Procedural Mixture Sets

Hendrik Rommeswinkel*

This version: March 2023

Abstract

The paper characterizes the Shannon (1948) and Tsallis (1988) entropies in a standard framework of decision theory, mixture sets. Procedural mixture sets are introduced as a variant of mixture sets in which it is not necessarily true that a mixture of two identical elements yields the same element. This allows the process of mixing itself to have an intrinsic value. The paper proves the surprising result that simply imposing the standard axioms of von Neumann-Morgenstern on preferences on a procedural mixture set yields the entropy as a representation of procedural value. An application to the relation between choice probabilities and decision times in decision processes elucidates the difficulty of extending the drift-diffusion model to multi-alternative choice.

Keywords: Decision theory, procedural value, decision processes, mixture sets, entropy, reduction of compound mixtures, reducibility, associativity

1 Introduction

Information theoretic measures are commonly employed in economic theory. They are used to study inequality (Shorrocks, 1980; Theil, 1967),

*Waseda Institute for Advanced Study, Waseda University, 1-chome-104 Totsukamachi, Shinjuku City, Tokyo 169-8050, Japan. Email: rommesiwnkel@aoni.waseda.jp. Many thanks to two anonymous referees, participants at the Society for the Advancement of Economic Theory conference 2019, Chun-ting Chen and Patrick DeJarnette for helpful comments.

segregation (Frankel & Volij, 2011), the utility of gambling (Luce et al., 2008a, 2008b), diversity (Nehring & Puppe, 2009), consumer demand (Theil, 1965), freedom of choice (Suppes, 1996), market concentration (Hennessy & Lapan, 2007; Herfindahl, 1950; Hirschman, 1980), and information costs (Caplin et al., 2017; Sims, 2003). The present paper provides an axiomatic foundation of two of the most commonly employed information measures, the Shannon (1948) and Tsallis (1988) entropies (the latter of which is a monotone transformation of the Rényi (1961) entropy). We show that these measures arise naturally from the standard von Neumann-Morgenstern axioms imposed on a variation of mixture sets that allows for procedural aspects to play a role.

The expected utility representation by von Neumann and Morgenstern (1944) was initially stated using an "algebra of combining" of compound lotteries. Herstein and Milnor (1953) simplified the axioms greatly by introducing "mixture sets". Mixture sets can be given the interpretation of nested binary lotteries; we play one lottery to determine which lottery is resolved next, which determines which lottery is resolved afterwards, and so on. From the Reduction of Compound Mixtures axiom of mixture sets and the von Neumann and Morgenstern (1944) axioms follows interiority; if $a \succsim b$, then $a \succsim \mu a \oplus (1-\mu)b \succsim b$.

One example where interiority is violated is when modeling the time it takes an individual to make decisions in the context of stochastic choice . The literature on decision processes (Bogacz et al., 2006; Usher et al., 2013) studies how long it takes a decision maker to choose from a set of alternatives and how this duration depends on properties of the decision problem such as the probabilities of alternatives being chosen. Suppose a and b are decision processes and $a \sim b$ denotes that these decision processes take an equal amount of time to complete. In this context the relation $\succsim$ has the interpretation of "takes as least as long as" and represents a procedural value (time) as opposed to a consequentialist value (utility). Suppose $\mu a \oplus (1-\mu)b$ is a decision process in which a decision maker chooses with probability μ to complete the sub-decision process a and complete the sub-decision process b with probability $1-\mu$. If choice is not instantaneous, then it is reasonable to assume that $\mu a \oplus (1-\mu)b \succ a \sim b$, i.e., that the decision to choose between the two sub-decision processes takes some additional time to complete.

This paper provides two ways in which the mixture set assumptions of Herstein and Milnor (1953) and the expected utility axioms of von Neumann and Morgenstern (1944) can be adjusted to relax interiority. In the first main result we change the Reduction of Compound Mixtures axiom of mixture sets into an Associativity condition. Associativity is a new axiom that is not a generalization of the aforemention Reduction of Compound Mixtures. Associativity states that the order of mixing does not matter but unlike Reduction of Compound Mixtures allows for $\mu a \oplus (1-\mu)a \not\sim a$. We call sets that fulfill these assumptions procedural mixture sets. The expected utility axioms — Weak Order, Continuity, and Independence — of von Neumann and Morgenstern (1944) then characterize the Shannon and Tsallis entropies.

In the second main result we restrict the Independence axiom of von Neumann and Morgenstern (1944) to mixtures with disjoint support but maintain the mixture set assumptions introduced by Herstein and Milnor (1953). We again obtain the Shannon and Tsallis entropy representation (for a subset of the relation). In both cases, the entropy can be interpreted as a function representing the procedural value of going through the motions of resolving a mixture. We call such a representation a mixture entropy value.

We apply this model of procedural value to decision times in stochastic decision processes. We show that decision processes that follow the Luce (1959) model of stochastic choice form a procedural mixture set. In particular, Luce (1959)'s IIA axiom of choice probabilities guarantees Associativity while the fact that choice is non-instantaneous motivates the violation of interiority.[1] In this context, the Independence axiom means that the time it takes to complete a decision process is increasing in the duration of each sub-decision process. The mixture entropy value then emerges as a representation of decision times for multi-alternative choice. This results in a corollary that jointly characterizes the Luce model of stochastic choice and an entropy representation of decision times.

The Luce model of stochastic choice is restrictive. In some cases, the associative structure generated by the Luce model only holds for a subset of the decisions. Using our second main characterization result (which weakens the Independence axiom instead of changing the mixture set assumptions), we obtain a model of decision times and choice probabilities for which the Luce stochastic choice model and the mixture entropy decision time representation hold whenever the alternatives have mutually exclusive choice relevant attributes. The weakening of the Independence

[1] In the context of decisions under risk, uncertainty effects may induce violations of interiority (Gneezy et al., 2006). This suggests that individuals also attach a "procedural" value to the uncertainty of lotteries.

axiom requires decision times to only increase in the decision times of sub-decision processes that contain alternatives sufficiently distinct from the remaining alternatives of the decision process. This allows the decision time between two similar alternatives to differ from the decision time between two very different alternatives.

The paper proceeds as follows. We present the axioms and the representation theorem for procedural mixtures in Section 3. Section 4 provides comparative statics results and interprets the parameters of the model. Section 5 discusses the representation results that maintain Reducibility but weaken Independence. The relation to the literature is given in Section 6. Section 7 concludes.

2 Introductory Example

Suppose an analyst observes a sample of decision makers choosing between items a and b. This choice is not instantaneous; the decision maker undergoes some thought process until a decision is made. The duration of this thought process will generally depend on how difficult the decision is. The literature on decision processes (Bogacz et al., 2006; Usher et al., 2013) has established that the more uncertain ex ante the choice between two alternatives is, the longer a decision takes. Thus, when two alternatives are equally likely to be chosen, then the decision process takes longer than when one alternative is much more likely to be chosen than the other.[2] The drift diffusion model (Ratcliff, 1978) captures this basic empirical fact but there is no agreed-upon generalization to multiple options (Ratcliff et al., 2016). The generalization to multiple options is complicated by the presence of similarity, attraction, and compromise effects (Roe et al., 2001) in stochastic choice. Generally, the availability of additional alternatives may influence the relative choice probabilities of two alternatives, complicating the extension of the relation between choice probabilities and decision times from two options to multiple options. However, we will show that even in the absence of such effects, it turns out that plausible assumptions lead to a very restrictive functional form for the relation between probabilities and decision times.

Take the classical example of a decision maker who faces a choice between taking an *airplane* , a *bus*, or a *car* to travel to another city. An

[2]In experiments, the decision time commonly refers to the average decision time of a sample of subjects and the choice probabilities refer to the relative frequency of choice.

analyst records the choice probabilities and decision times shown in Figure 1.

Option 1	Option 2	Option 3	Duration
Airplane 20%	Bus 20%	Car 60%	1.9s
Airplane 50%	Bus 50%	None	1.5s
Bus 50%	Bus 50%	None	1.5s
Airplane 25%	Car 75%	None	1.3s
Bus 25%	Car 75%	None	1.3s
Airplane 100%	None	None	.5s
Bus 100%	None	None	.5s
Car 100%	None	None	.5s

Figure 1: Choice Probability and Decision Time Data

The choices from singleton sets are trivial and always have a probability of 100%. They are also resolved the most quickly. The binary choice data exhibits the following pattern: the airplane and bus are equally likely to be chosen when compared with each other and both have the same choice probability and decision time when compared with the car. This property is an instance of the Independence of Irrelevant Alternatives (IIA) axiom of Luce (1959). Consistently, the probabilities of the three option decision problem in Figure 1 are such that conditioning on any subset yields the corresponding two option choice probabilities. The choice probability data therefore follows the Luce (1959) model.

We also observe that the (meaningless) decision between two buses takes as much time as the decision between an airplane and a bus. This is a particularly strong violation of Reducibility; even the choice between two identical options has a positive deliberation time. This is consistent with the blue bus - red bus paradox of Debreu (1960) according to which the introduction of (meaningless) distinctions can influence the relative choice probability to other alternatives.[3]

We now present a deliberation process that the decision maker might go through in order to make a decision between the three alternatives airplane, bus, and car. We assume that the decision maker goes through a sequence of binary decisions about attributes of goods. An attribute is a shared property of some of the goods and thus by iteratively choosing attributes, the decision maker narrows down the options sequentially until arriving at a final option. In the case of three alternatives, there are only three possible configurations for such a decision process, depicted in Figure 2. For example, the decision maker might initially decide to use a slow method of transportation and then solve the subsequent decision problem of choosing between the bus and the car.

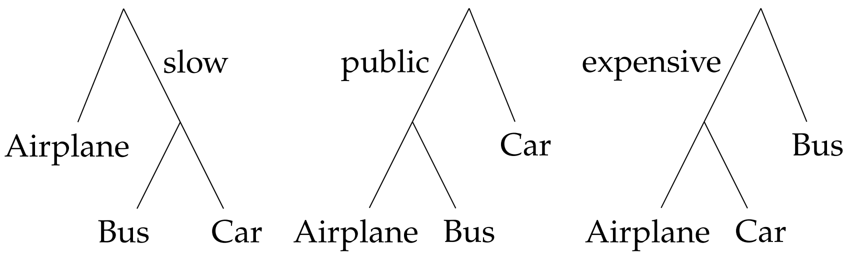

Figure 2: Choice Process with Attributes

The literature on decision processes poses that average decision times depend on the relative choice probabilities and parameters related to the difficulty of the decision problem. Consistent with the spirit of the decision process literature, we assume in the first step that the attributes do not influence the decision process beyond the probabilities of the alternatives being chosen. We can then reduce the data structure to Figure 3.

Given our data in Figure 1, we do not observe which of the choice processes in Figure 3 the decision maker follows. The data in Figure 1 only contain a single entry for the decision time with the given choice probabilities of the three alternatives. Thus, from the perspective of the analyst, all of the three configurations are observationally equivalent. This

[3]In Section 5 we formally introduce attributes of alternatives which allow the decision duration of the choice between two buses to differ from the choice duration between an airplane and a bus.

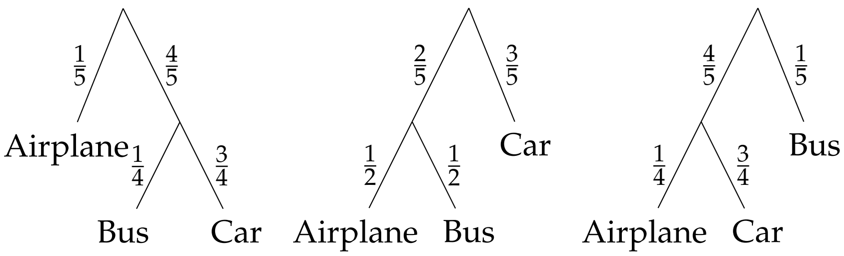

Figure 3: Choice Process with Probabilities

motivates a property we call Associativity. Associativity means that the three processes in Figure 3 are either indistinguishable for the analyst or actually yield identical decision times. Intuitively, Associativity effectively means that we can reduce the three choice processes of Figure 3 to what is shown in Figure 4.

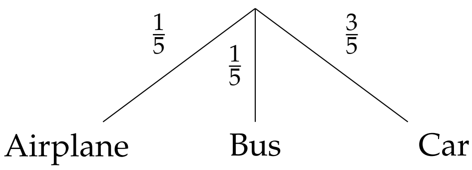

Figure 4: Decision Process without Branches

Finally, in decision processes (at least in their most idealized form) the exact choices (such as airplane, bus, or car) are irrelevant and decision times only depend on the choice probabilities. Thus, we can reduce the decision process further to Figure 5.

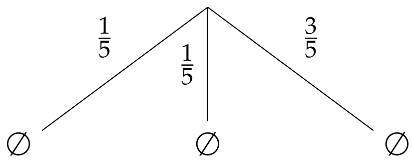

Figure 5: Decision Process without Alternative Details

Suppose now for all decision processes of the form shown in Figure 5 the analyst has gathered a decision time. That is, for any number of options and relative choice probabilities, a decision time has been recorded. We now assume that the decision time of a choice process is increasing in the decision time of any sub-decision process. Going back to Figure 3, we can identify the relevant sub-decision processes as the remaining decision once any subset of alternatives has been eliminated in the decision process. It turns out that this assumption (together with Continuity) allows us to characterize a sharp functional form for the average decision time; the Tsallis and Shannon entropies of the choice probabilities given in Figure 1.

To do so, the next section formally introduces a data structure, procedural mixture sets, and a representation theorem both of which have use outside of this particular example.

3 Procedural Mixtures

For ease of comparison, we first recapitulate the axioms of Herstein and Milnor (1953) which for a mixture set $\mathcal{M}$ are given as follows:

$$\mu a \oplus (1-\mu) b \in \mathcal{M} \tag{1}$$
$$1a \oplus (1-1) b = a, \tag{2}$$
$$\mu a \oplus (1-\mu) b = (1-\mu) b \oplus \mu a, \tag{3}$$
$$\lambda[\mu a \oplus (1-\mu) b] \oplus (1-\lambda) b = (\lambda\mu) a \oplus (1-\lambda\mu) b \tag{4}$$

where each axiom holds for all $a, b \in \mathcal{M}$ and all $\mu, \lambda \in [0,1]$. We may call these axioms, respectively, Closure, Connectedness, Commutativity, and Reduction of Compound Mixtures. λ and μ are the mixture weights. In the context of our example, these are probabilities and we will use these terms interchangeably. An element a is called an *outcome* if there do not exist distinct b and c such that $a = \mu b \oplus (1-\mu) c$ for some $\mu > 0$.

The Reduction of Compound Mixtures axiom is implied[4] by two economically distinct properties, Associativity and Reducibility, which are, respectively:

$$\lambda[\mu a \oplus (1-\mu) b] \oplus (1-\lambda) c = (\lambda\mu) a \oplus (1-\lambda\mu) \left[\frac{\lambda(1-\mu)}{1-\lambda\mu} b \oplus \frac{(1-\lambda)}{1-\lambda\mu} c\right] \tag{5}$$
$$\mu a \oplus (1-\mu) a = a. \tag{6}$$

for all $a, b \in \mathcal{M}$ and all $\mu \in [0,1]$ and $\lambda \in [0,1)$. Associativity states that the order of mixing does not matter. Reducibility expresses that the process of mixing is irrelevant.

[4]Notably, Associativity is not implied by Reduction of Compound Mixtures as Example 1 in Mongin (2001) shows. I thank an anonymous referee for pointing this out.

Example. The classical example of a mixture set is a set of lotteries on a set of alternatives $\mathcal{X}$. These can be formalized as $\{p : \mathcal{X} \to [0,1] | \sum_{x \in \mathcal{X}} p(x) = 1\}$ with a mixture operation fulfilling $(\alpha p \oplus (1-\alpha)q)(x) = \alpha p(x) + (1-\alpha)q(x)$ for all $x \in \mathcal{X}$. Notice that in our example, a decision process in which a decision maker chooses Bus with probability $1/2$ or another Bus is not the same as the trivial decision process in which the decision maker does not make a choice and receives a Bus with certainty. Lotteries and mixture sets would not reflect this since $\frac{1}{2}\text{Bus} \oplus \frac{1}{2}\text{Bus} = \text{Bus}$. In procedural mixture sets we therefore remove the Reducibility axiom to allow the procedural mixture set to distinguish between the decision process involving a choice, $\frac{1}{2}\text{Bus} \oplus \frac{1}{2}\text{Bus} \neq \text{Bus}$ and the trivial decision process involving no choice, Bus. Thus, the mixture operation $\mu a \oplus (1-\mu)b$ means in this context that the decision maker makes a time-consuming decision with probabilities μ and $1-\mu$ between alternatives a and b and that this decision is time-consuming even if the alternatives are effectively identical. *End of example.*

Fishburn (1982) generalized mixture sets by replacing the identity $=$ with an equivalence relation in the mixture axioms. We now remove the axiom of Reducibility and perform a generalization analogous to Fishburn (1982)

Definition 1 (Procedural Mixture Set)**.** A procedural mixture set $\langle \mathcal{S}, \oplus, \approx \rangle$ is a set $\mathcal{S}$ endowed with a mixture operator $\oplus : \mathcal{S} \times \mathcal{S} \times [0,1] \to \mathcal{S}$ and an equivalence relation $\approx$ which fulfills for all $a, b, c \in \mathcal{S}$ and all $\mu \in [0,1]$, $\lambda \in [0,1)$:

$$1a \oplus (1-1)b \approx a, \tag{7}$$

$$\mu a \oplus (1-\mu)b \approx (1-\mu)b \oplus \mu a, \tag{8}$$

$$\lambda[\mu a \oplus (1-\mu)b] \oplus (1-\lambda)c \approx (\lambda\mu)a \oplus (1-\lambda\mu)\left[\frac{\lambda(1-\mu)}{1-\lambda\mu}b \oplus \frac{(1-\lambda)}{1-\lambda\mu}c\right] \tag{9}$$

The Closure axiom of mixture sets is not needed given the definition of the mixture operator. Connectedness and Commutativity remain unchanged. Reduction of Compound Mixtures is replaced by Associativity.

Example. The data structure consisting of entries in the form of rows in Figure 3 is a procedural mixture set given our assumptions about choice probabilities and equivalent decision times discussed in Section 2. To see this, we now specify the set $\mathcal{S}$, the operation $\oplus$, the equivalence relation $\approx$, and the resulting equivalence classes $\mathcal{S}/\approx$.

As the set $\mathcal{S}$ we employ the set of binary weighted trees. A binary weighted tree $M = (N, \ll) \in \mathcal{S}$ is a set of nodes $N \cup \{0\}$ endowed with a relation $\ll \subset (N \cup \{0\}) \times N$ such that for all $n \in N$, there is exactly one $m \in N \cup \{0\}$ such that $m \ll n$ and either no or exactly two distinct $o \in N$ such that $n \ll o$. For each branch $(m, n) \in \ll$ there is a real number $w(m, n) \in [0, 1]$ such that $w(m, n) + w(m, o) = 1$ if $(m, o) \in \ll$. A decision process is a binary weighted tree with each node representing a decision over an unobservable attribute. The relative choice probabilities are given by the weights attached to each branch. Intermediate nodes have the interpretation of attributes while terminal nodes have the interpretation of alternatives.

The mixture operation $\mu a \oplus (1-\mu) b$ of binary trees a and b is the binary tree in which at the origin 0 there are two branches with weights μ and $1-\mu$ after which the subtrees a and b follow, respectively.

Let the terminal weight of a terminal node o be the product $w(0, k) \cdot w(k, l) \cdot \ldots \cdot w(o, n) \cdot w(n, m)$. The equivalence relation $\approx$ identifies binary trees a and b if the tuple of nonzero terminal weights of a is a permutation of the tuple of nonzero terminal weights of b. Thus, each equivalence class can be represented by by an (unordered) tuple $(\mu_1, \ldots \mu_n)$ of numbers $\mu_i \in (0, 1]$ such that $\sum_{i=1}^{n} \mu_i = 1$, as depicted in Figure 5.

It is straightforward to show that $(\mathcal{S}, \oplus, \approx)$ as defined fulfill the procedural mixture set axioms. Connectedness holds since only the nonzero terminal weights matter for the equivalence classes of $\approx$. In the context of our example, this means that alternatives with a zero probability of being chosen can be treated as being not available. Commutativity holds because we can permute the terminal weights within the equivalence classes of $\approx$. In the context of our example, this means that the order in which the alternatives are listed does not matter. Associativity holds because restructuring the tree in a way permitted by Associativity does not change the tuple of terminal weights. In the context of our example, this means that we cannot observe whether a decision maker uses a particular attribute to first narrow down the choices. We do not observe any potential intermediate choice but only the final choice probabilities and decision times across all available alternatives. The equivalence relation $\approx$ goes even further than just fulfilling the procedural mixture set axioms – it also imposes that the exact alternatives do not matter for decision times.

End of example.

We are interested in binary relations on procedural mixture sets. A function $U : \mathcal{S} \to \mathbb{R}$ is called a *representation* of $\succsim$ if $a \succsim b$ if and only if $U(a) \geq U(b)$.

Example. Decision times naturally induce a binary relation on the procedural mixture set discussed above. $a \succsim b$ holds if and only if the decision process a takes at least as long as the decision process b. In this context, the symbol $\succsim$ therefore does not have the common interpretation of "weakly preferred". *End of example.*

In particular, we are interested in the following representation of binary relations on procedural mixture sets.

Definition 2 (Mixture Entropy)**.** A binary relation $\succsim$ on $\mathcal{S}$ is a mixture entropy value if there exists a function $U : \mathcal{S} \to \mathbb{R}$ called a mixture entropy and parameters $q \in \mathbb{R}, r \in \mathbb{R}_{++}$ such that U represents $\succsim$ and for all $a, b \in \mathcal{S}$ and $\mu \in [0, 1]$,

$$\begin{aligned} U(\mu a \oplus (1-\mu)b) =& \mu^r U(a) + (1-\mu)^r U(b) + q \cdot H_r(\mu) \\ H_r(\mu) =& \begin{cases} -\mu \ln \mu - (1-\mu)\ln(1-\mu), & r = 1 \\ \frac{1-\mu^r-(1-\mu)^r}{r-1}. & r \neq 1 \end{cases} \end{aligned} \tag{10}$$

Example. If $\succsim$ represents the decision times, then the decision times must be a monotone transformation T of the representation U. For simplicity, assume for the moment that T is the identity function, i.e., $T(u) = u$.[5] For such a linear T, the parameters that lead to the decision times in Figure 1 are characterized as follows: First, each trivial decision (where only one option is available) takes a *reaction time* of 0.5 seconds. This can be seen as the non-decision component of a response time (Luce, 1986). Second, the most difficult binary decision (in which the choice probabilities are equal) takes 1.5 seconds. Let the *deliberation time* of this decision process be the decision time minus the reaction time, i.e. 1 second. Third, given any decision process, if we replace every final option by repeating the exact same decision process, then the deliberation time doubles. For example, four options that are equally likely to be chosen take twice the deliberation time, i.e., 2 seconds, as two equally likely options. Specifically, the parameters of the mixture entropy value are $r = 1$ and $k = 1/\ln 2$. In the remainder of this section, we provide a set of simple axioms that

[5]In Section 4, we will see in more detail how the parameters affect decision times when T is arbitrary.

characterize when the decision times are an increasing transformation of (10). *End of example.*

Let $\succsim$ be a relation on a procedural mixture set $\mathcal{S}$. We use the symbols $\sim$ and $\succ$ to denote the symmetric and asymmetric parts of $\succsim$. We assume the following classical axioms:

Axiom 1 (Weak Order)**.** $\succsim$ is complete and transitive.

A weak order is *nontrivial* if for some a, b, $a \succsim b$ but not $b \succsim a$.

Example. While transitivity is plausible in the context of decision times, completeness is a fairly strong assumption. It requires that we have sufficiently many alternatives that can be offered to decision makers such that any decision process with an arbitrary probability of choices can be compared with any other decision process in its duration. This might not hold in case the set of alternatives is not rich enough. *End of example.*

Axiom 2 (Continuity)**.** For any $a, b, c \in \mathcal{S}$, the sets $\{\mu | \mu a \oplus (1-\mu)b \succsim c\}$ and $\{\mu | c \succsim \mu a \oplus (1-\mu)b\}$ are closed.

Example. In the context of our example, a plausible violation of Continuity arises for sequences of decision processes where the probability of one alternative being chosen approaches zero. An alternative that is extremely unlikely to be chosen may nonetheless distract decision makers and its presence may increase the duration it takes them to make a decision. *End of example.*

Axiom 3 (Independence)**.** If $a, a', b \in \mathcal{S}$, $\mu \in (0,1)$ then $a \succsim a' \Leftrightarrow \mu a \oplus (1-\mu)b \succsim \mu a' \oplus (1-\mu)b$.

Our Independence axiom needs to be slightly stronger than that of Herstein and Milnor (1953). Reducibility allows them to generate our third axiom from a weaker assumption requiring only indifferences.

Example. In the context of our example, Independence means that the decision duration of a decision process is increasing in the duration of every sub-decision process. In a process that can be written as $\mu a \oplus (1-\mu)b$, the greater the decision time of a, the greater the decision time of $\mu a \oplus (1-\mu)b$. Specifically, consider the comparison of decision times of the decision process between an airplane and a car and between a bus and a car. These only differ on the sub-decision process in case a car is not chosen. If the reaction time of the trivial decision process offering the airplane is just as long as the reaction time of being offered the bus, Independence requires that also the decision process of the choice between an airplane and a car takes just as long as the decision process of a bus and a car. *End of example.*

Theorem 1. *Let $\succsim$ be a binary relation on a procedural mixture set $\langle \mathcal{S}, \oplus, \sim \rangle$. Then the following two statements are equivalent.*

1. *$\succsim$ fulfills axioms 1-3.*
2. *$\succsim$ is an entropy mixture value.*

If U^1 and U^2 are entropy mixture value representations of the same nontrivial weak order, then $r^1 = r^2$, and if $r^1 = 1$, then $U^1 = \phi U^2 + \psi$ and $q^1 = \phi q^2$ and if $r^1 \neq 1$, then $U^1 = \phi U^2 + \psi$ and $q^1 = \phi q^2 + \psi$ where $\phi \in \mathbb{R}_+$ and $\psi \in \mathbb{R}$.

We have (recursively) characterized two possible representations. Either we obtain the expected entropy mixture value of the mixed elements plus the Shannon (1948) entropy. Alternatively, we obtain the expected entropy mixture value under power-form probability distortions plus the Tsallis (1988) entropy. We delay the interpretation of the parameters of the model until discussing their comparative statics in Section 4. We first show that the characterization highlights a difficulty in finding plausible extensions of the drift-diffusion model of decision times.

Example. It is noteworthy that we have only characterized a representation of the decision times and not the exact decision times. Thus, the actual decision times can be any increasing transformation of U. Therefore, the representation with $r = 1$, $q > 0$ and $U(x) = 0$ for trivial decision processes x is for example consistent with any model of *binary* decision processes in which the decision duration of $\mu x \oplus (1 - \mu) y$ is strictly increasing in $\min(\mu, 1 - \mu)$. This holds for the drift-diffusion model of Ratcliff (1978) (but also an infinitude of other models). To obtain the drift-diffusion model, the monotone transformation $T : \mathbb{R} \to \mathbb{R}$ would be

$$T = f \circ (H_1)^{-1} \tag{11}$$

$$f(\mu) = k \cdot \frac{1 - \mu - \mu}{\ln(1 - \mu) - \ln(\mu)} \tag{12}$$

$$(H_1)^{-1}(u) = \min\{\mu \in [0, 1] : H_1(\mu) = u\}. \tag{13}$$

In other words, T first recovers the probability $\mu \leq 1-\mu$ from the entropy $H_1(\mu)$ and then applies the formula for the average decision time in a drift-diffusion model. This seems like forcing the issue but reveals the reason for the difficulty of finding extensions of the drift-diffusion model to multiple options: Suppose this extension fulfills Axioms 1-3 and agrees with the drift-diffusion model on binary choices. Since the decision times of choices among two and three alternatives overlap, the awkward form of T also applies to three-element choices and for such choices $(H_1)^{-1}$ does not recover a choice probability.

Extensions of the drift-diffusion model therefore face the following tradeoff: (1) they may give up on decision times of a process being continuously increasing in the decision times of sub-decision processes or (2) they may give up on choice probabilities fulfilling Luce's IIA axiom or (3) they accept the awkward form of T and try to find a stochastic process and boundary conditions that generate decision times that can be represented by an entropy. *End of example.*

Remark 1. The multi-mixture representations follow from compounding, for example if $r = q = 1$,

$$\begin{aligned} &U\left(\alpha a \oplus (1-\alpha)\left(\frac{\beta}{1-\alpha}b \oplus \frac{\gamma}{1-\alpha}c\right)\right) \\ =&\alpha U(a) - \alpha \ln \alpha + \beta U(b) - \beta \ln \beta + \gamma U(c) - \gamma \ln \gamma. \end{aligned} \tag{14}$$

which is the expectation of U plus the Shannon entropy over three components.

Similar results hold for $r \neq 1$.

In many contexts, the Shannon (1948) measure is the standard measure of entropy. The use of the Shannon entropy in the previous representation entails the following property.

Axiom 4 (Mixture Cancellation)**.** For all $a, a', b, b' \in \mathcal{S}$ and $\mu, \lambda \in (0,1)$,

$$\mu\left(\frac{\mu}{\mu+\lambda}a \oplus \frac{\lambda}{\mu+\lambda}a\right) \oplus (1-\mu-\lambda)b \sim \mu\left(\frac{\mu}{\mu+\lambda}a' \oplus \frac{\lambda}{\mu+\lambda}a'\right) \oplus (1-\mu-\lambda)b' \tag{15}$$

$$\Leftrightarrow \quad (\mu+\lambda)a \oplus (1-\mu-\lambda)b \sim (\mu+\lambda)a' \oplus (1-\mu-\lambda)b'. \tag{16}$$

It is straightforward to apply Mixture Cancellation to Theorem 1 to obtain the following corollary.

Corollary 1. *Let $\succsim$ be a binary relation on a procedural mixture set $\langle \mathcal{S}, \oplus, \sim \rangle$. Then the following two statements are equivalent.*

1. *$\succsim$ fulfills axioms 1-4.*
2. *$\succsim$ is an entropy mixture value with $r = 1$.*

Example. In the context of decision processes Mixture Cancellation has a straightforward interpretation: Let a be a more difficult decision than a' and b' be more difficult than b exactly such that (16) holds. The decision processes in (15) can be understood as identical to the ones in (16) except that before[6] the sub-decisions a and a' an additional decision with relative probability $\frac{\mu}{\mu+\lambda}$ is performed. The condition thus says that the different deliberation times of a and a' do not "interact" with the additional deliberation time from adding an additional decision. An example of such choices are given by the domain of choices that follow Hick's law (Hick, 1952). Hick observed that with remarkable precision the response time to press a button in response to a signal increases logarithmically in the number of buttons, similar to how the Shannon entropy of uniform variables increases logarithmically in the number of outcomes. This suggests that $r = 1$ and T being the identity are suitable to model exact decision times for choices where Hick's law applies.

However, this might not be the case if the decision makers become increasingly constrained in their decision making capability as the number of options increases. In this case, we would expect that if a takes longer than a', then

$$\begin{aligned} (\mu+\lambda)a \oplus (1-\mu-\lambda)b \quad &\sim \quad (\mu+\lambda)a' \oplus (1-\mu-\lambda)b' \\ \Rightarrow \quad & \mu\left(\frac{\mu}{\mu+\lambda}a \oplus \frac{\lambda}{\mu+\lambda}a\right) \oplus (1-\mu-\lambda)b \\ \succ \quad & \mu\left(\frac{\mu}{\mu+\lambda}a' \oplus \frac{\lambda}{\mu+\lambda}a'\right) \oplus (1-\mu-\lambda)b'. \end{aligned} \tag{17}$$

That is, the additional decision with relative probability $\mu/(\mu+\lambda)$ interacts with the decision processes a and a' such that decision times increase more when this decision precedes the more complicated decision process a. As we will see in Section 4, such behavior is closely linked to the parameter r. *End of example.*

[6]By Associativity, the additional decision could also be performed after a or a'.

We end this section with a corollary that applies the main theorem to stochastic choice models and that makes some of the informal discussion of the example application precise. Given the close link between Luce's IIA of decision probabilities and Associativity, it is natural to simultaneously characterize the Luce model choice probabilities and entropy mixture decision times.

Let $\mathcal{X}$ be a set of alternatives and $\mathcal{C}$ be the set of finite subsets of $\mathcal{X}$. A stochastic choice function is a function $p : \mathcal{C} \times \mathcal{X} \to [0,1]$ such that for all $C \in \mathcal{C}$, $x \notin C$, $p(x, C) = 0$ and $\sum_{x \in C} p(x, C) = 1$. For any $C \subseteq D \in \mathcal{C}$, we further define $p(C, D) = \sum_{x \in C} p(x, D)$. A decision time $\tau : \mathcal{C} \to \mathbb{R}_+$ is a function that tells us for every finite subset of alternatives how long it takes (on average) to make a decision.

We introduce the following joint model of choice probabilities and decision times:

Definition 3 (Luce-Hick Model)**.** A stochastic choice function p and a decision time τ form a Luce-Hick model if

1. there exists a function $v : \mathcal{X} \to \mathbb{R}$ such that for all $C \in \mathcal{C}$ and $x \in C$,
$$p(x, C) = \frac{\exp(v(x))}{\sum_{y \in C} \exp(v(y))}, \text{ and} \tag{18}$$
2. there exists a continuous, strictly monotone function T and $r \in \mathbb{R}_{++}$ such that for all $C \in \mathcal{C}$,
$$T \circ \tau(C) = \begin{cases} \frac{1}{r-1} \left(1 - \sum_{x \in C} p(x, C)^r\right) & r \neq 1 \\ \sum_{x \in C} p(x, C) \ln p(x, C) & r = 1 \end{cases} \tag{19}$$
and $\tau(\{x\}) = \tau(\{y\}) = T^{-1}(0)$ for all $x, y \in \mathcal{X}$.

That is, in the Luce-Hick model the choice probabilities follow the Luce model of stochastic choice and a monotone transformation of the decision times (of equiprobable decisions) follows Hick's law. Compared with the empirical results of Hick (1952), the above definition makes the stronger claim that also the decision times of non-equiprobable decision processes can be represented by an entropy but neither requires T to be linear nor the entropy to be in Shannon form, i.e., $r = 1$.

In order to characterize the Luce-Hick model via the procedural mixture set theorem, we require a sufficiently rich set of outcomes to generate decision processes with arbitrary choice probabilities.

Definition 4 (Richness of Outcomes)**.** The set of alternatives $\mathcal{X}$ fulfills richness if for every $x \in \mathcal{X}$ and every $\mu \in [0,1]$ there is a countable number of alternatives $\{y_1, y_2, \ldots\}$ such that $p(x, \{x, y_i\}) = \mu$.

We now introduce conditions that (given a rich set of alternatives) are necessary and sufficient to characterize the Luce-Hick model.

Definition 5 (Positivity)**.** A stochastic choice function p and a decision time τ fulfill positivity if for all $x, y \in \mathcal{X}$, $p(x, \{x, y\}) > 0$ and $\tau(\{x, y\}) > \tau(\{x\})$.

Thus, every element of an opportunity set has a nonzero probability of being chosen and there is a positive deliberation time for the choice between two items.

Definition 6 (Independence of Irrelevant Alternatives)**.** A stochastic choice function p fulfills IIA if for all $C \in \mathcal{C}$ we have that

$$p(x, \{x, y\}) / p(y, \{x, y\}) = p(x, C \cup \{x, y\}) / p(y, C \cup \{x, y\}) \tag{20}$$

for all $x, y \in C$.

Luce's choice axiom states that relative probabilities are unaffected by the addition of other options. We next impose that comparative decision times are unaffected by additional options.

Definition 7 (Independent Decision Times)**.** A stochastic choice function p and decision time τ fulfill independence of decision times if for all $C, D, E \in \mathcal{C}$ such that $(C \cup D) \cap E = \emptyset$ and $p(C, C \cup E) = p(D, D \cup E)$ it holds that

$$\begin{aligned} & \tau(C) \geq \tau(D) \\ \Leftrightarrow \quad & \tau(C \cup E) \geq \tau(D \cup E). \end{aligned} \tag{21}$$

This states that the decision time of a decision process is monotone in the decision time of its subprocesses, i.e., the time it would take to make a choice from a subset of the alternatives.

Definition 8 (Continuity of Decision Times)**.** A stochastic choice function p and decision time τ fulfill continuity of decision times if for all sequences of sets $(A^k \equiv \{a_1^k, \ldots, a_n^k\})_k$ and $A = \{a_1, \ldots, a_m\}$, if $p(A^k, a_i^k) \to p(A, a_i)$ for all $i \in \{1, \ldots, m\}$ and $p(A^k, a_i^k) \to 0$ for all $i \in \{m+1, \ldots, n\}$ then $\tau(A^k) \to \tau(A)$.

Continuity has two main implications. First, it imposes that decision times are continuous in the choice probabilities of alternatives and only these choice probabilities matter for decision times. Second, it imposes that given a limit n on the number of alternatives, as choice probabilities of some ($m - n$ many) alternatives converge to zero, they do not affect the decision times.

The following Corollary is now obvious:

Corollary 2 (Choice Probabilities and Decision Times)**.** *Suppose $\mathcal{X}$ fulfills richness in outcomes. Then the following statements are equivalent.*

1. *p and τ fulfill Positivity, IIA, Independence of Decision Times, and Continuity of Decision Times.*

2. *p and τ form a Luce-Hick model.*

Interestingly, for very different reasons Luce was aware of the importance of his choice axiom for the use of entropy in psychophysics, writing "[...] information theory implicitly presupposes the consequences of [IIA], which are relatively strong—specifically, when discrimination is imperfect, it means that choice behavior can be scaled by a ratio scale" Luce (1959, p.12).

4 Comparative Statics

In the value of a procedural mixture, $U(\mu a \oplus (1-\mu)b)$, the parameter q sets a threshold for $U(a)$ and $U(b)$ that determines whether mixing increases U or not. To make this precise, we introduce a positive and a negative value of mixing.

Definition 9 (Value of Mixing)**.** $\succsim$ exhibits a negative (positive) value of mixing at $a \in \mathcal{M}$ if $a \succ (\prec) \mu a \oplus (1-\mu)a$.

The following result is then straightforward:

Proposition 1 (Monotone Mixing)**.** *If $\succsim$ has a mixture entropy representation, then the following statements are equivalent:*

1. *$\succsim$ exhibits a negative (positive) value of mixing at a,*

2. *$\succsim$ exhibits a negative (positive) value of mixing at $\mu a \oplus (1-\mu)a$,*

3. $U(a)(r-1) > (<)q$.

If for some $a \in \mathcal{S}$ it is the case that $U(a)(r-1) \leq q$, then $U(\mu a \oplus (1-\mu)a) \geq U(a)$. From this result follows that if $r > 1$, then iteratively mixing an element with itself yields a sequence of elements for which U converges to q. If $r \leq 1$, then U diverges to ∞ or $-\infty$, depending whether for the initial element $U(a)(r-1) \gtrless q$.

Example. In our example data of Figure 1, the value of mixing is positive. If $r > 1$, then the mixture entropy U converges to q as the number of options increases. If $r \leq 1$, then additional options let U diverge to ∞. It is noteworthy that this does not mean that decision times diverge. Since the representation is ordinal, the actual decision times $T \circ U$ may still be bounded in case $\lim_{u\to\infty} T(u) < \infty$. Thus, limit behavior of decision times alone does not allow us to distinguish between $r < 1$ and $r > 1$. *End of example.*

In addition to setting a threshold for a positive value of mixing, q controls in a procedural mixture (e.g., $U(\mu a \oplus (1-\mu)b)$) the relative importance of value derived from the mixture weight ($H_r(\alpha)$) compared with the value from the mixed elements ($U(a)$ and $U(b)$). The relevant comparative statics results are relegated to Appendix D because these results are only relevant if there are outcomes $x \succ y$ as the following remark shows:

Remark 2. If $\mathcal{S}$ is generated from finite procedural mixtures of a set $\mathcal{X}$ and for all $x, y \in \mathcal{X}$, $U^1(x) = U^1(y)$, and $U^2(x) = U^2(y)$ then U^1 and U^2 represent the same relation if and only if the signs of $U^1(x)(r^1-1) - q^1$ and $U^1(x)(r^1-1) - q^1$ are identical and $r_1 = r_2$.

It follows from this remark that the magnitude of the parameter q only matters in comparison to a cardinal value of outcomes, (i.e., elements that are not generated from mixtures themselves). If there do not exist outcomes $x \succ y$, then by the uniqueness properties of the representation we can find an affine transformation of U such that the valuation of any existing outcomes is equal to zero. Any subsequent multiplication of q by a positive factor results in an increasing linear transformation of U (which does not change the represented relation).

Example. The previous remark is the underlying reason why the Luce-Hick model only has a parameter r and no parameter q. It is plausible that trivial decisions always have the same reaction time and that nontrivial decisions take longer than trivial decisions. Thus, if $\mathcal{X} \ni x, y$ refers to the set of trivial decisions and $\tau(x) = \tau(y)$ for all its elements, then the only meaningful parameter is r. *End of example.*

We now turn to the interpretation of the parameter r. The parameter r controls the degree of the effect of mixing on the value. That is, it controls how much the value increases (or decreases) by an additional mixing stage.

Definition 10 (Comparative Value of Mixing)**.** $\succsim_1$ yields a higher value of mixing than $\succsim_2$ if for all $\alpha, \beta, \gamma < 1/2$ and some $d \in \mathcal{S}$ at which $\succsim_1$ and $\succsim_2$ exhibit a positive value of mixing,

$$\alpha d \oplus (1-\alpha)d \quad \succsim_1 \quad \beta(\gamma d \oplus (1-\gamma)d) \oplus (1-\beta)(\gamma d \oplus (1-\gamma)d),$$

then

$$\alpha d \oplus (1-\alpha)d \quad \succsim_2 \quad \beta(\gamma d \oplus (1-\gamma)d) \oplus (1-\beta)(\gamma d \oplus (1-\gamma)d).$$

In words, if under $\succsim_2$ a binary mixture has a greater value than a mixture across four elements, then this must be also the case under $\succsim_1$.

Proposition 2. *Let $\succsim_1$ and $\succsim_2$ be a mixture entropy value with representations U_1 and U_2 and parameters r_1, q_1 and r_2, q_2, respectively. Then the following statements are equivalent.*

1. *$\succsim_1$ yields a higher value of mixing than $\succsim_2$.*

2. $r_1 \geq r_2$.

Example. The decision data of Hick (1952) suggest that (given a linear T), $r = 1$ is a plausible parameter for the decision of which one of a number of buttons n on a keyboard to press. This suggests that the decision times of let's say whether to press a button with the left or right hand does not affect the additional decision time from choosing whether to press with the index finger or the pinkie. In contrast, preferential choices such as the commonly in experiments studied snack choices may become increasingly complex as the number of alternatives rises. Choices between food items may be relatively simple between two items but may become increasingly complex as additional options are added. Preferential choices would then exhibit a higher value of mixing, leading to a different parameter r. *End of example.*

5 Disjoint Independence

Violations of interiority (and thus mixture entropy representations) do not necessarily need to arise from a violation of reducibility. Instead, they may also arise from a weakening of the Independence axiom to mixtures between elements with disjoint support. In this case, only a subset of the relation $\succsim$ can be represented by the functional form characterized previously. The present section makes this precise.

Example. In the introductory example, we assumed that choice relevant attributes are irrelevant for decision times once we control for the choice probabilities. This assumption may only be valid for a carefully chosen set of options. In our example data of Figure 1, in the choice between two buses each is chosen with equal probability just as in the choice of a bus and an airplane each is chosen with equal probability. However, the choice between the similar alternatives is perhaps made much faster than the choice between the more different alternatives. We may therefore want to construct a model that allows for the decision times of these two decision processes to differ.

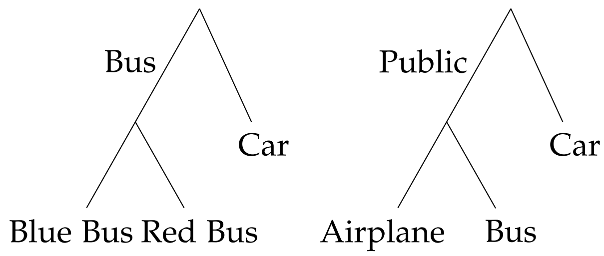

Figure 6: Blue Bus Red Bus Paradox

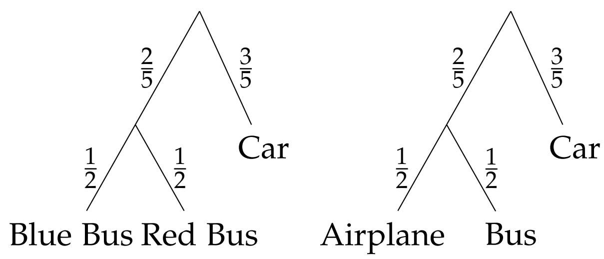

Figure 7: Blue Bus Red Bus Paradox

To exemplify this more clearly, Figure 7 depicts the well-known red bus/blue bus paradox of Debreu (1960). There are two major concerns. First, the IIA axiom of Luce (1959) may not hold if the substitutability between the blue and the red bus is higher than the substitutability between either bus and the car. For example, even if the blue bus and red bus are equally likely and either bus is equally likely to an airplane, the probability of choosing a car in the left decision process may be more likely than the probability of choosing a car in the right decision process.

Second, the decision time on the left may be lower than the decision time on the right. Making a choice between the blue bus and the red bus may be very fast once one realizes that the two buses take the same route. The choice between either colored bus and an airplane may involve more careful deliberation. Thus, decision times might not only depend on choice probabilities but also on the attributes by which alternatives are distinguished. Note that a faster decision time on the left than on the right also violates Independence in case that the reaction times of all trivial decision processes are identical.

To account for these two issues, we need to have a notion of how different two alternatives are. We can formalize this via attributes. Following Nehring and Puppe (2002) for a given universe of alternatives $\mathcal{X}$, an *attribute* is a subset of $\mathcal{X}$ with the interpretation that it contains all elements of $\mathcal{X}$ that share this attribute. We define $\mathcal{A} \subset 2^{\mathcal{X}}$ as a set of *choice-relevant attributes*. The set of attributes can be exogeneously given or is under some circumstances identifiable from choice data (Kovach & Tserenjigmid, 2022). *End of example.*

Like in Section 3, we first provide a formal result for (this time classical) mixture sets and then provide an application to decision times that uses the above introduced attribute structure. Let $\mathcal{Z}$ be a set. For a given mixture set $\mathcal{M}$, a *support* is a function $supp : \mathcal{M} \to 2^{\mathcal{Z}}$ that fulfills for all $a, b \in \mathcal{M}$ and all $\alpha \in (0,1)$: $supp(\alpha a \oplus (1-\alpha)b) = supp(a) \cup supp(b)$. We say that $\mathcal{M}$ is *rich* (with respect to its support) if the image $supp(\mathcal{M})$ is closed under nonempty intersections and under nonempty relative complements. A subset Z of $\mathcal{Z}$ is *essential* if there exist $a, b \in \mathcal{M}$ such that $supp(a) \subseteq supp(b) = Z$ and $a \not\sim b$.

Example. Suppose $\mathcal{A}$ is a set of choice-relevant attributes. We can then generate a support by letting $\mathcal{Z} = 2^{\mathcal{A}}$ be the power set of choice relevant attributes. The support $supp(x) = \{A \in \mathcal{A} | x \in A\}$ then assigns every alternative $x \in \mathcal{X}$ the set of its choice relevant attributes. Any mixture between two alternatives with nonzero mixture weights has the union of the attributes of the alternatives involved in the mixture, etc.. *End of example.*

Axiom 5 (Disjoint Independence)**.** A relation $\succsim$ on a mixture set $\mathcal{M}$ with a support $supp$ fulfills disjoint independence if for all $a, a', b \in \mathcal{M}$, $\mu \in (0,1)$, if $(supp(a) \cup supp(a')) \cap supp(b) = \emptyset$ then

$$\begin{aligned} & a \succsim a' \\ \Leftrightarrow \quad & \mu a \oplus (1-\mu) b \succsim \mu a' \oplus (1-\mu) b. \end{aligned} \tag{22}$$

We can embed a mixture set $\mathcal{M}$ with a relation $\succsim^*$ fulfilling Axioms 1,2, and 5 partly into a procedural mixture set $\mathcal{S}$ with a relation $\succsim$ fulfilling Axioms 1-3. If there are sufficiently many essential subsets in the support, the embedded part has the same uniqueness properties as the mixture entropy representation and we obtain the following theorem:

Theorem 2 (Procedural Mixture Set Embedding)**.** *Let $\mathcal{M}$ be a rich mixture set with support supp and $\succsim^*$ be a binary relation on $\mathcal{M}$ such that there exist at least three disjoint and essential subsets of $\mathcal{X}$.*

Then the relation $\succsim^*$ fulfills Axioms 1, 2, and 5 if and only if there exists a continuous function $U : \mathcal{M} \to \mathbb{R}$ representing $\succsim^*$ such that for some $q \in \mathbb{R}, r \in \mathbb{R}_{++}$

$$U(\mu a \oplus (1-\mu) b) = \mu^r U(a) + (1-\mu)^r U(b) + q \cdot H_r(\mu) \tag{23}$$

$$H_r(\mu) = \begin{cases} -\mu \ln \mu - (1-\mu)\ln(1-\mu), & r = 1 \\ \frac{1-\mu^r-(1-\mu)^r}{r-1}, & r \neq 1 \end{cases} \tag{24}$$

if $supp(a) \cap supp(b) = \emptyset$.

Chen and Rommeswinkel (2020) prove a similar result for four disjoint subsets using a different proof technique.

Example. Since a characterization requires a sufficiently rich set of decision processes, we still require a rich set of outcomes such that for any two given goods a and b with characteristics A and B, we still need to be able to induce all possible choice probabilities between 0 and 1. This is implausible for goods with a high degree of standardization (e.g., candy bars) but not implausible for goods that are less standardized (e.g., more or less visually appealing fruits). The perhaps easiest way to achieve this experimentally may be to bundle the goods of different attributes with different monetary amounts (or changing the likelihood/timing of receiving the good) to induce different choice probabilities. The remainder of this section makes this idea precise. *End of example.*

We now suppose that the decision maker chooses between combinations of one of a (possibly finite) set of material alternatives $x \in \mathcal{X}$ with choice-relevant attributes $\mathcal{A}$ and payoffs. If the decision maker chooses x from a finite set $C \in \mathcal{C}$, then she receives in addition the monetary payoff $m(x) \in \mathbb{R}_+$. $\mathbb{R}_+^{\mathcal{X}}$ denotes the set of possible payoff functions. Importantly, and in contrast to Section 3, we do not allow the same alternative $x \in \mathcal{X}$ to appear more than once among the possible options. Since we presume that we know the decision relevant set of attributes, we can model similar options by making their attributes identical.

A *stochastic choice function* is in this section a function $p : \mathcal{X} \times \mathcal{C} \times \mathbb{R}^{\mathcal{X}} \to [0, 1]$ that returns a choice probability $p(x, A, m)$ of an alternative x being chosen from a set X given payoffs m and fulfills firstly that $p(x, C, m) = 0$ if $x \notin C$ and secondly that $p(x, C, m) = p(x, C, m')$ if $\forall y \in C$, $m(y) = m'(y)$.

Definition 11 (Restricted Luce-Hick Model)**.** A stochastic choice function p and a decision time τ form a restricted Luce-Hick model if

1. for all $C \in \mathcal{C}$, $\tau(C, m) \equiv f(\{p(x, C, m)\}_{x \in C})$ is continuous, and
2. there exists a function $v : \mathcal{X} \to \mathbb{R}$ such that for all $C \subseteq D$ and $x \in C$,
$$p(x, C, m) = \frac{\exp(v(x), m(x))}{\sum_{y \in C} \exp(v(y), m(y))} \tag{25}$$
whenever all elements in D have mutually disjoint supports, and
3. there exists a continuous, strictly monotone function T and a parameter $r \in \mathbb{R}_{++}$ such that for all $C \subseteq D$,
$$\begin{aligned} T \circ \tau(D, m) =& p(C, D, m)^r \cdot T \circ \tau(C, m) + p(D - C, D, m)^r \cdot T \circ \tau(D - C, m) \\ &+ H_r(p(C, D, m)) \end{aligned} \tag{26}$$
whenever all elements in D have mutually disjoint supports.

To characterize the Restricted Luce-Hick model, we make the following axiomatic impositions.

Definition 12 (Positivity)**.** For all $C \in \mathcal{C}$, $x, y \in C$, $m \in \mathbb{R}^{\mathcal{X}}$, $p(x, C, m) > 0$ and $\tau(\{x, y\}, m) > \tau(\{x\}, m)$.

Thus, every alternative in any opportunity set has a strictly positive chance of being chosen and deliberation times of binary choices are positive.

Definition 13 (Monotonicity in Payoffs)**.** For all $C \in \mathcal{C}$ and $x \in C$, if $m(x) \geq m'(x)$ and for all $y \in C - \{x\}$, $m(y) \leq m'(y)$, then $p(x, C, m) \geq p(x, C, m')$ and if $\{m^k(x)\} \to \infty$ and $\{m^k(y)\} \to m(y)$, then $\{p(x, C, m^k)\} \to 1$.

Thus, for a fixed set of alternatives, the choice probability is non-decreasing in the payoff of that alternative and non-increasing in the payoffs of other alternatives. As the payoff of a single alternative goes to infinity, its probability to be chosen converges to 1.

Definition 14 (Continuity in Payoffs)**.** If $\{m^k\} \to m$, then $\{p(x, C, m^k)\} \to p(x, C, m)$ for all $C \in \mathcal{C}$ and all $x \in C$. Moreover, whenever for some sequence $\{m^k\}$ it holds that $\{p(x, C, m^k)\} \to p(x, C, m)$ for all $x \in C$, then $\{\tau(C, m^k)\} \to \tau(C, m)$.

Thus, for a fixed set of alternatives, if the payoff function converges, then the choice probabilities also converge. If the probabilities converge, then the decision times converge, too.

Definition 15 (Restricted IIA)**.** A stochastic choice function p fulfills restricted IIA if for all $D \in \mathcal{C}$ containing only elements with mutually disjoint support we have that

$$p(x, \{x, y\}) / p(y, \{x, y\}) = p(x, D) / p(y, D) \tag{27}$$

for all $x, y \in D$.

Restricted IIA effectively means that choice probabilities are only required to follow the Luce (1959) model in case their attributes do not overlap with those of other alternatives.

Definition 16 (Restricted Decision Time Independence)**.** A decision time τ fulfills restricted decision time independence if whenever all elements of $D \in \mathcal{C}$ have mutually disjoint supports, $A, B, C \subset D$ fulfill $(A \cup B) \cap C = \emptyset$, and $m, m' \in \mathbb{R}^{\mathcal{X}}$ are such that for all $y \in C$, $p(y, A \cup C, m) = p(y, B \cup C, m')$ it holds that

$$\begin{aligned} & \tau(A, m) \geq \tau(B, m') \\ \Leftrightarrow \quad & \tau(A \cup C, m) \geq \tau(B \cup C, m'). \end{aligned} \tag{28}$$

Thus, for elements with disjoint attributes, comparative decision times are independent of common alternatives that have the same probability.

Corollary 3 (Decision Time Representation)**.** *Suppose $\mathcal{X}$ is endowed with a support $supp$ and there are at least six alternatives with disjoint support from all other alternatives. Then the following statements are equivalent.*

1. (p, τ) *fulfill Positivity, Monotonicity in Payoffs, Continuity in Payoffs, Restricted IIA, and Restricted Decision Time Independence.*
2. (p, τ) *have a Restricted Luce-Hick representation.*

The assumption of six alternatives having no attributes in common with any other alternative is the simplest way of guaranteeing that the support of the mixture set induced by the attributes is coarse enough to uniquely determine r and q. Interesting variations of Corollary 3 are possible. For example, in the nested logit the IIA axiom holds whenever the two alternatives x and y have the same attributes as the alternatives in the sets C or whenever the alternatives both have distinct attributes from the alternatives in C (Kovach & Tserenjigmid, 2022). Thus, in the nested logit we have an associative structure not only in the disjoint attribute case but also in the identical attribute case and can use Theorem 1 to embed a procedural mixture set to characterize decision times which may have different parameters r and q for different subsets of alternatives.

6 Literature

There are three branches of literature related to the present paper; a literature on axiomatic characterizations of entropies, a literature on axiomatic characterizations of decision times, and a literature on mixture sets and relaxations of the Reduction of Compound Mixtures axiom.

6.1 Characterizations of Entropy Functions

For a general survey of the literature of the characterization of information measures, see Csiszár (2008).

Krantz et al. (1971, ch. 3.12) defined entropy structures and showed that a relation represented by H_r fulfills the assumptions of an entropy structure. However, they did not provide a characterization result of H_r or H_1. Their operation $\circ$ of an entropy structure captures the idea of

$a \circ b$ denoting the physical system consisting of two independent physical systems a and b. The mixture operation $\mu a \oplus (1-\mu)b$ instead better applies to the mixture of distinguishable gases or liquids a and b with proportion μ and thus our representation captures the so-called *entropy of mixing*.

Closely related to the present paper is Luce et al. (2008b) in which the utility of gambling is characterized as expected utility plus the entropy of the lottery. There are three technical improvements the present paper makes. First, Luce et al. (2008b) assume the existence of a status quo consequence and directly impose that the utility of a gamble between the status quo and some outcome ocurring in some event is separable. Second, Luce et al. (2008b) assume that outcomes and gambles are closed under an operation they call "joint receipt", interpreted as receiving two gambles simultaneously. They further assume that the utility over the two received gambles is additive, i.e., that the utility of the joint receipt of lotteries is the sum of the individual lotteries. Preferences over the gambles are therefore independent and thus the decision maker's risk attitude over one gamble may not be influenced by whether the second gamble is risky or not. Third, Luce et al. (2008b) assume the existence of kernel equivalents. The kernel equivalent of a gamble is an outcome that when received simultaneously with an event-resolving but payoff-irrelevant gamble leaves the decision maker indifferent. Overall, their axioms are somewhat nonstandard and lack the accessibility of the mixture sets introduced in Herstein and Milnor (1953).

We show that only small adjustments to the standard axioms need to be made to obtain entropy measures as a utility component. We do not assume the joint receipt of gambles or the existence of a status quo outcome. Additive separability instead naturally arises from the von Neumann-Morgenstern Independence axiom. While Luce et al. (2008b) assumes that certainty equivalents and kernel equivalents exist, our model and axiomatization are consistent with the nonexistence of certainty equivalents such as in the case when the mixture set is generated starting from mixtures of a finite set of alternatives, mixtures of these mixtures, etc..

The literature on rational inattention has provided characterizations of expected utility with entropy costs of attention (Caplin et al., 2017; de Oliveira et al., 2017). Ellis (2018), Lin (2020), and Lu (2016) characterize more general information cost functions. Most of this literature relies on observations of choices over menus or alternatives and treats choices over information structures as unknown.

6.2 Decision Processes and Axiomatizations of Response Times

With respect to our application to decision processes, there exists a large literature on decision processes (Bogacz et al., 2006; Usher et al., 2013) especially the drift-diffusion model (Ratcliff, 1978; Ratcliff et al., 2016). Several studies have proposed models for decision times in decision processes with multiple alternatives (Baldassi et al., 2020; Krajbich & Rangel, 2011; McMillen & Holmes, 2005; Tajima et al., 2019). Most closely related to the present paper are the axiomatic studies of Baldassi et al. (2020) and Echenique and Saito (2017).

Like the present paper, Baldassi et al. (2020) also employ the Luce model to characterize decision times. They obtain the decision times of the drift diffusion model for binary choice by imposing that across decisions the accuracy (in our notation $\max(\mu, 1-\mu) - 1/2$) is proportional to the product of the average decision time multiplied by the ease of comparison $\ln(\max(\mu, 1-\mu)) - \ln\min(\mu, 1-\mu)$. This ad hoc functional imposition generates decision times as in the drift-diffusion model for binary choices. Their multi-alternative extension of the DDM, the Metropolis-DDM algorithm, is not axiomatically characterized but made plausible by following certain stylized facts from eye-tracking data for multi-alternative choice. In contrast to the extremely parsimonious model of the present paper, this leads to a model rich in parameters.[7]

Echenique and Saito (2017) axiomatize response times for binary choice data. Similar to the present study, they obtain a representation of response times instead of a direct characterization of the functional form of response times. Different from the present study, they work with deterministic choices and distinguish between the response time to choose a over b and the response time to choose b over a. Most importantly, they address the issue of finite data while the present study requires a rich data set fulfilling (at least for a subset of the options) the IIA axiom.

[7] For every decision time limit, their model parameters include a measure across all offered options, a partition of the offered options into consideration sets, and for each consideration set an exploration matrix the size of the set.

6.3 Mixture Sets and the Reduction of Compound Mixtures Axiom

There is a vast literature of decision theoretic papers that employ the mixture sets introduced by Herstein and Milnor (1953). Commonly the axioms on preferences are being varied instead of the structure of the mixture set. In contrast, Mongin (2001) examines under which conditions mixture sets can be treated as convex subsets of a vector space. Ghirardato et al. (2003) and Ghirardato and Pennesi (2020) further increase the applicability of mixture set results by showing how (subjective) mixture sets can be constructed from preferences over acts.

The Reduction of Compound Mixtures axiom has received substantial attention. The literature on recursive utility models following Kreps and Porteus (1978) analyzes intertemporal decision problems in which within each time period the Reduction of Compound Mixtures axiom holds but between time periods it does not. A large literature following Segal (1990) remove the Reduction of Compound Mixtures assumption completely and study two-period mixtures under various axioms on preferences. In contrast to their work, this paper studies mixtures with an arbitrary, finite number of stages and maintains that the order of resolution of compounding is irrelevant while the process of mixing is not.

7 Conclusion

Our analysis provides a foundation for the study of violations of interiority due to procedural aspects. The entropy representation is obtained by relaxing the assumption that mixtures of an element with itself yields the same element or by weakening the Independence axiom to mixtures of sufficiently distinct elements. Entropy measures play an important role in a large number of applications and the simple axiomatization provided in this paper may prove useful in other contexts.

The application to decision processes provides a parsimonious model of the relation between choice probabilities and decision times. It is perhaps striking that the central prediction of the drift-diffusion model — that decision times are monotone in how even the choice probabilities are — is obtained from very simple assumptions about decision times for choices between multiple alternatives. However, the result can also be understood as an impossibility result; if one accepts that the decision time of a decision process should increase in the decision times of its subsets and the choice probabilities fulfill IIA, then one has to accept that decision times are related to choice probabilities in the somewhat restricted functional form of an entropy. While the representation is ordinally consistent with the binary drift-diffusion model, the very different functional form highlights the difficulty of extending the drift-diffusion model to multiple options.

The application opens up two interesting research avenues. First, it is interesting to consider variations of the decision times in the drift-diffusion model by varying either the stochastic process or the calculation of the average response time (e.g., using the mode or a generalized mean of the stopping time). Second, the result on Disjoint Independence provides a starting point for characterizations in which the entropy represents decision times only on the domain of choices where the Luce model is plausible. There is much recent interest in the literature on axiomatic foundations of variations of the Luce model, e.g., the mixed logit (Saito, 2018), the nested logit with subjective attributes (Kovach & Tserenjigmid, 2022), or the conditional logit (Breitmoser, 2020). Axiomatically studying decision times for such models would be an interesting avenue for future research given the prevalence of stochastic choice in empirical applications.

Acknowledgements

I thank two anonymous referees and the editor for extensive comments. This work was financially supported by the Center for Research in Econometric Theory and Applications (Grant No. 109-L900-203) from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan, and by the Ministry of Science and Technology (MOST), Taiwan, under Grant No. 107-2410-H-002-031 and 109-2634-F-002-045.

A Proof of Theorem 1

Proof. Neccessity is straightforward. We prove sufficiency.

Let $\mathcal{Q} = \mathcal{S}/\sim$ be the quotient set of $\mathcal{S}$ with respect to the equivalence relation $\sim$. Note that whenever $a, b \in \mathcal{S}$ and $a \sim b$, then any utility representation U must fulfill $U(a) = U(b)$. Note further that $\mathcal{Q}$ is a procedural mixture set when endowed with $\succsim^*$ such that $q \succsim^* r$ if and only if $a \succsim b$ for some $a \in q$ and $b \in r$. Thus, finding a utility on $\mathcal{Q}$ is equivalent to finding a utility on $\mathcal{S}$. We therefore assume for the remainder of the proof that $\mathcal{S} = \mathcal{Q}$.

Let the order topology on $\mathcal{S}$ be the topology generated by the subbase of upper and lower contour sets of the asymmetric part of $\succsim$.

Lemma 1. *$\mathcal{S}$ is topologically connected under the order topology.*

Proof. If $\mathcal{S}$ is not connected, then it is the union of two nonempty disjoint open sets $\mathcal{S}'$ and $\mathcal{S}''$. Take any elements $s' \in \mathcal{S}'$ and $s'' \in \mathcal{S}''$. The set $\mathcal{S}''' = \{a | \exists \mu : a = \mu s' \oplus (1-\mu)s''\}$ is disconnected in the subspace topology by the disjoint nonempty open sets $S' \cap S'''$ and $S'' \cap S'''$. Since the upper and lower contour sets of $\succsim$ form a subbase of $\mathcal{S}$, the upper and lower contour sets of $\succsim$ in $\mathcal{S}'''$ form a subbase of the subspace topology. By Axiom 2, the bijection $f : \mu \mapsto \mu s' \oplus (1-\mu)s''$ is then continuous. But then the preimages $f^{-1}(\mathcal{S}' \cap \mathcal{S}'')$ and $f^{-1}(\mathcal{S}'' \cap \mathcal{S}''')$ are open, disjoint, and disconnect the unit interval, a contradiction. $\square$

Lemma 2. *$\succsim$ is coseparable, i.e.,*

$$\mu a \oplus (1-\mu)b \sim \bar{\mu}\bar{a} \oplus (1-\bar{\mu})\bar{b} \tag{29}$$
$$\mu a' \oplus (1-\mu)b \sim \bar{\mu}\bar{a}' \oplus (1-\bar{\mu})\bar{b} \tag{30}$$
$$\mu a \oplus (1-\mu)b' \sim \bar{\mu}\bar{a} \oplus (1-\bar{\mu})\bar{b}' \tag{31}$$

jointly imply

$$\mu a' \oplus (1-\mu)b' \sim \bar{\mu}\bar{a}' \oplus (1-\bar{\mu})\bar{b}' \tag{32}$$

Proof. Using Commutativity and Associativity it is straightforward to show that

$$1/2[\mu a \oplus (1-\mu)b] \oplus 1/2[\mu a' \oplus (1-\mu)b'] \tag{33}$$
$$=1/2[\mu a' \oplus (1-\mu)b] \oplus 1/2[\mu a \oplus (1-\mu)b'] \tag{34}$$

for any μ, a, b, a', b'. Using Axiom 3 together with the assumptions stated above then guarantee the desired result. $\square$

Lemma 3. *$\succsim$ can be represented by continuous U, F such that*

$$U(\mu a \oplus (1-\mu)b) = F(a,\mu) + F(b, 1-\mu) \tag{35}$$

Proof. We either obtain the representation trivially, if $a \sim b$ for all $a, b \in \mathcal{S}$ or using the main theorem of Qin and Rommeswinkel (2018) which provides a representation theorem for weak orders on (open subsets of) $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ with the representation $f(x,z) + g(y,z)$. Here we choose $\mathcal{X} = \mathcal{S}$, $\mathcal{Y} = \mathcal{S}$, and $\mathcal{Z} = (0,1)$ and endow the space with the product topology of the order topologies and the subspace topology of the reals. Thus, we will first obtain the representation on $\mu \in (0,1)$ and then extend it to $[0,1]$ using Axiom 2. To apply the main theorem of Qin and Rommeswinkel (2018), we require the following conditions: essentiality, conditional independence of the $\mathcal{X}$ and $\mathcal{Y}$ dimension given $\mathcal{Z}$, coseparability of $\mathcal{X}$ and $\mathcal{Y}$ given $\mathcal{Z}$, continuity in the product topology, and topological connectedness of $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$.

Since we have a product space, the well-behavedness assumptions of Qin and Rommeswinkel (2018) are not needed and we also only need essentiality instead of strict essentiality. Essentiality requires that for at least some μ and some a, then there exist some b, b' such that $\mu a \oplus (1-\mu)b \not\sim \mu a \oplus (1-\mu)b'$ and for some a, b there exist some μ, μ' such that $\mu a \oplus (1-\mu)b \not\sim \mu' a \oplus (1-\mu')b$. The former is guaranteed by Axiom 3 and the exclusion of the case $a \sim b$ for all $a, b \in \mathcal{S}$. The latter is guaranteed by Axiom 2 and the exclusion of the case $a \sim b$ for all $a, b \in \mathcal{S}$. Next, we need conditional independence of the $\mathcal{X}$ and $\mathcal{Y}$ dimensions for fixed $\mathcal{Z}$ dimension. This holds by Axiom 3. Further, coseparability of the $\mathcal{X}$ and $\mathcal{Y}$ dimension given $\mathcal{Z}$ has been shown above. Continuity of $\succsim$ holds in the order topology on $\mathcal{S}$. However, we require continuity in the product topology on $\mathcal{S} \times \mathcal{S} \times (0,1)$. By Axioms 2 and 3 the product topology is finer than the order topology on $\mathcal{S}$, guaranteeing continuity in the product topology. Topological connectedness of the product topology follows from the connectedness of its components $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$. The interval $(0,1)$ is obviously connected and each component $\mathcal{S}$ is connected in the order topology.

From Qin and Rommeswinkel (2018) then follows the existence of functions F and E such that $\succsim$ can be represented by

$$U(\mu a \oplus (1-\mu)b) = F(a,\mu) + E(b,\mu) \tag{36}$$

Commutativity of the mixture set guarantees that we can redefine E and F such that $E(b,\mu) = F(b, 1-\mu)$. To see this, note that from $U(\mu a \oplus (1-\mu)b) = U((1-\mu)b \oplus \mu a)$ follows that $F(b, 1-\mu) + E(a, 1-\mu) = F(a,\mu) + E(b,\mu)$. Since this holds for all a, $E(b,\mu) = F(b, 1-\mu) + E(a^*, 1-\mu) - F(a^*, \mu)$ for some arbitrarily chosen a^*. Thus, $E(b,\mu) = F(b, 1-$

$\mu) + f(\mu)$ where $f(\mu) \equiv E(a^*, 1-\mu) - F(a^*, \mu)$. Substituting in the original representation, $U(\mu a \oplus (1-\mu)b) = F(a,\mu) + F(b, 1-\mu) + f(\mu)$. Redefining $\hat{F}(a,\mu) = F(a,\mu) + \begin{cases} f(\mu)/2 & \mu < 1/2 \\ f(1-\mu)/2 & \mu \geq 1/2. \end{cases}$, it follows that $U(\mu a \oplus (1-\mu)b) = \hat{F}(a,\mu) + \hat{F}(b, 1-\mu)$. □

Lemma 4. *$F(a,\mu) = A(\mu)U(a) + B(\mu)$ for all μ and all $a \in \mathcal{S}$.*

Proof. For fixed μ, $F(a,\mu)$ is a monotone transformation of U:

$$
\begin{aligned}
& F(a,\mu) \geq F(b,\mu) && (37)\\
\Leftrightarrow\ & F(a,\mu) + F(c, 1-\mu) \geq F(b,\mu) + F(c, 1-\mu) && (38)\\
\Leftrightarrow\ & \mu a \oplus (1-\mu)c \succsim \mu b \oplus (1-\mu)c && (39)\\
\Leftrightarrow\ & a \succsim b && (40)\\
\Leftrightarrow\ & U(a) \geq U(b) && (41)
\end{aligned}
$$

Therefore, we can write $F(a,\mu) = G(U(a),\mu)$. We obtain from Associativity:

$$
\begin{aligned}
& U(\mu a \oplus (1-\mu)[\lambda b \oplus (1-\lambda)c]) && (42)\\
=& G(U(a),\mu) + G(G(U(b),\lambda) + G(U(c), 1-\lambda), 1-\mu) && (43)\\
=& G(U(b), (1-\mu)\lambda) + G(G(U(a), \frac{\mu}{1-(1-\mu)\lambda}) + G(U(c), \frac{(1-\mu)(1-\lambda)}{1-(1-\mu)\lambda}), 1-(1-\mu)\lambda) && (44)\\
=& U((1-\mu)\lambda b \oplus (1-(1-\mu)\lambda)[\frac{\mu}{1-(1-\mu)\lambda} a \oplus (1-\lambda)\frac{(1-\mu)(1-\lambda)}{1-(1-\mu)\lambda} c]) && (45)
\end{aligned}
$$

Noting that we have two continuous additive representations over $\mathcal{S} \times \mathcal{S}$ (specifically here the elements a and b), by the uniqueness of additive representations, we have that $G(\cdot, 1-\mu)$ in (43) is positively affine in its first argument. Since b, c and μ, λ are arbitrary, this holds for all utility levels. Therefore $G(U(a),\mu) = A(\mu)U(a) + B(\mu)$ for all $a, b \in \mathcal{S}$ and $\mu \in [0,1]$. □

Lemma 5. *$A(\mu) = \mu^r$, $r \in \mathbb{R}_{++}$.*

Proof. We define $H(\mu) = H(1-\mu) = B(\mu) + B(1-\mu)$. Using Associativity, we can derive that

$$A(\lambda)\left[A(\mu)U(a) + A(1-\mu)U(b) + H(\mu)\right] + A(1-\lambda)U(c) + H(\lambda) \tag{46}$$

$$\begin{aligned}=&A(\lambda\mu)U(a) + H(\lambda\mu)\\ &+ A(1-\lambda\mu)\left[A\left(\frac{\lambda(1-\mu)}{1-\lambda\mu}\right)U(b) + A\left(\frac{1-\lambda}{1-\lambda\mu}\right)U(c) + H\left(\frac{\lambda(1-\mu)}{1-\lambda\mu}\right)\right]\end{aligned} \tag{47}$$

Consider a substitution a' for a under which the above condition needs to still hold. If $\Delta U = U(a) - U(a')$, then it follows that

$$A(\lambda)A(\mu)\Delta U = A(\lambda\mu)\Delta U \tag{48}$$

and therefore A is multiplicative. Using Cauchy's functional equation it is straightforward to derive that $A(\mu) = \mu^r$, $r \in \mathbb{R}$. By Axiom 3, $r > 0$. □

We now finish the proof. We obtain

$$\lambda^r H(\mu) + H(\lambda) = (1-\lambda\mu)^r\left[H\left(\frac{\lambda(1-\mu)}{1-\lambda\mu}\right)\right] + H(\lambda\mu) \tag{49}$$

and substitute: $\lambda = 1-x$ and $\lambda\mu = y$. Using $H(x) = H(1-x)$ we obtain:

$$(1-x)^r H\left(\frac{y}{1-x}\right) + H(x) = (1-y)^r H\left(\frac{x}{1-y}\right) + H(y) \tag{50}$$

with two types of solutions (Ebanks et al., 1987):

$$A(\mu) = \mu; \quad H(\mu) = -(\mu\ln\mu + (1-\mu)\ln(1-\mu))q + s \tag{51}$$

$$A(\mu) = \mu^r; \quad H(\mu) = -(\mu^r + (1-\mu)^r - 1)q + s \tag{52}$$

where $q, s \in \mathbb{R}$. From Axiom 2 and Connectedness, we also have that in both representations $s = 0$. We have therefore obtained the desired representation:

$$U(\mu a \oplus (1-\mu)b) = \mu^r U(a) + (1-\mu)^r U(b) + q \cdot H_r(\mu) \tag{53}$$

$$\text{with } H_r(\mu) = \begin{cases} -\mu\ln\mu - (1-\mu)\ln(1-\mu), & r = 1 \\ -\mu^r - (1-\mu)^r + 1, & r \neq 1 \end{cases} \tag{54}$$

Regarding uniqueness, note that if preferences are nontrivial, then we immediately have an additively separable preference $U(\mu x \oplus (1-\mu)y)$ over a continuum of x and y and thus U is unique up to affine transformations. The uniqueness properties of r and q follow immediately. □

B Proof of Corollary 2

Proof. We prove sufficiency: The characterization of the Luce model is standard. We fix some $y \in \mathcal{X}$ and define $v(y) = 1$ and $v(x) = \ln p(x, \{x,y\}) - \ln 1 - p(x, \{x,y\})$. It is straightforward to then show that (25) holds.

We form an equivalence relation $\approx$ on $\mathcal{C}$ such that $\mathcal{C} \approx \mathcal{D}$ if and only if there exists enumerations $C = \{x_1, \ldots, x_n\}$ and $D = \{y_1, \ldots, y_n\}$ such that $p(x_i, C) = p(y_i, D)$ for all $i \in \{1, \ldots, n\}$. From Continuity of Decision Times follows that if $C \approx D$, then $\tau(C) = \tau(D)$. Each element of $\mathcal{C}/\approx$ can be represented by a finite tuple $(\mu_1, \ldots, \mu_n)$ with $\sum_i \mu_i = 1$, $\mu_i \in (0,1]$, and the convention that $\mu_i \geq \mu_{i+1}$ for all $i \in \{0, \ldots, n\}$. Notice that this makes a statement such as $C \in (\mu_1, \ldots, \mu_n)$ meaningful; it means that the set C is an element of the equivalence class represented by $(\mu_1, \ldots, \mu_n)$. We endow the set $\mathcal{C}/\approx$ with an operation $\oplus$ such that $\mu(\mu_1, \ldots, \mu_n) \oplus (1-\mu)(\lambda_1, \ldots, \lambda_k)$ is the tuple obtained from rearranging $(\mu\mu_1, \ldots, \mu\mu_n, (1-\mu)\lambda_1, \ldots, (1-\mu)\lambda_k)$ into descending order. We further endow the mixture set $\langle \mathcal{C}/\approx, \oplus, = \rangle$ with the weak order induced by τ: $a \succsim b$ if there exist $C \in a$ and $D \in b$ such that $\tau(C) \geq \tau(D)$. By Continuity of Decision Times, indeed $a \succsim b$, $C \in a$ and $D \in b$ holds if and only if $\tau(C) \geq \tau(D)$.

Notice that if $C \cap E = \emptyset$, $p(C, C \cup E) = \mu$, $C \in a$, and $E \in b$, then $C \cup E \in \mu C \oplus (1-\mu)E$. From Independence of Decision Times then follows that the relation induced by τ on the procedural mixture set fulfills Independence.

Continuity of $\succsim$ follows straightforward from the fact that τ is continuous in the choice probabilities.

By Theorem 1 there exists a representation U of $\succsim$ on $\mathcal{C}/\approx$. Since U is continuous and τ is continuous, there must exist a continuous monotone transformation T such that if $C \in b$, then $T \circ \tau(C) = U(b)$. Now let $C \cap D = \emptyset$, $C \in a$, $D \in b$ and $\mu = p(C, C \cup D)$. Then $T \circ \tau(C \cup D) = U(\mu a \oplus (1-\mu)b) = p(C, C \cup D)^r T \circ \tau(C) + p(D, C \cup D)^r T \circ \tau(D) + qH_r(p(C, C \cup D))$. Since $\tau(\{x\}) = \tau(\{y\})$ for all x, y and U is unique up to affine transformations, we can assume without loss of generality that $U((1)) = 0$, i.e., $T \circ \tau(\{x\}) = 0$. If this is the case, then by Positivity and Remark 2 it is without loss of generality to assume $q = 1$. □

C Proof of Proposition 2

For notational convenience, we define $a \equiv \alpha d \oplus (1-\alpha)d$ and $b = \beta c \oplus (1-\beta)c$, where $c = \gamma d \oplus (1-\gamma)d$.

$a \succ (\prec) d$ if and only if $sgn(q_1 - U_1(d)(r_1-1)) = sgn(q_2 - U_2(d)(r_2 - 1)) > (<) 0$.

By the uniqueness properties of U, $a \succsim b$ if and only if

$$sgn(q - U(d)(r-1))H_r(\alpha) \geq sgn(q - U(d)(r-1))((\gamma^r + (1-\gamma)^r)H_r(\beta) + H_r(\gamma)) \tag{55}$$

Thus, $a \succsim_1 b$ implies $a \succsim_2 b$ and $a \succ b_1$ implies $a \succ_2 b$ if one of the following is true: $q_1 - U_1(d)(r_1-1) = q_2 - U_2(d)(r_2-1) = 0$, $q_1 - U_1(d)(r_1-1) > 0 < q_2 - U_2(d)(r_2-1)$ and

$$\begin{aligned} &H_{r_1}(\alpha) \geq (>) (\gamma^{r_1} + (1-\gamma)^{r_1})H_{r_1}(\beta) + H_{r_1}(\gamma) \\ \Rightarrow\ &H_{r_2}(\alpha) \geq (>) (\gamma^{r_2} + (1-\gamma)^{r_2})H_{r_2}(\beta) + H_{r_2}(\gamma) \end{aligned} \tag{56}$$

or $q_1 - U_1(d)(r_1-1) < 0 > q_2 - U_2(d)(r_2-1)$ and (C) holds with opposite inequalities.

Define the Rényi (1961) entropy $R_r(\alpha) = \ln(\alpha^r + (1-\alpha)^r)/(1-r)$. Notice that $R_r(\alpha) \geq R_r(\beta)$ if and only if $H_r(\alpha) \geq H_r(\beta)$ and $H_r(\alpha) \geq (\gamma^r + (1-\gamma)^r)H_r(\beta) + H_r(\gamma)$ if and only if $R_r(\alpha) \geq R_r(\beta) + R_r(\gamma)$. That is, the Rényi (1961) and Tsallis (1988) entropy are order-equivalent.

Lemma 6. *Suppose* $0 \leq r < s$, $\alpha, \beta, \gamma \leq 1/2$, *and*

$$R^r(\alpha) = R^r(\beta) + R^r(\gamma) \tag{57}$$

then,

$$R_s(\alpha) > R_s(\beta) + R_s(\gamma). \tag{58}$$

Proof. It is straightforward to show that $\alpha \geq \beta$ and $\alpha \geq \gamma$ since for $r > 0$, $R^r(\gamma) \geq 0$. We substitute: $x_1 = \alpha^{r-1}$, $x_2 = (1-\alpha)^{r-1}$, $y_1 = (\beta\gamma)^{r-1}$, $y_2 = (\beta(1-\gamma))$, $y_3 = ((1-\beta)\gamma)^{r-1}$, $y_4 = ((1-\beta)(1-\gamma))^{r-1}$, $w_{ij} = (x_i y_j)^{1/(r-1)}$ and exponentiate both sides to obtain that (58) is equivalent to:

$$sgn(1-s)\sum_{ij} w_{ij}(x_i^t - y_j^t) > 0 \tag{59}$$

where $t = (s-1)/(r-1)$.

Note that the vector y with weights $(w_{11} + w_{21}, \ldots, w_{14} + w_{24})$ is a mean-preserving spread of the vector x with weights $(w_{11} + \ldots + w_{14}, w_{21} + \ldots + w_{24})$ since by (57), we have that

$$\sum_{ij} w_{ij}(x_i - y_j) = 0 \tag{60}$$

Since y is a mean-preserving spread of x, we have by the properties of generalized means that $M^t(\vec{w}, \vec{x}) \equiv (\sum_{ij} w_{ij} x_i^t)^{1/t} > (\sum_{ij} w_{ij} y_j^t)^{1/t} \equiv M^t(\vec{w}, \vec{y})$ if $t < 1$ and the reverse inequality holds if $t > 1$. It follows that $\sum_{ij} w_{ij}(x_i^t - y_j^t)$ is negative if $t > 1$ or $t < 0$ and positive if $0 < t < 1$. Since $0 < t < 1$ holds if and only if $s < 1$, (59) holds. □

The lemma establishes that if at some r we have that (α, β, γ) are such that $a \sim_1 b$, then at a higher r it must be the case that $a \succ_2 b$. Since irrespective of the choice of s the LHS of (58) is increasing in $\alpha \leq 1$ and the RHS is increasing in $\beta \leq 1/2$, $\gamma \leq 1/2$, it follows that $\{(a,b) : a \succsim_1 b\} \subseteq \{(a,b) : a \succsim_2 b\}$

D Further Comparative Statics

The comparative statics of Section 4 focus with respect to the parameter q of the representation on the property whether $U(a)(r-1) - q \gtreqless 0$. The present section provides additional results that require the existence of nontrivial preference on a set of *consequences*, i.e. elements of $\mathcal{S}$ that cannot be written in the form $\mu a \oplus (1-\mu) b$ with $\mu \in (0,1)$. Let $\mathcal{X}$ be a set of consequences. The following definitions are the standard definitions of certainty equivalents and comparative risk aversion for mixture sets adjusted to the procedural mixture setting.

Definition 17 (Certainty Equivalent)**.** The certainty equivalent $c = ce(\alpha x \oplus (1-\alpha) y) \in \mathcal{X}$ of a procedural mixture of outcomes x and y is an outcome that fulfills $\alpha x \oplus (1-\alpha) y \sim \alpha c \oplus (1-\alpha) c$.

Definition 18 (Comparative Risk Aversion)**.** $\succsim_1$ is at least as risk averse as $\succsim_2$ if for all $\alpha \in (0,1)$ and all $x, y, z \in \mathcal{X}$, we have that

$$\alpha x \oplus (1-\alpha) y \succsim_1 \alpha z \oplus (1-\alpha) z \tag{61}$$

$$\Rightarrow \quad \alpha x \oplus (1-\alpha) y \succsim_2 \alpha z \oplus (1-\alpha) z \tag{62}$$

Provided with our adjusted definitions, we can prove the following standard result for decisions under risk which extends to the procedural case:

Proposition 3. *Let $\succsim_1$ and $\succsim_2$ be mixture entropy values with representations U_1 and U_2 and parameters r_1, q_1 and r_2, q_2, respectively. Let $U_1(\mathcal{X})$ and $U_2(\mathcal{X})$ be convex sets. Suppose there exist some $x, y \in \mathcal{X}$ such that $x \succ_1 y$. Then the following statements are equivalent.*

1. *$\succsim_1$ is at least as risk averse as $\succsim_2$.*
2. *The restriction of U_1 to $\mathcal{X}$ is a concave monotone transformation of U_2 and $r_1 = r_2$.*

Proof. $\Leftarrow$ is trivial, we prove $\Rightarrow$: It is straightforward to show that for all $z, w \in \mathcal{X}$, $z \succ_1 w$ if and only if $z \succ_2 w$. Since $H_r(1) = 0$ and $\succsim_1$ and $\succsim_2$ are continuous, it follows that utilities over outcomes must be continuous monotone transformations of another, i.e., $U_1 = T \circ U_2$ when restricted to $\mathcal{X}$.

Since the values of outcomes are a convex set, for every $x, y \in \mathcal{X}$ such that $x \succ_2 y$ we can find z such that $z = ce_1(1/2x \oplus 1/2y)$. Notice that by the definition of a certainty equivalent we have $U_1(x)/2 + U_1(y)/2 = U_1(z)$. If T is not concave, then for some such x and y, $U_2(x) + U_2(y) = T^{-1}(U_1(x))/2 + T^{-1}(U_1(y))/2 < T^{-1}(U_1(z)) = U_2(z)$. But then $1/2x \oplus 1/2y \succsim_1 1/2z$ but $1/2x \oplus 1/2y \prec_2 1/2z \oplus 1/2z$, contradicting that $\succsim_1$ is at least as risk averse as $\succsim_2$. We have thus established that T is concave.

Notice now that if $r_1 \neq r_2$, then $p_1(\alpha) \equiv \frac{\alpha^{r_1}}{\alpha^{r_1} + (1-\alpha)^{r_1}} \neq \frac{\alpha^{r_2}}{\alpha^{r_2} + (1-\alpha)^{r_2}} \equiv p_2(\alpha)$. Without loss of generality, assume that $p_1(\alpha) \geq p_2(\alpha)$ for $\alpha \geq 1/2$. Since T is monotone and continuous, it is differentiable almost everywhere. Without loss of generality, assume T is differentiable at $U_2(x)$ and $U_2(x) = U_1(x)$ and $\partial T(U_2(x))/\partial U = 1$. Then we can find outcomes $x' \succ x \succ x''$ such that $p_1(\alpha)T(U_2(x')) + (1 - p_1(\alpha))T(U_2(x'')) > U_2(x) > p_2(\alpha)U_2(x') + (1 - p_2(\alpha))U_2(x'')$ contradicting that $\succsim_1$ is at least as risk averse as $\succsim_2$. $\square$

Mixture entropy values which are equally risk averse can be compared by how much the value of consequences is compared to the value of mixing.

Definition 19 (Comparative Consequentialism). $\succsim_2$ is at least as consequentialist than $\succsim_1$ if for all $\alpha \in (0,1)$ and all $x, y \in \mathcal{X}$, we have that

$$\alpha x \oplus (1-\alpha)x \succsim_2 y \tag{63}$$
$$\Rightarrow \quad \alpha x \oplus (1-\alpha)x \succsim_1 y \tag{64}$$

Proposition 4. *Let $\succsim_1$ and $\succsim_2$ be equally risk averse mixture entropy values with representations U_1 and U_2 and parameters r_1, q_1 and r_2, q_2, respectively. Let $U_1(\mathcal{X})$ and $U_2(\mathcal{X})$ be convex sets. Suppose there exist some $x, y \in \mathcal{X}$ such that $x \succ_1 y$. Then the following statements are equivalent.*

1. *$\succsim_2$ is at least as consequentialist as $\succsim_1$.*
2. *There exist $s \in \mathbb{R}_{>0}$ and $t \in \mathbb{R}$ such that $U_2(x) = sU_1(x) + t$ for all $x \in \mathcal{X}$ and $sq_1 + t \geq q_2$.*

Proof. $\Leftarrow$ is trivial, we prove $\Rightarrow$: If $\succsim_1$ and $\succsim_2$ are equally risk averse, then $r_1 = r_2 = r$ and U_1 is an affine transformation of U_2 on outcomes. Assume for the moment that $U_1 = U_2$ on outcomes. Notice that since $H_r(\alpha) \geq 0$ for all r and all α, (63) holds if and only if $q_1 \geq q_2$. Since U_2 may be an affine transformation of U_1 on outcomes, by the uniqueness properties of U, the desired result follows. □

This provides a link between the parameters of our representation and the intensity of the value of mixing relative to the value of consequences. Notice that this comparison is only meaningful when $r_1 = r_2$ and the value of consequences are cardinally comparable.

E Proof of Theorem 2

Proof. The intuition for the result is simply that similar to a procedural mixture set, a mixture set in which Independence only applies to disjoint mixtures allows for $a \sim^* b$ and $\mu a \oplus (1-\mu)b \not\sim \mu a \oplus (1-\mu)a = a$ (since a and a do not have disjoint support). The key difficulty is that the Disjoint Independence axiom might not apply to sufficiently many elements of the mixture set to restrict preferences to the desired representation.[8]

To prove the result, we first extend $\succsim^*$ from $\Delta\mathcal{X}$ to a set $\Delta\mathcal{X}^\infty$ which is a mixture set generated from finite mixtures of countably many copies of

[8]For example, if there are only two outcomes, Disjoint Independence does not restrict preferences.

$\mathcal{X}$. The resulting relation is additively separable across the different copies of $\mathcal{X}$. $\Delta\mathcal{X}^\infty$ contains a subset that is isomorphic to a procedural mixture set and on which $\succsim^*$ fulfills the vNM axioms. We thus have by Theorem 1 the desired representation on the procedural mixture set. Because disjoint mixtures in $\mathcal{X}$ coincide with procedural mixtures in the procedural mixture set, the utility representation on $\Delta\mathcal{X}$ fulfills (23). The details of these steps follow below.

Let $\mathcal{X}^\infty = \sqcup_{i=0}^{\infty}\mathcal{X}$ be the disjoint union of countably many copies of $\mathcal{X}$. $x_i^k \in \mathcal{X}^\infty$ refers to the kth copy of $x_i \in \mathcal{X}$. Let ΔX^∞ be the mixture set generated from $\mathcal{X}^\infty$. A generic element of $\Delta\mathcal{X}^\infty$ can therefore be represented by $m = \{(x_i^k, \mu_i^k), \ldots, (x_j^l, \mu_j^l)\}$ such that $\sum_{i,k} \mu_i^k = 1$. Note that $\mathcal{M} \subset \Delta X^\infty$.

Let $\mathcal{J}$ be a partition of the support such that there are three elements that each contain an essential pair of outcomes. By standard results (Wakker, 1989, e.g.,), our axioms[9] together with the existence of three essential pairs of outcomes with mutually disjoint support guarantees that there exists an additive representation $U(m) = \sum_J u((x_i^1, \mu_i^1)_{i\in J})$ on $\Delta\mathcal{X}$ such that each component is normalized to zero if $\sum_{i\in J} \mu_i^1 = 0$. We can uniquely extend the relation $\succsim^*$ on $\mathcal{M}$ to $\Delta\mathcal{X}^\infty$ by summation of the utilities obtaining a utility representation unique up to affine transformations of the form $\sum_k \sum_{J\in\mathcal{J}} u((x_i^k, \mu_i^k)_{i\in J})$.

Let $\mathcal{P}$ be the closure of $\mathcal{X}$ under an operator $\oplus$, i.e., the minimal set such that $\mathcal{X} \subset \mathcal{P}$ and for all $p, q \in \mathcal{P}$ and all $\mu \in [0,1]$, we have that $\mu p \oplus (1-\mu)q \in \mathcal{P}$. Similarly, let $\approx$ be the minimal relation such that the procedural mixture axioms are fulfilled. Notice that the quotient set $\mathcal{P}/\approx$ is also a procedural mixture set with $=$ being the equivalence relation.

A generic element of $\mathcal{P}/\approx$ can be represented by a finite set $\{(x_1, \mu_1), \ldots, (x_n \mu_n)\}$ where all $x_i \in \mathcal{X}$ with $x_i = x_j$ permitted also for $i \neq j$. Thus, there is a natural mapping from $\mathcal{P}/\approx$ into $\Delta\mathcal{X}^\infty$ which we denote $\phi : \mathcal{P}/\approx \to \Delta\mathcal{X}^\infty$.

Let $\succsim$ be defined by $p \succsim q$ if $\phi(p) \succsim^* \phi(q)$. It is straightforward to see that $\succsim$ fulfills Weak Order and Continuity. Independence follows from the fact that $\succsim^*$ has an additively separable representation across different copies of $\mathcal{X}$ contained in $\mathcal{X}^\infty$ and fulfills Disjoint Independence.

It follows that $\succsim^*$ on $\mathcal{P}/\approx$ has an entropy adjusted expected utility representation $V : \mathcal{P}/\approx \to \mathbb{R}$. Notice that on $\phi(\mathcal{P}/\approx) \subset \Delta\mathcal{X}^\infty$ we have that $V \circ \phi^{-1}$ is additively separable across the partition of indexes J. It follows

[9] Disjoint Independence implies preference separability across disjoint subsets of the support. Continuity implies topological connectedness of each dimension.

that $V \circ \phi^{-1}$ is an affine transformation of U. The desired functional form of disjoint mixtures when U is restricted to $\Delta\mathcal{X}$ follows. $\square$

F Proof of Corollary 3

Proof. (Sketch)

Given Theorem 2, the result is quite straightforward and we only sketch the main ideas of the proof.

First, we show that p is a stochastic choice function following the Luce model when restricted to choices with disjoint support. We fix some $y, x \in \mathcal{X}$ that are disjoint from all other alternatives, and define for all m, $z \in \mathcal{X} - \{y\}$:

$$v(y, m(y)) = \ln \frac{p(y, \{x, y\}, m) / p(x, \{x, y\}, m)}{p(y, \{x, y\}, m_0) / p(x, \{x, y\}, m_0)} \tag{65}$$

$$v(z, m(z)) = \ln \frac{p(z, \{y, z\}, m)}{p(y, \{y, z\}, m)} v(y, m(y)) \tag{66}$$

where m_0 is such that $m_0(y) = 0$ and $m_0(z) = m(z)$ for all $z \in \mathcal{X} - \{y\}$. By Positivity and Restricted IIA, v is well defined; Positivity guarantees that all probabilities are nonzero and Restricted IIA guarantees that $v(y, m(y))$ does not depend on $m(x)$ and $v(z, m(z))$ does not depend on $m(y)$. Using Restricted IIA, it is then straightforward to show (by adding y) that if $D \in \mathcal{C}$ contains only elements disjoint from one another, then the choice probabilities take the desired form for all alternatives in D.

Second, we show that varying m in $p(\ldots, m)$ generates a mixture set. By Continuity, τ is a continuous function of the choice probabilities. Monotonicity in Payoffs and Restricted IIA guarantee that the probabilities form a convex set if all alternatives are disjoint. Since every alternative can appear at most once in a decision process and only the choice probabilities matter, every decision process (C, m) can be represented by a probability measure $p_{C,m}$ over the elements of C. Thus, the set $\mathcal{C} \times \mathbb{R}^{\mathcal{X}}$ equipped with the relation $\succsim$ induced by τ is isomorphic to a mixture set $\mathcal{M}$. We endow $\mathcal{M}$ with a support $supp : \mathcal{M} \to 2^{\mathcal{A}}$ such that $supp(p_{C,m}) = \{A \in \mathcal{A} | A \cap C \neq \emptyset\}$.

Third, it is straightforward to show that the weak order induced by τ fulfills Continuity and Disjoint Independence given our axioms. It follows that for any set D of mutually disjoint elements the desired representation holds but it may in principle be the case that two sets D and D' have mixture entropy representations that are incompatible with another, e.g., have different parameters r and r'. However, it is obvious that if $C \subseteq D$, then the representation for D is also valid for C. Given the six alternatives $\{d_1, \ldots, d_6\}$ that do not share attributes with any other alternative, we can fix a unique mixture entropy representation and uniquely extend it to $\{d_1, \ldots, d_6\} \cup D'$. $\square$

References

Baldassi, C., Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M., & Pirazzini, M. (2020). A behavioral characterization of the drift diffusion model and its multialternative extension for choice under time pressure. *Management Science*, *66*(11), 5075–5093.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*, 700–765.

Breitmoser, Y. (2020). An axiomatic foundation of conditional logit. *Economic Theory*, *72*, 245–261.

Caplin, A., Dean, M., & Leahy, J. (2017, August). *Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy* (w23652). National Bureau of Economic Research. Cambridge, MA. https://doi.org/10.3386/w23652

Chen, T.-Y., & Rommeswinkel, H. (2020). Measuring Consumer Freedom.

Csiszár, I. (2008). Axiomatic Characterizations of Information Measures. *Entropy*, *10*, 261–273.

Debreu, G. (1960). Review of rd luce, individual choice behavior: A theoretical analysis. *American Economic Review*, *50*, 186–188.

de Oliveira, H., Denti, T., Mihm, M., & Ozbek, K. (2017). Rationally Inattentive Preferences and Hidden Information Costs. *Theoretical Economics*, *12*(2), 621–654.

Ebanks, B. R., Kannappan, P., & Ng, C. T. (1987). Generalized fundamental equation of information of multiplicative type. *Aequationes Mathematicae*, *32*(1), 19–31. https://doi.org/10.1007/BF02311295

Echenique, F., & Saito, K. (2017). Response time and utility. *Journal of Economic Behavior and Organization*, *139*, 49–59.

Ellis, A. (2018). Foundations for optimal inattention. *Journal of Economic Theory*, *173*, 56–94.

Fishburn, P. C. (1982). *Foundations of expected utility*. Reidel.

Frankel, D., & Volij, O. (2011). Measuring school segregation. *Journal of Economic Theory*, *146*(1), 1–38.

Ghirardato, P., Maccheroni, F., Marinacci, M., & Siniscalchi, M. (2003). A subjective spin on roulette wheels. *Econometrica*, *71*(6), 1897–1908.

Ghirardato, P., & Pennesi, D. (2020). A general theory of subjective mixtures. *Journal of Economic Theory*, 43. https://doi.org/10.1016/j.jet.2020.105056

Gneezy, U., List, J. A., & Wu, G. (2006). The Uncertainty Effect: When a Risky Prospect Is Valued Less Than Its Worst Possible Outcome. *Quarterly Journal of Economics*, *121*(4), 27.

Hennessy, D. A., & Lapan, H. (2007). When different market concentration indices agree. *Economics Letters*, *95*(2), 234–240. https://doi.org/10.1016/j.econlet.2006.10.011

Herfindahl, O. C. (1950). *Concentration in the Steel Industry*. University Microfilms.

Herstein, I. N., & Milnor, J. (1953). An Axiomatic Approach to Measurable Utility. *Econometrica*, *21*(2), 291–297. https://doi.org/10.2307/1905540

Hick, W. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, *4*(1), 11–26.

Hirschman, A. O. (1980). *National power and the structure of foreign trade*. Univ of California Press.

Kovach, M., & Tserenjigmid, G. (2022). Behavioral foundations of nested stochastic choice and nested logit. *Journal of Political Economy*, *130*(9), 2411–2461.

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *PNAS*, *108*(33), 13852–13857.

Krantz, D., Luce, D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement volume 1: Additive and polynomial representations* (Vol. 1). Dover Books on Mathematics.

Kreps, D. M., & Porteus, E. L. (1978). Temporal Resolution of Uncertainty and Dynamic Choice Theory. *Econometrica*, *46*(1), 185. https://doi.org/10.2307/1913656

Lin, Y.-H. (2020). Stochastic Choice and Rational Inattention.

Lu, J. (2016). Random Choice and Private Information. *Econometrica*, *84*(6), 1983–2027.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. Wiley.

Luce, R. D. (1986). *Responce times: Their role in inferring elementary mental organization*. Oxford University Press.
Luce, R. D., Ng, C. T., Marley, A. A. J., & Aczél, J. (2008a). Utility of gambling I: Entropy modified linear weighted utility. *Economic Theory*, *36*(1), 1–33. https://doi.org/10.1007/s00199-007-0260-5
Luce, R. D., Ng, C. T., Marley, A. A. J., & Aczél, J. (2008b). Utility of gambling II: Risk, paradoxes, and data. *Economic Theory*, *36*(2), 165–187. https://doi.org/10.1007/s00199-007-0259-y
McMillen, T., & Holmes, P. (2005). The dynamics of choice among multiple alternatives. *Journal of Mathematical Psychology*, *50*(1), 30–57.
Mongin, P. (2001). A note on mixture sets in decision theory. *Decisions in Economics and Finance*, *24*(1), 59–69. https://doi.org/10.1007/s102030170010
Nehring, K., & Puppe, C. (2002). A Theory of Diversity. *Econometrica*, *70*(3), 1155–1198.
Nehring, K., & Puppe, C. (2009). Diversity. In *Handbook of Rational and Social Choice*. Oxford University Press.
Qin, W.-z., & Rommeswinkel, H. (2018). Quasi-separable Preferences.
Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108.
Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Science*, *20*(4), 260–281.
Rényi, A. (1961). On measures of information and entropy. *Proceedings of the Fourth Berkeley Symposium on Mathematics, Statistics and Probability 1960*, 547–561.
Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision making. *Psychological Review*, *108*(2).
Saito, K. (2018). *Axiomatization of the mixed logit model* [Available at https://authors.library.caltec
Segal, U. (1990). Two-Stage Lotteries without the Reduction Axiom. *Econometrica*, *58*(2), 349. https://doi.org/10.2307/2938207
Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, *27*(3), 379–423, 623–656.
Shorrocks, A. F. (1980). The Class of Additively Decomposable Inequality Measures. *Econometrica*, *48*(3), 613.
Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, *50*(3), 665–690. https://doi.org/10.1016/S0304-3932(03)00029-1

Suppes, P. (1996). The nature and measurement of freedom. *Social Choice and Welfare*, *13*(2), 183–200.

Tajima, S., Drugowitsch, J., Patel, N., & Pouget, A. (2019). Optimal policy for multi-alternative decisions. *Nature Neuroscience*, *22*, 1503–1511.

Theil, H. (1965). The Information Approach to Demand Analysis. *Econometrica*, *33*(1), 67–87.

Theil, H. (1967). *Economics and Information Theory*. North-Holland.

Tsallis, C. (1988). Possible generalization of boltzmann-gibbs statistics. *Journal of Statistical Physics*, *52*, 479–487.

Usher, M., Tsetsos, K., Yu, E., & Lagnado, D. A. (Eds.). (2013). *Dynamics of decision making: From evidence accumulation to preference and belief*. Frontiers.

von Neumann, J., & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press.

Wakker, P. P. (1989). *Additive representations of preferences: A new foundation of decision analysis* (Vol. 4). Springer Science & Business Media.