Entangled vs. Separable Choice*

Nail Kashaev ⓡ Martin Plávala ⓡ Victor H. Aguiar†

This version: July, 2024. First version: March, 2024

Abstract We investigate joint probabilistic choice rules describing the behavior of two decision makers, each facing potentially distinct menus. These rules are separable when they can be decomposed into individual choices correlated solely through their respective probabilistic choice rules. Despite its significant interest for the study of peer effects, influence, and taste variation, a complete characterization of these rules has remained elusive (Chambers, Masatlioglu, and Turansick, 2021). We fully characterize separable choices through a finite system of inequalities inspired by Afriat's theorem. Our results address the possibility of entangled choices, where decision makers behave as if they do not communicate, yet their choices are not separable. More generally, we establish that separable joint choice restrictions can be factored into individual choice restrictions if only if at least one decision maker's probabilistic choice rule uniquely identifies the distribution over deterministic choice rules. The no communication condition and the individual restrictions are no longer sufficient in the absence of this uniqueness. Our results offer robust tools for distinguishing between separable decision-making and behaviors beyond mere peer effects such as imitation and cheating.

JEL classification numbers: C10.

Keywords: correlated choice, stochastic choice, entanglement, partial identification.

---

*The paper subsumes some results from Kashaev, Aguiar, Plávala, and Gauthier (2023a). The "ⓡ" symbol indicates that the authors' names are in certified random order, as described by Ray and Robson (2018). We thank Roy Allen, Adam Brandenburger, Chris Chambers, Geoffroy de Clippel, David Freeman, Ricky Li, Toru Kitagawa, Kareen Rozen, Salvador Navarro, Krishna Pendakur, Roberto Serrano, Wilson Perez, Bruno Salcedo, John Quah, Erya Yang, and Lanny Zrill for useful discussions and encouragement. Aguiar and Kashaev gratefully acknowledge financial support from the Social Sciences and Humanities Research Council Insight Development Grant. Plávala acknowledges support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, project numbers 447948357 and 440958198), the Sino-German Center for Research Promotion (Project M-0294), the German Ministry of Education and Research (Project QuKuK, BMBF Grant No. 16KIS1618K), the DAAD, and the Alexander von Humboldt Foundation. Aguiar thanks USFQ School of Economics for kindly hosting him while writing this paper.

†Kashaev: Department of Economics, University of Western Ontario; nkashaev@uwo.ca. Plávala: Naturwissenschaftlich-Technische Fakultät, Universität Siegen; Institut für Theoretische Physik, Leibniz Universität Hannover, Hannover, Germany; martin.plavala@itp.uni-hannover.de. Aguiar: Department of Economics, Simon Fraser University; vaguiarl@sfu.ca.

1. Introduction

We study the behavior of decision makers (DMs) that are summarized by *separable* joint (probabilistic) choice rules. These separable choice rules describe the joint stochastic behavior of two DMs whose choices can be correlated only via their individual choice rules. That is, the DMs could know each other but make choices as if they are in separate rooms with no communication possible: each DM chooses according to her individual choice function from her individual menu.

Separable joint choice rules allow simple modeling of the joint behavior of several DMs as a straightforward interaction of individual behaviors, making these joint behaviors compositional and tractable (Kashaev et al., 2023a). They appear naturally in the study of peer effects (Sacerdote, 2011), twins educational choice data (Ashenfelter and Krueger, 1994, Miller, Mulvey, and Martin, 1995), influence in choice (Chambers, Cuhadaroglu, and Masatlioglu, 2023, Kashaev, Lazzati, and Xiao, 2023b), and dynamic taste variation, where each DM is interpreted as a time period (Frick, Iijima, and Strzalecki, 2019, Cherchye, De Rock, Griffith, O'Connell, Smith, and Vermeulen, 2017, Kashaev et al., 2023a). In addition, since separability requires DMs to choose autonomously, the lack of it may be classified as cheating, imitation, influence, or collusive behavior. Hence, characterizing separable behavior may be used to test for these non-autonomous factors.

We define entanglement of choice of two DMs and showcase why it poses conceptual and computational challenges to characterizing separable choice rules. In particular, all possible interactions of the individual choice rule restrictions imply the restrictions on the joint choice rule. We call these restrictions *separable*. One might expect these restrictions to characterize separable choice rules since there is no communication. However, this is not always the case. When the individual probabilistic choice rule, which is a mixture of deterministic choice rules, does not uniquely identify the distribution over individual deterministic choice rules, the separable restrictions fail to be sufficient, thus, leading to entanglement of choice. Entangled choices are joint choice rules that are not separable yet are consistent with the separable restrictions. Moreover, we show that the separable restrictions are necessary and sufficient to characterize separable choice if and only if the uniqueness property holds.

This result can be used in general finite environments to obtain the separable restrictions and full characterizations under the uniqueness restriction. We generalize recent results in stochastic choice with random utility by Chambers et al. (2021) and Li (2021), solving an open question posed in Strzalecki (2021) for the case of uniqueness. We also justify the importance of uniqueness (point identification) when working with separable rules: lack of uniqueness leads to the possibility of entangled choice that we argue is an undesirable property of models. A similar phenomenon appears in the analysis of finite games with multiple equilibria. The multiplicity of equilibria, similar to the lack of uniqueness in our setting, often leads to a correlation between the choices of players even after conditioning on available information (De Paula and Tang, 2012).

The solution for a general non-unique case is unknown and is hard to obtain: testing whether a system is separable or entangled is an NP-hard problem (Gurvits, 2003). Using Bell inequalities (Bell, 1964, Rosset, Bancal, and Gisin, 2014), we provide a characterization of separable joint probabilistic rules for a simple scenario solving an open question posed by Chambers et al. (2021).

We also provide a general characterization of separable rules via linear restrictions analogous to the Afriat (1967)'s characterization of static utility maximization. This answers the open problem posed in Chambers et al. (2021) asking for a finite characterization of separable choice rules. These restrictions can be interpreted as requiring that we can extend the separable joint probabilistic rule of two DMs to multiple virtual DMs. The extended joint probabilistic rule must rule out the communication between the virtual DMs and agree with the original choice rule when marginalized to the original two DMs. Our no communication condition requires that the random choices of one of the DMs conditional on the menu she faces does not depend on the menu faced by the other DM. We show that such an extension exists for any number of virtual DMs if and only if the joint stochastic rule is separable. Crucially, we show that it is enough to test a finite number of extensions in practice. Entangled choices pass a test of *no communication*– the key separable restriction but they cannot be extended in the sense described above preserving no communication.

There is also a connection with the testability of causal models with instrumental variables (Pearl, 1995, Gunsilius, 2021) that use Bell-type inequalities. In our case, the no communication condition is similar to an exclusion restriction where the menu of the one DM is irrelevant to the choices of the other DM conditional on the latter menu, even when an observer will see correlation between choices between them. However, our domain is different and we provide not just testable restrictions but a full characterization of our problem.

2. A Difficulty with Separable Choice: An Example of Entangled Choice

One DM. Consider a hypothetical experimental setting with four distinct alternatives: x, y, w, and z. A DM, Frodo ($t = 1$), has to pick a single alternative when presented with any of the two menus $\{xw\}$ and $\{yz\}$. The columns of matrix A^t below encode all possible deterministic choice rules or choice patterns of Frodo. The rows of A^t encode all possible pairs of choices and menus that could be observed. The entries of A^t are either 0 or 1. For example, the entry in the first row $(x, \{xw\})$ and the first column is equal to 1 and is interpreted as the deterministic choice rule that picks x from menu $\{xw\}$. Because choice rules are single-valued, the subvector of each column corresponding to the same menu must add up to 1.

$$A^t = \begin{array}{c} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \end{array} \begin{array}{||c} \text{x,\{xw\}} \\ \text{w,\{xw\}} \\ \text{y,\{yz\}} \\ \text{z,\{yz\}} \end{array} .$$

Frodo can exhibit stochastic behavior, in particular, consider a distribution over the columns on A^t, ν_t, such that the probabilistic choice rule of DM t, ρ_t, is given by

$$\rho_t = A^t \nu_t.$$

The probabilistic choice rule ρ_t is such that all its entries are nonnegative and are such that

$$\rho_t(x, \{xw\}) + \rho_t(w, \{xw\}) = \rho_t(y, \{yz\}) + \rho_t(z, \{yz\}) = 1.$$

We highlight that other than these nonnegativity and adding-up constraints, there are no more restrictions on ρ_t because Frodo randomizes over all possible deterministic choice rules in this setup. In other words, any probabilistic choice rule can be represented as a mixture of all deterministic choice rules.

Formalizing, let the grand choice set be a finite and nonempty set X^t. Let $\mathcal{X}^t$ be a nonempty collection of nonempty subsets of X^t, C^t be the set of all choice functions on $\mathcal{X}^t$. The probabilistic choice rule ρ_t is consistent with the stochastic choice model if there exists a probability measure over C^t, μ_t, such that

$$\rho_t(m, M) = \int \left[\mathbb{1} \left(c^t(M) = m \right) \right] d\mu_t(c^t)$$

for all menus $M \in \mathcal{X}^t$ and all $m \in M$.

Two DMs. Next, we consider two DMs. Frodo ($t = 1$) and Sam ($t = 2$) make choices from binary menus. The choices are separated because Frodo and Sam make their choices as if they are in separate isolated rooms with no communication before or during the experiment. However, Frodo and Sam grew up in the same village and are friends. Note that since rooms are isolated and no communication devices are permitted, Frodo and Sam cannot coordinate choices after the start of the experiment, nor can they see what menu the other DM is seeing. However, since Frodo and Sam are friends, their randomization devices over deterministic choice functions can be arbitrarily correlated.

Frodo and Sam face, in each trial, ordered pairs of menus or menu paths (e.g., $(\{xw\}, \{yz\}))$). The first menu will be available to Frodo and the second to Sam. We assume that all possible combinations of menus are presented to Frodo and Sam in sequential trials such that there are 4 menu paths. In each menu path, say $(\{xw\}, \{yz\})$, the experimenter observes the probability of each of the 4 choice paths such that

$$\rho(x, \{xw\}; y, \{yz\}) + \rho(x, \{xw\}; z, \{yz\}) + \rho(w, \{xw\}; y, \{yz\}) + \rho(w, \{xw\}; z, \{yz\}) = 1$$

Assuming that there are no restrictions on Sam's individual behavior (i.e., $A^2 = A^1$), we can encode all possible joint choice patterns of Frodo and Sam when faced with a menu path in the columns of matrix A below. For example, column 1 combines column 1 of A^1 and column 1 of A^2, each describing the individual choice rules. The 16 rows of A correspond to all choice and menu paths. For example, the entry of the first column and the first row equals 1. This means that both DMs pick the same alternative from the same menu because the choice rules that describe their deterministic behavior in this column are the same. Column 2 combines column 1 of A^1 and column 2 of A^2. Hence, its second entry equals 1 because Sam chooses y when faced with $\{yz\}$. In total, there are 16 possible deterministic choice patterns for the joint problem faced by the DMs.

$$A = \left(\begin{array}{cccccccccccccccc}
1&0&1&0 & &&& & 1&0&1&0 & &&& \\
0&1&0&1 & &\mathbf{0}&& & 0&1&0&1 & &\mathbf{0}&& \\
1&1&0&0 & &&& & 1&1&0&0 & &&& \\
0&0&1&1 & &&& & 0&0&1&1 & &&& \\
&&& & 1&0&1&0 & &&& & 1&0&1&0 \\
&\mathbf{0}&& & 0&1&0&1 & &\mathbf{0}&& & 0&1&0&1 \\
&&& & 1&1&0&0 & &&& & 1&1&0&0 \\
&&& & 0&0&1&1 & &&& & 0&0&1&1 \\
1&0&1&0 & 1&0&1&0 & &&& & &&& \\
0&1&0&1 & 0&1&0&1 & &\mathbf{0}&& & &\mathbf{0}&& \\
1&1&0&0 & 1&1&0&0 & &&& & &&& \\
0&0&1&1 & 0&0&1&1 & &&& & &&& \\
&&& & &&& & 1&0&1&0 & 1&0&1&0 \\
&\mathbf{0}&& & &\mathbf{0}&& & 0&1&0&1 & 0&1&0&1 \\
&&& & &&& & 1&1&0&0 & 1&1&0&0 \\
&&& & &&& & 0&0&1&1 & 0&0&1&1
\end{array}\right)\left\|\begin{array}{c}
x, \{xw\}, x, \{xw\} \\
x, \{xw\}, w, \{xw\} \\
x, \{xw\}, y, \{yz\} \\
x, \{xw\}, z, \{yz\} \\
w, \{xw\}, x, \{xw\} \\
w, \{xw\}, w, \{xw\} \\
w, \{xw\}, y, \{yz\} \\
w, \{xw\}, z, \{yz\} \\
y, \{yz\}, x, \{xw\} \\
y, \{yz\}, w, \{xw\} \\
y, \{yz\}, y, \{yz\} \\
y, \{yz\}, z, \{yz\} \\
z, \{yz\}, x, \{xw\} \\
z, \{yz\}, w, \{xw\} \\
z, \{yz\}, y, \{yz\} \\
z, \{yz\}, z, \{yz\}
\end{array}\right.$$

Note that the vector that collects the choice probabilities over all 4 possible choice paths, ρ, is given (up to permutation) by

$$\rho = A\nu \tag{1}$$

for some probability distribution over the columns of A, ν.

Equation (1) captures the structure of the thought experiment in matrix form for its simple domain. Next, we describe the thought experiment for the general case. Let $\mathcal{C} = C^1 \times C^2$ and $c = (c^1, c^2)$ be an element of $\mathcal{C}$.

Definition 1. The joint probabilistic choice rule ρ is consistent with the thought experiment if there exists a probability measure over $\mathcal{C}$, μ, such that

$$\rho(m_1, M_1; m_2, M_2) = \int \left[\mathbb{1}\left(c^1(M_1) = m_1\right) \mathbb{1}\left(c^2(M_2) = m_2\right)\right] d\mu(c)$$

for all menu paths (M_1, M_2) and all $m_1 \in M_1 \in \mathcal{X}^1$ and $m_2 \in M_2 \in \mathcal{X}^2$.

Under the thought experiment, ρ satisfies nonnegativity, adding-up constraints for each menu path, and a condition that we call *marginality*:

$$\rho(x,\{xw\};x,\{xw\}) + \rho(x,\{xw\};w,\{xw\}) = \rho(x,\{xw\};y,\{yz\}) + \rho(x,\{xw\};z,\{yz\})$$

and

$$\rho(x,\{xw\};x,\{xw\}) + \rho(w,\{xw\};x,\{xw\}) = \rho(y,\{yz\});x,\{xw\}) + \rho(z,\{yz\};x,\{xw\}).$$

In general, ρ satisfies marginality if

$$\sum_{m_t \in M_t} \rho(m_1, M_1; m_2, M_2)$$

does not depend on M_t for all m_{3-t}, M_{3-t}, and t.

Marginality follows from Frodo and Sam being in separate rooms and the fact that they randomize over the deterministic choice patterns using a distribution that does not depend on the particular menu path. This observation has been done before by Strzalecki (2021) and echoed in Chambers et al. (2021). In addition, this condition is known as *no signaling* or no communication in the theoretical physics literature for systems that are not related to choice (Rosset et al., 2014).

Since the probabilistic choice generated by Frodo by randomizing over his deterministic choice rules does not impose any additional restrictions on individual ρ_t, the observer has a reasonable expectation that there are no more restrictions on joint ρ beyond nonnegativity, adding-up, and marginality constraints. However, as we show in the next section, it is possible to construct ρ that satisfies all these separable restrictions, yet our thought experiment can not generate it. That is, the thought experiment generates new *emerging* restrictions on ρ. We call the discrepancy between the separable behavioral implications and the actual implications of the thought experiment *entanglement of choice*. We say that ρ is entangled if it satisfies the separable restrictions yet fails to be generated by the thought experiment.

2.1. A Characterization of the Thought Experiment

We use Bell's inequalities (Bell, 1964) to provide the emerging restrictions generated by the thought experiment. In particular, we use the CHSH (Clauser, Horne, Shimony, and Holt) inequalities (Rosset et al., 2014). Define 4 measures of choice coordination for 4 different menu paths as

$$
\begin{aligned}
E_{\{xw\},\{xw\}} &= \rho(x,\{xw\};x,\{xw\}) + \rho(w,\{xw\};w,\{xw\}) - \rho(x,\{xw\};w,\{xw\}) - \rho(w,\{xw\};x,\{xw\}),\\
E_{\{xw\},\{yz\}} &= \rho(x,\{xw\};y,\{yz\}) + \rho(w,\{xw\};z,\{yz\}) - \rho(x,\{xw\};z,\{yz\}) - \rho(w,\{xw\};y,\{yz\}),\\
E_{\{yz\},\{xw\}} &= \rho(y,\{yz\};x,\{xw\}) + \rho(z,\{yz\};w,\{xw\}) - \rho(z,\{yz\};x,\{xw\}) - \rho(y,\{yz\};w,\{xw\}),\\
E_{\{yz\},\{yz\}} &= \rho(y,\{yz\};y,\{yz\}) + \rho(z,\{yz\};z,\{yz\}) - \rho(y,\{yz\};z,\{yz\}) - \rho(z,\{yz\};y,\{yz\}).
\end{aligned}
$$

Note that $E_{\{xw\},\{xw\}} = 1$ if in menu path $(\{xw\},\{xw\})$ Frodo and Sam coordinate on picking the same alternative. At the same time, $E_{\{xw\},\{xw\}} = -1$ if DMs always pick different alternatives. Also, $\left|E_{\{xw\},\{xw\}}\right| \leq 1$ by construction. Hence, we can interpret E_{M_1,M_2} as a measure of cooperation between Frodo and Sam in menu path (M_1, M_2).

Now, we can define the following CHSH inequalities:

$$
\begin{aligned}
-2 &\leq E_{\{xw\},\{xw\}} + E_{\{yz\},\{xw\}} + E_{\{xw\},\{yz\}} - E_{\{yz\},\{yz\}} \leq 2,\\
-2 &\leq E_{\{xw\},\{xw\}} + E_{\{yz\},\{xw\}} - E_{\{xw\},\{yz\}} + E_{\{yz\},\{yz\}} \leq 2,\\
-2 &\leq E_{\{xw\},\{xw\}} - E_{\{yz\},\{xw\}} + E_{\{xw\},\{yz\}} + E_{\{yz\},\{yz\}} \leq 2,\\
-2 &\leq -E_{\{xw\},\{xw\}} + E_{\{yz\},\{xw\}} + E_{\{xw\},\{yz\}} + E_{\{yz\},\{yz\}} \leq 2.
\end{aligned}
$$

The CHSH inequalities provide lower and upper bounds on the cooperation between Frodo and Sam across *all* menu paths in the thought experiment. In particular, at least one of these inequalities is violated if Frodo and Sam fully cooperate in 3 out of 4 menu paths but defect in one. For example, the second inequality is violated if $E_{\{xw\},\{xw\}} = E_{\{yz\},\{xw\}} = E_{\{yz\},\{yz\}} = 1$ and $E_{\{xw\},\{yz\}} = -1$. This kind of coordination between Frodo and Sam is consistent with no communication/marginality but is not consistent with separability. Sam and Frodo may be cheating and in fact communicating, but we show it is possible to fake no communication by switching their behavior to anti-coordination to fool the experimenter. Yet the Bell inequalities catch the excess correlation in their choices detecting no separability.

Proposition 1. Probabilistic choice rule ρ is consistent with the thought experiment if and only if it satisfies marginality and the CHSH inequalities.

The sufficiency of Proposition 1 follows from Fine (1982) (Proposition 2). Necessity is trivial to verify in this case. The CHSH inequalities are not implied by the interaction of the individual behavioral restrictions, and thus are emerging. In particular, they are distinct from adding-up, nonnegativity, and marginality.

Table 1 depicts a continuum of nondegenerate examples of ρ that satisfy marginality, but do not obey the CHSH inequalities. In particular, $E_{\{xw\},\{xw\}} = E_{\{yz\},\{xw\}} = E_{\{xw\},\{yz\}} = 2(\alpha - \beta)$ and $E_{\{yz\},\{yz\}} = 2(\beta - \alpha)$. Hence, the first CHSH inequality is violated if and only if

$$E_{\{xw\},\{xw\}} + E_{\{yz\},\{xw\}} + E_{\{xw\},\{yz\}} - E_{\{yz\},\{yz\}} = 8(\alpha - \beta) > 2.$$

In other words, for $\frac{3}{8} < \alpha \leq \frac{1}{2}$ and $\beta = \frac{1}{2} - \alpha$ the implied by Table 1 ρ satisfies marginality but cannot be explained by the thought experiment. The violation of CHSH is maximal when $\alpha = 1/2$ and $\beta = 0$. That particular configuration was documented in Chambers et al. (2021).

When ρ violates the CHSH inequalities, we say ρ is entangled since the thought experiment cannot explain it, and there must be some form of unobserved communication between Frodo and Sam. The CHSH inequalities in our domain correspond to nonparametric analogues of the excess variance approach to detecting imitation in the peer effects literature (Sacerdote, 2011). In other words, they provide a threshold on correlation of Sam's and Frodo's choices above which choice is not separable and becomes entangled.

Frodo/Sam	$x, \{xw\}$	$w, \{xw\}$	$y, \{yz\}$	$z, \{yz\}$
$x, \{xw\}$	α	β	α	β
$w, \{xw\}$	β	α	β	α
$y, \{yz\}$	α	β	β	α
$z, \{yz\}$	β	α	α	β

Table 1 – Rows correspond to choices and menus for Frodo and columns correspond to choices and menus for Sam. Entries are the joint probabilities over the choice path formed by the corresponding row and column. The parameters are such that $\alpha + \beta = 1/2$ and $\frac{3}{8} < \alpha \leq \frac{1}{2}$.

Next, we describe a setting in which entangled choice arises. Consider the example in Table 1 corresponding to $\alpha = 1/2$ and $\beta = 0$. Sam and Frodo ate the big and the small cakes at a birthday celebration they had not been invited to. Eating a big cake is a severe crime in Hobbiton, while eating a small cake is a misdemeanor. After being caught, they can only be charged for one crime at a time. When either Frodo or Sam are facing the charge of eating the big cake they face the menu of actions $\{xw\}$: plead innocent (x) or plead guilty (w). When either Frodo or Sam are facing the charge of eating the small cake, the analogous actions $\{yz\}$ with y meaning to plead innocent. With equal probabilities, they face either a soft or strict judge. They both get acquitted if they coordinate pleading innocent when the judge is soft. Similarly, they get a guilty plea deal when the judge is strict if they coordinate on pleading guilty. Frodo and Sam cannot communicate, but they can pay a lawyer to reveal to them the type of the judge they will face. Frodo is willing to pay the lawyer enough only if he faces the charge of eating the big cake. The same is true for Sam. When the lawyer gets paid enough by at least one of the them, she reveals the type of the judge to both of them truthfully. This means that they coordinate pleading at the same time innocent whenever the judge is soft and they both get acquitted. They plead guilty and get a guilty plea deal when the judge is strict. When they both face the charge of eating the small cake, they do not pay and the lawyer makes them anti-coordinate, maximizing their penalties. In this example, the joint behavior of Sam and Frodo is not separable. Entanglement arises because the randomization over actions is menu-dependent because of the information provided by a hidden agent, the lawyer.

Next, we explain this unexpected behavior in the thought experiment and provide a necessary and sufficient condition for it. The condition is based on recent results from mathematical quantum physics that have studied a similar mathematical structure in other domains (Aubrun, Lami, Palazuelos, and Plávala, 2021).

3. Separable Restrictions over Separable Choice

Recall that the thought experiment produces a probabilistic choice rule that satisfies $\rho = A\nu$ for some distribution over columns of A. Every column of A corresponds to a composite type that describes the behavior of each DM. That is, separability in the thought experiment is captured by interactions of different individual types. Formally,

$$A = A^1 \otimes A^2,$$

where $\otimes$ is the Kronecker product of matrices.[1]

When we look at the individual behavioral restrictions implied by $\rho_t = A^t \nu$, we find that the restrictions on ρ_t can be captured in the form of inequality constraints $H^t \rho_t \geq 0$, where

$$H^t = \begin{pmatrix} -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

That is, $\rho_t = A^t \nu$ for some distribution ν if and only if $H^t \rho_t \geq 0$. Rows of H^t capture the restrictions over ρ_t we discussed previously. In particular, rows 1 and 2 combine to provide the adding-up constraint. Rows 3-6 are the nonnegativity restrictions.

Since the behavior in the thought experiment is generated from the interaction of individual types, we can think of separable restrictions over ρ as those coming from the interaction of the individual restrictions captured by H^t. In other words, the correlation of choice cannot obscure the requirement that each DM chooses from its own menu that is separated from the other DM and that the joint randomization does not depend on the particular menu path. Formally, we define the interaction of such restrictions by

$$(H^1 \otimes H^2)\rho \geq 0,$$

where H^1 and H^2 are the individual restrictions for DM 1 and 2, respectively.

Rows of H^t are restrictions over the behavior captured by ρ_t. The Kronecker product of H^1 and H^2

[1] If C is an m-by-n matrix and D is a p-by-q matrix, then the Kronecker product $C \otimes D$ is the pm-by-qn block matrix:

$$C \otimes D = \begin{pmatrix} C_{1,1}D & \dots & C_{1,n}D \\ \vdots & \ddots & \vdots \\ C_{m,1}D & \dots & C_{m,n}D \end{pmatrix}.$$

requires ρ to satisfy the combinations of these restrictions: every restriction of Frodo is combined with every restriction of Sam. These are new restrictions on joint behavior. Importantly, all separable restrictions on joint behavior must arise from the restrictions on individual behavior.

Direct computation of joint restrictions produces adding-up, nonnegativity, and marginality constraints in our thought experiment. In particular, marginality arises from the interactions of the individual adding-up constraints. In this formal sense, we label marginality as a separable restriction over the behavior implied by the thought experiment. Using this formalization, we can now say that ρ in Table 1 satisfies the separable restrictions yet it fails to be consistent with the behavior implied by the thought experiment.

Remark 1. In the appendix, we show that separable restrictions are always necessary. They are always implied by the thought experiment and any extension of it for multiple agents and richer finite domains.

4. A Necessary and Sufficient Condition for (Lack of) Entanglement of Choice

Now we consider a slightly more general thought experiment for Frodo and Sam with matrices $A^{1,\Diamond}$ and A^2 describing their behavior. Matrix $A^{1,\Diamond}$ is constructed from columns of A^1. It collects all the *allowable* deterministic choice rules for Frodo in our setup. The experimenter controls the allowable behavior. For example, the experimenter can introduce a dominance relation among alternatives to restrict the behavior of Frodo.

We restrict $A^{1,\Diamond}$ to be *generating*. That is, we require the system of equations $A^{1,\Diamond}\nu = \rho_1$ to have a solution (possibly with negative entries) for every probabilistic choice rule ρ_1. In other words, all signed measures over the columns of $A^{1,\Diamond}$ should be able to generate all possible probabilistic choice rules ρ_1. In the thought experiment, $A^{1,\Diamond}$ is generating if and only if it consists of at least 3 different columns of A^1.

Since $A^{1,\Diamond}$ is a submatrix of A^1, Frodo's individual probabilistic choice is (weakly) more restricted. We collect the restrictions of the behavior of DM 1 in matrix $H^{1,\Diamond}$ such that $H^{1,\Diamond}\rho_1 \geq 0$ if and only if $\rho_1 = A^{1,\Diamond}\nu_1$ for some distribution over columns of $A^{1\Diamond}$, ν_1. The *thought experiment*$^\Diamond$ is defined completely analogous to the thought experiment but with A^1 replaced by $A^{1,\Diamond}$.

We also need two more definitions. We say that $A^{1,\Diamond}$ generates a *unique representation* when the system $\rho_1 = A^{1,\Diamond}\nu_1$ has a unique solution for all probabilistic choice rules ρ_1. This is a restriction on the columns of $A^{1,\Diamond}$. In particular, it is satisfied if and only if the columns of $A^{1,\Diamond}$ are linearly independent (i.e., $A^{1,\Diamond}$ has full column rank). For a generating $A^{1,\Diamond}$ the latter happens if and only if $A^{1,\Diamond}$ has exactly 3 out of 4 columns of A^1. That is, A^1 is generating but does not generate a unique representation.

Finally, we say that the thought experiment$^\Diamond$ produces *only separable restrictions* on ρ if

$$\rho = \left(A^{1,\Diamond} \otimes A^2\right)\nu \text{ for some distribution } \nu \iff (H^{1,\Diamond} \otimes H^2)\rho \geq 0.$$

In other words, the thought experiment$^\Diamond$ produces only separable restrictions if it cannot generate an entangled ρ. As we demonstrated in the previous section, when $A^{1,\Diamond} = A^1$, the thought experiment$^\Diamond$ does not produce only separable restrictions. The next theorem provides a necessary and sufficient condition for entanglement of choice.

Theorem 1. *Thought experiment$^\Diamond$ produces only separable restrictions on ρ if and only if $A^{1,\Diamond}$ generates a unique representation.*

In the appendix, using the results in Aubrun et al. (2021), we generalize Theorem 1 to all finite choice sets, arbitrary menu structures, and any finite number of DMs.

Entanglement of choice happens in the thought experiment$^\Diamond$ with $A^{1,\Diamond} = A^1$ because of the lack of uniqueness. The latter leads to emerging restrictions in behavior that allow for the existence of ρ that satisfies the separable restrictions on joint behavior yet cannot come from separable choice.

Our results can be applied to other menu structures or other restrictions on $A^{1,\Diamond}$ (e.g., random utility and beyond), as long as uniqueness is guaranteed for one of the DMs. For example, in the literature of mathematical psychology, researchers have studied the random interval and random semiorder models of stochastic choice introduced in Davis-Stober, Doignon, Fiorini, Glineur, and Regenwetter (2018) that generalize random utility. These models have conditions guaranteeing uniqueness (Doignon and Saito, 2023).

Nonuniqness of representation means that the underlying distribution over deterministic choice rules (i.e., columns of $A^{1,\Diamond}$) is partially- or set-identified (see, for instance, Kline and Tamer, 2023). When these distributions for both DMs are set-identified, then DMs may potentially coordinate on the individual distributions they can use. This coordination goes beyond correlation in choice. As a result, partial identification leads to behaviors that are not consistent with the thought experiment. However, when the distribution over deterministic choice rules of at least one DM is point identified, that DM will stick to this distribution, and there is no reason to coordinate with the other DM.

Chambers et al. (2021) provides a characterization of a version of our thought experiment with the random utility model (Block and Marschak, 1960, Falmagne, 1978, McFadden and Richter, 1990) restrictions on $A^{1,\Diamond}$ and A^2. They, however, only show the sufficiency of the unique representation property of $A^{1,\Diamond}$. Theorem 1 shows that in their setting, uniqueness is necessary while the random utility restriction is not. Li (2021) considers a setting similar to Chambers et al. (2021) but with finitely many DMs and at most three alternatives. As a result, the unique representation property of $A^{1,\Diamond}$ is satisfied. Because of that, the axiomatization in Li (2021) corresponds to the separable conditions generated by the interactions of the individual conditions. Our main result is a generalization of the characterizations in Chambers et al. (2021) and Li (2021). Marginality is equivalent to a version of the separable choice model with negative probabilities (this result was shown for a general setup by Abramsky and Brandenburger, 2011 and for stochastic choice by Chambers et al., 2021).

4.1. An Example of Necessary and Sufficient Separable Restrictions for Separable Choice

Consider the matrix of allowable behavior of Frodo:

$$A^{1,\Diamond} = \left(\begin{array}{ccc} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{array}\right) \begin{array}{||c} \text{x,\{xw\}} \\ \text{w,\{xw\}} \\ \text{y,\{yz\}} \\ \text{z,\{yz\}} \end{array}.$$

In this thought experiment$^{\Diamond}$, a dominance restriction is induced by ruling out the case when y is objectively better than z and w is objectively better than x for Frodo. For example, this may happen if, when faced with all 4 alternatives, Frodo never picks w or y. It is direct to verify that the matrix above is generating. Crucially, $A^{1,\Diamond}$ generates a unique representation. For simplicity, we let the behavior of Sam be explained by the same matrix as in the thought experiment. The restrictions on the behavior of Frodo are given by

$$H^{1,\Diamond} = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & -1 & 0 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Sam's behavior is the same as in the original thought experiment with behavior captured by H^2. Note that compared to the original thought experiment, $H^{1,\Diamond}$ contains additional rows (rows 1 and 2). These rows correspond to two monotonicity restrictions: $\rho_1(x, \{xw\}) \geq \rho_1(y, \{yz\})$ and $\rho_1(z, \{yz\}) \geq \rho_1(w, \{xw\})$. They appear because of the dominance relation we imposed by removing one of the columns from A^1. Interactions of individual monotonicity restrictions with the nonnegativity restrictions in H^2 lead to monotonicity restrictions on ρ:

$$\rho\,(x, \{xw\}; m_2, M_2) - \rho\,(y, \{yz\}; m_2, M_2) \geq 0,$$
$$\rho\,(z, \{yz\}; m_2, M_2) - \rho\,(w, \{xw\}; m_2, M_2) \geq 0,$$

for all M_2 and $m_2 \in M_2$. In this example, the above monotonicity restrictions are not the only separable restrictions beyond marginality. However, these extra restrictions are implied by monotonicity and marginality (i.e., they are redundant).

Proposition 2. The probabilistic choice rule ρ is generated by the above thought experiment$^\Diamond$ if and only if ρ satisfies marginality and monotonicity.

This proposition is a corollary of Theorem 1 and its proof is omitted for brevity. This proposition demonstrates that marginality and monotonicity are separable restrictions and are necessary and sufficient to describe the joint behavior of Sam and Frodo. This happens because Frodo's stochastic behavior is unique at the individual level. We highlight that in this case, entanglement of choice does not happen because any ρ that satisfies the separable restrictions can be generated by the thought experiment$^\lozenge$. Note that the CHSH inequalities for the thought experiment$^\lozenge$ are valid. However, they are implied by marginality and monotonicity.

5. A General Characterization of Separable Choice

Uniqueness is the key condition for the separable restrictions to fully characterize the thought experiment. In this section, we characterize the unrestricted (nonunique) thought experiment via Afriat (1967)'s like inequalities (i.e. a finite system of linear inequalities).

Our characterization relies on the fact that separable choice can be extended to a counterfactual setup where Sam choice experiment is replicated $k \geq 1$ times. By replication we mean that the collection of choice sets faced by the additional $k+1$ DMs coincide with the one faced by the original Sam. Formally, for $k \geq 1$, we consider a situation when the DMs face a menu path $M^{\text{ext},k} = (M_1, M_2, \dots, M_{k+1})$ such that $M_t \in \mathcal{X}^t = \{\{xw\}, \{yz\}\}$, $t = 1, \dots, k+1$. Similarly to ρ, we can define the *extended* probabilistic choice rule $\rho^{\text{ext},k}$ as a collection of joint probability distribution on extended menu paths. Note that when $k = 1$, we return to our original setting with Frodo and Sam. That is, $\rho^{\text{ext},1} = \rho$.

Next, we define the notion of marginality for extended probabilistic choice rules. (When $k = 1$ this definition coincides with the one given in Section 2.)

Definition 2 (Marginality)**.** We say that $\rho^{\text{ext},k}$, $k \geq 1$, satisfies *marginality* if its marginal distributions do not depend on the menu it was summed over. That is, for all $t = 1, \dots, k+1$, menu path $M^{\text{ext},k}$, and choices in it $(x_t)_{t=1\dots,k+1}$,

$$\sum_{x_t \in M_t} \rho^{\text{ext},k}(x_1, M_1; \dots; x_{k+1}, M_{k+1})$$

does not depend on M_t.

For any extended probabilistic choice rule $\rho^{\text{ext},k}$ of size $k \geq 1$ and for any $2 \leq j \leq k+1$ define the virtual joint probabilistic rule for any pair of menus (M_1, M_j) and any $x_1 \in M_1$ and $x_j \in M_j$ as

$$\rho_j^{\text{ext},k}(x_1, M_1; x_j, M_j) = \sum_{t \in \{2,\dots,k+1\} \setminus \{j\}} \frac{1}{|\mathcal{X}^t|} \sum_{M_t \in \mathcal{X}^t} \sum_{x_t \in M_t} \rho^{\text{ext},k}\left(x_1, M_1; \dots; x_{k+1}, M_{k+1}\right).$$

The virtual rule marginalizes Frodo $(t = 1)$ and the j-th replica of Sam $(t = j)$. When $\rho^{\text{ext},k}$ satisfies marginality, then the marginal distributions are menu-independent and

$$\rho_j^{\text{ext},k}(x_1, M_1; x_j, M_j) = \sum_{t \in \{2,\dots,k+1\} \setminus \{j\}} \sum_{x_t \in M^t} \rho^{\text{ext},k}\left(x_1, M_1; \dots; x_{k+1}, M_{k+1}\right).$$

Hence, marginality ensures that $\rho_j^{\text{ext},k}(\cdot, M_1; \cdot, M_j)$ is a well-defined distribution over $M_1 \times M_j$. Notice that we have k possible virtual joint probabilistic rules. Define also the average of these virtual rules as $\rho^{\text{v},k}$, where

$$\rho^{\text{v},k}(x_1, M_1; x_2, M_2) = \frac{1}{k} \sum_{j=2}^{k+1} \rho_j^{\text{ext},k}(x_1, M_1; x_j, M_j).$$

Next, we introduce two notions that provide a characterization of the thought experiment.

Definition 3. The stochastic choice rule ρ is k-marginalizable and k-marginalizable on average if there exists its marginalizable extension $\rho^{\text{ext},k}$ such that $\rho = \rho_j^{\text{ext},k}$ for all $j = 2, \dots, k$, and $\rho = \rho^{\text{v},k}$, respectively.

The notion of k-marginality requires the existence of symmetric marginalizable extension that leads to the original stochastic choice rule independently of which virtual Sam is not averaged over. At the same time, k-marginality on average is implied by k-marginality, but does not require any of virtual rules to agree with the original ρ.

Now we are ready to state our main result for the thought experiment. Recall that $|\mathcal{X}^2| = 2$ is the number of menus faced by Sam.

Theorem 2. *The following are equivalent:*

(i) ρ is consistent with the thought experiment.

(ii) ρ is k-marginalizable for any finite $k \geq 1$.

(iii) ρ is $|\mathcal{X}^2|$-marginalizable on average.

Theorem 2 is an Afriat (1967)'s like characterization of separable choice since marginality can be represented as a set of finite linear constraints. Theorem 2 provides two characterizations of the thought experiment. Theorem 2(ii) intuitively means that separable ρ admits a marginalizable extension that coincides with ρ when projected to any virtual Sam. Theorem 2(iii) instead implies that we can just check the average marginalization for $k = |\mathcal{X}^2|$ thus making the verification of the consistency of ρ with the thought experiment computationally feasible.

Remark 2. We highlight that Bell-type inequalities for a generalization of the thought experiment for any finite number of menus and alternatives, and any finite number of DMs remains an open question. The CHSH inequalities are particular to our original thought experiment. In contrast, k-marginality characterizes the thought experiment for the aforementioned extensions as formalized in the appendix.

6. Conclusions

Our result can be applied to obtain separable necessary conditions for any joint stochastic choice model expressed as a separable finite mixture of choice functions. When the model is generating and unique, we provide a full characterization via Bell (1964)'s type inequalities that depend only on the joint probabilistic choice rule. Absent uniqueness we provide an Afriat (1967)'s like characterization of separable choice via a finite system of restrictions.

References

Abramsky, S. and A. Brandenburger (2011): "The sheaf-theoretic structure of non-locality and contextuality," *New Journal of Physics*, 13, 113036. 4

Afriat, S. N. (1967): "The construction of utility functions from expenditure data," *International economic review*, 8, 67–77. 1, 5, 5, 6

Ashenfelter, O. and A. Krueger (1994): "Estimates of the economic return to schooling from a new sample of twins," *The American economic review*, 1157–1173. 1

Aubrun, G., L. Lami, C. Palazuelos, and M. Plávala (2021): "Entangleability of cones," *Geometric and Functional Analysis*, 31, 181–205. 2.1, 4, A.2

Aubrun, G., A. Müller-Hermes, and M. Plávala (2022): "Monogamy of entanglement between cones," *arXiv preprint arXiv:2206.11805*. A.3

Bell, J. S. (1964): "On the einstein podolsky rosen paradox," *Physics Physique Fizika*, 1, 195. 1, 2.1, 6

Bertsekas, D., A. Nedic, and A. Ozdaglar (2003): *Convex analysis and optimization*, vol. 1, Athena Scientific. A.3

Block, H. and J. Marschak (1960): "Random Orderings and Stochastic Theories of Responses". in I. Olkin, S. Ghurye, W. Hoeffding, W. Madow, and H. Man (eds) Contributions to Probability and Statistics, Stanford University Press," . 4

Chambers, C. P., T. Cuhadaroglu, and Y. Masatlioglu (2023): "Behavioral influence," *Journal of the European Economic Association*, 21, 135–166. 1

Chambers, C. P., Y. Masatlioglu, and C. Turansick (2021): "Correlated Choice," *arXiv preprint arXiv:2103.05084*. (document), 1, 2, 2.1, 4

Cherchye, L., B. De Rock, R. Griffith, M. O'Connell, K. Smith, and F. Vermeulen (2017): "A new year, a new you? Heterogeneity and self-control in food purchases," *Heterogeneity and Self-Control in Food Purchases (December 2017)*. 1

Davis-Stober, C. P., J.-P. Doignon, S. Fiorini, F. Glineur, and M. Regenwetter (2018): "Extended formulations for order polytopes through network flows," *Journal of mathematical psychology*, 87, 1–10. 4

De Paula, A. and X. Tang (2012): "Inference of signs of interaction effects in simultaneous games with incomplete information," *Econometrica*, 80, 143–172. 1

Doignon, J.-P. and K. Saito (2023): "Adjacencies on random ordering polytopes and flow polytopes," *Journal of Mathematical Psychology*, 114, 102768. 4

Falmagne, J.-C. (1978): "A representation theorem for finite random scale systems," *Journal of Mathematical Psychology*, 18, 52–72. 4

Fine, A. (1982): "Hidden variables, joint probability, and the Bell inequalities," *Physical Review Letters*, 48, 291. 2.1

Frick, M., R. Iijima, and T. Strzalecki (2019): "Dynamic random utility," *Econometrica*, 87, 1941–2002. 1

Gunsilius, F. F. (2021): "Nontestability of instrument validity under continuous treatments," *Biometrika*, 108, 989–995. 1

Gurvits, L. (2003): "Classical deterministic complexity of Edmonds' problem and quantum entanglement," in *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, 10–19. 1

Kashaev, N., V. H. Aguiar, M. Plávala, and C. Gauthier (2023a): "Dynamic and stochastic rational behavior," *arXiv preprint arXiv:2302.04417*. ∗, 1

Kashaev, N., N. Lazzati, and R. Xiao (2023b): "Peer Effects in Consideration and Preferences," *arXiv preprint arXiv:2310.12272*. 1

Kline, B. and E. Tamer (2023): "Recent developments in partial identification," *Annual Review of Economics*, 15, 125–150. 4

Li, R. (2021): "An Axiomatization of Stochastic Utility," *arXiv preprint arXiv:2102.00143*. 1, 4

McFadden, D. and M. K. Richter (1990): "Stochastic rationality and revealed stochastic preference," *Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz, Westview Press: Boulder, CO*, 161–186. 4

Miller, P., C. Mulvey, and N. Martin (1995): "What do twins studies reveal about the economic returns to education? A comparison of Australian and US findings," *The American Economic Review*, 85, 586–599. 1

Pearl, J. (1995): "On the testability of causal models with latent and instrumental variables," in *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, 435–443. 1

Ray, D. and A. Robson (2018): "Certified random: A new order for coauthorship," *American Economic Review*, 108, 489–520. ∗

Rosset, D., J.-D. Bancal, and N. Gisin (2014): "Classifying 50 years of Bell inequalities," *Journal of Physics A: Mathematical and Theoretical*, 47, 424022. 1, 2, 2.1

Sacerdote, B. (2011): "Peer effects in education: How might they work, how big are they and how much do we know thus far?" in *Handbook of the Economics of Education*, Elsevier, vol. 3, 249–277. 1, 2.1

Strzalecki, T. (2021): *Stochastic Choice*, Mimeo. 1, 2

A. Proof of Theorem 1

We prove Theorem 1 at a more general level. We consider $1 \leq T < \infty$ DMs indexed by $t \in \mathcal{T} = \{1, \cdots, T\}$. Let X^t be a nonempty finite choice set. In each $t \in \mathcal{T}$, there are $J^t < \infty$ distinct menus denoted by

$$M_j^t \in 2^{X^t} \setminus \{\emptyset\}, \quad j \in \mathcal{J}^t = \{1, \ldots, J^t\}.$$

Since X^t is a finite set, we denote the i-th element of menu $j \in \mathcal{J}^t$ as $x_{i|j}^t$. That is, $M_j^t = \{x_{i|j}^t\}_{i \in \mathcal{I}_j^t}$, where $\mathcal{I}_j^t = \{1, 2, \ldots, I_j^t\}$ and I_j^t is the number of elements in menu j.

Define a menu path as an ordered collection of indexes $\mathbf{j} = (j_t)_{t \in \mathcal{T}}$, $j_t \in \mathcal{J}^t$. Menu paths encode menus that are seen by each of the corresponding DMs. Let $\mathbf{J}$ be the set of all observed menu paths. Given $\mathbf{j} \in \mathbf{J}$, a choice path is an array of alternatives $x_{\mathbf{i}|\mathbf{j}} = \left(x_{i_t|j_t}^t\right)_{t \in \mathcal{T}}$ for some collection of indexes $\mathbf{i} = (i_t)_{t \in \mathcal{T}}$ such that $i_t \in \mathcal{I}_{j_t}^t$ for all t. Similar to a menu path, a choice path encodes the choices of DMs in a given sequence of menus that DMs have faced. The set of all possible choice path index sets $\mathbf{i}$, given a menu path $\mathbf{j}$, is denoted by $\mathbf{I_j}$.

Note that every $\mathbf{j} \in \mathbf{J}$ encodes the Cartesian product of menus $\times_{t \in \mathcal{T}} M_{j_t}^t \subseteq \times_{t \in \mathcal{T}} X^t$. Then, for every $\mathbf{j}$, let $\rho_{\mathbf{j}}$ be a probability measure on $\times_{t \in \mathcal{T}} M_{j_t}^t$. That is, $\rho_{\mathbf{j}}\left(x_{\mathbf{i}|\mathbf{j}}\right) \geq 0$ for all $\mathbf{i} \in \mathbf{I_j}$ and $\sum_{\mathbf{i} \in \mathbf{I_j}} \rho_{\mathbf{j}}\left(x_{\mathbf{i}|\mathbf{j}}\right) = 1$. The primitive in our framework is the joint probabilistic choice rule $\rho = (\rho_{\mathbf{j}})_{\mathbf{j} \in \boldsymbol{J}}$.

Given ρ we can define the generalized thought experiment in an analogous way to our original thought experiment but such that

$$\rho = (\otimes_{t=1}^{T} A^t)\nu$$

for some $\nu \geq 0$, and where columns of A^t collects all allowable choice functions of DM t. We assume that A^t is generating for all $t \in \mathcal{T}$.

$\mathcal{H}$- and $\mathcal{V}$-representations

For a single DM we can check if the probabilistic choice rule is consistent with the thought experiment at the individual level by checking that the stochastic choices of DM t belong to the cone

$$\left\{A^t v \,:\, v \geq 0\right\}.$$

This is called the $\mathcal{V}$-representation of the cone. The Weyl-Minkowski theorem states that there exists an equivalent representation of the cone (the $\mathcal{H}$-representation) via some matrix H_t:

$$\left\{z \,:\, H^t z \geq 0\right\}.$$

The $\mathcal{V}$-representation of the cone associated with the thought experiment provides an interpretation of the former as the observed distribution over choices is a finite mixture of deterministic types. The $\mathcal{H}$-representation of the cone associated with the thought experiment corresponds to what is usually called an axiomatization via linear inequalities. Importantly, these inequalities represent facets of the cone.

Proposition 3. If

$$\left\{K^t v \,:\, v \geq 0\right\} = \left\{z \,:\, L^t z \geq 0\right\}$$

for all $t \in \mathcal{T}$, then

$$\left\{\left(\otimes_{t\in\mathcal{T}} K^t\right) v \,:\, v \geq 0\right\} \subseteq \left\{z \,:\, \left(\otimes_{t\in\mathcal{T}} L^t\right) z \geq 0\right\}.$$

We say that the generalized experiment produces only separable restrictions on ρ whenever there exists a $\nu \geq 0$ such that $\rho = (\otimes_{t=1}^{T} A^t)\nu$ if and only if $(\otimes_{t=1}^{T} H^t)\rho \geq 0$ and ρ satisfies marginality.

Theorem 3. *The generalized experiment produces only separable restrictions on ρ if and only if ρ*

satisfies marginality and A^t is associated with a unique representation for all $t \in \mathcal{T}$ except at most one.

A.1. Proof of Proposition 3

For completeness we provide here the proof of Proposition 3. Let $L_T = \otimes_{t=1}^{T} L^t$ and $K_T = \otimes_{t=1}^{T} K^t$. Note that for any v, z and $\otimes_{t=1}^{T} K^t$ such that $\left(\otimes_{t=1}^{T} K_t\right) v = z$ is well-defined, we can construct V and Z such that columns of V and Z are subvectors[2] of v and z and

$$\left(\otimes_{t=1}^{T} K^t\right) v = z \iff K^T V \left(\otimes_{t=1}^{T-1} K^t\right)' = Z.$$

Recall that by definition, $L^t K^t v \geq 0$ for all $v \geq 0$. Hence,

$$\begin{aligned}
&\forall v \geq 0,\ L^1 K^1 v \geq 0 \implies \forall V \geq 0,\ L^2 K^2 V (L^1 K^1)' \geq 0 \iff \\
&\forall v \geq 0,\ (L^1 K^1 \otimes L^2 K^2) v \geq 0 \implies \forall V \geq 0,\ L^3 K^3 V (L^1 K^1 \otimes L^2 K^2)' \geq 0 \iff \\
&\forall v \geq 0,\ (\otimes_{t=1}^{3} L^t K^t) v \geq 0 \implies \forall V \geq 0,\ L^4 K^4 V (\otimes_{t=1}^{3} L^t K^t)' \geq 0 \implies \\
&\cdots \implies \forall v \geq 0,\ (\otimes_{t=1}^{T} L^t K^t) v \geq 0 \iff \forall v \geq 0,\ L_T K_T v \geq 0.
\end{aligned}$$

Hence,

$$\{K_T v \,:\, v \geq 0\} \subseteq \{z \,:\, L_T z \geq 0\}.$$

A.2. Proof of Theorem 3

First we show necessity of marginality. By definition the generalized thought experiment, there exists a distribution over $\mathcal{C}$, the collection of all choice function c that are mappings from the collection of menus in each t to alternatives, μ, such that

$$\rho\left(\left(x_{i_t|j_t}\right)_{t\in\mathcal{T}}\right) = \int \prod_{t\in\mathcal{T}} \mathbb{1}\left(c(M_{j_t}^t) = x_{i_t|j_t}^t\right) d\mu(c)$$

[2] A vector x is a subvector of $y = (y_j)_{j\in J}$, if $x = (y_j)_{j\in J'}$ for some $J' \subseteq J$.

for all $\mathbf{i},\mathbf{j}$. Fix some $t' \in \mathcal{T}$, $x_{\mathbf{i}|\mathbf{j}}$, and $j_{t'} \in \mathcal{J}^{t'}$. Note that

$$
\begin{aligned}
&\sum_{i \in \mathcal{I}^{t'}_{j_{t'}}} \rho\left(x_{\mathbf{i}|\mathbf{j}}\right) = \\
&\sum_{i \in \mathcal{I}^{t'}_{j_{t'}}} \int \mathbb{1}\left(c(M^{t'}_{j_{t'}}) = x^{t'}_{i|j_{t'}}\right) \prod_{t \in \mathcal{T}\setminus\{t'\}} \mathbb{1}\left(c(M^{t}_{j_t}) = x^{t}_{i_t|j_t}\right) d\mu(c) = \\
&\int \sum_{i \in \mathcal{I}^{t'}_{j_{t'}}} \mathbb{1}\left(c(M^{t'}_{j_{t'}}) = x^{t'}_{i|j_{t'}}\right) \prod_{t \in \mathcal{T}\setminus\{t'\}} \mathbb{1}\left(c(M^{t}_{j_t}) = x^{t}_{i_t|j_t}\right) d\mu(c) = \\
&\int \prod_{t \in \mathcal{T}\setminus\{t'\}} \mathbb{1}\left(c(M^{t}_{j_t}) = x^{t}_{i_t|j_t}\right) d\mu(c),
\end{aligned}
$$

where the last equality follows from $c(M^{t'}_{j_{t'}})$ being a singleton and $\{x^{t'}_{i|j_{t'}}\}_{i \in \mathcal{I}^{t'}_{j_{t'}}}$ being a partition. The right-hand side of the last expression does not depend on the choice of $j_{t'}$. Marginality follows from t' and $x_{\mathbf{i}|\mathbf{j}}$ being arbitrary.

Next, we show that any ρ that satisfies marginality belongs to a linear span of columns of $A_T = (\otimes_{t=1}^{T} A^t)$. That is, the system $A_T v = \rho$ always has a solution and the cone generated by A_T is proper when restricted to ρ that satisfies marginality. Hence, Theorem 3 follows from Theorem A in Aubrun et al. (2021).

Consider the following modification of A^t, $t \in \mathcal{T}$. From every menu, except the first one, we pick the last alternative and remove the corresponding row from A^t. Let A^{t*} denote the resulting matrix. Thus, matrix A^t can be partitioned into A^{t*} and A^{t-}, where rows of A^{t-} correspond to alternatives removed from A^t. Consider the first row of A^{t-}. It corresponds to the last alternative from the second menu at time t. Note that the sum of all rows that correspond to the same menu is equal to the row of ones. Hence, the first row of A^{t-} is equal to the sum of the rows that correspond to menu 1 minus the sum of the remaining rows in menu 2. That is, the first row of A^{t-} can be written as

$$(1, \ldots, 1, -1, \ldots, -1, 0, \ldots, 0)A^{t*}.$$

Similarly, the second row of A^{t-} can be written as

$$(1, \ldots, 1, 0, \ldots, 0, -1, \ldots, -1, 0, \ldots, 0)A^{t*}.$$

In matrix notation, $A^{t-} = G^t A^{t*}$, where G^t is the matrix with the k-th row having the elements that correspond to the alternatives from the first menu at time t are equal to 1, the elements that correspond to the alternatives from the k-th menu are equal to -1, and the rest of elements are equal to 0.

Next note that, up to a permutation of rows, $A_{\mathcal{T}}$ can be partitioned into $A^*_{\mathcal{T}} = \otimes_{t\in\mathcal{T}} A^{t*}$ and matrices of the form $\otimes_{t\in\mathcal{T}} C^t$, where $C^t \in \{A^{t*}, A^{t-}\}$, with $C^t = A^{t-}$ for at least one t. We will stack all these matrices into $A^-_{\mathcal{T}}$. Next, let ρ^* denote the subvector of ρ that corresponds to choice paths that do not contain any of the alternatives removed from A^t, $t \in \mathcal{T}$. Thus, $\rho = (\rho^{*\prime}, \rho^{-\prime})'$, where ρ^- corresponds to all elements of ρ that contain at least one of the removed alternatives. As a result, we can split the original system into two: $A^*_{\mathcal{T}} v = \rho^*$ and $A^-_{\mathcal{T}} v = \rho^-$.

Consider the system $A^*_{\mathcal{T}} v = \rho^*$. Since we only removed one row from each menu except the first one and A^t can generate any ρ_t, A^{t*} has full row rank for all t. Then $A^*_{\mathcal{T}}$ is also of full row rank and, hence, $A^*_{\mathcal{T}} A^{*\prime}$ is invertible and $v^* = A^{*\prime}\left(A^*_{\mathcal{T}} A^{*\prime}\right)^{-1} \rho^*$ solves the system. If, we show that

$$A^-_{\mathcal{T}} v^* = \rho^-,$$

then we prove that $A_{\mathcal{T}} v = \rho$ always has a solution, which will complete the proof.

Note that $A^-_{\mathcal{T}}$ consists of the blocks of the form $\otimes_{t\in\mathcal{T}} C^t$, where $C^t \in \{A^{t*}, A^{t-}\}$ and $C^t = A^{t-}$ for at least one t. Next note that for any A, B, and C, we have that

$$A \otimes (BC) = diag(B)(A \otimes C),$$

where $diag(B)$ is the block-diagonal matrix constructed from B.

Let W^t (with inverse $W^{t,-1}$, which pushes the last element of $\mathcal{T}$ to t-th position) be a transformation that recomputes all objects for the time span where t is pushed to the end. Transformation W^t satisfies the following three properties: $W^t[C] = C$ if C does not depend on $\mathcal{T}$; $W^t[CD] = W^t[C]W^t[D]$ for any matrices C and D; and $W^t[\otimes_{t'\in\mathcal{T}} A^{t'*}] = \otimes_{t'\in\mathcal{T}\setminus\{t\}} A^{t'*} \otimes A^{t*}$. Let Y^t be an operator such that $Y^t[\cdot] = W^{t,-1}\left[diag(G^t)W^t[\cdot]\right]$.

Consider $\otimes_{t\in\mathcal{T}} C^t$, where $C^t \in \{A^{t*}, A^{t-}\}$ and $C^t = A^{t-}$ for only one t. Hence,

$$\otimes_{t'\in\mathcal{T}} C^{t'} v^* = W^{t,-1}\left[diag(G^t)W^t\left[\rho^*\right]\right] = Y^t[\rho^*].$$

Note that because ρ satisfies marginality, $diag(G^T)\rho^*$ is the subvector of ρ^- that corresponds to choice paths that contain one of the removed alternatives from the last DM only. So, $W^t[\rho^*]$ first pushes DM t to the very end, then $diag(G^t)W^t[\rho^*]$ computes the elements of ρ^-, and finally $W^{t,-1}[diag(G^t)W^t[\rho^*]]$ moves DM t back to her place.

Next, consider $\otimes_{t\in\mathcal{T}}C^t$, where $C^t \in \{A^{t*}, A^{t-}\}$ and $C^t = A^{t-}$ and $C^{t'} = A^{t'-}$ for two distinct t, t'. Similarly to the previous case,

$$\otimes_{t'\in\mathcal{T}} C^{t'}v^* = W^{t,-1}\left[diag(G^t)W^t\left[W^{t',-1}\left[diag(G^{t'})W^{t'}[\rho^*]\right]\right]\right] = Y^t[Y^{t'}[\rho^*]] = Y^t \circ Y^{t'}[\rho^*],$$

where $Y^t \circ Y^{t'}$ denotes the composite operator. Again, $W^{t',-1}\left[diag(G^{t'})W^{t'}[\rho^*]\right]$ computes the subvector of ρ^- that corresponds to choice paths where an alternative from only one time t' was missing. Applying to the resulting vector $W^{t,-1}[diag(G^t)W^t[\cdot]]$ computes the subvector of ρ^- with alternatives missing from t and t' only. Repeating the arguments for all possible rows of $A^-_{\mathcal{T}}$, we obtain that

$$\otimes_{t'\in\mathcal{T}} C^{t'}v^* = \circ_{t':C^{t'}=A^{t'-}}Y^{t'}[\rho^*]$$

and, by marginality, $A^-_{\mathcal{T}}v^* = \rho^-$. Hence, v^* is a solution to $A_{\mathcal{T}}v = \rho$.

A.3. Proof of Theorem 2

(i) $\implies$ (ii). Suppose that ρ is consistent with the thought experiment. Then it follows from the proof of Theorem 1 that ρ satisfies marginality (i.e. it is 1-marginalizable).

Next, fix any $k > 1$. The corresponding finite-mixture representation of the extension of ρ is captured by $A^1 \otimes A^{2,\otimes(k)}$, where $A^{2,\otimes(k)} = \otimes_{l=1}^k A^2$. Since ρ is consistent with the thought experiment, there exists a distribution over columns of $A^1 \otimes A^2$, ν, such that $\rho = (A^1 \otimes A^2)\nu$. Take any nonzero component of ν, ν_l, and take the corresponding l^{th} column of $(A^1 \otimes A^2)$. This column corresponds to a pair of individual preference profiles (r_1, r_2). Consider the matrix $A^1 \otimes A^{2,\otimes(k)}$. Find the column of it that corresponds to $(r_1, r_2, r_2, \ldots, r_2)$ and assign weight ν_l to it. Repeat this procedure for all nonzero components of ν. As a result, we construct a distribution over columns of $A^1 \otimes A^{2,\otimes(k)}$, ν^*.

Define

$$\rho^{\text{ext},k} = (A^1 \otimes_{t=2}^{k+1} A^2)\nu^*.$$

By construction, $\rho^{\text{ext},k}$ is marginalizable, since ρ is, and $\rho = \rho_j^{\text{ext},k}$ for any $j = 2, \ldots, k$. Hence, ρ is k-marginalizable. The fact that the choice of k was arbitrary completes the proof.

(ii) $\implies$ (iii). If ρ is k-marginalizable for any finite k, then it is $|\mathcal{X}^2|$-marginalizable. If ρ is $|\mathcal{X}^2|$-marginalizable, then it is trivially $|\mathcal{X}^2|$-marginalizable on average.

(iii) $\implies$ (i). First, we provide some preliminary results. Let $|M_1|$ denote the cardinality of M_1 and $\mathbf{1}$ denote the vector of ones.

Lemma 1. *For any t, the set of (individual) choice rules is a Cartesian product of simplices. That is,*

$$\{A^t\nu \,:\, \nu \geq 0, \nu'\mathbf{1} = 1\} = \times_{M_t \in \mathcal{X}^t} \Delta^{|M_t|-1}.$$

As a result, any (individual) choice rule can be generated by A^t.

Proof. Note that columns of A^t (all rationalizable deterministic choice functions) are Cartesian products of columns of $|\mathcal{X}^t|$ identity matrices with the identity matrix corresponding to menu M_t being of the size $|M_t|$-by-$|M_t|$. Hence, $\{A^t\nu \,:\, \nu \geq 0, \nu'\mathbf{1} = 1\}$ is a convex hull of the Cartesian product of several sets. The result then follows from the fact that the convex hull of the Cartesian product of two sets is equal to the Cartesian product of the convex hull of each set (Bertsekas, Nedic, and Ozdaglar, 2003). ■

Let

$$C_t = \left\{A^t\nu \,:\, \nu \geq 0\right\}, \qquad C = \left\{\left(A^1 \otimes A^2\right)\nu \,:\, \nu \geq 0\right\}.$$

Also, let H^t be a matrix such that

$$C_t = \{\rho_t \,:\, H^t\rho_t \geq 0\}.$$

Since A^t can generate any choice rule, there are only 2 types of restrictions captured in H^t: nonnegativity and adding-up constraints. The former restrict every component of ρ_t to be nonnegative;

the latter require $\sum_{x\in M_t}\rho_{M_t}(x)$ to be the same for all $M_t \in \mathcal{X}^t$. Note that since C_t is a cone, $\sum_{x\in M_t}\rho_{M_t}(x)$ is allowed to be different from 1.

The next lemma establishes that the cone generated by the thought experiment is proper when restricted to marginalizable vectors, which is a linear vector subspace. It follows from the proof of Theorem 3.

Lemma 2. *The cone C and, thus, C_t are proper when restricted to marginalizable ρ.*

Recall that $B^{\otimes k} = \otimes_{m=1}^{k} B$ for any matrix B.

Lemma 3.

$$\rho_v \geq 0 \text{ and is marginalizable} \iff \left[\left(H^1 \otimes H^{2,\otimes\left|\mathcal{X}^2\right|}\right)\rho_v \geq 0\right]$$

Proof. We need to show that $\left(H^1 \otimes H^{2,\otimes\left|\mathcal{X}^2\right|}\right)$ contains only nonnegativity constraints and the constraints implied by marginality. First, note that H^t can be partitioned as

$$H^t = \begin{pmatrix} H^{t,\mathrm{m}} \\ I \end{pmatrix},$$

where $H^{t,\mathrm{m}}$ correspond to adding-up constraints and the identity matrix I corresponds to the nonnegativity constraints. As a result, for any t, t', we have that

$$H^t \otimes H^{t'} = \begin{pmatrix} H^{t,\mathrm{m}} \otimes H^{t',\mathrm{m}} \\ H^{t,\mathrm{m}} \otimes I \\ I \otimes H^{t',\mathrm{m}} \\ I \otimes I \end{pmatrix}.$$

The first block in the above matrix corresponds to the constraints implied by marginality (one sums over two different menus rather than one), the second and the third blocks correspond to marginality, while the last one corresponds to the nonnegativity constraint. Repeating this argument finitely many times, we observe that one will eventually obtain all the nonnegativity and marginality constraints. ■

Take $\phi = \mathbf{1}/\left|\mathcal{X}^t\right|$. Note that since the rows of H^t correspond either to equality constrains or to nonnegativity constraints, the sum of rows of H^t is equal to $\mathbf{1}'$. Hence, $\phi' = \mathbf{1}/\left|\mathcal{X}^t\right| H^t$ is the linear combination of rows of H^t, and belongs to the cone dual to C_t.

Next, define

$$K_\phi = \{x \in C_t \;:\; \phi' x = 1\}.$$

By definition, for every $x \in C_t$, there exists $\nu \geq 0$ such that $x = A^t \nu$. Hence, $1 = \phi' x = \phi' A^t \nu = \left|\mathcal{X}^t\right| \mathbf{1}'\nu / \left|\mathcal{X}^t\right| = \mathbf{1}'\nu$. Thus, by Lemma 1,

$$K_\phi = \{A^t \nu \;:\; \nu \geq 0, \nu'\mathbf{1} = 1\} = \times_{M_t \in \mathcal{X}^t} \Delta^{|M_t|-1}$$

is a Cartesian product of $\left|\mathcal{X}^t\right|$ simplices. Thus, since by Lemma 2 C is proper when restricted to marginalizable vectors, by Theorem 2 of Aubrun, Müller-Hermes, and Plávala (2022)

$$C = \left\{ \left(I \otimes \gamma^{\phi}_{|\mathcal{X}^2|}\right) x \;:\; \left(H^1 \otimes H^{2,\otimes|\mathcal{X}^2|}\right) x \geq 0 \right\},$$

where

$$\gamma^{\phi}_k = \frac{1}{k} \sum_{j=1}^{k} (\phi')^{\otimes j-1} \otimes I \otimes (\phi')^{\otimes k-j}.$$

Next, applying Lemma 3, we can conclude that

$$C = \left\{ \left(I \otimes \gamma^{\phi}_{|\mathcal{X}^2|}\right) x \;:\; x \geq 0 \text{ and is marginalizable} \right\}.$$

Given that the matrix $\left(I \otimes \gamma^{\phi}_{|\mathcal{X}^2|}\right)$ marginalizes x over all extensions and leaves only menus in $t = 1$ and $t = 2$, we obtain that a stochastic choice function ρ is separable if and only if it is $|\mathcal{X}^2|$-marginalizable on average.

As a result, (i) $\implies$ (ii) $\implies$ (iii) $\implies$ (i).