A Complexity Hierarchy for Stochastic Choice*

Erya Yang † Adam Brandenburger ‡

Preliminary Version

July 4, 2024

Abstract

We develop a hierarchy of families of stochastic choice rules in terms of increasing complexity based on the concepts of a conditional probability space (Rényi, 1955) and a dimensionally-ordered system of measures (Rényi, 1956). The levels of our hierarchy are: single-valued or point conditional probability spaces (PCPS's); conditional probability spaces; probabilistic mixtures of PCPS's; mutually absolutely continuous mixtures of PCPS's; absolutely continuous mixtures of PCPS's; and signed probability mixtures of PCPS's. We show at a general measure-theoretic level that the first five levels are strictly nested. In the finite case, the sixth level nests all the other levels (the general case here appears to be open). Our hierarchy organizes various well-known axioms for stochastic choice and identifies some new relationships among them. It also offers a precise definition of contextuality in stochastic choice and, from this, a new classification of some leading behavioral effects in choice.

1 Introduction

The appearance of unbounded measures in applications of probability theory to areas including Bayesian statistics, statistical and quantum mechanics, and probabilistic number theory motivated Rényi (1955) to introduce the concept of a conditional probability space (CPS). A CPS begins with a measurable space and a family of nonempty conditioning events. It then associates with each such event a probability measure that concentrates on the event, to be interpreted as the probability given the event. Different probability measures in the CPS satisfy a chain-rule consistency property. If the conditioning events are finite or compact or controlled in some other way, the problem of unbounded measures can be avoided. Rényi developed his theory in a series of publications; see Rényi (1955, 1956, 1970a, 1970b).

In game theory, CPS's (where they are usually called conditional probability systems) have proved tailor-made for describing how players update beliefs in a game tree. Here, the conditioning events are the information sets and the player at a particular information set possesses a probability measure over the strategy choices by other players that are compatible with the event that the information set has been reached. Early work in this direction was undertaken by Myerson (1986a, 1986b). Battigalli and Siniscalchi (1999, 2002) developed the fundamental epistemic game theory of trees using CPS's.

In this paper, we use the concept of a CPS in decision theory – specifically, to build a complexity hierarchy for stochastic choice. The use of CPS's in this area was foreshadowed by Luce (1959) and, recently, they have employed by Cerreia-Vioglio et al. (2021) to formulate an independence of irrelevant alternatives axiom for stochastic choice. We return to these connections later.

Stochasticity in choice has been justified in several ways. It may reflect a decision maker's preference for randomization (e.g., Machina, 1989, Cerreia-Vioglio et al., 2019, and Agranov and Ortoleva, 2017). Or, it may arise from an irreducible "trembling-hand" in choice (cf. Selten, 1975). Also, cognitive neuroscience has identified a utility-like process in the brain (Platt and Glimcher, 1999) that has a stochastic relationship to choice (Webb, 2019, Steverson, Brandenburger, and Glimcher, 2019).

We next preview our complexity hierarchy. At the lowest level is the family of points CPS's (PCPS's), which are CPS's whose associated probability measures are all Dirac (degenerate). We show that this level characterizes the family of point stochastic choice rules (SCR's) that satisfy the Weak Axiom of Stochastic Revealed Preference or WASRP (Bandyopadhyay, Dasgupta, and Pattanaik 1999, 2002). The next level is the family of CPS's. This level characterizes the family of SCR's satisfying a Generalized

*We thank Samson Abramsky, Stefan Bucher, Vered Kurtz-David, Kenway Louie, and Stuart Zoble for important input, and Kai Steverson for early inspiration for this project. Financial support from NYU Stern School of Business, NYU Shanghai, and J.P. Valles is gratefully acknowledged.

†Business School, Sun Yat-sen University, Shenzhen, 518107, China, yangery@mail.sysu.edu.cn; corresponding author

‡Stern School of Business, Tandon School of Engineering, NYU Shanghai, New York University, New York, NY 10012, U.S.A., adam.brandenburger@stern.nyu.edu

Independence of Irrelevant Alternatives (GIIA) axiom. We also establish an equivalence between GIIA and a dimensional ordering property of CPS's (Rényi, 1956). The third level of our hierarchy consists of the family of probabilistic mixtures of PCPS's. We build a measurable space of PCPS's (using results in Cohen, 1980) and define an SCR from a probability measure on this space. We show that this level characterizes SCR's satisfying the Block-Marschak condition (Block and Marschak, 1960). Above this level, our hierarchy splits into two branches. On one branch, the next level consists of signed mixtures of PCPS's, where we build the family of SCR's that arise from a signed probability measure on the space of PCPS's. On the other branch, we first consider the family of SCR's that satisfy a mutual absolute continuity (m.a.c.) condition: An SCR belongs to this level if there is probabilistic mixture of PCPS's which, for each conditioning event, is m.a.c. with respect to the SCR. The next level on this branch requires only absolutely continuity with respect to the SCR. When the measurable space is finite, this level nests under the level consisting of signed mixtures of PCPS's. Indeed, this latter level is the top of the hierarchy – a result that follows from Dogan and Yildiz (2023). (The situation appears to be an open question in the infinite case.) Figure 1 depicts our complete complexity hierarchy.

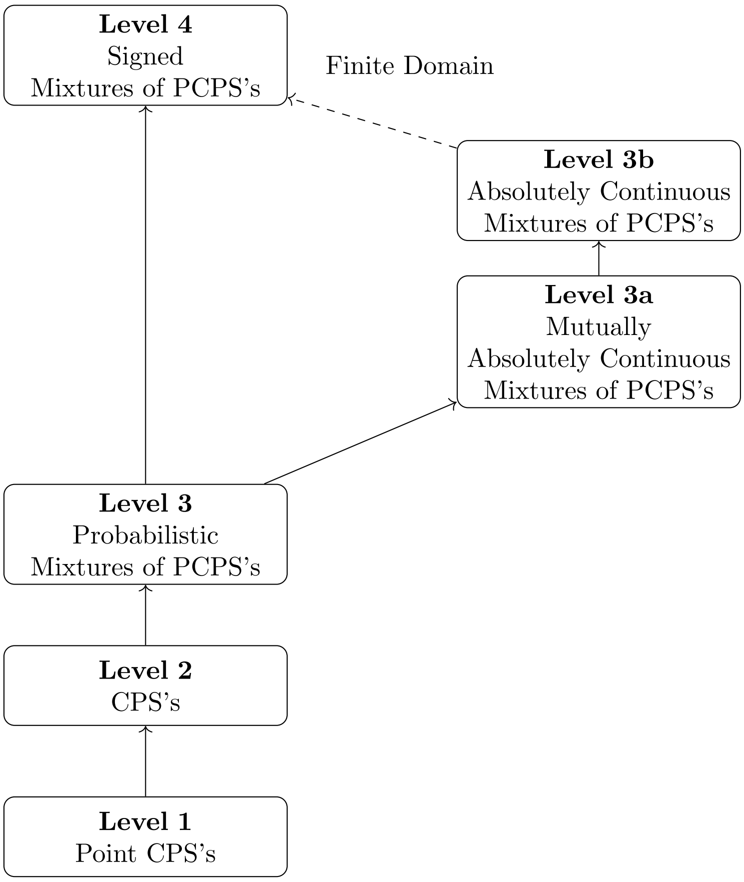

Figure 1: A Complexity Hierarchy Based on Conditional Probability Spaces

We believe that our CPS-based hierarchy gives a unified picture of various well-known axioms for stochastic choice. Thus, WASRP, which is a natural stochastic analog to the classic Weak Axiom of Revealed Preference in demand theory, our GIIA axiom, which extends the classic (Luce, 1959) axiom for stochastic choice, and the BM condition, which is necessary and sufficient for a random-utility representation of preferences (Falmagne, 1978), look quite different a priori. Yet, our results show that they can be understood in a unified way as all being characterized in terms of variants of one concept from probability theory, namely, that of a CPS. In the main text, we fit other axioms for stochastic choice into this same framework. For surveys of stochastic choice theory, see, e.g., Chambers and Echenique (2016), Saito (2019), and Strzalecki (2022).

Our hierarchy also suggests a precise definition of the notion of contextuality in stochastic choice. It is common in the literature to refer to violations of the Luce Rule (IIA) as due to context effects. (See Davis-Stober et al., 2023 for a careful review.) In particular, the Similarity Effect (Debreu, 1960), the Compromise Effect (Simonson, 1989, Simonson and Tversky, 1992), the Attraction Effect (Huber, Payne, and Puto, 1982, Simonson, 1989), and the Repulsion Effect (Aaker, 1991) are often described this way. In our complexity hierarchy, the first two effects arise at Level 3. The third and fourth effects – understood as violations not just of GIIA but of the Regularity axiom of stochastic choice – first become possible at Level 4. We see all behavior up to and including Level 3 as non-contextual, and contextual behavior as strictly higher up. Our argument begins with the fact that a Level-3 SCR is, by definition, a probabilistic mixture of PCPS's and is therefore, by our first-level characterization, a probabilistic mixture of point SCR's satisfying WASRP. Next WASRP and GIIA are easily seen to coincide for point SCR's. Thus, Level-3 behavior can violate GIIA, but only in the sense that the decision maker – or an observer – randomizes over point SCR's satisfying GIIA. (Under an epistemic interpretation, the decision maker or observer is uncertain as to which point SCR is in effect.) We contend that one should call behavior contextual only once it is not explainable in terms of probabilistic mixing over classical objects. This is why we propose using the term "contextuality" only for behavior above Level 3. Aside from this point about classification, we also achieve a unified treatment of all four behavioral effects. We produce all of them in a common format, which is one or other kind of mixture of PCPS's.

The overall architecture of our complexity hierarchy for stochastic choice was inspired by a contextual hierarchy for quantum systems developed by Abramsky and Brandenburger (2011). We return to this connection later in the paper.

2 Framework

Given a measurable space $(\Omega, \mathcal{F})$, we write $\Delta(\Omega)$ for the set of all probability measures on $(\Omega, \mathcal{F})$. Let $\mathcal{G}$ be a sub-family of $\mathcal{F}$ with $\emptyset \notin \mathcal{G}$.

Definition 1. *A **stochastic choice rule (SCR)** (relative to $\mathcal{G}$) is a map $c : \mathcal{G} \to \Delta(\Omega)$, which we write as $G \mapsto c_G(\cdot)$, satisfying:*

1. *$c_G(G) = 1$ for every $G \in \mathcal{G}$.*

Definition 2. *A **conditional probability space (CPS)** (relative to $\mathcal{G}$) is a map $p : \mathcal{G} \to \Delta(\Omega)$, which we write as $G \mapsto p_G(\cdot)$ for $G \in \mathcal{G}$, satisfying:*

1. *$p_G(G) = 1$ for every $G \in \mathcal{G}$;*
2. *$p_G(E) = p_G(F)p_F(E)$ for every $E \subseteq F \subseteq G$ with $E \in \mathcal{F}$ and $F, G \in \mathcal{G}$.*

Observe that a CPS is an SCR satisfying an extra chain rule requirement (Condition 2 of Definition 2). In Rényi (1955), Condition 2 takes the form: If $E, F \in \mathcal{F}$, $G \in \mathcal{G}$, and $F \cap G \in \mathcal{G}$, then $p_G(E \cap F) = p_G(F)p_{F \cap G}(E)$. This is readily seen to be equivalent to our Condition 2.

Luce (1959) already noted the equivalence between CPS's and SCR's satisfying IIA, for the case of finite choice sets and strictly positive probabilities. Cerreia-Vioglio et al. (2021) extend this equivalence by dropping the strict positivity condition for finite choice sets. One of our Level 2 results further extends the equivalence to arbitrary choice sets.

A basic result in Rényi (1956) gives a condition under which every CPS can be generated from what he calls a dimensionally ordered system of measures. The next definition is from Rényi (1956, p.63). In what follows, a total order is an order that is complete, transitive, and antisymmetric.

Definition 3. *Fix a measurable space $(\Omega, \mathcal{F})$ and a sub-family $\mathcal{G}$ of $\mathcal{F}$ with $\emptyset \notin \mathcal{G}$. Let $(\Gamma, \prec)$ denote an arbitrary totally ordered index set, and suppose that for every $\gamma \in \Gamma$ there is an associated measure μ_γ on $(\Omega, \mathcal{F})$. Call the system of measures **dimensionally ordered** if:*

1. *for every $G \in \mathcal{G}$, there is a $\gamma \in \Gamma$ such that $0 < \mu_\gamma(G) < +\infty$;*
2. *if $\mu_\alpha(G) < +\infty$ and $\alpha \prec \beta$, then $\mu_\beta(G) = 0$.*

Note that for every $G \in \mathcal{G}$, there is a unique $\gamma \in \Gamma$ such that $0 < \mu_\gamma(G) < +\infty$. Also, if $\beta \prec \gamma$, then $\mu_\beta(G) = +\infty$. Now, for $G \in \mathcal{G}$ and $F \in \mathcal{F}$, we define:

$$p_G(F) = \frac{\mu_\gamma(F \cap G)}{\mu_\gamma(G)}, \tag{1}$$

where γ is this unique index for G. The next result is implied by the proof of Theorem 14 in Rényi (1955), but we give a self-contained treatment.

Theorem 1. *The family of maps $G \mapsto p_G(\cdot)$ defined in Equation (1) is a CPS relative to $\mathcal{G}$.*

Proof. The proof of Condition 1 of Definition 2 is immediate. For Condition 2, we need to show:

$$\frac{\mu_\gamma(E \cap G)}{\mu_\gamma(G)} = \frac{\mu_\gamma(F \cap G)}{\mu_\gamma(G)} \times \frac{\mu_\delta(E \cap F)}{\mu_\delta(F)}, \tag{2}$$

where $\gamma \in \Gamma$ is the unique index such that $0 < \mu_\gamma(G) < +\infty$ and $\delta \in \Gamma$ the unique index such that $0 < \mu_\delta(F) < +\infty$. Rewrite Equation (2) as:

$$\frac{\mu_\gamma(E)}{\mu_\gamma(G)} = \frac{\mu_\gamma(F)}{\mu_\gamma(G)} \times \frac{\mu_\delta(E)}{\mu_\delta(F)}, \tag{3}$$

or

$$\mu_\gamma(E) \times \mu_\delta(F) = \mu_\gamma(F) \times \mu_\delta(E). \tag{4}$$

Now note that since $F \subseteq G$, we have $\mu_\delta(G) > 0$ and therefore $\delta \preceq \gamma$. Case (i): If $\delta = \gamma$, then Equation (4) is immediately satisfied. Case (ii): If $\delta \prec \gamma$, then $\mu_\gamma(F) = 0$. But $\mu_\gamma(E) = 0$, since $E \subseteq F$, and then both sides of Equation (4) are equal to 0, completing the proof. □

Definition 4. *The family of maps $G \mapsto p_G(\cdot)$ will be called the CPS* ***generated by the dimensionally ordered set of measures*** *$\{\mu_\gamma : \gamma \in \Gamma\}$, or, for short, a* ***dimensionally generated*** *(or* ***DGEN****) CPS.*

Definition 5. *Using Equation (1) to define maps $G \mapsto c_G(\cdot)$, we can similarly define the concept of an SCR* ***generated by a dimensionally ordered set of measures*** *$\{\mu_\gamma : \gamma \in \Gamma\}$, or, for short, refer to a* ***dimensionally generated*** *(or* ***DGEN****) SCR.*

Definition 6. *Fix a measurable space $(\Omega, \mathcal{F})$ and a sub-family $\mathcal{G}$ of $\mathcal{F}$ with $\emptyset \notin \mathcal{G}$. Call $\mathcal{G}$ an* ***additive*** *family if $G, H \in \mathcal{G}$ implies $G \cup H \in \mathcal{G}$.*

The next result is Theorem 1 in Rényi (1956), which gives a sufficient condition for a CPS to be dimensionally generated.

Theorem 2. *If the family of maps $G \mapsto p_G(\cdot)$ is a CPS relative to an additive family $\mathcal{G}$, then it is DGEN.*

We will make important use of this result, so we give a sketch of the proof here. Given two events $G, H \in \mathcal{G}$, define the indicator of the two events by:

$$i(G, H) = \frac{p_{G\cup H}(G)}{p_{G\cup H}(H)}. \tag{5}$$

If $p_{G\cup H}(H) = 0$, set $i(G, H) = +\infty$. Also, note that nominator and denominator cannot both be 0. Define a relation on events in $\mathcal{G}$ by setting $G \sim H$ if $0 < i(G, H) < +\infty$, and say G and H are of the same dimension. Lemma 2 in Rényi (1956) establishes that $\sim$ is an equivalence relation, which is used to partition $\mathcal{G}$ into mutually disjoint classes $\mathcal{G}_\gamma$. For two distinct indices β and γ, write $\beta \prec \gamma$ or $\gamma \prec \beta$ if $i(G, H) = 0$ or $+\infty$, respectively, for $G \in \mathcal{G}_\beta$ and $H \in \mathcal{G}_\gamma$. (This definition can be shown to be independent of the representatives G and H chosen.) The relation $\prec$ is a total order.

The next step in the proof is to build measures μ_γ on $(\Omega, \mathcal{F})$, one for each equivalence class $\mathcal{G}_\gamma$, so that, if $G \in \mathcal{G}_\gamma$, then Equation 1 is satisfied for this μ_γ. The final step is to show that the system of measures $\{\mu_\gamma\}$ is dimensionally ordered, i.e., satisfies our Definition 3.

Császár (1955) weakens the requirement of additivity in Theorem 2 to quasi-additivity, which requires that if $G, H \in \mathcal{G}$, then there is $J \in \mathcal{G}$ such that $G \cup H \subseteq J$ and $p_J(G) + p_J(H) > 0$. (It is easy checked that additivity implies quasi-additivity.) We do not make further use of this condition in the current paper.

3 Hierarchy Level 1

In this level, we establish an equivalence between point CPS's and point SCR's that satisfy a stochastic analog to the Weak Axiom of Revealed Preference.

3.1 Preliminaries

Fix a measurable space $(\Omega, \mathcal{F})$ and let $\mathcal{G}$ be a subfamily of $\mathcal{F}$ with $\emptyset \notin \mathcal{G}$. Let s be a **selection** on $\mathcal{G}$. Thus $s : \mathcal{G} \to \Omega$ and for each $G \in \mathcal{G}$, $s(G) \in G$.

Definition 7. *Fix a CPS $p : \mathcal{G} \times \mathcal{F} \mapsto [0, 1]$. Suppose there is a selection s, such that for each $G \in \mathcal{G}$ and $F \in \mathcal{F}$:*

$$p_G(F) = \mathbf{1}_F(s(G)). \tag{6}$$

That is, $p_G(\cdot)$ is the Dirac measure concentrated on $s(G)$. We call such a CPS a ***point CPS (PCPS)****.*

Definition 8. *Fix a SCR $c : \mathcal{G} \to \Delta(\Omega)$. Suppose there is a selection s such that for each $G \in \mathcal{G}$:*

$$c_G(F) = \mathbf{1}_F(s(G)). \tag{7}$$

That is, $c_G(\cdot)$ is the Dirac measure concentrated on $s(G)$. We call such an SCR a ***point SCR (PSCR)****.*

Bandyopadhyay, Dasgupta, and Pattanaik (1999, 2002) and Dasgupta and Pattanaik (2007) introduced an extension of the Weak Axiom of Revealed Preference (WARP) of classic demand theory to stochastic demand theory and then to general stochastic choice rules.

Definition 9. *A stochastic choice rule $c : \mathcal{G} \to \Delta(\Omega)$ satisfies the* ***Weak Axiom of Stochastic Revealed Preference (WASRP)*** *if for every $G, H \in \mathcal{G}$ and every $F \in \mathcal{F}$ with $F \subseteq G \cap H$:*

$$c_H(F) - c_G(F) \leq c_G(G \backslash H). \tag{8}$$

To justify WASRP, start with the classical case of a deterministic choice rule (DCR) and the Weak Axiom of Revealed Preference (WARP). We do not need to define a DCR separately, since we can interpret a selection s as a DCR. As usual, we say a DCR s satisfies WARP if for every $G, H \in \mathcal{G}$: If $\omega, \omega' \in G \cap H$ and $s(G) = \omega$, then $s(H) \neq \omega'$.

Next notice that a PSCR naturally induces a DCR, via its selection s, and vice versa. We then have: If a PCPS satisfies WASRP, then the induced DCR satisfies WARP, and, in fact, if a DCR satisfies WARP, then the induced PCPS satisfies WASRP. See Dasgupta and Pattanaik (2007, Remark 2.9). Thus, WASRP generalizes WARP to SCR's.

For further justification, observe that while WASRP does not require equality of the choice probabilities $c_H(F)$ and $c_G(F)$, which might be the obvious extension of WARP but would be too restrictive, it limits how big the difference can be. This difference must lie in the range $[-c_H(H\backslash G), +c_G(G\backslash H)]$. Bandyopadhyay, Dasgupta, and Pattanaik (1999) and Dasgupta and Pattanaik (2007) provide the following argument for the upper bound. When the choice set changes from G to H, this rules out "competing" choices in $G\backslash H$ to the alternatives in F. This effect may raise the probability $c_H(F)$ that the chosen alternative lies in F. But this increase, viz. the difference $c_H(F) - c_G(F)$, should be bounded above by the original choice probability for $G\backslash H$, viz. $c_G(G\backslash H)$. The argument for the lower bound is analogous.

3.2 Characterization

Here we establish the equivalence between PCPS's and PSCR's satisfying WASRP. The reason we state the equivalence in terms of PSCR's and WASRP, rather than in terms of DCR's and WARP, is that it makes immediate the comparison with the implications of WASRP in Section 4, where we examine general (non-point) SCR's.

Theorem 3. *Fix a measurable space $(\Omega, \mathcal{F})$ and a sub-family $\mathcal{G}$ of $\mathcal{F}$ with $\emptyset \notin \mathcal{G}$. The family of PCPS's $p_G(\cdot)$ relative to $\mathcal{G}$ coincides with the family of PSCR's relative to $\mathcal{G}$ satisfying WASRP.*

Proof. Fix a PCPS $p_G(\cdot)$ relative to $\mathcal{G}$. We need to show that $p_G(\cdot)$, viewed as an SCR, is a PSCR and satisfies WASRP. It is immediate that $p_G(\cdot)$ is a PSCR, so we show that WASRP is satisfied. Fix $G, H \in \mathcal{G}$ and $F \in \mathcal{F}$ with $F \in G \cap H$. Case (i): $s(G) \in G\backslash H$. Then $p_G(G\backslash H) = 1$, from which Inequality 8 is satisfied. Case (ii): $s(G) \in F$. Then $p_G(F) = 1$, from which Inequality 8 is again satisfied. Case (iii): $s(G) \in (G \cap H)\backslash F$. Then $p_G(F) = 0$ and $p_G(G\backslash H) = 0$, so, in light of Inequality 8, we need to show that $p_H(F) = 0$. Suppose not, so that $s(H) \in F$ and $p_H(F) = 1$. Then $p_H(G \cap H) = 1$. Using $p_H(F) = p_H(G \cap H)\, p_{G\cap H}(F)$, we find $p_{G\cap H}(F) = 1$. Using $p_G(F) = p_G(G \cap H)\, p_{G\cap H}(F)$, $p_G(F) = 0$ (as already shown), and $p_G(G \cap H) = 1$ (since $s(G) \in G \cap H$), we find $p_{G\cap H}(F) = 0$, a contradiction. This establishes one direction of the proof.

For the reverse direction, fix an SCR $c_G(\cdot)$ relative to $\mathcal{G}$ satisfying WASRP. We need to show that the chain rule holds: $c_G(E) = c_G(F)c_F(E)$ for every $E \subseteq F \subseteq G$ with $E \in \mathcal{F}$ and $F, G \in \mathcal{G}$. Write WASRP in the form:

$$c_G(E) - c_F(E) \leq c_F(F\backslash G). \tag{9}$$

Case (i): $s(G) \in E$ and $s(F) \in E$. Then $c_G(E) = 1$, and therefore $c_G(F) = 1$, and also $c_F(E) = 1$, so that the chain rule is satisfied. Case (ii): $s(G) \in E$ and $s(F) \notin E$. Then $c_G(E) = 1$ and $c_F(E) = 0$. But from Inequality 9, using $c_F(F\backslash G) = c_F(\emptyset) = 0$, we then get $1 - 0 \leq 0$, a contradiction. Case (iii): $s(G) \notin F$. Then $c_G(E) = 0$ and $c_G(F) = 0$, so that the chain rule is again satisfied. □

4 Hierarchy Level 2

In this level, we establish the equivalence between CPS's and SCR's satisfying a Generalized Independence of Irrelevant Alternatives axiom.

4.1 Preliminaries

Definition 10. *A stochastic choice rule $c : \mathcal{G} \to \Delta(\Omega)$ satisfies* ***Generalized Independence of Irrelevant Alternatives (GIIA)*** *if for every $G, H \in \mathcal{G}$, and every $E, F \in \mathcal{F}$ with $E \subseteq G \cap H$ and $F \subseteq G \cap H$, we have:*

$$c_G(E) \times c_H(F) = c_H(E) \times c_G(F). \tag{10}$$

It is clear that GIIA implies the usual statement of IIA in terms of probability ratios, in the case that $c_H(E) \times c_H(F) \neq 0$. Our generalization is the same as the definition of Odds Independence in Cerreia-Vioglio et al. (2021), except that we extend the definition from finite choice sets to a general sub-family of choice sets.

4.2 Characterization

Theorem 4. *Fix a measurable space $(\Omega, \mathcal{F})$ and a sub-family $\mathcal{G}$ of $\mathcal{F}$ with $\emptyset \notin \mathcal{G}$. The family of CPS's $G \mapsto p_G(\cdot)$ relative to $\mathcal{G}$ coincides with the family of SCR's $G \mapsto c_G(\cdot)$ relative to $\mathcal{G}$ satisfying Generalized Independence of Irrelevant Alternatives (GIIA).*

Proof. Fix a CPS $G \mapsto p_G(\cdot)$. Also fix $G, H \in \mathcal{G}$ and $E, F \in \mathcal{F}$, where $E \subseteq G \cap H$ and $F \subseteq G \cap H$. Using the equivalent Rényi (1955) form of Condition 2 of a CPS, we can write:

$$p_G(E) = p_G(E \cap H) = p_G(H) \times p_{G\cap H}(E), \tag{11}$$
$$p_G(F) = p_G(F \cap H) = p_G(H) \times p_{G\cap H}(F), \tag{12}$$
$$p_H(E) = p_H(E \cap G) = p_H(G) \times p_{G\cap H}(E), \tag{13}$$
$$p_H(F) = p_H(F \cap G) = p_H(G) \times p_{G\cap H}(F). \tag{14}$$

Multiplying the left sides of Equations 11 and 14, and the left sides of Equations 12 and 13, we obtain:

$$p_G(E) \times p_H(F) = p_H(E) \times p_G(F), \tag{15}$$

establishing the forward direction of the proof. (This direction extends the proof of Theorem 6 in Rényi (1955).) For the reverse direction, suppose that Equation 15 holds and $H \subseteq G$, and set $F = H$. We obtain

$$c_G(E) \times c_H(H) = c_H(E) \times c_G(H), \tag{16}$$

from which, since $c_H(H) = 1$, our Condition 2 of a CPS is satisfied. □

4.3 Relationships Among Axioms

In this section, we establish the relationships among stochastic-choice axioms depicted in Figure 2. (The abbreviations REG and SST refer to Regularity and Strong Stochastic Transitivity, respectively, and will be defined as we proceed.)

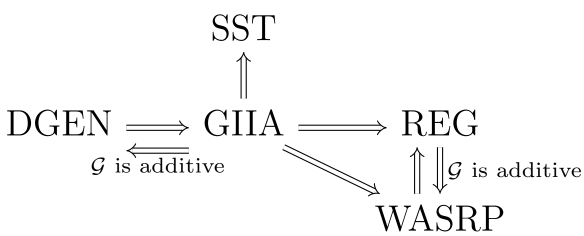

Figure 2: Relationships Among Axioms

Definition 11. *A stochastic choice rule $c : \mathcal{G} \to \Delta(\Omega)$ satisfies* ***Regularity (REG)*** *if for every $G, H \in \mathcal{G}$ with $G \subseteq H$, and every $F \in \mathcal{F}$ with $F \subseteq G$, we have $c_H(F) \leq c_G(F)$.*

Regularity is a basic condition postulating that the probability of choosing a given set of alternatives is decreasing in the size of the choice set as measured by set inclusion. By the chain rule, a CPS satisfies Regularity, since $p_H(F) = p_H(G)\, p_G(F) \leq p_G(F)$.

It is easily seen that WASRP, GIIA, and REG all coincide in the case of point SCR's. (In this case, the selection s for REG satisfies Chernoff's Condition α; see Chernoff, 1954.) In general, the three conditions differ, as we now explore.

Theorem 5. *If an SCR satisfies GIIA, then it satisfies REG.*

Proof. Set $G \subseteq H$ and $F = G$ in Equation 10 to get:

$$c_G(E) \times C_H(G) = c_H(E) \times c_G(G) = c_H(E), \tag{17}$$

from which $c_G(E) \geq c_H(E)$. □

Theorem 6. *If an SCR satisfies WASRP, then it satisfies REG.*

Proof. Set $G \subseteq H$ in Equation 8 to get:

$$c_H(F) - c_G(F) \leq c_G(\emptyset) = 0, \tag{18}$$

as required. □

Theorem 7. *If an SCR satisfies GIIA, then it satisfies WASRP.*

Proof. Fix $G, H \in \mathcal{G}$ and $F \in G \cap H$. By GIIA, we can write:

$$c_G(G \cap H) \times c_H(F) = c_H(G \cap H) \times c_G(F). \tag{19}$$

Suppose $c_G(F) \neq 0$, so that:

$$\frac{c_H(F)}{c_G(F)} = \frac{c_H(G \cap H)}{c_G(G \cap H)} = \lambda, \tag{20}$$

where λ is independent of F. Set $c_G(G \cap H) = k$, where $0 \leq k \leq 1$, so that $c_G(G \backslash H) = 1 - k$. Write:

$$c_H(F) - c_G(F) = (\lambda - 1)\, c_G(F), \tag{21}$$

from which we need to show that:

$$(\lambda - 1)\, c_G(F) \leq 1 - k. \tag{22}$$

If $\lambda \leq 1$, then this inequality is immediate, so suppose $\lambda > 1$. Since $c_G(F) \leq c_G(G \cap H) = k$, it suffices to show:

$$(\lambda - 1)\, k \leq 1 - k, \tag{23}$$

or:

$$\lambda\, k \leq 1. \tag{24}$$

But using Equation 20, we find:

$$\lambda\, k = c_H(G \cap H) \leq 1, \tag{25}$$

as desired.

The remaining case is that $c_G(F) = 0$. If $c_H(F) \neq 0$, then we can redo the proof with G and H interchanged. If $c_H(F) = 0$, then WASRP is immediately satisfied. □

Theorem 8. *If an SCR is DGEN, then it satisfies GIIA. If $\mathcal{G}$ is an additive family, then, if an SCR satisfies GIIA, it is DGEN.*

Proof. For the forward direction, note that by Theorem 1, the family $G \mapsto c_G(\cdot)$ satisfies Condition 2 of a CPS. Therefore, by the forward direction of Theorem 4, it satisfies GIIA. For the reverse direction, observe that if the family $G \mapsto c_G(\cdot)$ satisfies GIIA, then, by the reverse direction of Theorem 4, it is a CPS, so that, using additivity and Theorem 2, it is DGEN. □

We give an example of how a family of dimensionally ordered measures can be used to build an SCR.

Example 1. *Let $\Omega = \{x, y, z, w\}$, $\mathcal{F} = 2^\Omega$, and $\mathcal{G} = 2^\Omega \backslash \{\emptyset\}$. Consider the family of measures $\{\mu_\alpha, \mu_\beta, \mu_\gamma\}$, where $\alpha \prec \beta \prec \gamma$, as depicted in Table 1.*

	x	y	z	w
μ_α	κ	$+\infty$	$+\infty$	$+\infty$
μ_β	0	λ	μ	$+\infty$
μ_γ	0	0	0	ν

Table 1: A Dimensionally Ordered Family of Measures

Here, all of κ, λ, μ, and ν lie strictly between 0 and $+\infty$. It can be checked that Conditions 1 and 2 of Definition 3 are satisfied, so that, by Theorem 1, this family of measures generates a CPS. By Theorem 4, we also get an SCR satisfying GIIA. For instance:

$$c_{\{x,y,z\}}(\{x\}) = \frac{\mu_\beta(\{x\})}{\mu_\beta(\{x,y,z\})} = \frac{0}{\lambda+\mu}, \tag{26}$$

$$c_{\{x,y,z\}}(\{y\}) = \frac{\mu_\beta(\{y\})}{\mu_\beta(\{x,y,z\})} = \frac{\lambda}{\lambda+\mu}, \tag{27}$$

$$c_{\{x,y\}}(\{x\}) = \frac{\mu_\beta(\{x\})}{\mu_\beta(\{x,y\})} = \frac{0}{\lambda}, \tag{28}$$

$$c_{\{x,y\}}(\{y\}) = \frac{\mu_\beta(\{y\})}{\mu_\beta(\{x,y\})} = \frac{\lambda}{\lambda}, \tag{29}$$

so that:

$$0 \times 1 = c_{\{x,y,z\}}(\{x\}) \times c_{\{x,y\}}(\{y\}) = c_{\{x,y\}}(\{x\}) \times c_{\{x,y,z\}}(\{y\}) = 0 \times \frac{\lambda}{\lambda+\mu}, \tag{30}$$

as required by GIIA (Definition 10).

We can interpret $\{\mu_\alpha, \mu_\beta, \mu_\gamma\}$ as a family of (unnormalized) "choice measures" that the decision maker (DM) holds on the full choice domain. In making choices, the DM prioritizes these measures in reverse order to $\prec$. Thus, the DM begins with a primary measure μ_γ on the full set Ω. Conditional on the complement of the support $\mathrm{supp}\mu_\gamma$, the DM turns to the secondary measure μ_β. Conditional on the union of supports $\mathrm{supp}\mu_\gamma \cup \mathrm{supp}\mu_\beta$, the DM turns to the tertiary measure μ_α. This way of looking at the family $\{\mu_\alpha, \mu_\beta, \mu_\gamma\}$ actually coincides with how CPS's can be defined in general for finite Ω (Myerson, 1986a, 1986b; Blume, Brandenburger, and Dekel, 1991). The virtue of the Rényi (1956) approach (Definition 3) is that it applies to the general measure-theoretic case.

The next example establishes that WASRP does not imply GIIA.

Example 2. *Let $\Omega = \{x, y, z, w\}$, $\mathcal{F} = 2^\Omega$, and $\mathcal{G} = \{\Omega, \{x, y, z\}\}$. Set $G = \Omega$ and $H = \{x, y, z\}$. Notice that $\mathcal{G}$ is additive. We consider the SCR fully defined by the following conditions:*

$$c_G(\{x\}) = \frac{1}{2},\ c_G(\{y\}) = c_G(\{z\}) = 0, \tag{31}$$

$$c_H(\{x\}) = \frac{1}{2},\ c_H(\{y\}) = 0,\ c_H(\{z\}) = \frac{1}{2}, \tag{32}$$

$$c_G(\{x,y\}) = \frac{1}{2},\ c_G(\{y,z\}) = 0,\ c_G(\{z,x\}) = \frac{1}{2}, \tag{33}$$

$$c_H(\{x,y\}) = \frac{1}{2},\ c_H(\{y,z\}) = \frac{1}{2},\ c_H(\{z,x\}) = 1, \tag{34}$$

$$c_G(\{x,y,z\}) = \frac{1}{2}. \tag{35}$$

We find $c_G(G\backslash H) = \frac{1}{2}$, and it can be verified that $c_H(F) - c_G(F) \leq \frac{1}{2}$ for all $F \subseteq H$. Hence WASRP is satisfied. However, GIIA is violated. To see this, note, for example, that:

$$c_G(\{x,y\}) \times c_H(\{x,z\}) = \frac{1}{2} \times 1 \neq \frac{1}{2} \times \frac{1}{2} = c_H(\{x,y\} \times c_G(\{x,z\}). \tag{36}$$

This same example shows that REG does not imply GIIA, even under additivity. This can be checked directly, or by using the fact that WASRP implies REG.

We also see that WASRP does not imply DGEN, even under additivity. This is because Theorem 8 says DGEN implies GIIA, and we have just seen that GIIA does not hold. It is instructive to show directly that this example fails DGEN. Define indices γ and η by:

$$0 < \mu_\gamma(G) < +\infty, \tag{37}$$

$$0 < \mu_\eta(H) < +\infty. \tag{38}$$

Monotonicity yields $\mu_\gamma(H) \leq \mu_\gamma(G)$, so that $\mu_\gamma(H) < +\infty$ and therefore $\eta \preceq \gamma$. Suppose $\eta \prec \gamma$. Then $\mu_\gamma(H) = 0$, and so:

$$c_G(H) = \frac{\mu_\gamma(H \cap G)}{\mu_\gamma(G)} = \frac{\mu_\gamma(H)}{\mu_\gamma(G)} = 0 \neq \frac{1}{2}. \tag{39}$$

Alternatively, suppose $\eta = \gamma$. Then:

$$0 = c_G(\{z\}) = \frac{\mu_\gamma(\{z\} \cap G)}{\mu_\gamma(G)}, \tag{40}$$

so that $\mu_\gamma(\{z\}) = 0$. From this, we have $\mu_\eta(\{z\}) = 0$, and therefore:

$$c_H(\{z\}) = \frac{\mu_\eta(H \cap \{z\})}{\mu_\eta(H)} = 0 \neq \frac{1}{2}. \tag{41}$$

Finally, we know that Condition 2 of a CPS does not hold in this example, because then Theorem 4 would imply GIIA holds. We can again see this directly by writing:

$$c_G(\{z\}) = 0 \neq \frac{1}{2} \times \frac{1}{2} = c_G(H)\, c_H(\{z\}). \tag{42}$$

We conclude from Example 2 that WASRP does not characterize Level 2 of our hierarchy, even under additivity.

It remains to examine whether REG implies WASRP. This is answered in the negative by Example 3.2 in Dasgupta and Pattanaik (2007). However, we do have:

Theorem 9. *If $\mathcal{G}$ is an additive family, then, if an SCR satisfies REG, then it satisfies WASRP.*

Proof. We extract a proof from Dasgupta and Pattanaik (2007, Proposition 4.2). Fix $G, H \in \mathcal{G}$ and $F \in \mathcal{F}$ with $F \subseteq G \cap H$. Suppose, for a contradiction, that Inequality 8 does not hold, that is:

$$c_H(F) - c_G(F) > c_G(G \backslash H). \tag{43}$$

By Regularity, $c_G(F) \geq c_{G \cup H}(F)$ and $c_G(G \backslash H) \geq c_{G \cup H}(G \backslash H)$, so that from 43 we get:

$$c_H(F) > c_{G \cup H}(F) + c_{G \cup H}(G \backslash H). \tag{44}$$

Now write $G \cup H$ as the disjoint union $F \cup G \backslash H \cup H \backslash F$, so that:

$$c_{G \cup H}(F) + c_{G \cup H}(G \backslash H) + c_{G \cup H}(H \backslash F) = 1. \tag{45}$$

Combining 44 and 45 yields:

$$c_H(F) + c_{G \cup H}(H \backslash F) > 1, \tag{46}$$

from which, using Regularity again:

$$1 = c_H(F) + c_H(H \backslash F) > 1, \tag{47}$$

a contradiction. □

Definition 12. *An SCR $c : \mathcal{G} \to \Delta(\Omega)$ satisfies* ***Strong Stochastic Transitivity (SST)****,* ***Moderate Stochastic Transitivity (MST)****,* ***Weak Stochastic Transitivity (WST)*** *if $c_{\{x,y\}}(\{x\}) \geq 1/2$ and $c_{\{y,z\}}(\{y\}) \geq 1/2$ implies, respectively:*

$$c_{\{z,x\}}(\{x\}) \geq \begin{cases} \max\{c_{\{x,y\}}(\{x\}), c_{\{y,z\}}(\{y\})\}, \\ \min\{c_{\{x,y\}}(\{x\}), c_{\{y,z\}}(\{y\})\}, \\ \frac{1}{2}. \end{cases} \tag{48}$$

Theorem 10. *If an SCR satisfies GIIA, it satisfies Strong (therefore Moderate and Weak) Stochastic Transitivity.*

Proof. From GIIA, we can write:

$$c_{\{x,y\}}(\{x\}) \times c_{\{x,y,z\}}(\{y\}) = c_{\{x,y,z\}}(\{x\}) \times c_{\{x,y\}}(\{y\}), \tag{49}$$

$$c_{\{y,z\}}(\{y\}) \times c_{\{x,y,z\}}(\{z\}) = c_{\{x,y,z\}}(\{y\}) \times c_{\{y,z\}}(\{z\}), \tag{50}$$

$$c_{\{z,x\}}(\{x\}) \times c_{\{x,y,z\}}(\{z\}) = c_{\{x,y,z\}}(\{x\}) \times c_{\{z,x\}}(\{z\}). \tag{51}$$

Case (i): Suppose that $c_{\{x,y,z\}}(\{y\}) > 0$ and $c_{\{x,y,z\}}(\{z\}) > 0$. Then, using the above and Regularity (implied by Theorem 5), we get:

$$\frac{c_{\{x,y\}}(\{x\})}{c_{\{x,y\}}(\{y\})} = \frac{c_{\{x,y,z\}}(\{x\})}{c_{\{x,y,z\}}(\{y\})}, \tag{52}$$

$$\frac{c_{\{y,z\}}(\{y\})}{c_{\{y,z\}}(\{z\})} = \frac{c_{\{x,y,z\}}(\{y\})}{c_{\{x,y,z\}}(\{z\})}, \tag{53}$$

$$\frac{c_{\{z,x\}}(\{x\})}{c_{\{z,x\}}(\{z\})} = \frac{c_{\{x,y,z\}}(\{x\})}{c_{\{x,y,z\}}(\{z\})}. \tag{54}$$

Multiplying Equations 52 and 53 together and using Equation 54, we get:

$$\frac{c_{\{z,x\}}(\{x\})}{c_{\{z,x\}}(\{z\})} = \frac{c_{\{x,y\}}(\{x\})}{c_{\{x,y\}}(\{y\})} \times \frac{c_{\{y,z\}}(\{y\})}{c_{\{y,z\}}(\{z\})}. \tag{55}$$

From $c_{\{x,y\}}(\{x\}) \geq 1/2$ and $c_{\{y,z\}}(\{y\}) \geq 1/2$, it follows that:

$$\frac{c_{\{x,y\}}(\{x\})}{c_{\{x,y\}}(\{y\})} \geq 1 \text{ and } \frac{c_{\{y,z\}}(\{y\})}{c_{\{y,z\}}(\{z\})} \geq 1. \tag{56}$$

Therefore:

$$\frac{c_{\{z,x\}}(\{x\})}{c_{\{z,x\}}(\{z\})} \geq \max\left\{\frac{c_{\{x,y\}}(\{x\})}{c_{\{z,y\}}(\{y\})}, \frac{c_{\{y,z\}}(\{y\})}{c_{\{y,z\}}(\{z\})}\right\}, \tag{57}$$

from which the first condition in Inequality 47 holds.

Case (ii): Suppose that $c_{\{x,y,z\}}(\{y\}) = 0$. Then Equation 50 and the premise of SST imply $c_{\{x,y,z\}}(\{z\}) = 0$. By Equation 51, either $c_{\{x,y,z\}}(\{x\}) = 0$ or $c_{\{z,x\}}(\{z\}) = 0$. The first case contradicts $c_{\{x,y,z\}}(\{x, y, z\}) = 1$. In the second case, $c_{\{z,x\}}(\{x\}) = 1$, so that SST must hold.

Case (iii): Suppose that $c_{\{x,y,z\}}(\{z\}) = 0$. Again, we have $c_{\{x,y,z\}}(\{x\}) = 0$ or $c_{\{z,x\}}(\{z\}) = 0$. In the second case, we repeat the argument above. In the first case, we also repeat the argument above, to obtain $c_{\{x,y,z\}}(\{y\}) = 1$, But then Equation 49 implies $c_{\{x,y\}}(\{x\}) = 0$, contradicting the premise of SST. □

5 Hierarchy Level 3

In this level, we consider probabilistic mixtures of PCPS's. We begin with two well-known behavioral effects that can be modeled this way. We then build a measurable space of PCPS's, using results from Cohen (1980). We go on to provide a characterization of SCR's that can be represented as probabilistic mixtures of PCPS's. This result also establishes that CPS's (Level 2) indeed lie below Level 3 as developed in this section.

5.1 The Similarity and Compromise Effects

The Similarity Effect was introduced by Debreu (1960) and named by Tversky (1972a, 1972b). We review the effect couched in the McFadden (1974) "red bus-blue bus" language.

Example 3. *The underlying set of choices is* $\Omega = \{\text{red bus}, \text{blue bus}, \text{taxi}\}$. *The decision maker (DM) is indifferent between taking a bus or a taxi and flips a coin to decide between these two modes of travel. Within the bus category, the DM does not care about the color of the bus and therefore flips a coin to decide on red vs. blue if both choices are available. Set* $\mathcal{G} = \{\Omega, \{\text{blue bus}, \text{taxi}\}\}$ *and consider the following three PCPS's* p^{r}, p^{b}, p^{t}:

$$p^r_\Omega(\{\text{red bus}\}) = 1,\ p^r_{\{\text{blue bus},\text{taxi}\}}(\{\text{blue bus}\}) = 1, \tag{58}$$

$$p^b_\Omega(\{\text{blue bus}\}) = 1,\ p^b_{\{\text{blue bus},\text{taxi}\}}(\{\text{blue bus}\}) = 1, \tag{59}$$

$$p^t_\Omega(\{\text{taxi}\}) = 1,\ p^t_{\{\text{blue bus},\text{taxi}\}}(\{\text{taxi}\}) = 1. \tag{60}$$

The DM puts probability $\frac{1}{4}$ *on each of the first and second PCPS's* p^r *and* p^b, *and probability* $\frac{1}{2}$ *on the third PCPS* p^t. *If the choice set is* Ω, *then the DM averages over the PCPS's to put probability* $\frac{1}{2}$ *on a bus and probability* $\frac{1}{2}$ *on a taxi. If the choice set is* $\{\text{blue bus}, \text{taxi}\}$, *then the DM averages to get again probability* $\frac{1}{2}$ *on a bus and probability* $\frac{1}{2}$ *on a taxi. This is the intuitive set of choice probabilities. But, as first pointed out by Debreu (1960), such a pattern violates IIA. In the first case, the probability ratio of blue bus to taxi is* $1 : 2$, *while in the second case it is* $1 : 1$. *Using Theorem 4 from the previous section, we see that this example also shows that a probabilistic mixture of PCPS's need not be a CPS.*

The next example shows how to encompass the Compromise Effect (Simonson, 1989, Simonson and Tversky, 1992) in the same manner.

Example 4. *The underlying set of choices is $\Omega = \{$l-camera, m-camera, h-camera$\}$, where l-camera is the low-quality/low-price option, m-camera is the medium-quality/medium-price option, and h-camera is the high-quality/high-price option. Set $\mathcal{G} = \{\{$l-camera, m-camera$\}, \Omega\}$. Faced with the choice set $\{$l-camera, m-camera$\}$, the DM selects each camera with equal probability. Faced with the choice set Ω, the DM selects each of the l-camera and the h-camera with probability $\frac{1}{4}$, and selects the m-camera with probability $\frac{1}{2}$. This is a violation of IIA. The usual story is that the addition of the h-camera to the choice set emphasizes the inferiority of the l-camera, which in turn makes the m-camera stand out as a good compromise between the low quality of the l-camera and the high price of the h-camera.*

This scenario can again be represented via a probabilistic mixture of PCPS's:

$$p^l_\Omega(\{\text{l-camera}\}) = 1,\ p^l_{\{\text{l-camera, m-camera}\}}(\{\text{l-camera}\}) = 1, \tag{61}$$

$$p^h_\Omega(\{\text{h-camera}\}) = 1,\ p^h_{\{\text{l-camera, m-camera}\}}(\{\text{l-camera}\}) = 1, \tag{62}$$

$$p^m_\Omega(\{\text{m-camera}\}) = 1,\ p^m_{\{\text{l-camera, m-camera}\}}(\{\text{m-camera}\}) = 1. \tag{63}$$

The DM puts probability $\frac{1}{4}$ on each of the first and second PCPS's p^l and p^h, and probability $\frac{1}{2}$ on the third PCPS p^m. The key here is that when the choice set expands to include the h-camera, then, with probability $\frac{1}{4}$, the DM will switch from the l-camera to the h-camera, reflecting the inferiority of the former.

A notable feature of our treatment of the Similarity and Compromise Effects is that, although they are usually described quite differently in behavioral terms (our language in the two examples reflected this), they have the same formal structure within our framework. While, arguably, some descriptive detail is lost under our approach, we think its parsimony is a virtue.

5.2 Preliminaries

Our goal in this section is to define probabilistic mixtures of PCPS's in the general case. We fix a general set Ω, let the family $\mathcal{G}$ of conditioning events consisting of all finite nonempty subsets of Ω. We will sometimes call $\mathcal{G}$ **finitary** in this case. Let $\mathcal{F}$ be the σ-field generated by $\mathcal{G}$. Note that $\mathcal{F}$ contains all singletons $\{\omega\}$.

Let Λ be the set of all total orders $\rhd$ on Ω and let Π be the set of all PCPS's (relative to the fixed family $\mathcal{G}$) on $(\Omega, \mathcal{F})$. We want to build a map $f : \Lambda \to \Pi$. To do so, start with a total order $\rhd \in \Lambda$, and define a selection $s : \mathcal{G} \to \Omega$ by:

$$s(G) = !\,\alpha \text{ such that } \alpha \rhd G \backslash \{\alpha\}, \tag{64}$$

where we write $\alpha \rhd G\backslash\{\alpha\}$ to mean $\alpha \rhd \beta$ for all $\beta \in G\backslash\{\alpha\}$.

Theorem 11. *The selection s defines a PCPS $p : \mathcal{G} \to \Delta(\Omega)$ on $(\Omega, \mathcal{F})$.*

Proof. We need to show that if $E \subseteq F \subseteq G$ with $E \in \mathcal{F}$ and $F, G \in \mathcal{G}$, then $p_G(E) = p_G(F)p_F(E)$. Case (i): $p_G(E) = 1$, that is, $s(G) \in E$. Then $p_G(F) = 1$ by monotonicity, and since $F \subseteq G$, we certainly have:

$$s(G) \rhd \beta \ \forall\ \beta \in F\backslash\{s(G)\}, \tag{65}$$

so that $p_F(E) = 1$. Case (ii): $p_G(E) = 0$, that is, $s(G) \in F\backslash E$ or $s(G) \in G\backslash F$. In the first case, we find $p_F(E) = 0$, and in the second case we find $p_G(F) = 0$. □

Let $f : \Lambda \to \Pi$ be the map that takes an order $\rhd$ to the PCPS p just constructed. Next, define the set:

$$\mathcal{A} = \{(\alpha, G) : \alpha \in G \text{ and } G \in \mathcal{G}\}, \tag{66}$$

and, for $(\alpha, G) \in \mathcal{A}$, let:

$$[\alpha, G]_\Lambda = \{\rhd \in \Lambda : \alpha \rhd G\backslash\{\alpha\}\}, \tag{67}$$

$$[\alpha, G]_\Pi = \{p \in \Pi : p_G(\alpha) = 1\}. \tag{68}$$

Observe that, by construction, if $\rhd \in [\alpha, G]_\Lambda$, then $f(\rhd) \in [\alpha, G]_\Pi$.

Theorem 12. *Suppose:*

$$\sum_{i=1}^{M} \mathbf{1}_{[\alpha_i, G_i]_\Pi} \leq \sum_{j=1}^{N} \mathbf{1}_{[\alpha_j, G_j]_\Pi}. \tag{69}$$

for integers M and N. Then:

$$\sum_{i=1}^{M} \mathbf{1}_{[\alpha_i, G_i]_\Lambda} \leq \sum_{j=1}^{N} \mathbf{1}_{[\alpha_j, G_j]_\Lambda}. \tag{70}$$

Proof. Fix an order $\rhd \in \Lambda$ such that (the numbering is without loss of generality):

$$\rhd \in [\alpha_i, G_i]_\Lambda \text{ for } i \leq m, \tag{71}$$

$$\rhd \notin [\alpha_i, G_i]_\Lambda \text{ for } i > m, \tag{72}$$

for some $m \leq M$. We need to show that there is a subset $\{j_1, j_2, \ldots, j_m\}$ of the index set $\{1, 2, \ldots, N\}$ such that:

$$\rhd \in [\alpha_{j_i}, G_{j_i}]_\Lambda \text{ for } i \leq m. \tag{73}$$

From Equation 64 and the definition of f, the PCPS $p = f(\rhd)$ satisfies:

$$p \in [\alpha_i, G_i]_\Pi \text{ for } i \leq m. \tag{74}$$

By Inequality 69, there are indices $j_1, j_2, \ldots, j_m$ (there is no loss of generality in using the same indices as above) such that:

$$p \in [\alpha_{j_i}, G_{j_i}]_\Pi \text{ for } i \leq m. \tag{75}$$

By the construction of p from $\rhd$, Inequality 75 holds if and only if:

$$\alpha_{j_i} \rhd G_{j_i} \backslash \{\alpha_{j_i}\}, \tag{76}$$

for each $i \leq m$, from which we have:

$$\rhd \in [\alpha_{j_i}, G_{j_i}]_\Pi \text{ for } i \leq m. \tag{77}$$

We conclude that $\rhd$ lies in at least m sets on the right-hand side of Inequality 70, which establishes the theorem. □

Cohen (1980, Section 4) builds a field $\mathcal{B}_\Lambda$ on Λ with the property that all sets $[\alpha, G]_\Lambda$ lie in this field. We want to build a field $\mathcal{D}_\Pi$ on Π that contains all sets $[\alpha, G]_\Pi$ and on which we can define a finitely additive probability measure Q, starting from a finitely additive probability measure P on $(\Lambda, \mathcal{B}_\Lambda)$. To do so, we use some results from the theory of charges (i.e., finitely additive measures). The following definition is from Rao and Rao (1983, Definition 3.2.2).

Definition 13. *Fix an arbitrary set Ω and let $\mathcal{E}$ be a collection of subsets of Ω. A function $\mu : \mathcal{E} \to \mathbb{R}^+$ is a **positive real partial charge on** $(\Omega, \mathcal{E})$ if:*

$$\sum_{i=1}^{M} \mathbf{1}_{C_i} \leq \sum_{j=1}^{N} \mathbf{1}_{D_j} \tag{78}$$

implies:

$$\sum_{i=1}^{M} \mu(C_i) \leq \sum_{j=1}^{N} \mu(D_j), \tag{79}$$

for any $C_1, \ldots, C_M, D_1, \ldots, D_N \in \mathcal{E}$.

The next observation is part of Rao and Rao (1983, Theorem 3.1.9).

Theorem 13. *Let $\mathcal{E}$ be a field and let μ be a finitely additive probability measure on $(\Omega, \mathcal{E})$. Then μ is a positive real partial charge on $(\Omega, \mathcal{E})$.*

We now fix a finitely additive probability measure P on $(\Lambda, \mathcal{B}_\Lambda)$ and consider the following collection of subsets of Π:

$$\mathcal{H} = \{[\alpha, G]_\Pi : (\alpha, G) \in \mathcal{A}\} \cup \{\Pi\}. \tag{80}$$

Define a function $Q : \mathcal{H} \to \mathbb{R}^+$ by setting:

$$Q([\alpha, G]_\Pi) = P([\alpha, G]_\Lambda) \text{ and } Q(\Pi) = 1. \tag{81}$$

Theorem 14. *The function Q is a positive real partial charge on $\mathcal{H}$.*

Proof. Suppose Inequality 69 holds. Then, by Theorem 12, Inequality holds. By Theorem 13, we know that P is a positive real partial charge on $(\Lambda, \mathcal{B}_\Lambda)$. Therefore, applying Inequality 79:

$$\sum_{i=1}^{M} P([\alpha_i, G_i]_\Lambda) \leq \sum_{j=1}^{N} P([\alpha_j, G_j]_\Lambda), \tag{82}$$

from which

$$\sum_{i=1}^{M} Q([\alpha_i, G_i]_\Pi) \leq \sum_{j=1}^{N} Q([\alpha_j, G_j]_\Pi). \tag{83}$$

This establishes that Q is a positive real partial charge on $\mathcal{H}$, except if, instead of Inequality 69, we have:

$$\mathbf{1}_\Pi + \sum_{i=1}^{M} \mathbf{1}_{[\alpha_i, G_i]_\Pi} \leq \sum_{j=1}^{N} \mathbf{1}_{[\alpha_j, G_j]_\Pi}, \tag{84}$$

or:

$$\sum_{i=1}^{M} \mathbf{1}_{[\alpha_i, G_i]_\Pi} \leq \sum_{j=1}^{N} \mathbf{1}_{[\alpha_j, G_j]_\Pi} + \mathbf{1}_\Pi. \tag{85}$$

But then the proof of Theorem 12 can be re-run to show that:

$$\mathbf{1}_\Lambda + \sum_{j=1}^{N} \mathbf{1}_{[\alpha_j, G_j]_\Pi} \leq \sum_{j=1}^{N} \mathbf{1}_{[\alpha_j, G_j]_\Lambda}, \tag{86}$$

or:

$$\sum_{j=1}^{N} \mathbf{1}_{[\alpha_j, G_j]_\Pi} \leq \sum_{j=1}^{N} \mathbf{1}_{[\alpha_j, G_j]_\Lambda} + \mathbf{1}_\Lambda. \tag{87}$$

The rest of the argument proceeds as before, with Inequality 86 or 87 replacing Inequality 70. If $\mathbf{1}_\Pi$ is added to both sides of Inequality 69, this term cancels out. □

We are now ready to apply an extension theorem for charges (Rao and Rao, 1983, Theorem 3.2.10).

Theorem 15. *Fix an arbitrary set Ω, let $\mathcal{E}$ be a collection of subsets of Ω with $\Omega \in \mathcal{E}$, and suppose μ is a positive real partial charge on $(\Omega, \mathcal{E})$. Then, for any field $\mathcal{D}$ containing $\mathcal{E}$, there is a finitely additive probability measure $\tilde{\mu}$ on $(\Omega, \mathcal{D})$ that extends μ.*

To apply Theorem 15 to our setting, let $\mathcal{D}_\Pi$ denote the field generated by the collection of sets $\mathcal{H}$ in Equation 80. (In the statement to follow, we do not distinguish the extension of Q from Q itself. No confusion should result.)

Theorem 16. *Fix the finitely additive probability space $(\Lambda, \mathcal{B}_\Lambda, P)$. There is a finitely additive probability measure Q on $(\Pi, \mathcal{D}_\Pi)$ such that:*

$$Q([\alpha, G]_\Pi) = P([\alpha, G]_\Lambda), \tag{88}$$

for all $(\alpha, G) \in \mathcal{A}$.

Proof. The proof is immediate from putting together Theorem 14, Theorem 15, and the definition of Q in Equation 81. □

5.3 Characterization

In this section, we provide the desired characterization of SCR's that are realizable as probabilistic mixtures of PCPS's. We begin with the classic Block and Marschak (1960) condition, in the form of Cohen (1980, Definition 2.4).

Definition 14. *The **Block-Marschak (BM) condition** associated with an SCR $c : \mathcal{G} \to \Delta(\Omega)$ is the requirement that the function $q : \Omega \times \mathcal{G} \times \mathcal{G} \to \mathbb{R}$ defined by:*

$$q(\alpha, G, H) = \sum_{j=0}^{|G|} (-1)^{|G|-j} \sum_{\{E \subseteq G : |E| = j\}} c_{H \setminus E}(\{\alpha\}), \tag{89}$$

for $G \subsetneq H$ and $\alpha \in H \backslash G$, is everwhere non-negative.

Definition 15. *An SCR $c : \mathcal{G} \to \Delta(\Omega)$ is **realizable as a Λ-based probability mixture** if there is a finitely additive probability measure P on $(\Lambda, \mathcal{B}_\Lambda)$ such that for all $(\alpha, G) \in \mathcal{A}$:*

$$c_G(\{\alpha\}) = P([\alpha, G]_\Lambda). \tag{90}$$

Cohen (1980, Theorem 4.1) establishes that an SCR is realizable as a probabilistic mixture of total orders if and only if the Block-Marschak condition is satisfied. This theorem extends the classic equivalence result of Falmagne (1978) to the infinite choice context. Cohen (1980) also identifies sufficient conditions for a realization in terms of a countably additive probability measure, but we will use his result with finitely additivity in order to obtain necessary and sufficient conditions in our setting.

Definition 16. *An SCR $c : \mathcal{G} \to \Delta(\Omega)$ is **realizable as a Π-based probability mixture** if there is a finitely additive probability measure Q on $(\Pi, \mathcal{D}_\Pi)$ such that for all $(\alpha, G) \in \mathcal{A}$:*

$$c_G(\{\alpha\}) = Q([\alpha, G]_\Pi). \tag{91}$$

Theorem 4.1 in Cohen (1980) and Theorem 16 above immediately imply:

Theorem 17. *An SCR is realizable as a Π-based probability mixture if and only if the Block-Marschak condition is satisfied.*

5.4 Relationships Among Axioms

The first result establishes that the Level-3 SCR's of this section indeed lie above Level 2 as defined in the previous section. We need a preliminary result, which is Theorem 3 in Falmagne (1978). Throughout, we assume that $\mathcal{G}$ is finitary.

Theorem 18. *For all $G, H \in \mathcal{G}$:*

$$\sum_{\alpha \in H \backslash G} q(\alpha, G, H) = \sum_{\beta \in G} q(\beta, G \backslash \{\beta\}, H). \tag{92}$$

Theorem 19. *If an SCR satisfies GIIA, then it satisfies the BM condition.*

Proof. We will first treat the case that G is a singleton and then treat the general case by successive substitution. We make essential use of the fact that $\mathcal{G}$ is an additive family.

Suppose $G = \{\omega\}$, so that:

$$q(\alpha, \{\omega\}, H) = (-1) \times c_H(\{\alpha\}) + (+1) \times c_{H \backslash \{\omega\}}(\{\alpha\}). \tag{93}$$

By Theorem 8 and Equation 1, we have:

$$c_H(\{\alpha\}) = \frac{\mu_{\gamma(H)}(\{\alpha\})}{\mu_{\gamma(H)}(H)}, \tag{94}$$

where, for any $F \in \mathcal{G}$, $\gamma(F)$ is the unique index for which $0 < \mu_\gamma(F) < +\infty$. Similarly:

$$c_{H \backslash \{\omega\}}(\{\alpha\}) = \frac{\mu_{\gamma(H \backslash \{\omega\})}(\{\alpha\})}{\mu_{\gamma(H \backslash \{\omega\})}(H \backslash \{\omega\})}. \tag{95}$$

Case (i): $\gamma(H) = \gamma(H \backslash \{\omega\}) = \gamma$. Then:

$$q(\alpha, \{\omega\}, H) = (-1) \times \frac{\mu_\gamma(\{\alpha\})}{\mu_\gamma(H)} + (+1) \times \frac{\mu_\gamma(\{\alpha\})}{\mu_\gamma(H \backslash \{\omega\})} \geq 0, \tag{96}$$

since $\mu_\gamma(H) \geq \mu_\gamma(H \backslash \{\omega\})$ by monotonicity. Case (ii): $\gamma(H \backslash \{\omega\}) \prec \gamma(H)$. Then:

$$\mu_{\gamma(H)}(H \backslash \{\omega\}) = 0, \tag{97}$$

so that, again using monotonicity, $\mu_{\gamma(H)}(\{\alpha\}) = 0$ and therefore $q(\alpha, \{\omega\}, H) \geq 0$, as required.

Now consider a general (finite) G. We can write:

$$q(\alpha, G, H) = \sum_{j=0}^{|G|} (-1)^{|G|-j} \sum_{\{E \subseteq G : |E| = j\}} c_{H \backslash E}(\{\alpha\}) \tag{98}$$

$$= \sum_{j=0}^{|G|} (-1)^{|G|-j} \sum_{\{E \subseteq G : |E| = j\}} \frac{\mu_{\gamma(H \backslash E)}(\{\alpha\})}{\mu_{\gamma(H \backslash E)}(H \backslash E)}. \tag{99}$$

Observe that since $\{\alpha\} \subseteq H\backslash E$, we have $\gamma(\{\alpha\}) \preceq \gamma(H\backslash E)$. If $\gamma(\{\alpha\}) \prec \gamma(H\backslash E)$ for some E, then $\mu_{\gamma(H\backslash E)}(\{\alpha\}) = 0$. This means we can re-write Equation 99 as:

$$q(\alpha, G, H) = \mu_{\gamma(\{\alpha\})}(\{\alpha\}) \times \sum_{j=0}^{|G|}(-1)^{|G|-j} \sum_{\substack{\{E\subseteq G:|E|=j\}\\ \{E\subseteq G:\gamma(H\backslash E)=\gamma(\{\alpha\})\}}} \frac{1}{\mu_{\gamma(\alpha)}(H\backslash E)}. \tag{100}$$

From this, we can calculate:

$$\sum_{\alpha\in H\backslash G} q(\alpha, G, H) = \sum_{\alpha\in H\backslash G} \mu_{\gamma(\{\alpha\})}(\{\alpha\}) \times \sum_{j=0}^{|G|}(-1)^{|G|-j} \sum_{\substack{\{E\subseteq G:|E|=j\}\\ \{E\subseteq G:\gamma(H\backslash E)=\gamma(\{\alpha\})\}}} \frac{1}{\mu_{\gamma(\alpha)}(H\backslash E)} \tag{101}$$

$$= \mu_{\gamma(\{\alpha\})}(H\backslash G) \times \sum_{j=0}^{|G|}(-1)^{|G|-j} \sum_{\substack{\{E\subseteq G:|E|=j\}\\ \{E\subseteq G:\gamma(H\backslash E)=\gamma(\{\alpha\})\}}} \frac{1}{\mu_{\gamma(\alpha)}(H\backslash E)}. \tag{102}$$

If $q(\alpha, G, H) = 0$, there is nothing to prove, so we can assume $q(\alpha, G, H) \neq 0$. This tells us that the double sum on the right side of Equation 100 is nonzero, so that, using $\mu_{\gamma(\{\alpha\})}(H\backslash G) > 0$ and Equation 102, we know that $\sum_{\alpha\in H\backslash G} q(\alpha, G, H) \neq 0$. We can therefore divide Equation 100 by Equation 102 to get:

$$\frac{q(\alpha, G, H)}{\sum_{\alpha\in H\backslash G} q(\alpha, G, H)} = \frac{\mu_{\gamma(\{\alpha\})}(\{\alpha\})}{\mu_{\gamma(\{\alpha\})}(H\backslash G)}. \tag{103}$$

From this:

$$q(\alpha, G, H) = \frac{\mu_{\gamma(\{\alpha\})}((\{\alpha\})}{\mu_{\gamma(\{\alpha\})}(H\backslash G)} \times \sum_{\alpha\in H\backslash G} q(\alpha, G, H) \tag{104}$$

$$= \frac{\mu_{\gamma(\{\alpha\})}(\{\alpha\})}{\mu_{\gamma(\{\alpha\})}(H\backslash G)} \times \sum_{\beta\in G} q(\beta, G\backslash\{\beta\}, H) \tag{105}$$

$$= \frac{\mu_{\gamma(\{\alpha\})}(\{\alpha\})}{\mu_{\gamma(\{\alpha\})}(H\backslash G)} \times \sum_{\beta\in G}\left[\frac{\mu_{\gamma(\{\beta\})}(\{\beta\})}{\mu_{\gamma(\{\beta\})}(H\backslash(G\backslash\{\beta\}))} \times \sum_{\beta\in H\backslash(G\backslash\{\beta\})} q(\beta, G\backslash\{\beta\}, H)\right] \tag{106}$$

$$= \frac{\mu_{\gamma(\{\alpha\})}(\{\alpha\})}{\mu_{\gamma(\{\alpha\})}(H\backslash G)} \times \sum_{\beta\in G}\left[\frac{\mu_{\gamma(\{\beta\})}(\{\beta\})}{\mu_{\gamma(\{\beta\})}(H\backslash(G\ \backslash\{\beta\}))} \times \sum_{\theta\in G\backslash\{\beta\}} q(\theta, G\backslash\{\beta, \theta\}, H)\right] \tag{107}$$

$$= \cdots. \tag{108}$$

Here, Equality 105 uses Theorem 18. Equality 106 comes from substituting in the analog, using β and $G\backslash\{\beta\}$ in place of α and G, to Equation 104. Equality 107 is another appeal to Theorem 18. Noting that G is finite, we repeat this substitution process a finite number of times until the only term involving $q(\cdot,\cdot,\cdot)$ on the right side is of the form $q(\xi, \{\omega\}, H)$ for some singleton $\{\omega\} \subseteq G$ and $\xi \in H\backslash\{\omega\}$. We know from the first part of the proof that this term is non-negative. The rest of the right side will reduce to nested weights and sums, where all weights are positive, so we can conclude that $q(\alpha, G, H) \geq 0$, as required. □

Theorem 20. *If an SCR satisfies BM, then it satisfies REG.*

Proof. Consider $G, H \in \mathcal{G}$ with $G \subseteq H$, $\alpha \in G$, and two PCPS's $p_G(\cdot)$ and $p_H(\cdot)$. Using the chain rule $p_H(\{\alpha\}) = p_H(G) \times p_G(\{\alpha\})$, we have $p_H(\{\alpha\}) = 1$ implies $p_G(\{\alpha\}) = 1$. Thus:

$$[\alpha, H]_\Pi = \{p \in \Pi : p_H(\{\alpha\}) = 1\} \subseteq \{p \in \Pi : p_G(\{\alpha\}) = 1\} = [\alpha, G]_\Pi, \tag{109}$$

from which, using monotonicity:

$$c_H(\{\alpha\}) = Q([\alpha, H]_\Pi) \leq Q([\alpha, G]_\Pi) = c_G(\{\alpha\}), \tag{110}$$

as required. □

Figure 3 depicts the implications we have now established.

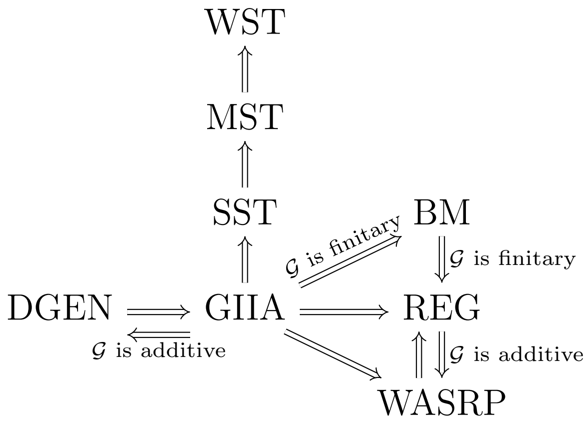

Figure 3: Relationships Among Axioms Contd.

We next show that BM does not imply SST, and MST does not imply BM. The first example is based on the Condorcet Paradox and is adapted from Strzalecki (2022, Example 3.17).

Example 5. *Let $\Omega = \{x, y, z\}$, $\mathcal{F} = 2^\Omega$, and $\mathcal{G} = 2^\Omega \backslash \{\emptyset\}$. Consider the three PCPS's defined by:*

$$p^1_\Omega(\{x\}) = 1,\ p^1_{\{y,z\}}(\{y\}) = 1, \tag{111}$$

$$p^2_\Omega(\{y\}) = 1,\ p^2_{\{z,x\}}(\{z\}) = 1, \tag{112}$$

$$p^3_\Omega(\{z\}) = 1,\ p^3_{\{x,y\}}(\{x\}) = 1. \tag{113}$$

The DM puts probability $\frac{1}{3}$ on each PCPS, from which we find:

$$c_{\{x,y\}}(\{x\}) = c_{\{y,z\}}(\{y\}) = c_{\{z,x\}}(\{z\}) = \frac{2}{3}, \tag{114}$$

which violates WST and therefore also MST and SST.

Example 6. *Marschak (1960) proposed the triangle condition $c_{\{x,y\}}(\{x\}) + c_{\{y,z\}}(\{y\}) + c_{\{z,x\}}(\{z\}) \leq 2$ as a form of transitivity and conjectured that it is necessary and sufficient (on a finite domain) for the BM condition to hold. McFadden and Richter (1971) presented a counter-example to sufficiency. This example is also presented as Figure 1 in Cohen and Falmagne (1990) and reproduced as Figure 4 here. In the diagram, an arrow from node x to node y means $c_{\{x,y\}}(\{x\}) = 1$ and $c_{\{x,y\}}(\{y\}) = 0$, while the absence of an arrow between nodes x and y means $c_{\{x,y\}}(\{x\}) = c_{\{x,y\}}(\{y\}) = \frac{1}{2}$. It can be checked that the example satisfies MST, so that it demonstrates that MST does not imply BM.*

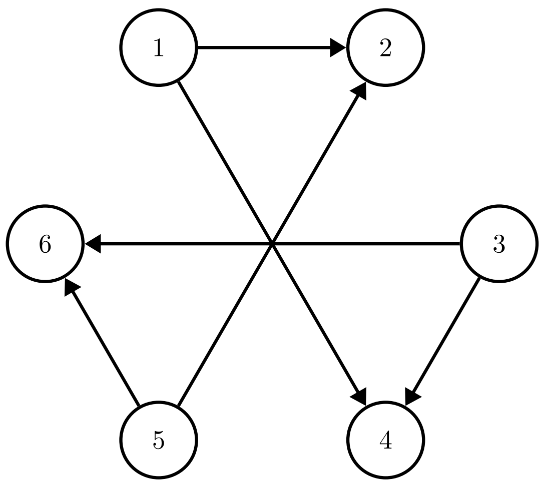

Figure 4: MST without BM

Overall, we see that while GIIA implies both BM and SST, the hierarchy we build in this paper diverges from a hierarchy built on successive weakenings of stochastic transitivity.

6 Hierarchy Levels 3a and 3b

In this section, we consider probability mixtures of PCPS's that are mutually absolutely continuous (m.a.c.) or absolutely continuous (a.c.) with respect an SCR. By definition, these two levels – which we call Levels 3a and 3b – lie above Level 3.

6.1 Preliminaries

To define Levels 3a and 3b, fix, as in Section 5.1, a general set Ω, let the family $\mathcal{G}$ of conditioning events be finitary, and let $\mathcal{F}$ be the σ-field generated by $\mathcal{G}$. We again write Π for the set of all PCPS's (relative to the fixed family $\mathcal{G}$ on $(\Omega, \mathcal{F})$. We build the same field $\mathcal{D}_\Pi$ on Π as in Section 5.3.

Definition 17. *An SCR $c : \mathcal{G} \to \Delta(\Omega)$ is* ***mutually absolutely continuously (m.a.c.) realizable as a Λ-based probability mixture*** *if there a finitely additive probability measure Q on $(\Pi, \mathcal{D}_\Pi)$ such that for all $(\alpha, G) \in \mathcal{A}$:*

$$c_G(\{\alpha\}) > 0 \text{ if and only if } Q([\alpha, G]_\Pi) > 0. \tag{115}$$

Using Theorem 17, it is immediate that we can also write Definition 17 as: An SCR $c : \mathcal{G} \to \Delta(\Omega)$ lies in Level 3a if and only if there another SCR $c' : \mathcal{G} \to \Delta(\Omega)$ that satisfies the Block-Marschak condition and is such that for all $(\alpha, G) \in \mathcal{A}$:

$$c_G(\{\alpha\}) > 0 \text{ if and only if } c'_G(\{\alpha\}) > 0. \tag{116}$$

Definition 18. *An SCR $c : \mathcal{G} \to \Delta(\Omega)$ is* ***absolutely continuously (a.c.) realizable as a Λ-based probability mixture*** *if there a finitely additive probability measure Q on $(\Pi, \mathcal{D}_\Pi)$ such that for all $(\alpha, G) \in \mathcal{A}$:*

$$c_G(\{\alpha\}) = 0 \text{ implies } Q([\alpha, G]_\Pi) = 0. \tag{117}$$

Again using Theorem 17, we can rewrite Definition 18 as: An SCR $c : \mathcal{G} \to \Delta(\Omega)$ lies in Level 3b if and only if there another SCR $c' : \mathcal{G} \to \Delta(\Omega)$ that satisfies the Block-Marschak condition and is such that for all $(\alpha, G) \in \mathcal{A}$:

$$c_G(\{\alpha\}) = 0 \text{ implies } c'_G(\{\alpha\}) = 0. \tag{118}$$

6.2 Choice Correspondences and Level-3a Characterization

Definition 19. *A* ***choice correspondence (CC)*** *is a map $C : \mathcal{G} \to \mathcal{F}$ satisfying $\emptyset \neq C(G) \subseteq G$ for every $G \in \mathcal{G}$.*

Clearly, an SCR $c_G(\cdot)$, induces a choice correspondence by setting $C(G) = \text{supp} c_G(\cdot)$ for every $G \in \mathcal{G}$. Indeed, we can view a CC as an equivalence class of SCR's that, for each $G \in \mathcal{G}$, share the same support. We will say that a CC is **Level-3a** if there is a Level-3a SCR that induces it.

We now offer a characterization of Level-3a CC's and, therefore, a characterization of equivalence classes of Level-3a SCR's. Our characterization works on finite domains Ω. Let $|\Omega|! = n$ and enumerate all total orders on Ω as $\rhd_1, \rhd_2, \ldots, \rhd_n$. Now fix a Level-3 SCR $c'_G(\cdot)$. By Falmagne (1978) (and the finite case of Theorem 4.1 in Cohen, 1980), we know there is a probability measure P on $\{\rhd_1, \rhd_2, \ldots, \rhd_n\}$ such that:

$$c'_G(\{\alpha\}) = \sum_{\{i : \alpha \rhd_i G \backslash \{\alpha\}\}} P(\rhd_i). \tag{119}$$

Writing $\Sigma = \text{supp}\, P$, we have:

$$\text{supp}\, c'_G = \{\alpha \in G : \alpha \rhd_i G \backslash \{\alpha\} \text{ for some } i \in \Sigma\}. \tag{120}$$

Definition 20. *A choice correspondence $C : \mathcal{G} \to \mathcal{F}$ is* ***multi-order justified*** *if there is a set of total orders $\{\rhd_1, \rhd_2, \ldots, \rhd_n\}$ such that:*

$$C(G) = \{\alpha \in G : \alpha \rhd_i G \backslash \{\alpha\} \text{ for some } i \in \Sigma\}. \tag{121}$$

Now, consider any Level-3a SCR $c_G(\cdot)$ that shares the same support as $c'_G(\cdot)$, i.e., that satisfies Equation 116, and let $C(\cdot)$ be the induced CC. Then we know that $C(\cdot)$ is multi-order justified. The following characterization of such choice correspondences is Theorem 3 in Aizerman and Malishevski (1981).

Theorem 21. *A choice correspondence is multi-order justified if and only if whenever $G, H \in \mathcal{G}$, with $G \subseteq H$:*

1. $C(H) \cap G \subseteq C(G)$;

2. $C(H) \subseteq G$ *implies* $C(G) \subseteq C(H)$.

Conditions 1 and 2 of Theorem 21 are, respectively, Postulate 4 (also called Condition α) and Postulate 5* (also called the AM or Reduction Condition) in Chernoff (1954). Plott (1973) introduced the **path independence** condition on choice correspondences: $C(G \cup H) = C(C(G) \cup H)$. Aizerman and Malishevski (1981) note that path independence is equivalent to the conjunction of Chernoff's Postulate 4 and 5*.

Summarizing, we have a characterization of equivalence classes of Level-3a SCR's – where an equivalence class is defined relative to common supports – in terms of properties of the (common) induced choice correspondence. Alternatively put, to check whether an SCR lies in Level 3a of our hierarchy it is necessary and sufficient to check that its induced choice correspondence satisfies Chernoff's Postulates 4 and 5*, or, equivalently, Plott's path-independence condition.

6.3 Level-3b One-Way Characterization

Repeating the arguments at the beginning of Section 6.2, we see that an SCR $c_G(\cdot)$ is Level 3b if the following holds: Let $c'_G(\cdot)$ be a Level-3 SCR that satisfies Equation 118. Suppose, for some $G \in \mathcal{G}$, that $\alpha \rhd G \backslash \{\alpha\}$ for some $\rhd$ with $P(\rhd) > 0$, where P realizes $c'_G(\cdot)$ as in Equation 119. Then, if $C(\cdot)$ is the choice correspondence associated with $c_G(\cdot)$, we have $\alpha \in C(G)$.

Definition 21. *A choice correspondence $C : \mathcal{G} \to \mathcal{F}$ is* ***order sub-justified*** *if there is a total order $\rhd$ such that for any $G \in \mathcal{G}$:*

$$\alpha \rhd G \backslash \{\alpha\} \text{ implies } \alpha \in C(G). \tag{122}$$

The following one-way characterization is Lemma 5 in Moulin (1985).

Theorem 22. *If a choice correspondence satisfies Chernoff's Condition α, then it is order sub-justified.*

In words, a sufficient condition for an SCR to lie in Level 3b of our hierarchy is that its induced choice correspondence satisfies Condition α.

There are deterministic choice models that do not belong to Level 3b, such as the menu-dependent incentivized choice patterns studied in Kopylov and Yang (2024). Their model accommodates the scenario:

$$C(\{a, b\}) = \{a\},\ C(\{a, b, c\}) = \{b\},\ C(\{a, b, c, d\}) = \{a\}, \tag{123}$$

which is readily seen to violate Chernoff's Condition α. Moreover, there is no total order $\rhd$ on $\{a, b, c, d\}$ that sub-justifies C.

7 Hierarchy Level 4

In this level, we consider signed probability mixtures of PCPS's. We begin by examining the Attraction and Repulsion Effects in this framework. We then define the general case of a finitely-additive signed probability measure on the field of PCPS's we introduced in Section 5. If the choice domain Ω is finite, then a key result of Dogan and Yildiz (2023) implies that any SCR can be represented as a signed mixture of PCPS's, so that Level 4 becomes the top of our complexity hierarchy. It follows that, in the case of a finite domain, Levels 3a and 3b are nested under Level 4.

7.1 The Attraction and Repulsion Effects

The Attraction Effect (Huber, Payne, and Puto, 1982, Simonson, 1989), and the Repulsion Effect (Aaker, 1991) are canonical examples of behavior that is described as contextual, i.e., as not conforming to classic axioms such as Luce IIA. We said in the Introduction that a violation of IIA is not per se an indication of contextuality in choice. In particular, we asserted that Level 3 behavior is non-contextual. But the Attraction and Repulsion Effects are contextual in our framework, because they lie further up.

Example 7. *The underlying set of choices is $\Omega = \{x, y, z\}\}$ and $\mathcal{G} = \{\Omega, \{x, y\}\}$. We consider an SCR satisfying:*

$$c_{\{x,y\}}(\{x\}) < c_\Omega(\{x\}), \tag{124}$$

which is a violation of Regularity. In one scenario for the Attraction Effect (Simonson and Tversky, 1992), item x is a nice pen, item y is a certain sum of money, and item z is a plain pen. The addition of item z highlights the attractiveness of item x, which is then chosen with higher probability. In a scenario for the Repulsion Effect (Simonson, 2014, Kruis et al., 2020), item x is candy, item y is an orange, and item z is *a spoiled clementine. The addition of item z casts doubt on the freshness of item y, so that, again, item x is chosen with higher probability.*

We now show how to produce the the Attraction and Repulsion Effects via a signed mixture of PCPS's. Specifically, consider four PCPS's relative to $\mathcal{G}$:

$$p^1_{\{x,y\}}(\{x\}) = 1,\, p^1_{\{x,y,z\}}(\{x\}) = 1, \tag{125}$$

$$p^2_{\{x,y\}}(\{y\}) = 1,\, p^2_{\{x,y,z\}}(\{y\}) = 1, \tag{126}$$

$$p^3_{\{x,y\}}(\{x\}) = 1,\, p^3_{\{x,y,z\}}(\{z\}) = 1, \tag{127}$$

$$p^4_{\{x,y\}}(\{y\}) = 1,\, p^4_{\{x,y,z\}}(\{z\}) = 1. \tag{128}$$

We put a signed probability measure (q^1, q^2, q^3, q^4), where each $q^i \in \mathbb{R}$ and $\sum_{i=1}^{4} q_i = 1$, on these four PCPS's and obtain:

$$c_{\{x,y\}}(\{x\}) = q^1 + q^3, \tag{129}$$

$$c_{\{x,y,z\}}(\{x\}) = q^1. \tag{130}$$

In order to create a violation of Regularity, we must have $q^3 < 0$. This makes intuitive sense. The PCPS p^3 selects item x – the nice pen or the candy – from the choice set $\{x, y\}$. But it selects item z – the plain pen or spoiled clementine – from the larger set $\{x, y, z\}$. We expect the DM to want to avoid choosing according to this PCPS. (We do not rule out that $q^4 < 0$ as well.) Our analysis here is similar to the treatment of the Attraction Effect by Dogan and Yildiz (2023, Example 2) in terms of orders.

We can be more precise about the idea that the DM wants to avoid the PCPS p^3. In Brandenburger et al. (2024), negative subjective probabilities on events are given an axiomatic basis and are interpreted as a negative willingness-to-bet on an event. More formally still, we can capture the Attraction-Repulsion Effect in a principal-agent manner. The principal is the DM and there are four agents, each with one of the four PCPS's above. The principal has preferences over which agent gets to make choices and dislikes the case that agent 3 – with PCPS p^3 – is the one to choose. A negative willingness-to-bet on this agent's being chosen is the formal counterpart to this attitude on the part of the principal.

We note that, at the formal level, it is possible to encompass the Attraction and Repulsion Effects at Level 3a of our hierarchy. This will still be a contextual analysis, since we are above Level 3. Indeed, since $c_{\{x,y\}}(\{x\}), c_{x,y,z}(\{x\}) \in (0, 1)$, we can easily find $0 < c'_{\{x,y,z\}}(\{x\}) < c'_{x,y}(\{x\}) \in (0, 1)$, where c' is absolutely continuous with respect to c.

But, we do not think it captures the essence of these effects. The reason is that the Level 3a analysis involves exhibiting a second SCR that is mutually absolutely continuous with respect to the one defined in Equations 129 and 130, where this second SCR must satisfy the Block-Marschak condition. By Theorem 20, it therefore satisfies Regularity. The essence of the Attraction and Repulsion Effects is the violation of Regularity, which is why we place them at Level 4 in our hierarchy.

Furthermore, if we consider an "extreme" version of these effects, where $0 = c_{\{x,y\}}(\{x\}) < c_\Omega(\{x\})$, this cannot be obtained at Level 3a (nor at Level 3b), while it can again be obtained at Level 4.

7.2 Characterization

We again fix a general set Ω, let the family $\mathcal{G}$ of conditioning events be finitary, and let $\mathcal{F}$ be the σ-field generated by $\mathcal{G}$. As before, we let $(\Pi, \mathcal{D}_\Pi)$ be our measurable space of PCPS's.

Definition 22. *An SCR $c : \mathcal{G} \to \Delta(\Omega)$ is **realizable as a Λ-based signed probability mixture** if there a finitely additive signed probability measure Q on $(\Pi, \mathcal{D}_\Pi)$ such that for all $(\alpha, G) \in \mathcal{A}$:*

$$c_G(\{\alpha\}) = Q([\alpha, G]_\Pi). \tag{131}$$

Level 4 of our hierarchy consists of SCR's satisfying Definition 22. We do not have a characterization of Level 4 in the general case, but Theorem 2 in Dogan and Yildiz (2023) gives a complete answer in the case of finite Ω.

Theorem 23. *Suppose the choice domain Ω is finite. Then every SCR $c : \mathcal{G} \to \Delta(\Omega)$ is realizable as a Π-based signed probability mixture.*

Proof. The argument transfers Theorem 2 in Dogan and Yildiz (2023) from total orders to PCPS's, mirroring the same move we made in Section 5.3. In our notation, their Theorem 1 states that given an SCR c, there is a signed probability measure P on the set Λ of strict linear orders such that for all $(\alpha, G) \in \mathcal{A}$:

$$c_G(\{\alpha\}) = P([\alpha, G]_\Lambda). \tag{132}$$

Paralleling our Theorem 17, we can use this partial information about the measure P and the extension theorem for charges to find a signed probability measure Q on the set Π of PCPS's that yields the same SCR c. $\square$

Theorem 23 says that when the choice domain Ω is finite, Level 4 is the top of our complexity hierarchy for stochastic choice. In particular, then, Levels 3a and 3b next under Level 4. Saito (2018) proves a more general result on realization of SCR's as signed mixtures of orders, where the family $\mathcal{G}$ of choice sets can be an arbitrary (nonempty) subset of $2^{\Omega}\backslash\{\emptyset\}$. Corollary 6(ii) and Footnote 22 in Saito (2018) demonstrate how Theorem 2 in Dogan and Yildiz (2023) follows.

The proof of Theorem 2 in Dogan and Yildiz (2023) works via a (significant) extension of the Ford-Fulkerson Theorem (Ford and Fulkerson, 2015) from combinatorial matrix theory to allow for negative row and column sums. But, as far as we can see, this method only works for a finite choice domain. We are therefore led to state what we believe to be two open issues for the case of an infinite choice domain Ω:

- We do not know if there are SCR's that lie above Level 4 of our complexity hierarchy.
- We do not know if Levels 3a and 3b nest under Level 4 of our complexity hierarchy.

8 Physical Systems

We mentioned in the Introduction that our complexity hierarchy for stochastic choice was inspired by a contextual hierarchy for quantum mechanics developed by Abramsky and Brandenburger (2011). Contextuality in quantum systems (Kochen and Specker, 1967), as formalized by Abramsky and Brandenburger (2011) and others, encompasses the famous quantum phenomenon of non-locality (Bell, 1964) and other non-classical features (e.g., Hardy, 1993, and Greenberger, Horne, and Zeilinger, 1989). While traditionally viewed as paradoxical aspects of the quantum world, the modern perspective on non-locality and contextuality is that these are valuable information-processing resources (Howard et al., 2014, Raussendorf, 2013).

The quantum hierarchy in Abramsky and Brandenburger (2011) is built using the language of sheaf theory, which is a tool for examining the possibility of "gluing together" local objects in some structure into a global object. In this section, we recast the Abramsky-Brandenburger hierarchy in probability-theoretic terms, to allow comparison with our hierarchy for stochastic choice.

8.1 Measurement-Outcome Frameworks

To describe a physical system, we fix a finite set X of **measurements** and a finite set O of possible **outcomes** for each measurement. A **measurement context** C is a subset of X consisting of a set of measurements that can be carried out together. (An important aspect of quantum systems is that not all measurements are necessarily compatible.) A **measurement cover** $\mathcal{M}$ is a family of measurement contexts C with $\cup_{C\in\mathcal{M}}C = X$.

To describe this set-up in the framework of the present paper, we set:

$$\Omega = X \times \bigsqcup_{C\in\mathcal{M}} O^C, \tag{133}$$

$$\mathcal{G} = \{C \times O^C : C \in \mathcal{M}\}, \tag{134}$$

where $\sqcup$ denotes disjoint union. For each $G = C \times O^C \in \mathcal{G}$, we build a probability measure $c_G(\cdot)$ on Ω by setting $c_G = \mu_C \otimes e_C$ where:

$$\mu_C \in \Delta(X) \text{ with } \operatorname{supp}\mu_C = C, \tag{135}$$

$$e_C \in \Delta(\bigsqcup_{C'\in\mathcal{M}} O^{C'}) \text{ with } e_C(O^C) = 1. \tag{136}$$

In the interpretation, one or more experimenters choose a compatible set C of measurements. The probabilities μ_C with which the measurements are selected plays no role in the physical theory, provided only that they are chosen "freely" by the experimenter(s). (Free choice of measurements is a basic assumption in the physical theories being modeled here; see Dickson, 1998.) This requirement is captured via the assumption that μ_C is an arbitrary full-support measure and c_G is built from it as a product. The other component measure e_C comes from the physical theory which specifies, for each measurement context C, probabilities on each of the possible joint outcomes. Thus, for a map $f \in O^C$, specifying an outcome in O associated with each measurement undertaken in C, the probability of this joint outcome is $e_C(f)$. These are the observable probabilities – as given by quantum mechanics, for example.

Definition 23. *A family* $\{e_C : C \in \mathcal{M}\}$ *of probability measures is called an* ***empirical model****.*

Example 8.1. *In the famous Bell-type scenario in quantum mechanics (Bell, 1964), there are two experimenters, Alice and Bob. The set of measurements is* $X = \{a, a', b, b'\}$ *and the set of outcomes is* $O = \{0, 1\}$*. Alice can choose one of the measurements* a *or* a'*, and Bob can choose one of the measurements* b *or* b'*, so that* $\mathcal{M} = \{\{a, b\}, \{a', b\}, \{a, b'\}, \{a', b'\}\}$*. The empirical model is depicted in Table 1 (where each row is a probability measure).*

	$(0,0)$	$(1,0)$	$(0,1)$	$(1,1)$
(a,b)	$1/2$	0	0	$1/2$
(a',b)	$3/8$	$1/8$	$1/8$	$3/8$
(a,b')	$3/8$	$1/8$	$1/8$	$3/8$
(a',b')	$1/8$	$3/8$	$3/8$	$1/8$

Table 2: The Bell Model

With the identifications we made in Equations 133- 136, we see that a measurement-outcome framework is naturally associated with an SCR $c : \mathcal{G} \to \Delta(\Omega)$. Moreover, all members of $\mathcal{G}$ are disjoint, so this SCR trivially satisfies GIIA. It follows that all SCR's constructed from our physical setting lie in Level 2 of our complexity hierarchy.

8.2 No Signaling

There is important inner structure to a measurement-outcome framework which is central in the physical setting. To extract this inner structure, consider two measurement contexts $C, C' \in \mathcal{M}$ with $C \cap C' \neq \emptyset$. Let $\mathrm{Res} : O^C \to O^{C \cap C'}$ be the restriction map, that is, for $f \in O^C$, $\mathrm{Res}(f) = f_{|C \cap C'}$. Define the restriction map $\mathrm{Res}' : O^{C'} \to O^{C \cap C'}$ analogously. Notice that for any $C \in \mathcal{M}$, we can view the probability measure e_C as living in $\Delta(O^C)$. Let $d_{C,C'}$ be the image measure of e_C under Res, that is, for $h \in O^{C \cap C'}$:

$$d_{C,C'}(h) = e_C(\{f \in O^C : \mathrm{Res}(f) = h\}). \tag{137}$$

Likewise, let $d_{C',C}$ be the image measure of $e_{C'}$ under Res'.

Definition 24. *An empirical model satisfies* ***no signaling*** *if for any* $h \in O^{C \cap C'}$*, we have:*

$$d_{C,C'}(h) = d_{C',C}(h). \tag{138}$$

In words, the no-signaling condition says that the outcome data (i.e., frequencies) obtained from a measurement m in measurement context C must be the same as that obtained from the same measurement m when undertaken in a different but compatible context C'. Thus, in the case of the Bell scenario (Example 2), the no-signaling condition is the statement that if, for example, Alice measures a, then the frequencies of 0's and 1's that Alice obtains are the same whether Bob measures b or b'. This condition was introduced by Popescu and Rohrlich (1994) and is usually justified by appeal to relativistic causality, which forbids superluminal transmission of information between Alice and Bob.

8.3 Phase-Space Characterization

A characterization of empirical models that satisfy no signaling can be obtained via the physical concept of phase space.

Definition 25. *The* ***phase space*** *for an empirical model is the set of maps* O^X*.*

Given a finite set Y, let $\Delta^{\mathrm{s}}(Y)$ denote the set of all signed probability measures on Y. For a measurement context $C \in \mathcal{M}$, let $\mathrm{RES}_C : O^X \to O^C$ be the restriction map. Given a signed probability measure $q \in \Delta^{\mathrm{s}}(O^X)$, we build the image measure under RES_C. That is, for $f \in O^C$, we build a signed probability measure $r_C \in \Delta^{\mathrm{s}}(O^C)$ by

$$r_C(f) = q(\{g \in O^X : \mathrm{RES}_C(g) = f\}). \tag{139}$$

Definition 26. *An empirical model* $\{e_C : C \in \mathcal{M}\}$ *is* ***realized by a phase-space signed probability measure*** $q \in \Delta^{\mathrm{s}}(O^X)$ *if for each* $C \in \mathcal{M}$*, we have* $r_C(f) = e_C(f)$ *for every* $f \in O^C$*.*

The next result is Theorem 5.4 in Abramsky and Brandenburger (2011). (In their set-up, the measurement cover $\mathcal{M}$ is an anti-chain, but this is immaterial for this theorem.)

Theorem 24. *An empirical model satisfies no signaling if and only if it is realized by a phase-space signed probability measure.*

This result and Theorem 2 in Dogan and Yildiz (2023), used in the previous section to prove our Theorem 23, exhibit a clear similarity. Indeed, if we take no signaling to be a basic physical requirement, then Theorem 24 says that signed probabilities on phase space constitute the top of the quantum-inspired hierarchy, while Dogan and Yildiz (2023, Theorem 2) and our Theorem 23 say that signed mixtures constitute the top of the stochastic-choice hierarchy (in the case of a finite domain). Examination of the exact relationship here may prove fruitful.

8.4 Quantum Hierarchy

Abramsky and Brandenburger (2011) build a multi-level hierarchy for quantum mechanics (QM). The classical level requires that there is a phase-space probability measure that realizes an empirical model is unsigned (non-negative). Level 3 in the present paper is the analog. The Bell (1964) model of QM cannot be realized this way. The next level requires that there is an unsigned phase-space probability measure that is mutually absolutely continuous with respect to an empirical model. That is, for each $C \in \mathcal{M}$, we have $r_C(f) > 0$ if and only if $e_C(f) > 0$. Level 3a in the present paper is the analog. The Bell model can be realized at this level, but the Hardy model of QM (Hardy, 1993) cannot be. The next level requires that there is an unsigned phase-space probability measure that is only absolutely continuous with respect to an empirical model. That is, for each $C \in \mathcal{M}$, we have $e_C(f) = 0$ implies $r_C(f) = 0$. Level 3b in the present paper is the analog. The Hardy model can be realized at this level, but the GHZ model of QM (Greenberger, Horne, and Zeilinger, 1989) cannot be.

Finally, it is a theorem of QM that all empirical models that describe quantum systems are no signaling (e.g., Horodecki and Ramanathan, 2019). It follows – recall Theorem 24 above – that the quantum levels defined in Abramsky and Brandenburger (2011) all nest under the top level composed of empirical models realized by phase-space signed probability measures. This observation completes the analogy between the stochastic-choice and quantum hierarchies. (Recall the finiteness assumption in Theorem 23. Abramsky and Brandenburger, 2011 also make finiteness assumptions – specifically, that the measurement and outcome sets are finite.)

Contextuality emerges in exactly parallel fashion in the two hierarchies. For stochastic choice, we argued that contextual behavior first arises at Level 3a. In particular, we said that Level 3, even though it violates IIA, is non-contextual. For QM, contextuality is non-classicality and first arises at the level of the Bell model.

It is striking that, by articulating the inner structure of a measurement-outcome framework, the levels of our hierarchy for stochastic choice can be recapitulated. This is despite the fact that, ignoring inner structure, measurement-outcome frameworks are all Level 2. In fact, and as mentioned in the Introduction, the direction of work was the opposite – the quantum hierarchy in Abramsky and Brandenburger (2011) inspired the hierarchy we have built in the current paper.

Two papers that relate to this section are Zeng and Zahn (2015) and Chambers, Masatlioglu, and Turansick (2023). The first establishes a connection between the Abramsky-Brandenburger quantum hierarchy and a part of deterministic choice theory. The second studies SCR's on choice sets that are product spaces and proves an equivalence between a marginality condition, which is a type of separability in choice, and a signed-probability representation. While the two frameworks are different, there is an obvious resemblance between Theorem 5.4 in Abramsky and Brandenburger (2011) and this result.

References

Aaker, J., "The Negative Attraction Effect? A Study of the Attraction Effect Under Judgment and Choice," *Advances in Consumer Research* 18, 1991, 462-469.

Abramsky, S., and A. Brandenburger, "The Sheaf-Theoretic Structure of Non-Locality and Contextuality," *New Journal of Physics*, 13, 2011, 113036.

Agranov, M., and P. Ortoleva, "Stochastic Choice and Preferences for Randomization," *Journal of Political Economy*, 125, 2017, 40-68.

Aizerman, M., and A. Malishevski, "General Theory of Best Variants Choice: Some Aspects," *IEEE Transactions on Automatic Control* 26, 1981, 1030-1040.

Bandyopadhyay, T., I. Dasgupta, and P. Pattanaik, "Stochastic Revealed Preference and the Theory of Demand," *Journal of Economic Theory*, 84, 1999, 95-110.

Bandyopadhyay, T., I. Dasgupta, and P. Pattanaik, "Demand Aggregation and the Weak Axiom of Stochastic Revealed Preference," *Journal of Economic Theory*, 107, 2002, 483-489.

Battigalli, P., and M. Siniscalchi, "Hierarchies of Conditional Beliefs and Interactive Epistemology in Dynamic Games," *Journal of Economic Theory*, 88, 1999, 188-230.

Battigalli, P., and M. Siniscalchi, "Strong Belief and Forward Induction Reasoning," *Journal of Economic Theory* 106, 2002, 356-391.

Bell, J., "On the Einstein-Podolsky-Rosen Paradox," *Physics*, 1, 1964, 195-200.

Block, H., and J. Marschak, "Random Orderings and Stochastic Theories of Responses," in Olkin, I., S. Ghurye, W. Hoeffding, W. Madow, and H. Mann (eds.), *Contributions to Probability and Statistics*, Stanford University Press, 1960, 97-132.

Blume, L., A. Brandenburger, and E. Dekel, "Lexicographic Probabilities and Choice Under Uncertainty," *Econometrica*, 59, 1991, 61-79.

Brandenburger, A., P. Ghirardato, D. Pennesi, and L. Stanca, "Event Valence and Subjective Probability," unpublished, 2024, at adambrandenburger.com.

Cerreia-Vioglio, S., D. Dillenberger, P. Ortoleva, and G. Riella, "Deliberately Stochastic", *American Economic Review*, 109, 2019, 2425-2445.

Cerreia-Vioglio, S., P.O. Lindberg, F. Maccheroni, and M. Marinacci, "A Canon of Probabilistic Rationality," *Journal of Economic Theory*, 196, 2021, 105289.

Chambers, C., and F. Echenique, "Stochastic Choice," in Chambers, C., and F. Echenique (eds.), *Revealed Preference Theory*, Econometric Society Monograph Number 56, Cambridge University Press, 2016, Chapter 7, 95-112.

Chambers, C., Y. Masatlioglu, and C. Turansick, "Correlated Choice," *Theoretical Economics*, 2023, at https://econtheory.org/ojs/index.php/te/article/viewForthcomingFile/5184/37331/1.

Chernoff, H., "Rational Selection of Decision Functions," *Econometrica*, 22, 1954, 422-443.

Cohen, M., "Random Utility – The Infinite Case," *Journal of Mathematical Psychology*, 22, 1980, 1-23.

Cohen, M., and J. Falmagne, "Random Utility Representation of Binary Choice Probabilities: A New Class of Necessary Conditions," *Journal of Mathematical Psychology*, 34, 1990, 88-94.

Császár, Á., "Sur la Structure des Espaces de Probabiliteé Conditionnelle," *Acta Math. Academiae Scientiarum Hungarica*, 6, 1955, 337-361.

Dasgupta, I., and P. Pattanail, "'Regular' Choice and the Weak Axiom of Stochastic Revealed Preference," *Economic Theory*, 31, 2007, 35-50.

Davis-Stober, C., A. Marley, W. McCausland, and B. Turner, "An Illustrated Guide to Context Effects," *Journal of Mathematical Psychology*, 115, 2023, 102790.

Debreu, G., "Review of 'Individual Choice Behavior' by R.D. Luce," *American Economic Review*, 1960, 50, 186-188.

Dickson, W., *Quantum Chance and Non-Locality*, Cambridge University Press, 1998.

Dogan, S., and K. Yildiz, "Every Choice Function Is Pro-Con Rationalizable," *Operations Research*, 71, 2023, 1857-1870.

Falmagne, J., "A Representation Theorem for Finite Random Scale Systems," *Journal of Mathematical Psychology*, 18, 1978, 52-72.

Ford, L., and and D. Fulkerson, *Flows in Networks*, Princeton Landmarks in Mathematics and Physics, 56, Princeton University Press, 2015.

Greenberger, D., M. Horne, and A. Zeilinger, "Going Beyond Bell's Theorem," in Kafatos, M. (ed.), *Bell's Theorem, Quantum Theory and Conceptions of the Universe*, Kluwer, 1989, 69-72.

Hardy, L., "Nonlocality for Two Particles Without Inequalities for Almost All Entangled States," *Physical Review Letters*, 71, 1993, 1665-1668.

Horodecki, P., and R. Ramanathan, "The Relativistic Causality versus No-Signaling Paradigm for Multi-Party Correlations," *Nature Communications*, 10, 2019, 1701.

Howard, M., J. Wallman, V. Veitch, and J. Emerson, "Contextuality Supplies the 'Magic' for Quantum Computation," *Nature*, 520, 2014, 351-355.

Huber, J., J. Payne, and C. Puto, "Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis," *Journal of Consumer Research*, 9, 1982, 90-98.

Kochen, S., and E. Specker, "The Problem of Hidden Variables in Quantum Mechanics," *Journal of Mathematics and Mechanics*, 17, 1967, 59-87.

Kopylov, I., and E. Yang, "Revealed Incentives and Consideration Sets," unpublished, 2024.

Kruis, J., G. Maris, M. Marsman, M. Bolsinova, H. van der Maas, "Deviations of Rational Choice: An Integrative Explanation of the Endowment and Several Context Effects," *Scientific Reports*, 10, 2020, at https://doi.org/10.1038/s41598-020-73181-2.

Luce, R.D., *Individual Choice Behavior: A Theoretical Analysis*, Wiley, 1959.

Machina, M., "Dynamic Consistency and Non-Expected Utility Models of Choice under Uncertainty," *Journal of Economic Literature*, 27, 1989, 1622-1668.

Marschak, J., "Binary Choice Constraints on Random Utility Indicators, in Arrow (ed.), *Stanford Symposium on Mathematical Methods in the Social Sciences*, Stanford University Press, 1960, 312-329.

McFadden, D., "Conditional Logit Analysis of Qualitative Choice Behavior," in Zarembka, P. (ed.), *Frontiers in Econometrics*, Academic Press, 1974, 105-142.

McFadden, D., and M. Richter, "On the Extension of a Set Function to a Probability on the Boolean Algebra Generated by a Family of Events, with Applications," working Paper No. 14, Mathematical Social Science Board Workshop on the Theory of Markets under Uncertainty, 1971, Department of Economics, UC Berkeley, 1971.

Moulin, H., "Choice Functions over a Finite Set: A Summary," *Social Choice and Welfare*, 2, 1985, 147-160.

Myerson, R., "Axiomatic Foundations of Bayesian Decision Theory," discussion paper, 1986a.

Myerson, R., "Multistage Games with Communication," *Econometrica*, 54, 1986b, 323-358.

Platt, M., and P. Glimcher, "Neural Correlates of Decision Variables in Parietal Cortex," *Nature*, 400, 1999, 233-238.

Plott, C., "Path Independence, Rationality, and Social Choice," *Econometrica*, 41, 1973, 1075-1091.

Popescu, S., and D. Rohrlich, "Quantum Nonlocality as an Axiom," *Foundations of Physics*, 24, 1994, 379-385.

Rao, B., and B. Rao, *Theory of Charges: A Study of Finitely Additive Measures*, Academic Press, 1983.

Raussendorf, R., "Contextuality in Measurement-Based Quantum Computation," *Physical Review A*, 88, 2013, 022322.

Rényi, A., "On a New Axiomatic Theory of Probability," *Acta Math. Academiae Scientiarum Hungarica*, 6, 1955, 285-335.

Rényi, A., "On Conditional Probability Spaces Generated by a Dimensionally Ordered Set of Measures," *Teor. Veroyatnost. i Primenen.*, 1956, 1, 61-71.

Rényi, A., *Foundations of Probability*, Holden-Day, 1970a.

Rényi, A., *Probability Theory*, North-Holland, 1970b.

Saito, K., "Axiomatizations of the Mixed Logit Model," 2018, at https://saito.caltech.edu/documents/3687/revised_mixed_logit11.pdf.

Saito, K., "Lecture Note on the Theory of Random Utility Models," 2019, at https://saito.caltech.edu/documents/7827/Discrete_RU_note2_sample.pdf.

Selten, R., "Reexamination of the Perfectness Concept of Equilibrium Points in Extensive Games," *International Journal of Game Theory*, 4, 1975, 25-55.

Simonson, I., "Choice Based on Reasons: The Case of Attraction and Compromise Effects," *Journal of Consumer Research*, 16, 1989, 158-174.

Simonson, I., "Vices and Virtues of Misguided Replications: The Case of Asymmetric Dominance," *Journal of Marketing Research*, 51, 2014, 514-519.

Simonson, I., and A. Tversky, "Choice in Context: Tradeoff Contrast and Extremeness Aversion," *Journal of Marketing Research*, 29, 1992, 281-295.

Steverson, K., A. Brandenburger, and P. Glimcher, "Choice-Theoretic Foundations of the Divisive Normalization Model," *Journal of Economic Behavior & Organization*, 164, 2019, 148-165.

Strzalecki, T., *Stochastic Choice Theory*, 2022, at https://scholar.harvard.edu/sites/scholar.harvard.edu/files/tomasz/files/manuscript_01.pdf.

Tversky, A., "Elimination by Aspects: A Theory of Choice," *Psychological Review*, 79, 1972a, 288-299.

Tversky, A., "Choice by Elimination," *Journal of Mathematical Psychology*, 9, 1972b, 341-367.

Webb, R., "The (Neural) Dynamics of Stochastic Choice," *Management Science*, 65, 2019, 230-255.

Zeng, W., and P. Zahn, "Contextuality and the Weak Axiom in the Theory of Choice," November 2015, at https://arxiv.org/abs/1512.02668.