that, at some fundamental level, *lexicographic probability systems* (LPSs) and *finitary* CPSs are merely different ways of encoding the same probabilistic information.

KEYWORDS: Conditional probability systems, equivalence, lexicographic probability systems, belief revision

*Rotman School of Management, University of Toronto, 105 Saint George Street, Toronto, Ontario, Canada M5S 3E6. Email: byungsoo.lee@utoronto.ca

1 Introduction

We begin by reproducing the definition of conditional probability systems in Rényi (1955). In this paper, we fix a measurable space $\langle X, \mathscr{X} \rangle$, where X is a nonempty set of states and $\mathscr{X}$ is a σ-algebra on X. Elements of $\mathscr{X}$ are called *events*. The set of all σ-additive probability measures on $\langle X, \mathscr{X} \rangle$ is denoted by Π.

Definition 1.1. Let $\pi\colon \mathscr{X} \to \Pi$ be a partial function. We denote its domain by $\mathscr{D}(\pi)$ and adopt the convention of writing $\pi(A \mid B)$ for $\pi(B)(A)$. We say that π is a *conditional probability system* (or simply *CPS*) on $\langle X, \mathscr{X} \rangle$ if

$$\pi(C \mid C) = 1 \tag{1}$$

for all $C \in \mathscr{D}(\pi)$; and

$$\pi(A \mid B)\pi(B \mid C) = \pi(A \mid C) \tag{2}$$

for all $A, B, C \in \mathscr{X}$ such that $B, C \in \mathscr{D}(\pi)$ and $A \subseteq B \subseteq C$. The set of all conditional probability systems on $\langle X, \mathscr{X} \rangle$ is denoted by Σ. Elements of the domain $\mathscr{D}(\pi)$ are called *conditioning events* (or simply *conditions*) of π. The set of all $\pi \in \Sigma$ such that $\mathscr{D}(\pi) = \mathscr{B}$ is denoted by $\Sigma_{\mathscr{B}}$. We also adopt the abbreviation $D(\pi) \coloneqq \bigcup \mathscr{D}(\pi)$.[1]

What we might call—for lack of a better word—the "classical" approach to conditional probability takes the prior belief as its primitive and defines conditional probability as a formula of prior probabilities. Rényi's approach to conditional probability takes conditional probabilities as the primitives and imposes consistency with the classical formula among those primitives as an axiom. While there are many interesting technical consequences of this approach, it has proven to be an extremely fruitful one in applied epistemology. CPSs have facilitated the modeling of agents' reasoning about events that have zero prior probability, which has resulted in what are now standard refinements of the ubiquitous Nash equilibrium solution concept (e.g., Myerson 1986). Since then, game theorists have provided epistemic foundations for important non-equilibrium concepts in game theory by representing beliefs about beliefs using CPSs (e.g., Battigalli 1996; Battigalli and Siniscalchi 2002; Catonini and De Vito 2024).

Rényi's approach also allows the expression of the distinction between "updates" and "revision" as Gärdenfors (1988) defined them. One might argue that the prior-based classical approach only allows for revisions since Bayesian updating reflects "a change of in knowledge about a *static* world but not for recording changes in an *evolving* world" (Kern-Isberner 2001, p. 393) because it simply rescales posterior beliefs after knowledge received rules out some states of the world. By taking conditional probabilities as primitives and allowing arbitrary

[1] The only requirement imposed on $\mathscr{D}(\pi)$ by Definition 1.1 is that $\emptyset \notin \mathscr{D}(\pi)$. See Rényi (1955).

limitations on the events that one can condition on, it is possible to model conditional beliefs that would be irreconcilable with the classical approach.

Example 1.1 (Contextuality in quantum mechanics)**.** Let $X = \{\uparrow, \downarrow\}^3$, where $(x_1, x_2, x_3) \in X$ represents three properties of an electron. It is sometimes the case that it is impossible to measure the second and third coordinates together. Furthermore, the measurement of the first coordinate may depend on whether one measures it together with the second or third coordinate. One can define a CPS π that reflects such *contextuality* by limiting the collection $\mathscr{D}(\pi)$ of conditioning events:

$$\pi(\{(\uparrow, x_2, \uparrow), (\uparrow, x_2, \downarrow)\} \mid \{\uparrow, \downarrow\} \times \{x_2\} \times \{\uparrow, \downarrow\}) = 1$$
$$\pi(\{(\uparrow, \uparrow, x_3), (\uparrow, \downarrow, x_3)\} \mid \{\uparrow, \downarrow\} \times \{\uparrow, \downarrow\} \times \{x_3\}) = 0$$

On the other hand, even if a CPS does not explicitly define a conditional probability, it may be inferred using the classical approach.

Example 1.2. Let $X = \{1, 2, 3\}$ and $\mathscr{D}(\pi) = \{X\}$. Define π as follows.

$$\pi(\{1\} \mid X) = \pi(\{2\} \mid X) = 1/3$$

One may infer that $\pi(\{1\} \mid \{1, 2\})$ *should be* $1/2$ even if it is not explicitly defined.

This leads to a question of what is the "correct" notion of equivalence between conditional probability systems. In section 2, we develop—and prove some virtues of—one useful notion of equivalence, which admits the existence of a canonical form for each CPS.

In section 3, we apply this notion of equivalence to finish an exercise partially completed in Brandenburger et al. (2023) by demonstrating that *lexicographic probability systems* (LPSs, see Blume et al. 1991) and *finitary* CPSs are different ways of encoding the same information. LPSs, like CPSs, have also been applied to foundations of game theory, most notably to solving epistemic puzzles relating to admissibility (e.g., Brandenburger et al. 2008; Dekel et al. 2016; Lee 2016; Keisler and Lee 2023; Catonini and De Vito 2024). Brandenburger et al. (2023) illustrate the relationship between two important epistemic operators in game theory (*strong belief* and *assumption*) through the use of conversion mappings between CPSs and LPSs. However, because these maps are not bijective and only partially specified in the LPS–to-CPS direction, a gap was left in the argument that they encode the same information. While the gap does not hinder their main goal of identifying the relationship between strong belief and assumption, we nevertheless believe that it is of independent interest. Section 4 concludes with a discussion of a potential extension.

2 Equivalence

In this paper, we adopt the following conventions so that the nature of each object can be identified to some extent by the typeface of the symbol that is used to represent it.

states (lower case):	$a, b, c, d \ldots$
sets of states (upper case):	$A, B, C, D \ldots$
sets of sets of states (calligraphic):	$\mathscr{A}, \mathscr{B}, \mathscr{C}, \mathscr{D} \ldots$
beliefs (Greek lower case):	$\xi, \pi, \lambda, \xi, \ldots$
sets of beliefs (Greek upper case):	$\Sigma, \Pi, \Lambda, \Xi, \ldots$

Definition 2.1. Let $A \in \mathscr{X}$ and $\pi \in \Sigma$. We say that A is *null* under π (or π-null) if $A \subseteq D(\pi)$ and $\pi(A \mid C) = 0$ for all $C \in \mathscr{D}(\pi)$. The set of all π-null events is denoted by $\mathscr{N}(\pi)$.

Notice that A cannot be π-null if there is some part of A that does not meet any condition of π, i.e., if there is some event $B \subseteq A$ such that $B \cap C = \emptyset$ for all $C \in \mathscr{D}(\pi)$. To put it another way, events that do not meet any condition of π cannot be π-null and, furthermore, events that contain such events cannot be π-null either. In the preference-based approach of Savage (1972) to subjective probability, an event A is defined to be null to the decision maker if he is indifferent between all choices when A occurs.

Our definition is congruent with the spirit of Savage's approach. Suppose that Ann, an equity analyst, models various scenarios but not all. We might represent the content of her research as follows: A scenario is some subset of the states of the world. A model of a scenario assigns probabilities conditional on the true state being in the scenario. Thus, the content of her research is some CPS π such that each condition C is a scenario and $\pi(C)$ is a model of that scenario. The conditional expected value of an investment is its valuation in the financial model that corresponds to the condition.

Suppose that, in each model, the valuation of a portfolio is unaffected by changes only to its returns in states of the world with central bank interest rate hikes, i.e., rate hikes are Savage-null according to each model. Should Ann tell her clients that rate hikes are irrelevant to portfolio returns? Perhaps the first instinct is to say yes, but consider the case that none of her scenarios include the possibility of rate hikes. Then the irrelevance of rate hikes in any specific model is not a conclusion of Ann's analysis but rather just an assumption of the model. In other words, Ann's report π tells us that rates hikes are Savage-null conditional on rate hikes being Savage-null.

This is materially different from saying that she considered scenarios that do include the possibility of rate hikes and concluded that rate hikes are impossible. Ann's clients may read her report and find it satisfactory if they already believe that rate hikes are impossible. If they do not, then they may ask her to add scenarios that allow for the possibility of rate hikes to the domain of the CPS that represents her research. This justifies our requirement that π-null events be composed entirely of states that are in at least one condition of π.

Another point of note is that our definition of π-null events requires that even an uncountable union of π-null events is π-null if and only if that union has null probability conditional on every event in the domain of π. Given that we informally take the view that members of $\mathscr{N}(\pi)$ are the Savage-null events, this means that we take the position that π pins down what is irrelevant to decision. Under such an interpretation, it may be the case that $\pi, \mu \in \Sigma$ agree with each other on the intersection of their domains yet contain contradictory information about making decisions. See the following example.

Example 2.1. Let X be the real line and $\mathscr{X}$ its Lebesgue-measurable subsets. Let $\mathscr{D}(\pi) = \{[0,1] \cup \{x\} \mid x \in X\}$ and $\mathscr{D}(\mu) = \mathscr{D}(\pi) \cup \{[0,2]\}$. Let $\pi(C) = \mu(C)$ be the uniform measure on $[0,1]$ for all $C \in \mathscr{D}(\pi)$ and let $\mu([0,2])$ be the uniform measure on $[0,2]$. Note that $\{x\} \in \mathscr{N}(\pi) \cap \mathscr{N}(\mu)$ for all $x \in [1,2]$ but $[1,2] \in \mathscr{N}(\pi) \setminus \mathscr{N}(\mu)$.

Such interpretations are possible, and even desirable, because Rényi (1955) takes conditional beliefs and conditioning events to be the *primitives* rather than deriving conditional beliefs from a prior belief. Another consequence of this interpretation of the CPS approach to probability is that null events are irrelevant to decisions but non-null events are not necessarily relevant. Thus, instead of defining non-null events, we define *potent* events, which are the non-null events such that every part meets some conditioning event.

Definition 2.2. Let $A \in \mathscr{X}$ and $\pi \in \Sigma$. We say that A is *potent* under π (or π-potent) if $A \subseteq D(\pi)$ and there exists $C \in \mathscr{D}(\pi)$ such that $\pi(A \mid C) \neq 0$. The set of all π-potent is denoted by $\mathscr{P}(\pi)$.

We summarize some frequently invoked properties of null and potent events below in the following lemma. The proofs are omitted because all parts follow trivially from definitions.

Lemma 2.1. Let $\pi \in \Sigma$.

(i) $\mathscr{D}(\pi) \subseteq \mathscr{P}(\pi)$;
(ii) $D(\pi) = \bigcup \mathscr{P}(\pi)$;
(iii) $\mathscr{P}(\pi) \cap \mathscr{N}(\pi) \neq \emptyset$;
(iv) $\mathscr{P}(\pi) \cup \mathscr{N}(\pi) = \{A \in \mathscr{X} \mid A \subseteq D(\pi)\}$;

It is important to recall that, in Rényi's definition, conditional beliefs are *primitives* and therefore *directly* defined instead of being *indirectly* encoded in a prior and extracted by means of the classical conditional probability formula. In the prior-based approach to conditional probability, consistency requirements for conditional beliefs can be quite strong because all non-null events are conditioning events. In the CPS-based approach, the consistency requirements can be quite weak because each CPS explicitly defines its own set of conditions, which can be sparse. Equation (2) in Definition 1.1 only requires pairwise consistency across the entries of the partial function $\pi\colon \mathscr{X} \to \Pi$; that is, if π is a CPS, then the entries $\pi(B)$ and $\pi(C)$ must satisfy (2) for every pair $B, C \in \mathscr{D}(\pi)$ such that $B \subsetneq C$. The following is an extreme case in which (2) need not be checked because the conditions in $\mathscr{D}(\pi)$ are incomparable in the subset relation.

Example 2.2. Let $X = \{1, 2, 3\}$ and $\mathscr{X} = 2^X$. Let $\mathscr{D}(\pi) = \{\{1, 2\}, \{2, 3\}, \{1, 3\}\}$. Let $\pi(\{x\} \mid \{x, (x + 1) \mod 3\}) = 1$.

Furthermore, there may be some conditional beliefs that are not directly defined by a CPS but nevertheless can be calculated using the classical approach. Naturally, there will be cases—such as Example 2.3 below—where our intuition strongly suggests that π and μ are encoding the same information even though $\pi \neq \mu$.

Example 2.3. Let $X = \{1, 2, 3\}$ and $\mathscr{X} = 2^X$. Let $\mathscr{D}(\pi) = \{\{1, 2\}, \{2, 3\}\}$ and $\mathscr{D}(\mu) = \{\{1, 2, 3\}\}$. Define π and μ as follows.

$$\pi(\{1\} \mid \{1, 2\}) = 1/2 = \pi(\{2\} \mid \{2, 3\})$$
$$\mu(\{1\} \mid \{1, 2, 3\}) = 1/3 = \mu(\{2\} \mid \{1, 2, 3\})$$

But how can we formalize such intuition about equivalence? We do by formalizing the direct and indirect implications of a CPS. In the remainder, let us identify each CPS π with its graph, which is a subset of $\mathscr{P} \times \Pi$. This allows Σ to be partially ordered as follows.

Definition 2.3. Let $\pi, \mu \in \Sigma$. We say that μ *extends* π (to $\mathscr{D}(\mu)$) if $\mu \supseteq \pi$.

The extension μ implies the beliefs conditional on the conditions of π in the very direct sense that $\pi(C) = \mu(C)$ for all $C \in \mathscr{D}(\pi)$. See Example 2.4 below.

Example 2.4. Let $X = \{1, 2, 3\}$ and $\mathscr{X} = 2^X$. Let $\mathscr{D}(\pi) = \{\{1, 2\}\}$ and $\mathscr{D}(\mu) = 2^X \setminus \emptyset$. Define π and μ as follows so that $\mu \supseteq \pi$.

$$\pi(\{1\} \mid \{1, 2\}) = 1/2 = \pi(\{2\} \mid \{1, 2\})$$
$$\mu(A \mid B) = \pi(A \cap B \mid \{1, 2, 3\})/\pi(B \mid \{1, 2, 3\})$$

On the other hand, extensions fail to capture indirect implications of the sort that leads us to intuit equivalence of beliefs in Example 2.3. This leads us to formalize the indirect implications of a CPS in two steps. We begin by giving a label to cases when one CPS can be extended to another using the classical conditional probability formula without changing what is relevant to decisions.

Definition 2.4. Let $\pi, \beta \in \Sigma$. We say that β *tightly extends* π and write $\beta \sqsupseteq \pi$ if β is the unique extension of π to $\mathscr{D}(\beta)$ and $\mathscr{P}(\pi) = \mathscr{P}(\beta)$.[2]

[2] i.e., $\beta = \nu$ if and only if $\nu \supseteq \pi$ and $\mathscr{D}(\nu) = \mathscr{D}(\beta)$ and $\mathscr{P}(\nu) = \mathscr{P}(\pi)$.

What can we say about the content of π and μ when π tightly extends μ? First of all, μ does not change which events are decision-relevant because $\mathscr{P}(\pi) = \mathscr{P}(\mu)$. Secondly, π exactly determines μ on the domain of μ via (2). However, due to the aforementioned fact that consistency checks in (2) are only applied to pairs of conditions that can be ordered by the subset relation, only *some* of the information encoded in π pins down μ. More specifically, the values of π on

$$\{E \in \mathscr{D}(\pi) \mid \exists F \in \mathscr{D}(\mu)\colon E \subseteq F \vee F \subseteq E\}$$

pin down μ. That is, different parts of π constrain the extension depending on the domain to which it is being extended. Given that π can itself contain unchecked inconsistencies of the sort in Example 2.2, it may be the case that π tightly extends both μ and ν that each directly imply contradictory information about beliefs conditional on some C. See Example 2.5 below.

Example 2.5. Let $X = \{1, 2, 3, 4\}$ and $\mathscr{X} = 2^X$. Let

$$\begin{aligned} \mathscr{D}(\pi) &= \{C \in \mathscr{X} \mid |C| = 2\}; \\ \mathscr{D}(\mu) &= \mathscr{D}(\pi) \cup \{\{1, 2, 3\}, \{1, 3\}\}; \\ \text{and} \quad \mathscr{D}(\nu) &= \mathscr{D}(\pi) \cup \{\{1, 3, 4\}, \{1, 3\}\}. \end{aligned}$$

Let π, μ, ν be defined as follows.

$$\pi(\{x\} \mid \{x, (x+1) \mod 4\}) = 1$$
$$\mu(\{1\} \mid \{1, 2, 3\}) = 1 \qquad \mu(\{1\} \mid \{1, 3\}) = 1$$
$$\nu(\{3\} \mid \{1, 3, 4\}) = 1 \qquad \nu(\{1\} \mid \{1, 3\}) = 0$$

We see that $\mu \sqsupseteq \pi$ and $\nu \sqsupseteq \pi$ but $\mu(\{1, 3\}) \neq \nu(\{1, 3\})$.

In Example 2.5, adding conditions to the domain of CPS π permits the calculation of the belief conditional on $\{1, 3\}$ via the consistency constraints that correspond to the addition of the conditions. On the other hand, depending on consistency checks imposed (i.e., the other conditions being added to the domain), the calculated belief conditional on $\{1, 3\}$ may be different! As such, we cannot say that $\mu(\{1, 3\})$ is *the* belief conditional on $\{1, 3\}$ implied by π just because μ tightly extends π.

Thus, further refinement is needed to formalize a notion of implied conditional beliefs that avoids such contradictions. We call this *imputation*.

Definition 2.5. Let $\pi, \mu \in \Sigma$. We say that π *imputes* μ and write $\pi \to \mu$ if, for each $\varrho \sqsupseteq \pi$, there exists $\nu \sqsupseteq \varrho$ such that $\nu \supseteq \mu$.

Informally, μ can be "calculated" from π if π pins down (i.e., is tightly extend by) some $\varrho \supseteq \mu$. When π imputes μ, μ can be calculated from π in this way. Furthermore, μ can be "calculated" from any ϱ that tightly extends π. As a consequence, if π imputes a belief conditional on some event C, that must be the *only* belief conditional on C that π imputes.

As such, it is reasonable to view imputations of π as implications of π. We therefore define equivalence to be equality of imputations as follows:

Definition 2.6. Let $\pi, \mu \in \Sigma$. We say that π is *equivalent* to μ and write $\pi \equiv \mu$ when

$$\{\nu \in \Sigma \mid \pi \to \nu\} = \{\nu \in \Sigma \mid \mu \to \nu\}. \tag{3}$$

The equivalence class of π is denoted by $[\pi] \coloneqq \{\mu \in \Sigma \mid \pi \equiv \mu\}$.

Its definition makes it obvious that $\equiv$ is an equivalence relation. The definition also suggests a way to define the canonical representation of each equivalence class.

Definition 2.7. Let $\pi \in \Sigma$. We define $\bar{\pi} \coloneqq \bigcup\{\mu \in \Sigma \mid \pi \to \mu\}$ and say that it is the *canonical representation* of π.

Proposition 2.1 justifies the use of canonical representations as defined by showing that each CPS is equivalent to its canonical representation. Lemma 2.2 that follows is a technical result used to prove Proposition 2.1.

Proposition 2.1. Let $\pi \in \Sigma$. Then $\pi \equiv \bar{\pi}$.

Proof. Lemma 2.2 says that $\{\zeta \in \Sigma \mid \zeta \sqsupseteq \pi\}$ has at least one $\sqsupseteq$-maximal element. Fix one such $\sqsupseteq$-maximal τ.

If $\mu \leftarrow \pi$, then, for each $\varrho \sqsupseteq \pi$, there exists $\nu \sqsupseteq \varrho$ such that $\nu \supseteq \mu$. It follows that, because $\tau \sqsupseteq \pi$, there exists $\nu \sqsupseteq \tau$ such that $\nu \supseteq \mu$. However, $\tau = \nu$ in that case because τ is maximal among the tight extensions of π. It follows that $\tau \supseteq \mu$ for all $\mu \leftarrow \pi$. Therefore, $\tau \supseteq \bar{\pi}$ and $\pi \to \bar{\pi}$. It is obvious that $\bar{\pi} \to \pi$ since $\bar{\pi} \supseteq \pi$. It follows that $\pi \equiv \bar{\pi}$. □

Lemma 2.2. Let $\pi \in \Sigma$. There exists $\mu \sqsupseteq \pi$ such that $\mu = \nu$ for all $\nu \sqsupseteq \mu$.

Proof. The set $\{\zeta \in \Sigma \mid \zeta \sqsupseteq \pi\}$ is partially ordered by $\sqsupseteq$.

Let Ξ be a chain in $\{\zeta \in \Sigma \mid \zeta \sqsupseteq \pi\}$. Equation (2) checks consistency of conditional beliefs across *pairs* of conditions. It follows that, if $\xi \cup \xi'$ is a CPS for every $\xi, \xi' \in \Xi$, then (2) is satisfied by $\bigcup \Xi$. Since Ξ is a chain, every pair in Ξ is comparable in $\sqsupseteq$ and therefore also in $\supseteq$. It follows that $\xi \cup \xi'$ is a CPS for every $\xi, \xi' \in \Xi$. Therefore, $\bigcup \Xi$ is a CPS.

We also have $\mathscr{P}(\bigcup \Xi) = \mathscr{P}(\pi)$ because $\xi \sqsupseteq \pi$ for all $\xi \in \Xi$. Furthermore, $\bigcup \Xi$ is the unique extension of each $\xi \in \Xi$ to $\mathscr{D}(\bigcup \Xi) = \bigcup_{\xi \in \Xi} \mathscr{D}(\xi)$ because it would otherwise be the case that the restriction of $\bigcup \Xi$ to the domain of some $\xi' \in \Xi$ does not coincide with ξ'.

Therefore, $\bigcup \Xi \sqsupseteq \xi$ for all $\xi \in \Xi$. We have shown that every $\sqsupseteq$-chain in $\{\zeta \in \Sigma \mid \zeta \sqsupseteq \pi\}$ has an upper bound. By Zorn's Lemma, it follows that $\{\zeta \in \Sigma \mid \zeta \sqsupseteq \pi\}$ has at least one $\sqsupseteq$-maximal element, which is the desired result. □

We conclude this section with a useful property, which we make use of in the subsequent section, of any CPS μ that provides ways to exactly calculate probabilities conditional on any π-potent event, i.e., μ that tightly extends π to $\mathscr{P}(\pi)$. It says that such a μ *must be* the canonical representation of π. When π is extended to $\mathscr{P}(\pi)$, conditional probabilities take on a more "classical" flavor. Recall that, in the classical approach to conditional probability, we can condition on any event that has a non-zero prior probability. While there is no prior in the CPS approach, such events may be viewed as analogous to potent events. Thus, Proposition 2.2 below says that, if there is only one way to extend a CPS to allow for such quasi-classical conditioning on all potent events, that extension is the canonical representation.

Proposition 2.2. Let $\pi, \mu \in \Sigma$. If $\mathscr{D}(\mu) = \mathscr{P}(\pi)$ and $\mu \sqsupseteq \pi$, then $\mu = \bar{\pi}$.

Proof. By assumption, μ is the unique extension of π to $\mathscr{P}(\pi)$. It follows that $\mu \sqsupseteq \nu$ for all $\nu \sqsupseteq \pi$ due to the fact that $\mathscr{D}(\mu) = \mathscr{P}(\pi) = \mathscr{P}(\nu) \supseteq \mathscr{D}(\nu)$ for such ν. It follows that $\pi \to \varrho$ if and only if $\mu \supseteq \varrho$. Therefore, $\bar{\pi} = \bigcup\{\varrho \in \Sigma \mid \mu \supseteq \varrho\} = \mu$. □

3 Lexicographic Probability Systems

In this section, we apply our definition of equivalent CPSs toward refining some existing results of Brandenburger et al. (2023) on the relationship between lexicographic probability systems and finitary conditional probability systems.

Definition 3.1. We say that CPS $\pi \in \Sigma$ is *finitary* if $\{\emptyset\} \cup \mathscr{D}(\pi)$ is a finite subalgebra of $\mathscr{X}$. The set of all finitary conditional probability systems is denoted by Φ.

Definition 3.2. We say that finite sequence $\lambda = (\lambda_0, \ldots, \lambda_{n-1}) \in \bigcup_{m \geq 1} \Pi^m$ of σ-additive probability measures on $\langle X, \mathscr{X} \rangle$ is a *lexicographic probability system* (or simply *LPS*) if it is *mutually singular*, i.e., there exist pairwise disjoint events $U_0, \ldots, U_{n-1} \in \mathscr{X}$ such that $\lambda_i(U_i) = 1$ for all $i \in \{0, \ldots, n-1\}$. The set of all LPSs on $\langle X, \mathscr{X} \rangle$ is denoted by Λ. The length n of such λ is denoted by $\ell(\lambda)$.[3]

Definition 3.3. Let $\lambda = (\lambda_0, \ldots, \lambda_{n-1}) \in \Lambda$. Define the following.

$$\mathscr{N}(\lambda) \coloneqq \{E \in \mathscr{X} \mid \forall i \quad \lambda_i(E) = 0\} \tag{4}$$

$$\mathscr{P}(\lambda) \coloneqq \{E \in \mathscr{X} \mid \exists i \quad \lambda_i(E) > 0\} \tag{5}$$

$$m(\lambda, E) \coloneqq \inf\{i \mid \lambda_i(E) > 0\} \tag{6}$$

[3] Blume et al. (1991) use the term LPS to refer to any finite sequence of probability measures and label mutually singular sequences as lexicographic *conditional* probability systems (LCPSs). On the other hand, Brandenburger et al. (2008) and Brandenburger et al. (2023) use the terminology in Definition 3.2.

Events in $\mathscr{N}(\lambda)$ and $\mathscr{P}(\lambda)$ are respectively called null and potent, which parallels the analogous terminology for CPSs. Note that $E \in \mathscr{P}(\lambda)$ if and only if $0 \leq m(\lambda, E) < \infty$.[4]

Brandenburger et al. (2023) define maps for conversions between LPSs and finitary CPSs. We restate those definitions below with some immaterial modifications in presentation. For a given set $\mathscr{B} \subseteq \mathscr{X}$ of events, $g_{\mathscr{B}}$ converts LPSs that assign nonzero lexicographic probability to all members of $\mathscr{B}$ into finitary CPSs.

Definition 3.4 (Brandenburger et al. (2023))**.** Let $\mathscr{B} \subseteq \mathscr{X}$ such that $\emptyset \notin \mathscr{B}$. Then $g_{\mathscr{B}} \colon \{\lambda \in \Lambda \mid \mathscr{B} \subseteq \mathscr{P}(\lambda)\} \to \Sigma$ is defined as follows.

$$\mathscr{D}(g_{\mathscr{B}}(\lambda)) \coloneqq \mathscr{B} \tag{7}$$

$$g_{\mathscr{B}}(\lambda)(A \mid B) \coloneqq \frac{\lambda_{m(\lambda,B)}(A \cap B)}{\lambda_{m(\lambda,B)}(B)} \quad \text{for all } B \in \mathscr{B} \tag{8}$$

The map f goes in the opposite direction and converts finitary CPSs into an LPSs. Although it has been reformulated here for maximum compactness, Definition 3.5 below is equivalent to that which is given in Brandenburger et al. (2023).

Definition 3.5 (Brandenburger et al. (2023))**.** Let $\phi \in \Phi$. Inductively define $W_m(\phi)$ and $\ell(\phi)$ then $f \colon \Phi \to \Lambda$ as follows.

$$W_m(\phi) \coloneqq \bigcap \{C \in \mathscr{D}(\phi) \mid \phi(C \mid X \setminus \textstyle\bigcup_{i<m} W_i(\phi)) = 1\} \tag{9}$$

$$\ell(\phi) \coloneqq \min\{m \mid \textstyle\bigcup_{i<m} W_i(\phi) = X\} \tag{10}$$

$$f(\phi) \coloneqq (f_i(\phi))_{0 \leq i < \ell(\phi)} \quad \text{where} \quad f_i(\phi) \coloneqq \phi(W_i(\phi)) \tag{11}$$

Brandenburger et al. (2023) show that these maps satisfy a number of meaningful properties:

- The map f is *surjective but not injective*.
- The map f is piecewise invertible in the sense that $g_{\mathscr{B}}(f(\phi)) = \phi$ for all $\phi \in \Phi$ with domain $\mathscr{B}$.

A *partial* case is thus made that finitary CPSs are LPSs and finitary CPSs are LPSs in an informal sense because the fact that $\phi \in \Phi \cap \Sigma_{\mathscr{B}}$ can be re-encoded as an LPS and back to itself suggests that no information content is destroyed by the conversions.

However, the domain of $g_{\mathscr{B}}$ is $\{\lambda \in \Lambda \mid \mathscr{B} \subseteq \mathscr{P}(\lambda)\} \neq \Lambda$. Furthermore, there exist $\lambda, \mathscr{A}, \mathscr{B}$ such that $g_{\mathscr{A}}(\lambda) \neq g_{\mathscr{B}}(\lambda)$ even if $g_{\mathscr{A}}(f(\phi)) = g_{\mathscr{B}}(f(\phi))$. In other words, $g_{\mathscr{B}}(\lambda)$ is an *arbitrary* conversion of LPS λ into a CPS to some extent via the choice of $\mathscr{B}$. To make the full case that finitary CPSs are LPSs and finitary

[4] On the other hand, $E \in \mathscr{N}(\lambda)$ if and only if $m(\lambda, E) = -\infty$, but we do not make use of such cases.

CPSs are LPSs, we would like to construct a meaningful bijection between Λ and Σ. Everything needed for such a construction is already in Brandenburger et al. (2023). We merely need to apply our notion of equivalence.

Definition 3.6. Let $\lambda \in \Lambda$. Define $g(\lambda) \coloneqq g_{\mathscr{P}(\lambda)}(\lambda)$ so that $\mathscr{D}(g(\lambda)) = \mathscr{P}(\lambda)$.

The map g in Definition 3.6 is a specific instance of the object $g_{\mathscr{B}}$ in Definition 3.4. The domain of the CPS that results from the conversion of LPS λ is made to be $\mathscr{P}(\lambda)$. Thus, the conversion of λ to $g(\lambda)$ does not depend on an arbitrary choice of the collection $\mathscr{B} \subseteq \mathscr{X}$.

Brandenburger et al. (2023) uses $g_{\mathscr{B}}$ in its analysis only when $\mathscr{B} \cup \{\emptyset\}$ is a finite subalgebra of $\mathscr{X}$. This is to ensure that the CPS $g_{\mathscr{B}}(\lambda)$ will be a finitary CPS. In many, perhaps even most, cases of interest, $\mathscr{P}(\lambda) \cup \{\emptyset\}$ will not even be finite and, therefore, $g(\lambda)$ will not be a finitary CPS. If the goal is to identify LPSs with finitary CPSs, this is not ideal.

This is where the notion of equivalence developed in this paper closes the gap. Proposition 3.2—which is shown using Proposition 3.1—demonstrates that

- each LPS converts to the canonical form of some finitary CPS; and
- for each finitary CPS, there is an LPS that converts to its canonical form;

which makes the case that LPSs and finitary CPSs are different encodings of the same information.

Proposition 3.1. Let $\phi \in \Phi$. Then $\mathscr{D}(\bar{\phi}) = \mathscr{P}(\phi)$ and the following holds for all $A \in \mathscr{X}$ and $B \in \mathscr{P}(\phi)$.

$$\bar{\phi}(A \mid B) = \frac{\phi(A \cap B \mid W_{m(\phi,B)}(\phi))}{\phi(B \mid W_{m(\phi,B)}(\phi))} \tag{12}$$

$$\text{where} \quad m(\phi, B) \coloneqq \inf\{i \mid \phi(B \mid W_i(\phi)) > 0\} \tag{13}$$

Proof. In this proof, we will abbreviate $W_{m(\phi,B)}(\phi)$ as $W_{m(\phi,B)}$ for legibility given that there is no risk of confusion. Let $\psi(A \mid B)$ be defined by the right-hand side of (12). For all $A, B, C \in \mathscr{X}$, such that $A \subseteq B \subseteq C$ and $B, C \in \mathscr{P}(\phi)$, we want $\psi(A \mid B)\psi(B \mid C) = \psi(A \mid C)$.

$$\frac{\phi(A \mid W_{m(\phi,B)})}{\phi(B \mid W_{m(\phi,B)})} \frac{\phi(B \mid W_{m(\phi,C)})}{\cancel{\phi(C \mid W_{m(\phi,C)})}} = \frac{\phi(A \mid W_{m(\phi,C)})}{\cancel{\phi(C \mid W_{m(\phi,C)})}}$$
$$\phi(A \mid W_{m(\phi,B)})\phi(B \mid W_{m(\phi,C)}) = \phi(A \mid W_{m(\phi,C)})\phi(B \mid W_{m(\phi,B)})$$

If $\phi(B \mid W_{m(\phi,C)}) = 0$, then the equation reduces to holds because $\phi(A \mid W_{m(\phi,C)}) = 0$ owing to $B \supseteq A$. If $\phi(B \mid W_{m(\phi,C)}) > 0$, then $m(\phi, C) = m(\phi, B)$ because $C \supseteq B$. Then the same expression is on both sides of the equality. It follows that the equation holds in every case so that ψ as defined in (12) is a CPS.

Furthermore, for every $B \in \mathscr{D}(\psi) = \mathscr{P}(\phi)$, $\psi(A \mid B)$ is exactly pinned down by (2) because $\phi(B \mid W_{m(\phi,B)}) > 0$:

$$\psi(A \mid B)\overbrace{\phi(B \mid W_{m(\phi,B)})}^{>0} = \phi(A \cap B \mid W_{m(\phi,B)})$$

Therefore, ψ is the unique extension of ϕ to $\mathscr{P}(\phi)$. Furthermore, it is immediate that $\mathscr{P}(\psi) = \mathscr{P}(\phi)$ because $\mathscr{P}(\psi) \subseteq \mathscr{P}(\phi)$ as defined and $\phi \subseteq \psi$. Thus, we have $\psi \sqsupseteq \phi$. Proposition 2.2 then implies that $\psi = \bar{\phi}$. □

Proposition 3.2. Let $\phi \in \Phi$. Then $(g \circ f)(\phi) = \bar{\phi}$ and $(g \circ f)(\phi) \equiv \phi$.

Proof. If $(g \circ f)(\phi) = \bar{\phi}$ then it is immediate that $(g \circ f)(\phi) \equiv \phi$ because $\phi \equiv \bar{\phi}$.

$$\begin{aligned} ((g \circ f)(\phi))(A \mid B) &= \frac{(f_{m(f(\phi),B)}(\phi))(A \cap B)}{(f_{m(f(\phi),B)}(\phi))(B)} = \frac{(\phi(W_{m(f(\phi),B)}(\phi)))(A \cap B)}{(\phi(W_{m(f(\phi),B)}(\phi)))(B)} \\ &= \frac{\phi(A \cap B \mid W_{m(f(\phi),B)}(\phi))}{\phi(B \mid W_{m(f(\phi),B)}(\phi))} \overset{\downarrow}{=} \frac{\phi(A \cap B \mid W_{m(\phi,B)}(\phi))}{\phi(B \mid W_{m(\phi,B)}(\phi))} \\ &= \bar{\phi}(A \mid B) \end{aligned}$$

The marked equality $\overset{\downarrow}{=}$ follows from $m(\phi, A) = m(f(\phi), A)$, which is itself a consequence of $(f_i(\phi))(A) = \phi(W_i(\phi))$ being true by definition. The final equality is due to Proposition 3.1. □

Because we already know from Brandenburger et al. (2023) that $f \colon \Phi \to \Lambda$ is surjective, Proposition 3.2 tells us that

- for each $\lambda \in \Lambda$, $g(\lambda) = \bar{\phi}$ for some $\phi \in \Phi$; and
- for each $\phi \in \Phi$, there is some $\lambda \in \Lambda$ such that $g(\lambda) = \bar{\phi}$; and

which—after identifying each CPS with its equivalence class—makes it possible to view g as a bijection $\Lambda \to \{[\phi] \mid \phi \in \Phi\}$ and f as its inverse $\{[\phi] \mid \phi \in \Phi\} \to \Lambda$. Thus, we are able to reach the desired goal of identifying LPSs with finitary CPSs and vice versa.

Elsewhere in the literature, Hammond (1994) and Halpern (2010) also explore—among other topics—the relationship between CPSs and LPSs, but there are a few key differences.[5] In the case of Hammond (1994), the state space X is required to be finite and he considers conversions to and from complete CPSs (See Myerson 1986), which allow conditioning on all events. In Halpern (2010), a refinement of CPSs called Popper spaces are used.

To conclude this section, we note a minor error in Brandenburger et al. (2023) that is made immaterial by Proposition 3.2. Proposition 4.1 in Brandenburger et al. (2023) "suggests a sense in which $g_{\mathscr{B}}$ can be seen as an inverse of $f_{\mathscr{B}}$".[6]

[5] We do not list all differences here because some–such as the coverage of non-standard probabilities—are topically related but orthogonal to the goal of demonstrating that finitary CPSs and LPSs encode the same information.

[6] $f_{\mathscr{B}}$ is the restriction of $f \colon \Phi \to \Lambda$ to $\Sigma_{\mathscr{B}} \subseteq \Phi$.

Part (ii) of the proposition says that if $\mathscr{A}$ is a finite subalgebra of $\mathscr{X}$, then $f(g_{\mathscr{B}}(\lambda))$ is a subsequence of λ for all $\lambda \in \Lambda$ such that $\mathscr{A} \setminus \{\emptyset\} = \mathscr{B} \subseteq \mathscr{P}(\lambda)$. The following counterexample demonstrates that this is not true.

Example 3.1. Let $X = \{a, b, c, d\}$ and $\mathscr{X} = \wp(X)$. Fix the finite algebra $\mathscr{A}$ on X whose atoms are $\{a\}$, $\{b, c\}$, and $\{d\}$. Let $\mathscr{B} = \mathscr{A} \setminus \{\emptyset\}$ and define $\lambda = (\lambda_0, \lambda_1) \in \Lambda$ as follows.

$$1/2 = \lambda_0(\{a\}) = \lambda_0(\{b\}) = \lambda_1(\{c\}) = \lambda_1(\{d\})$$

Then $\mathscr{B} \subseteq \mathscr{P}(\lambda) = \mathscr{X}$. The CPS $g_{\mathscr{B}}(\lambda)$ is defined as follows.

$$\begin{aligned} 1 &= g_{\mathscr{B}}(\lambda)(\{b\} \mid \{b, c\}) = g_{\mathscr{B}}(\lambda)(\{b\} \mid \{b, c, d\}) = g_{\mathscr{B}}(\lambda)(\{a\} \mid \{a, d\}) \\ 1/2 &= g_{\mathscr{B}}(\lambda)(\{x\} \mid \{a, b, c\}) = g_{\mathscr{B}}(\lambda)(\{x\} \mid \{a, b, c, d\}) \quad \text{for } x = a, b \end{aligned}$$

It follows that $f(g_{\mathscr{B}}(\lambda))$ is a length-2 LPS, but is *not* a subsequence of λ:

$$\begin{aligned} 1/2 &= f_0(g_{\mathscr{B}}(\lambda))(\{a\}) = f_0(g_{\mathscr{B}}(\lambda))(\{b\}) \\ 1 &= f_1(g_{\mathscr{B}}(\lambda))(\{d\}) \end{aligned}$$

A correction of Proposition 4.1(ii) in Brandenburger et al. (2023) should say that, if $\mathscr{A}$ is a finite subalgebra of $\mathscr{X}$, then $f_0(g_{\mathscr{B}}(\lambda)) = \lambda_0$ for all $\lambda \in \Lambda$ such that $\mathscr{B} = \mathscr{A} \setminus \{\emptyset\} \subseteq \mathscr{P}(\lambda)$. Nevertheless, the implicit goal of "[suggesting] a sense in which $g_{\mathscr{B}}$ can be seen as an inverse of $f_{\mathscr{B}}$" is to show that CPS are LPS and vice versa in some fundamental way. This is a goal we have accomplished concretely because Proposition 3.2 essentially establishes f and $\lambda \mapsto g_{\mathscr{P}(\lambda)}(\lambda)$ as mutually inverse bijections between Λ and the quotient space $\Phi/\equiv$.

4 Further Work

One avenue for extension of this work may be the characterization of when the canonical form of a CPS has a domain equal to its potent sets. We have already shown that CPS representations of LPSs always have this property. We would argue that this property is of interest because it captures to some extent when a CPS is consistent with a fully "revision"-based approach.

The intuition is that when the canonical form of a CPS has a domain equal to its potent sets, it cannot contain implicit inconsistencies with the conditional probability formula. If CPS π has events A and B in its domain but π-potent $A \cap B$ is not in the domain of its canonical form then π must contain inconsistencies about probabilities conditional on $A \cap B$ that are not checked by (2) for π. That, suggests in turn that one could arrive at different implicit[7] conditional beliefs given $A \cap B$ depending on the order in which information is observed (i.e., A

[7] We obviously could not arrive at multiple beliefs given a condition *in* the domain since those are explicitly specified.

then B versus B then A).[8] Such contextuality conflicts with the "revision"-only view that belief changes reflect only "a change of in knowledge about a *static* world".

Another related question of interest is when extensions of a CPS π specify probabilities conditional on added events in the expanded domain in a way that does not conflict with any of the implicit information already encoded in π. Given that such conflicts are not always checked by equation (2), we conjecture that the comparison of π's canonical form versus that of the extension may shed light on when such unchecked conflicts do or do not exist.

References

Battigalli, Pierpaolo (Apr. 1996). "Strategic Rationality Orderings and the Best Rationalization Principle". In: *Games and Economic Behavior* 13.2, pp. 178–200. ISSN: 0899-8256. DOI: 10.1006/game.1996.0033.

Battigalli, Pierpaolo and Marciano Siniscalchi (Oct. 2002). "Strong Belief and Forward Induction Reasoning". In: *Journal of Economic Theory* 106.2, pp. 356–391. ISSN: 00220531. DOI: 10.1006/jeth.2001.2942.

Blume, Lawrence, Adam Brandenburger, and Eddie Dekel (1991). "Lexicographic Probabilities and Choice Under Uncertainty". In: *Econometrica* 59.1, pp. 61–79. ISSN: 0012-9682. DOI: 10.2307/2938240.

Brandenburger, Adam, Amanda Friedenberg, and H. Jerome Keisler (Mar. 2008). "Admissibility in Games". In: *Econometrica* 76.2, pp. 307–352. ISSN: 0012-9682, 1468-0262. DOI: 10.1111/j.1468-0262.2008.00835.x.

— (May 2023). "The relationship between strong belief and assumption". en. In: *Synthese* 201.5, p. 175. ISSN: 1573-0964. DOI: 10.1007/s11229-023-04167-6.

Catonini, Emiliano and Nicodemo De Vito (Feb. 2024). "Cautious belief and iterated admissibility". In: *Journal of Mathematical Economics* 110, p. 102918. ISSN: 03044068. DOI: 10.1016/j.jmateco.2023.102918.

Dekel, Eddie, Amanda Friedenberg, and Marciano Siniscalchi (May 2016). "Lexicographic beliefs and assumption". In: *Journal of Economic Theory* 163, pp. 955–985. ISSN: 00220531. DOI: 10.1016/j.jet.2016.03.003.

Gärdenfors, P. (1988). *Knowledge in Flux. Modelling the Dymanics of Epistemic States*. MIT Press.

Halpern, Joseph Y. (Jan. 2010). "Lexicographic probability, conditional probability, and nonstandard probability". en. In: *Games and Economic Behavior* 68.1. Number: 1, pp. 155–179. ISSN: 08998256. DOI: 10.1016/j.geb.2009.03.013.

Hammond, Peter J. (1994). "Elementary Non-Archimedean Representations of Probability for Decision Theory and Games". en. In: *Patrick Suppes: Scientific Philosopher: Volume 1. Probability and Probabilistic Causality*. Ed.

[8] Some (e.g., Pothos and Busemeyer 2013) have proposed the use of quantum probability to model such cognitive effects, but it is apparent that CPSs can accommodate at least some aspects of such contextuality.

by Paul Humphreys. Synthese Library. Dordrecht: Springer Netherlands, pp. 25–61. ISBN: 978-94-011-0774-7. DOI: 10.1007/978-94-011-0774-7_2.

Keisler, H. Jerome and Byung Soo Lee (2023). "Common assumption of rationality". In: *Journal of Mathematical Economics* 109.C. DOI: 10.1016/j.jmateco.2023.10.

Kern-Isberner, Gabriele (2001). "Revising and Updating Probabilistic Beliefs". In: *Frontiers in Belief Revision*. Ed. by Mary-Anne Williams and Hans Rott. Dordrecht: Springer Netherlands, pp. 393–408. ISBN: 978-94-015-9817-0. DOI: 10.1007/978-94-015-9817-0_20.

Lee, Byung Soo (May 2016). "Admissibility and assumption". In: *Journal of Economic Theory* 163, pp. 42–72. ISSN: 00220531. DOI: 10.1016/j.jet.2016.01.006.

Myerson, Roger B. (1986). "Multistage Games with Communication". In: *Econometrica* 54.2, pp. 323–358. ISSN: 0012-9682. DOI: 10.2307/1913154.

Pothos, Emmanuel M. and Jerome R. Busemeyer (2013). "Can quantum probability provide a new direction for cognitive modeling?" In: *Behavioral and Brain Sciences* 36.3, pp. 255–274. DOI: 10.1017/S0140525X12001525.

Rényi, Alfréd (Sept. 1955). "On a new axiomatic theory of probability". In: *Acta Mathematica Academiae Scientiarum Hungaricae* 6.3, pp. 285–335. ISSN: 0001-5954, 1588-2632. DOI: 10.1007/BF02024393.

Savage, Leonard J. (1972). *The Foundations of Statistics*. New York: Dover Publications. ISBN: 978-0-486-61610-7.