Axiomatic Foundations for Entropic Costs of Attention *

Henrique de Oliveira†

October 1, 2014

Abstract

I characterize the preferences over sets of acts that result from the Rational Inattention Model of Sims [2003]. Each act specifies a consequence depending on the state of the world. The decision-maker can pay attention to information about the state of the world in order to choose a better act, but doing so is costly. Specifically, the cost equals the reduction in the entropy of beliefs resulting from observing information.

The model is shown to be essentially equivalent to three properties of choice: (1) choices are probabilistic, in the sense that the only feature of a state of the world that matters for the decision is the likelihood that the decision-maker ascribes to that state occurring; (2) decision problems depending on independent events may be solved separately of each other; (3) an option of conditioning the decision on events that are independent of the payoff-relevant events is worthless.

1 Introduction

When making a choice, it is valuable to have access to information, but when information is abundant, assimilating all of it may not be worth the effort. A decision-maker must first decide how to allocate attention to the different sources of information. In other words, she may deliberately choose not to pay attention to part of the information—she is rationally inattentive.

*I am deeply grateful to Eddie Dekel, Marciano Siniscalchi and Todd Sarver for their continuous advice and patience and to Luciano Pomatto, for many fruitful discussions. I am also grateful to Nabil Al-Najjar, Tommaso Denti, Peter Klibanoff, Maximiliam Mihm, Scott Ogawa, Alessandro Pavan and Bruno Strulovici for their helpful suggestions.

†Princeton University. Email: hrdo@princeton.edu

The allocation of attention can be understood as an *information acquisition problem.* The decision-maker faces a menu of options, each yielding a consequence in each state of the world: a set F of acts f associating the consequence $f(\omega) \in X$ to each state ω. Learning information about the state is valuable because it allows her to choose the act f which is best suited to the actual state, but it does not come for free, since paying attention is costly. Upon observing the information, a bayesian decision-maker updates her prior $\overline{p}$ to a posterior p. When choosing what information to pay attention to, the resulting posterior is not known yet; only the distribution of possible posteriors is known. Formally, the decision maker solves

$$\max_{\pi \in \Pi(\overline{p})} \left[\int_{\Delta(\Omega)} \max_{f \in F} \left(\sum_{\omega \in \Omega} u(f(\omega)) \, p(\omega) \right) \pi(dp) - c(\pi) \right]. \tag{1.1}$$

where $u : X \to \mathbb{R}$ is a utility function, $\Pi(\overline{p})$ is the set of distributions over posteriors consistent with the given prior $\overline{p}$ and $c : \Pi(\overline{p}) \to \mathbb{R}$ is an information cost function.

Sims [2003] proposed to use information theory to specify the cost of information. The basic idea is to use as a cost of information the *reduction in uncertainty* resulting from the observation of that information, where the uncertainty of a probability distribution p is given by its Shannon Entropy:

$$H(p) = -\sum_{\omega} p(\omega) \log_2 p(\omega).$$

The cost of information is then given by the *mutual information*:[1]

$$\mathcal{I}(\pi) = \int_{\Delta(\Omega)} \left(H(\overline{p}) - H(p) \right) \pi(dp). \tag{1.2}$$

I will refer to the maximization problem (1.1) with the cost of information given by the mutual information (1.2) as the *Rational Inattention Model.*

Sims's seminal contribution was followed by applications in various settings. It is therefore natural to ask when it provides a good approximation of behavior. However, in the usual interpretation, not only is the information cost c not directly observable, but neither

[1] To be precise, Sims's original formulation was to consider a constraint on how much information the agent can learn. This may be modeled as a cost which is zero if the amount of information learned is below a certain threshold and infinity otherwise. Both specifications are used in the subsequent literature, but here I will not treat the formulation with the constraint. However, it should be noted that by using the Lagrangian, both specifications are equivalent for local statements.

is the solution π (the information acquired). The connection between observable choices and the specification of the cost function c is then rather indirect; it results from the unobserved maximization in (1.1).

This paper shows an axiomatic characterization of the preferences over menus given by the Rational Inattention Model. The literature on rational inattention is usually focused on its necessary implications in richer environments, where it is difficult to disentangle the particular implications of rational inattention alone. In contrast, I focus on a simple choice environment and obtain axioms on preferences that are necessary and sufficient for the behavior to be consistent with the Rational Inattention Model. Because they are necessary, the axioms provide new testable implications of the model. Their sufficiency is useful for two reasons. First, the suitability of the model for a particular application may be easier to assess by evaluating the suitability of the axioms. Second, if an axiom is found to be unreasonable for an application, investigating the consequences of relaxing it can lead to a more appropriate model, which still enjoys the properties stemming from the axioms that are deemed reasonable. Future work in Decision Theory that explores such relaxations can thereby be useful for applied work.

1.1 Preview of Axioms

The preferences over menus represented by (1.1), where the cost of information is general, are characterized by De Oliveira et al. [2013]. In that paper, we show that the cost function can be fully identified from the preferences over menus, up to a normalization. In this paper, I introduce new axioms that will determine the particular cost of information given by (1.2). I discuss the three axioms that are most particular to the entropy specification below.

1.1.1 Symmetry

Suppose $\Omega = \{1, 2, 3, 4\}$ and the decision-maker's prior assigns equal probability to states 2 and 3.[2] Symmetry states that, when evaluating a menu, it is inconsequential to switch the roles of 1 and 2. For example, if we consider a menu of two acts $F = \{(a, b, c, d), (x, y, w, z)\}$, the symmetry axiom states that it must be indifferent to the menu $G = \{(a, c, b, d), (x, w, y, z)\}$:

[2] The prior is itself subjective; it can be derived from preferences over menus of a single act. The use of the prior to discuss the axiom is an expository device. The actual statement is solely in terms of the preferences.

equally likely states are exchangeable.[3]

Symmetry's implications for the costs of attention can be understood with the following example. When facing the menu $F = \{(1,1,0,0), (0,0,1,1)\}$ (the act $(1,1,0,0)$ pays 1\$ in states 1 and 2 and 0\$ in states 3 and 4) it is useful to know which cell of the partition $\mathcal{P} = \{\{1,2\}, \{3,4\}\}$ contains the state. When facing $G = \{(1,0,1,0), (0,1,0,1)\}$, it is useful to know which cell of the partition $\mathcal{Q} = \{\{1,3\}, \{2,4\}\}$ contains the state. The indifference of the decision-maker between F and G reveals that both types of information are equally hard to learn. Thus, Symmetry rules out situations where the two types of information are available in disparate formats, which could make learning about some events harder than about other events, even though they may be regarded as equally likely to occur.

1.1.2 Separability in Orthogonal Decisions

Consider $\Omega = \{1,2,3,4\}$, with all states equally likely, and the partitions $\mathcal{P} = \{\{1,2\}, \{3,4\}\}$ and $\mathcal{Q} = \{\{1,3\}, \{2,4\}\}$. If the decision maker learns which cell of the partition $\mathcal{P}$ contains the state, she would still assign equal probability to each cell of $\mathcal{Q}$ containing the state: the posterior is equal to the prior. In this case we say that the partitions $\mathcal{P}$ and $\mathcal{Q}$ are *orthogonal.*

Now consider the following menus

$$\begin{aligned} F &= \{(4,4,0,0), (0,0,4,4)\} \quad & G &= \{(5,0,5,0), (0,5,0,5)\} \\ F' &= \{(2,2,1,1), (1,1,2,2)\} \quad & G' &= \{(4,2,4,2), (2,4,2,4)\} \end{aligned}$$

When choosing from F or F', only information about $\mathcal{P}$ matters (they are $\mathcal{P}$-measurable). When choosing from G or G', only information about $\mathcal{Q}$ matters. Suppose that a coin is flipped: if it comes up heads the resulting consequence comes from the act of choice from F; otherwise it will come from the choice from G. We may represent this as $\frac{1}{2}F + \frac{1}{2}G$. Since the decision-maker does not know which menu will obtain, both information about $\mathcal{P}$ and information about $\mathcal{Q}$ matter. But the information that is useful to improve the choice from F (information about $\mathcal{P}$) is not useful to improve the choice from G.

[3] With a finite set of states, Symmetry trivially holds whenever no two states are equally likely. In the paper, the state space will be infinite, and the prior non-atomic, so that Symmetry will always be consequential. Its interpretation is not affected by this change.

Separability in Orthogonal Decisions states that the decision of which $\mathcal{P}$-measurable menu to face is separable from whatever $\mathcal{Q}$-measurable menu the decision-maker might be facing. In the example above, if the preference satisfies

$$\frac{1}{2}F+\frac{1}{2}G\succsim\frac{1}{2}F'+\frac{1}{2}G,$$

then the axiom requires that the same must be satisfied if G is switched with G'.

1.1.3 Irrelevance of Orthogonal Flexibility

Consider again the menu F from the previous example.

$$F=\{f_1,f_2\}=\{(4,4,0,0),(0,0,4,4)\}.$$

Learning which cell of the partition $\mathcal{Q}=\{\{1,3\},\{2,4\}\}$ contains the state should not be useful information, since the posterior belief about which act yields a better payoff remains the same. Learning the cell in $\mathcal{Q}$ would allow the agent to condition the choice of f_1 or f_2 depending on whether $\{1,3\}$ or $\{2,4\}$ contains the state. This can be represented by the menu

$$\tilde{F}=\{f_1,f_2,f_3,f_4\}=\{(4,4,0,0),(0,0,4,4),(4,0,0,4),(0,4,4,0)\}.$$

The act $f_3=(4,0,0,4)$ corresponds to the choice of f_1 if $\{1,3\}$ realizes and f_2 if $\{2,4\}$ realizes; the act $f_4=(0,4,4,0)$ reverses these choices. The statement that learning $\mathcal{Q}$ is not useful implies that $F\sim\tilde{F}$.

Independence of Orthogonal Flexibility extends this idea as follows. Suppose the decision maker, who faces the menu $\frac{1}{2}F+\frac{1}{2}G$, is now offered the opportunity to condition her choice from F on the partition $\mathcal{Q}$. The axiom states that this extra flexibility should still have no value:

$$\frac{1}{2}F+\frac{1}{2}G\sim\frac{1}{2}\tilde{F}+\frac{1}{2}G.$$

1.2 Related Literature

The Rational Inattention Model has been applied to various settings. In finance, Yang [2011] shows that the flexibility of choosing the type of information to be acquired can increase the degree of common knowledge of the fundamentals in a coordination game. In macroeconomics, Matejka and McKay [2012] analyze a market with inattentive consumers.

Mackowiak and Wiederholt [2012] show that an inattentive agent acquires less information under limited liability than under unlimited liability. Paciello and Wiederholt [2011] characterize the optimal monetary policy when firms are inattentive, with a cost of information which is an increasing convex function of the mutual information. Finally, Martin [2013] characterizes equilibria in a game where buyers pay attention before purchasing goods from sellers.

Many other applications follow Sims's original formulation, where the decision maker's information acquisition problem is constrained by an upper bound on the mutual information of the channel (while incurring no cost). The two approaches are "locally" equivalent in the following sense. When solving for the decision maker optimal choice of information using Sims' formulation, the Lagrangian multiplier effectively turns the constraint into a linear representation. Therefore, the two models behave similarly for small variations in the menu.

Two other papers examine the Rational Inattention Model from the perspective of choice theory. Matejka and McKay [2013] shows an equivalence between the behavior of a rationally inattentive decision maker and the stochastic choice described by a generalized multinomial logit model. Caplin and Dean [2013] also consider stochastic choices, but from menus of acts. Assuming that the prior belief of the agent can be controlled by an experimenter, they establish testable implications and test them in an experiment. However, those implications are not shown to be equivalent to Rational Inattention Model.

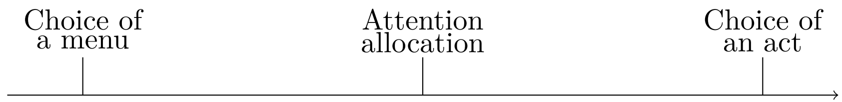
Cabrales et al. [2013] consider an ordering over information structures in the context of a portfolio choice problem, under the assumption that the utility function over monetary outcomes satisfy increasing relative risk aversion. They show that the order can be represented by the mutual information. In contrast, I impose no restrictions on risk aversion and no restrictions on the type of problems the decision-maker may face and ask when this decision-maker subjectively evaluates information using the mutual information formula.

2 Preliminaries

The set of states of the world is $\Omega = [0, 1]$, endowed with the Borel σ-algebra $\mathcal{B}$. Let X be a convex set of consequences (for example, X could be the set of lotteries over a fixed set of prizes). An *act* is a function $f : \Omega \to X$ that is measurable and takes finitely many values. A finite set of acts will be called a *menu* and denoted by F, G, H etc. The set of all acts is denoted by $\mathcal{F}$ and the set of all menus is $\mathbb{F}$.

2.1 Preferences

The primitive is a preference $\succsim$ defined over menus (elements of $\mathbb{F}$), which is interpreted according to the following timeline.



First, the decision-maker chooses among menus, revealing her preference. Then she has access to information and decides how much attention to pay to this information, and how to allocate the attention to different aspects of the available information. Finally, she picks an act from the menu chosen earlier.

2.2 Mixtures

Mixtures of acts are defined pointwise, as in Anscombe and Aumann [1963]. Given two acts f, g and a scalar $\alpha \in [0,1]$, denote by $\alpha f + (1-\alpha) g$ the act that in each state ω delivers the outcome $\alpha f(\omega) + (1-\alpha) g(\omega)$. For $\alpha \in [0,1]$, the mixture of two menus is defined as in Dekel et al. [2001]:

$$\alpha F + (1-\alpha) G = \{\alpha f + (1-\alpha) g : \; f \in F, g \in G\}.$$

We can interpret $\alpha F + (1-\alpha) G$ as a lottery over what menu the agent will be facing.

Given two acts f and g and an event E, the act fEg is identical to f in the event E and to g in its complement, as in

$$(fEg)(\omega) = \begin{cases} f(\omega) & \text{if } \omega \in E \\ g(\omega) & \text{if } \omega \notin E \end{cases}$$

and likewise for menus: $FEG = \{fEg : \; f \in F, \; g \in G\}$.

Definition 1. An event E is said to be *null* if $fEg \sim g$ for all acts f and g.

2.3 Partitions

A *partition of* Ω is a finite set of disjoint events $\mathcal{P} = \{E_1, \ldots, E_n\}$ whose union is Ω. An act f is $\mathcal{P}$-measurable if it is constant in each event of $\mathcal{P}$, and a menu is $\mathcal{P}$-measurable if each act in it is $\mathcal{P}$-measurable.

Definition 2. A partition $\mathcal{P} = \{E_1, \ldots, E_n\}$ is called an *equipartition* if $xE_iy \sim xE_jy$ for all i, j and $x, y \in X$.

Under expected utility, a partition is an equipartition if and only if all its events are regarded as equally likely.

Definition 3. Two equipartitions $\mathcal{P} = \{E_1, \ldots, E_n\}$ and $\mathcal{Q} = \{D_1, \ldots, D_m\}$ are said to be *orthogonal* if, for all $\mathcal{P}$-measurable acts f, g and h, we have

$$fD_ih \succsim gD_ih \implies fD_jh \succsim gD_jh$$

When the decision-maker evaluates acts using expected utility (which will be the case under the axioms that will follow), the condition $fD_ih \succsim gD_ih$ holds for some h if and only if it holds for all of them (this is Savage's postulate P2). Therefore, the *conditional preference* $\succsim_{D_i}$ over acts $f : D_i \to X$ is well defined by $f \succsim_{D_i} g$ if and only if $fD_ih \succsim gD_ih$ for some act h. This preference is an expected utility preference, and its probability distribution is obtained by conditioning the prior of the original preference to the event D_i. Thus, $\mathcal{P}$ and $\mathcal{Q}$ are orthogonal if and only if the preference $\succsim_{D_i}$ does not depend on i, that is, if and only if the conditional distribution over the cells of $\mathcal{P}$ is always the same. In other words, learning that the state of the world lies in $D_i \in \mathcal{Q}$ does not convey any useful information about what cell of $\mathcal{P}$ contains the state.

2.4 Information

Denote by $\Delta(\Omega)$ the set of all countably additive probability distributions over Ω and by $\Delta_d[0,1]$ the set of distributions that admit densities (with respect to the Lebesgue measure on $[0,1]$). An *information channel* is a distribution π over $\Delta(\Omega)$ with finite support[4], which can be understood as the probability that the decision-maker observes a signal (out of a finite set) together with the posterior belief obtained by Bayesian updating, after the observation of the signal. The set of information channels consistent with a given prior $\overline{p} \in \Delta(\Omega)$ is denoted by $\Pi(\overline{p})$.

Blackwell [1951, 1953] defines an important raking of informativeness of an information channel.

[4] Information channels are also known by the terms "information structures" and "experiments". The term "information channel" is common in Information Theory.

Definition 4. Let $\pi, \rho \in \Pi(\overline{p})$ be two information channels. Then π *is more informative than* ρif, for every menu F and utility function u,

$$\int_{\Delta(\Omega)} \max_{f\in F} \left(\int_{\Omega} u(f(\omega))\, p(d\omega) \right) \pi(dp) \geqslant \int_{\Delta(\Omega)} \max_{f\in F} \left(\int_{\Omega} u(f(\omega))\, p(d\omega) \right) \rho(dp).$$

In other words, the channel π is more informative than the channel ρ if the value of information given by π is always higher than the value of information given by ρ. This order is used in the definition of information cost of De Oliveira et al. [2013].

Definition 5. A function $c : \Pi(\overline{p}) \to [0, \infty]$ is an *information cost function* if it is weakly lower semicontinuous and satisfies the following properties:

1. No information is free: $c(\pi) = 0$ if $\pi(\overline{p}) = 1$;
2. Convexity: $c(\alpha\pi + (1-\alpha)\rho) \leqslant \alpha c(\pi) + (1-\alpha) c(\rho)$;
3. Blackwell Monotonicity: $c(\pi) \geqslant c(\rho)$ whenever π is more informative than ρ.

The mutual information (1.2) satisfies the properties above, and is therefore an information cost function.

2.5 A general representation

De Oliveira et al. [2013] provide axioms for the preference over menus that result from the general information acquisition problem. To be precise, the preference is represented by the utility function V, given by

$$V(F) = \max_{\pi\in\Pi(\bar{p})} \left[\int_{\Delta(\Omega)} \max_{f\in F} \left(\int_{\Omega} u(f(\omega))\, p(d\omega) \right) \pi(dp) - c(\pi) \right],$$

where $u : X \to \mathbb{R}$ is an affine utility function with unbounded range, $\overline{p} \in \Delta(\Omega)$ is the prior and $c : \Pi(\overline{p}) \to [0, \infty]$ is an information cost function. The original axioms of De Oliveira et al. [2013] are listed in the Appendix. The result extending the representation to the infinite state space $\Omega = [0, 1]$ is also provided there.

3 Axioms

This section shows the behavioral axioms. Axioms 1 and 2 allow the well-behaved extension of the result of De Oliveira et al. [2013] to the infinite state space $\Omega = [0,1]$. Axioms 3-5 are sufficient to represent the cost function as

$$c(\pi) = \int_{\Delta(\Omega)} \psi(p)\,\pi(dp) \tag{3.1}$$

where $\psi : \Delta(\Omega) \to \mathbb{R}$ is convex, $\psi(\overline{p}) = 0$ and $\psi \geqslant 0$. Because such a cost function c is linear in π, I refer to them as the Linearity Axioms. Axioms 6-8 are the main contribution of this paper. They show the extra assumptions that are required, beyond the formula in (3.1), for the preference to be represented by the Rational Inattention Model.

3.1 Extension Axioms

Axiom 1 (Monotone Continuity). *Let F, G and H be menus such that $G \succ H$. If $E_1 \supseteq E_2 \supseteq E_3 \ldots$ is a sequence of events such that $\bigcap_n E_n$ is null, there exists an $N \in \mathbb{N}$ large enough, such that*

$$n \geqslant N \implies F[E_n]\,G \succ F[E_n]\,H.$$

This axiom is a variation of a standard assumption on preferences over acts that guarantees countable additivity of the prior.

Axiom 2 (Absolute Continuity). *Let E be an event such that $\lambda(E) = 0$ (λ is the Lebesgue measure on $[0,1]$). Then for all outcomes x and y, $xEy \sim x$.*

This axiom allows us to restrict attention to beliefs that admit densities (with respect to the Lebesgue measure).

3.2 Linearity Axioms

Axiom 3 (Independence of Irrelevant Alternatives). *If $F \sim F \cap G \sim G$ then $F \sim F \cup G$.*

If the condition holds, we may think of the menus F and G as adding some options to $F \cap G$ which turn out to be irrelevant. The axiom states that these extra options are still irrelevant it they are simultaneously added to the menu $F \cap G$.

Axiom 4 (Linearity). *For every menu F, there exists an act h such that $h \sim F \sim F \cup \{h\}$*

This axiom states that there exists an act which is, at the same time, irrelevant and indifferent to the menu F.

Axiom 5 (Unbounded Attention). *Let $E \subseteq \Omega$ be non-null and y be an outcome. Then, for every $\alpha \in (0,1)$, there exist $x \succ y$ such that*

$$\{xEy, yEx\} \succ \alpha x + (1-\alpha)\, y.$$

If the decision-maker could freely learn whether the true state lies or not in E, she could always pick the better outcome x, so the menu would be indifferent to x. The axiom states that this situation can be approximated by choosing an outcome $x \in X$ that is sufficiently good. In other words, if the incentive to learn some information is high enough, the decision maker can do so with arbitrarily high precision.

3.3 Entropy Axioms

This section shows the main three axioms of the representation: Symmetry, Separability in Orthogonal Decisions, and Irrelevance of Orthogonal Flexibility.

3.3.1 Symmetry

To discuss the Symmetry Axiom, we first need to define a notion of relabeling of states.

Definition 6. A measurable bijection $\sigma : \Omega \to \Omega$ is called a *rearrangement* if $h \sim h \circ \sigma$ for every act h.

Here $h \circ \sigma$ is the act obtained from the composition of the act h with the rearrangement σ. To understand what a rearrangement means, it is helpful to think of the case when Ω is finite. In that case, σ being a bijection means that it is a permutation of the states. When the decision maker evaluates acts using expected utility, σ is a rearrangement if it only permutes states that are regarded as equally likely. Indeed, taking $x \succ y$, the act $h = xEy$ may be seen as a bet on the event E. The act $h \circ \sigma$, on the other hand, gives x when the state ω is such that $\sigma(\omega) \in E$ and y otherwise. In other terms, $h \circ \sigma = x \left[\sigma^{-1}(E)\right] y$. Therefore, the act $h \circ \sigma$ is a bet on the event $\sigma^{-1}(E)$; for a decision maker that evaluates acts using expected utility, the preference $h \sim h \circ \sigma$ is equivalent to E and $\sigma^{-1}(E)$ being equally likely.

Using the decision-maker's subjective prior $\overline{p}$, we may regard each act $f : \Omega \to X$ as a random variable that returns an outcome in X. A rearrangement then preserves the distribution of the random variable. Likewise, we may think of a menu as a set of random variables. A rearrangement of F, given by

$$F \circ \sigma = \{f \circ \sigma : \ f \in F\}$$

not only preserves the distribution of each act $f \in F$, but also preserves their *joint distribution.*

Axiom 6 (Symmetry)**.** *If σ is a rearrangement, then $F \circ \sigma \sim F$ for every menu F.*

Using expected utility, this may be rephrased as stating that events that are equally likely can have their roles exchanged without affecting the preference. For example, consider the partitions $\mathcal{P} = \{D, D^c\}$ and, given a rearrangement σ, let $E = \sigma^{-1}(D)$, so that the events D and E have the same subjective probability. Given a menu $F = \{xDy, yDx\}$, then $F \circ \sigma = \{xEy, yEx\}$. If the decision-maker finds it easier to learn information about D than to learn about E, her preference should be $F \succ F \circ \sigma$. Symmetry rules this out: events that are equally likely are also exchangeable for attention purposes.

Symmetry rules out situations where the information is only available in disparate formats, which may differ in how much effort they require to understand. For example, it may be that learning about the partition $\mathcal{P} = \{D, D^c\}$ can be done by reading a newspaper in English, while to learn about the partition $\mathcal{Q} = \{E, E^c\}$ one would have to read a french newspaper. A decision-maker who can only read English might not regard the two exchangeably, even though she might regard D and E as equally likely.

There are, however, at least two situations where the assumption of symmetry is plausible. First, in many models the set of "states of the world" can be understood in a narrow sense and it is plausible to assume that the information is presented in a uniform fashion. For example, if the states of the world are prices for some goods and the prices and those are all presented in dollars, then symmetry seems plausible. However, if prices are presented in different currencies, it might not be. For example, the event D might mean "good A is more expensive than good B" while the event E might mean "good A is more expensive than good C". If the price of A and B are displayed in the same currency, while the price of C is displayed in a different one, then it is possible that D and E are a priori equally likely, but learning about D is easier than learning about E.

Second, the decision-maker may have available resources that allow for relatively cheap conversion of information from one format to another.[5] For example, while displaying prices in different currencies may be a considerable hindrance for a small customer, it should be relatively unimportant for a large financial firm.

Axiom 7 (Separability in Orthogonal Decisions)**.** *Let $\mathcal{P}$ be a partition and E be an event orthogonal to $\mathcal{P}$. Suppose that F and F' are $\mathcal{P}$-measurable and G and G' are $\{E, E^c\}$-measurable. Then*

$$\alpha F + (1-\alpha) G \succsim \alpha F' + (1-\alpha) G \Rightarrow \alpha F + (1-\alpha) G' \succsim \alpha F' + (1-\alpha) G'.$$

When facing $\alpha F + (1-\alpha) G$, the attention that the decision maker pays can be useful both for F and for G, but if $\mathcal{P}$ and E are orthogonal, learning which cell of $\mathcal{P}$ contains the true state does not say anything about how likely it is that E occurred. In this sense, information that is specifically helpful for the decision in F cannot improve the decision for G, and vice-versa. The axiom states that, in this case, the decision of which between F and F' is preferred is independent of G—the problems are *separable*.

The axiom also rules out situations where the cost of learning useful information about F goes up as more effort is spent learning the information that is useful to G. To see this, consider the case where F and G' are general menus whereas F' and G are menus of a single constant act. When the decision maker faces $\alpha F + (1-\alpha) G$ she has no decision to take in case G occurs, so that when paying attention to the state, she is free to worry only about learning information that is relevant for F. However, when facing $\alpha F + (1-\alpha) G'$ she has to worry both about which act to pick in case F occurs and which act to pick in case G occurs. If spending this attention to decide from G makes paying attention to decide from F more expensive, then her preference could revert, violating the axiom.

Axiom 8 (Irrelevance of Orthogonal Flexibility)**.** *Let $\mathcal{P}$ be a partition and E be an event orthogonal to $\mathcal{P}$ and let F be $\mathcal{P}$-measurable and G be $\{E, E^c\}$-measurable. Then $\alpha F + (1-\alpha) G \sim \alpha F [E] F + (1-\alpha) G$.*

To understand the axiom, suppose first that $\alpha = 1$. We can interpret the menu $F [E] F$ as the situation where the decision-maker learns in advance whether E has occurred or not.

[5] This is an important assumption in the theorems on "efficient coding" in Information Theory. The idea there is that the bottleneck lies in the transmission of information, not in the compression or decompression that may happen at both ends.

Since E is orthogonal to $\mathcal{P}$, learning about the occurrence of E does not help in deciding which act in F to pick, so this information is useless; this can be expressed as $F \sim F[E]F$.

Notice that $F \subseteq F[E]F$ and therefore $F \precsim F[E]F$, by preference for flexibility. The axiom implies that for an E which is orthogonal to $\mathcal{P}$, the decision-maker does not value this extra flexibility. But it says more: the same holds even if there's also a decision G which depends only on E but not on $\mathcal{P}$.

4 Representation

Before stating the main result, we generalize some notions for the infinite state space $\Omega = [0,1]$. Recall that $\Delta_d[0,1]$ denotes the set of probability distributions over $[0,1]$ that admit a density. Denote by $\Pi(\overline{p})$ the set of distributions π over $\Delta_d(\Omega)$ that have finite support and satisfy $\int p(\omega)\pi(dp) = \overline{p}(\omega)$ for every p in the support of π . The entropy of a density p is defined by

$$H(p) = -\int_0^1 p(\omega)\log p(\omega)\,d\omega.$$

We can now state the main theorem.

Theorem 1. *The following statements are equivalent:*

1. *The preference relation $\succeq$ satisfies axioms 1-8 and the axioms of De Oliveira et al. [2013] (listed in appendix A.1).*

2. *There exist an affine utility function $u : X \to \mathbb{R}$ with unbounded range and a prior $\overline{p} \in \Delta_d[0,1]$, such that the preference $\succeq$ is represented by the function $V : \mathbb{F} \to \mathbb{R}$ defined as*

$$V(F) = \max_{\pi\in\Pi(\bar{p})}\left[\int_{\Delta_d[0,1]}\max_{f\in F}\left(\int_0^1 u(f(\omega))\,p(\omega)d\omega\right)\pi(dp) - c(\pi)\right].$$

where

$$c(\pi) = \int H(\overline{p}) - H(p)\,\pi(dp).$$

Moreover, the representation is unique up to the addition of a constant to u.

The uniqueness result is not stated in terms of all affine transformations of u, only the addition of constants. This may be regarded as a normalization caused by the choice of c.

If instead we wrote the representation with a constant multiplying c, the uniqueness result would be stated in terms of affine transformations.

5 Discussion

Theorem 1 can be used, as was discussed in the introduction, in discussing the suitability of the Rational Inattention Model for specific applications. It can also be useful to discuss alternative models, which may violate some of the axioms. We can then ask what is the contextual meaning of such violations.

A promising direction for future research is relaxing or modifying the axioms and examining the implications for the representation. Notably, some of the axioms I presented are not satisfied by Sims's original formulation of the problem, where the decision-maker maximizes the value of information subject to a constraint on the mutual information: $\mathcal{I}(\pi) \leqslant \kappa$. Characterizing the preferences resulting from this representation could help understand when the constrained or the linear model is more appropriate.

However, both the Rational Inattention Model analyzed here and Sims's formulation with a constraint represent extreme situations. The constraint reflects the decision-maker's inherent capacity for absorbing information, which she cannot control. This can be understood as an assumption that the model really incorporates everything that the decision-maker can pay attention to, so that fully utilizing her capacity is optimal. In contrast, the Rational Inattention Model reflects the idea of a constant cost for each unit of capacity. This may be reasonable, for example, if we understand that the model refers only to a small part of all the decisions that the decision-maker faces, so that the decision-maker may reallocate more attention from unmodeled tasks to the modeled task as the latter becomes more important. In that sense, the cost could come from the resulting payoff loss in the unmodeled tasks. It also incorporates the intuitive idea that the total amount of attention being used at a given time is under control, not just their qualitative feature (where the attention is directed).

Ideally, the model would incorporate both the idea that attention is scarce and that there is some flexibility in its supply; that as more attention is paid, it becomes more expensive. An approach that has been used in applications is to consider a distortion of the cost: $c(\pi) = f(\mathcal{I}(\pi))$, where f is increasing and convex (see Paciello and Wiederholt [2011]). The preferences implied by this model violate Separability in Orthogonal Decisions (see the discussion following the axiom). Relaxing SOD is therefore a natural step in unifying the constrained and linear models, even if it does not lead directly to a characterization of this distorted cost. In fact, the axiomatic approach would be even more appealing in that case, for it could show that the more natural generalization (in terms of its preference implications) may not be as immediately derived from the functional form.

A Extension of DDMO

This section extends the result of DDMO to the larger state-space $\Omega = [0, 1]$.

A.1 DDMO Axioms

The axioms of De Oliveira et al. [2013] are listed below for completeness. For a discussion, the reader should refer to that paper.

Axiom 9 (Weak Order). *The binary relation $\succeq$ is complete and transitive.*

Axiom 10 (Continuity). *Consider three menus F, G and H. The sets*

$$\{\alpha \in [0,1] : \alpha F + (1-\alpha)G \succeq H\} \quad \text{and} \quad \{\alpha \in [0,1] : H \succeq \alpha F + (1-\alpha)G\}$$

are closed.

Axiom 11 (Unboundedness). *There are consequences x and y with $x \succ y$ such that for all $\alpha \in (0, 1)$ there is a consequence z satisfying either $y \succ \alpha z + (1-\alpha)x$ or $\alpha z + (1-\alpha)y \succ x$.*

Axiom 12 (Weak Singleton Independence). *Consider a pair of menus F and G and a pair of acts h and h'. For each $\alpha \in (0, 1)$*

$$\alpha F + (1-\alpha)h \succeq \alpha G + (1-\alpha)h \quad \Rightarrow \quad \alpha F + (1-\alpha)h' \succeq \alpha G + (1-\alpha)h'.$$

Axiom 13 (Aversion to Randomization). *Consider a pair of menus F and G. For each $\alpha \in (0, 1)$*

$$F \sim G \quad \Rightarrow \quad F \succeq \alpha F + (1-\alpha)G.$$

Axiom 14 (Dominance). *Consider a pair of menus F and G such that for each act $g \in G$ there is an act $f \in F$ with $f(\omega) \succeq g(\omega)$ for each $\omega \in \Omega$. Then $F \succeq G$.*

A.2 Representation

A.2.1 Preliminaries

We first introduce notation and definitions that will be used in the proof.

- The state space $\Omega = [0, 1]$ is endowed with the Borel σ-algebra, denoted by $\mathcal{B}$.

- We denote by$\Delta(\Omega)$ the set of Borel finitely additive probabilities on Ω. Let $B(\Omega)$ be the set of Borel measurable and bounded real functions. We endow $\Delta(\Omega)$ with the weak* topology $\sigma(\Delta(\Omega), B(\Omega))$. Recall that in this topology a net (p_α) in $\Delta(\Omega)$ converges to $p \in \Delta(\Omega)$ if and only if $\int \xi dp_\alpha \to \int \xi dp$ for every $\xi \in B(\Omega)$. The topology makes $\Delta(\Omega)$ a compact Hausdorff space. Endow $\Delta(\Omega)$ with the Borel sigma algebra $\mathcal{B}(\Delta(\Omega))$ generated by this topology.

- Let $C(\Delta(\Omega))$ be the set of continuous functions real valued functions on $\Delta(\Omega)$, endowed with the sup-norm.

- We denote by $\Delta_\sigma(\Delta(\Omega))$ the set of countably additive regular Borel measures on $\Delta(\Omega)$. Endow $\Delta_\sigma(\Delta(\Omega))$ with the weak* topology $\sigma(\Delta_\sigma(\Delta(\Omega)), C(\Delta(\Omega)))$. A net (π_α) in $\Delta_\sigma(\Delta(\Omega))$ converges to $\pi \in \Delta_\sigma(\Delta(\Omega))$ if and only if $\int \eta d\pi_\alpha \to \int \eta d\pi$ for every $\eta \in C(\Delta(\Omega))$. In this topology, $\Delta_\sigma(\Delta(\Omega))$ is compact Hausdorff.

- Fix $\overline{p} \in \Delta$. We denote by$\Pi(\overline{p})$ the set of $\pi \in \Delta_\sigma(\Delta(\Omega))$ such that

$$\int_{\Delta(\Omega)} p(E)\,\pi(dp) = \overline{p}(E)$$

  for every $E \in \mathcal{B}$. Call $\pi \in \Pi(\overline{p})$ a *channel.* The integral is well defined. To see this, notice that for every event E the function $p \mapsto p(E)$ is $\sigma(\Delta(\Omega), B(\Omega))$-continuous. Hence, it is measurable. The set $\Pi(\overline{p}) \subseteq \Delta_\sigma(\Delta(\Omega))$ is convex and closed (hence compact).

- $\mathbb{F}$ denotes the set of all menus.

To each $F \in \mathbb{F}$, associate the function $\varphi_F : \Delta(\Omega) \to \mathbb{R}$ defined as

$$\varphi_F(p) = \max_{f \in F} \int_\Omega u(f(\omega))\,p(d\omega) \quad \forall p \in \Delta(\Omega).$$

Because φ_F is the supremum of a finite number of continuous functions, it is continuous. Let $\Phi_{\mathbb{F}} = \{\varphi_F : F \in \mathbb{F}\}$.

A.2.2 Niveloids

Let Φ be a subset of $C(\Delta(\Omega))$. A function $W : \Phi \to \mathbb{R}$ is *normalized* if $W(\alpha) = \alpha$ for every constant function $\alpha \in \Phi$. It is *translation invariant* if $W(\varphi + \alpha) = W(\varphi) + \alpha$ for all $\varphi \in \Phi$

and $\alpha \in \mathbb{R}$ such that $\varphi + \alpha \in \Phi$. It is a *niveloid* if $W(\varphi) - W(\psi) \leq \sup_{p \in \Delta(\Omega)} (\varphi(p) - \psi(p))$ for all $\varphi, \psi \in \Phi$. A niveloid is normalized, monotone and translation invariant (See Proposition 2 in Cerreia-Vioglio at al. (2012)).

Proposition 1. *Let $\Phi \subseteq C(\Delta(\Omega))$ be convex. A convex niveloid $W : \Phi \to \mathbb{R}$ can be extended to a convex niveloid $\bar{W} : C(\Delta(\Omega)) \to \mathbb{R}$.*

Proof. Proposition 4 in Cerreia-Vioglio et al. (2012). □

A.2.3 Information

Definition 7. Fix $\bar{p} \in \Delta(\Omega)$. Let π and ρ be two channels in $\Pi(\bar{p})$. We say that π is *more informative* than ρ, denoted $\pi \trianglerighteq \rho$, if

$$\int_{\Delta(\Omega)} \varphi(p)\, \pi(dp) \geqslant \int_{\Delta(\Omega)} \varphi(p)\, \rho(dp)$$

for every $\varphi \in \Phi_{\mathbb{F}}$.

Definition 8. Given a prior $\bar{p} \in \Delta(\Omega)$, a function $c : \Pi(\bar{p}) \to [0, \infty]$ is an *information cost function* if it is lower semicontinuous and satisfies the following properties:

1. *Grounded*: $c(\pi) = 0$ whenever $\pi(\{\bar{p}\}) = 1$.
2. *Convex*: $c(\alpha\pi + (1-\alpha)\rho) \leq \alpha c(\pi) + (1-\alpha)c(\rho)$ whenever $\pi, \rho \in \Pi(\bar{p})$ and $\alpha \in (0, 1)$.
3. *Blackwell Monotone*: $c(\rho) \leq c(\pi)$ whenever $\pi, \rho \in \Pi(\bar{p})$ and π is more informative than ρ.

A.2.4 Theorem

Theorem 2. *The following statements are equivalent:*

1. *The preference relation $\succeq$ satisfies axioms 8-13.*
2. *There exist an affine utility function $u : X \to \mathbb{R}$ with unbounded range, a prior $\bar{p} \in \Delta(\Omega)$, and an information cost function $c : \Pi(\bar{p}) \to [0, \infty]$, such that the preference $\succeq$ is represented by the function $V : \mathbb{F} \to \mathbb{R}$ defined as*

$$V(F) = \max_{\pi \in \Pi(\bar{p})} \left[\int_{\Delta(\Omega)} \max_{f \in F} \left(\int_{\Omega} u(f(\omega))\, p(d\omega) \right) \pi(dp) - c(\pi) \right] \quad \forall F \in \mathbb{F}.$$

A.2.5 Proof

The result is based on the proof of Theorem 1 in DDMO.

Claim 1. There exists an affine function $u : X \to \mathbb{R}$ with unbounded range, a prior $\overline{p} \in \Delta(\Omega)$ such that the function

$$U(f) = \int_{\Omega} u(f(\omega))\, d\overline{p}(\omega)$$

represents the preference $\succeq$ over $\mathcal{F}$. Moreover, if $(u, \overline{p})$ and $(u\prime, \overline{p}\prime)$ both represent $\succeq$ over $\mathcal{F}$ then $\overline{p} = \overline{p}\prime$ and there exists $\alpha > 0$ and β in $\mathbb{R}$ such that $u\prime = \alpha u + \beta$.

Proof. The proof in DDMO can be replicated verbatim. □

Claim 2. For every menu F there exists an outcome $x_F \in X$ such that $x_F \sim F$.

Proof. Because each $f \in F$ takes finitely many values, by monotonicity there exists outcomes x and y such that $x \succeq f \succeq y$. Because F is finite, we can choose outcomes x and y such that $x \succeq f \succeq y$ for every $f \in F$. By dominance, $x \succeq F \succeq y$. By continuity there exists $\alpha \in [0, 1]$ such that $\alpha x + (1 - \alpha) y \sim F$. Let $x_F = \alpha x + (1 - \alpha) y$. □

Now we define a functional $W : \Phi_{\mathbb{F}} \to \mathbb{R}$ such that

$$W(\varphi_F) = u(x_F)$$

for every menu F. To show that W is well-defined, we need to prove that $\varphi_F = \varphi_G$ implies $F \sim G$ hence $u(x_F) = u(x_G)$. As in DDMO, the next two claims accomplish this.

Claim 3. Consider a pair of menus F and G. If $\varphi_F \geqslant \varphi_G$, then for each $g \in G$ there exists $f \in coF$ (where coF is the convex hull of F) such that $f(\omega) \succsim g(\omega)$ for each $\omega \in \Omega$.

Proof. There exists a finite partition $\mathcal{P}$ such that both all acts in F and G are measurable with respect to $\mathcal{P}$. Restricting attention to $\mathcal{P}$-measurable acts, the same finite-dimensional argument in DDMO applies here. □

Claim 4. Consider a pair of menus F and G. If $G \subseteq coF$ then $F \succsim G$.

Proof. Let $G = \{g_1, ..., g_K\}$ and $G \subseteq coF$. The act g_1 can be written as the convex combination

$$g_1 = \sum_{f \in F} \alpha_f f$$

Let $G_1 = \sum_{f \in F} \alpha_f F$. Then $g_1 \in G_1$ and $F \subseteq G_1$. Aversion to randomization implies $F \sim G_1$. Suppose that for $k < K$ there exists a menu G_k such that $g_1, ..., g_k \in G_k$, $F \subseteq G_k$ and $F \sim G_k$. The act g_{k+1} can be written as the convex combination

$$g_{k+1} = \sum_{g \in G_k} \beta_g g$$

Let $G_{k+1} = \sum_{g \in G_k} \beta_g G_k$. Then $F \sim G_k \sim G_{k+1}$. By induction, there exists a menu G_K such that $G \subseteq G_K$, $F \subseteq G_K$ and $F \sim G_K$. By dominance, $G_K \succsim G$. Then $F \succsim G$. $\square$

We now show that if if $\varphi_F = \varphi_G$ then $F \sim G$. By claim 3, we can find a subset $H \subset coF$ such that for each $g \in G$ there exists $h \in H$ such that $h(\omega) \succsim g(\omega)$ for all $\omega \in \Omega$. By Claim 4, $F \succeq H$. By Dominance, $H \succeq G$. Hence $F \succeq G$. Similarly, $G \succeq F$. Hence $F \sim G$. The result shows that W is well-defined and monotone.

Claim 5. The functional W is a monotone, normalized, convex niveloid.

Proof. The proof in DDMO can be repeated verbatim. $\square$

Claim 6. *For each menu F,*

$$W(\varphi_F) = \max_{\pi \in \Pi(\bar{p})} \langle \varphi_F, \pi \rangle - c(\pi)$$

where $c : \Pi(\bar{p}) \to (-\infty, \infty]$ is such that

$$c(\pi) = \sup_{F \in \mathbb{F}} \langle \varphi_F, \pi \rangle - W(\varphi_F) \quad \forall \pi \in \Pi(\bar{p}).$$

Proof. By Proposition (1), the functional W extends to a convex niveloid on $C(\Delta(\Omega))$. Without ambiguity we denote the extension by W. Because W is a niveloid, then it is Lipschitz continuous. By Lemma 25 in Ergin and Sarver [2010], the subdifferential of W is nonempty at every $\varphi \in C(\Delta(\Omega))$. That is, for every $\varphi \in C(\Delta(\Omega))$ there exists a signed measure $m = \alpha\pi - \beta\nu$ where $\alpha, \beta \in \mathbb{R}_+$and $\pi, \nu \in \Delta_\sigma(\Delta(\Omega))$ such that

$$\langle \varphi, m \rangle - W(\varphi) \geq \langle \psi, m \rangle - W(\varphi) \quad \forall \psi \in C(\Delta(\Omega))$$

Because W is monotone and translation invariant, then we can take $m = \pi$ for some $\pi \in \Delta_\sigma (\Delta (\Omega))$ (see Ruszczyński and Shapiro [2006]). Now let

$$W^* (\pi) = \sup_{F \in \mathbb{F}} \langle \varphi_F, \pi \rangle - W(\varphi_F) \quad \forall \pi \in \Delta_\sigma (\Delta (\Omega))$$

Fix $\varphi_F \in \Phi_{\mathbb{F}}$. If π is in the subdifferential of W at φ_F then $\langle \varphi_F, \pi \rangle - W(\varphi_F) \geq \langle \varphi_G, \pi \rangle - W(\varphi_G)$ for every $G \in \mathbb{F}$. Hence $W^* (\pi) = \langle \varphi_F, \pi \rangle - W(\varphi_F)$. Therefore

$$W(\varphi_F) = \max_{\pi \in \Delta(\Delta(\Omega))} \langle \varphi_F, \pi \rangle - W^*(\pi)$$

We now prove that $W(\varphi_F) = \max_{\pi \in \Pi(\bar{p})} \langle \varphi_F, \pi \rangle - W^*(\pi)$. To this end, let $W^* (\pi) < \infty$. Fix an act $f = xEy$ such that $u(y) = 0$. Then

$$\langle \varphi_f, \pi \rangle - W^* (\pi) = u(x) \int_{\Delta(\Omega)} p(E) \, \pi(dp) - W^* (\pi) \leq W(f) = u(x) \, \bar{p}(E)$$

hence

$$\int_{\Delta(\Omega)} p(E) \, \pi(dp) \leq \bar{p}(E)$$

for every event E. Hence $\int_{\Delta(\Omega)} p(E) \, \pi(dp) = \bar{p}(E)$ for every event E. □

Claim 7. The function c is an information cost function.

Proof. Repeat verbatim. □

B Proof of Theorem 1

This section shows the sufficiency part of Theorem 1. The necessity of the axioms is left to the reader.-

B.1 Some Technical results

B.1.1 Finite support channels

We now show that it is without loss of generality to restrict our attention to channels with finite support.

Lemma 1. *Let F be a menu and $\pi \in \Pi(\bar{p})$ a channel. There exists a channel $\pi_F \in \Pi(\bar{p})$ with finite support such that:*

1. $\pi \trianglerighteq \pi_F$ *and*
2. $\langle \varphi_F, \pi \rangle = \langle \varphi_F, \pi_F \rangle$

Proof. Fix a menu F. It is without loss of generality to assume that there are no act f, g in F such that $f(\omega) \succeq g(\omega)$ for all $\omega \in \Omega$. For each $f \in F$, define the set of probabilities for which f is the best choice:

$$P_f = \left\{ p \in \Delta(\Omega) : \int u(f)\, dp \geqslant \int u(g)\, dp \, \forall g \in F \right\}.$$

P_f is a non-empty convex set; it is weak* closed, hence weak* compact. Let $\pi \in \Pi(\bar{p})$ be a channel. For each $f \in F$ such that $\pi(P_f) > 0$, let $p_f \in \Delta(\Omega)$ be defined as

$$p_f(E) = \int_{P_f} p(E)\, d\pi(p|P_f)$$

for every event E. For every $g \in F$, we have

$$\int u(f)\, dp_f = \int_{P_f} \left(\int u(f)\, dp \right) d\pi(p|P_f) \geq \int_{P_f} \left(\int u(g)\, dp \right) d\pi(p|P_f) = \int u(g)\, dp_f$$

hence $p_f \in P_f$.

Define a measure $\pi_F \in \Delta_\sigma(\Delta(\Omega))$ by $\pi_F(p_f) = \pi(P_f)$ for every $f \in F$. Let $F_+ = \{f \in F : \pi(P_f) > 0\}$. For every event E,

$$\sum_{f \in F_+} p_f(E)\, \pi_f(p_f) = \sum_{f \in F_+} \left(\int_{P_f} p(E)\, d\pi(p|P_f) \right) \pi(P_f) = \int_{\Delta(\Omega)} p(E)\, \pi(dp) = \bar{p}(E)$$

Thus, $\pi_F \in \Pi(\bar{p})$. Given any convex function φ, we have

$$\sum_{f \in F} \varphi(p_f)\, \pi_F(p_f) = \sum_{f \in F_+} \varphi\left(\frac{1}{\pi(P_f)} \int_{P_f} q\pi(dq) \right) \pi(P_f) \leqslant \sum_{f \in F_+} \int_{P_f} \varphi(q)\, \pi(dq) = \int_{\Delta(\Omega)} \varphi(q)\, \pi(dq).$$

This shows that $\pi \trianglerighteq \pi_F$. On the other hand, by the definition of π_F,

$$\langle \varphi_F, \pi \rangle = \sum_{f \in F_+} \left(\int_{P_f} \left(\int_\Omega u(f)\, dp \right) d\pi(p|P_f) \right) \pi(P_f) = \sum_{f \in F} \left(\int_\Omega u(f)\, dp_f \right) \pi(p_f) = \langle \varphi_F, \pi_F \rangle .$$

□

Corollary 1. *For every menu F, there exists a discrete channel $\pi \in \Pi(\bar{p})$ which is optimal for F, that is, $\partial V(F) \cap \Pi(\overline{p}) \neq \emptyset$.*

Proof. By Blackwell monotonicity of c, $c(\pi_F) \leqslant c(\pi)$. It follows from lemma 1 that

$$\langle \varphi_F, \pi \rangle - c(\pi) \leqslant \langle \varphi_F, \pi_F \rangle - c(\pi_F) .$$

□

Let $\partial V(F) = \partial_{Fin} V(F) \cap \Pi(\overline{p})$.

B.1.2 Countable additivity and absolute continuity of posteriors

Axiom 15 (Monotone Continuity)**.** *Let (E_n) be a sequence of events such that $E_n \downarrow \emptyset$. Then for all outcomes $x \succ y$ and z there exists N such that for every $n \geq N$,*

$$zE_nx \succ y \text{ and } x \succ zE_ny$$

The axiom implies that the prior $\bar{p}$ is countably additive.

Proposition 2. *Let $\pi \in \Pi(\bar{p})$ have finite support. Fix $p \in supp(\pi)$. If $\bar{p}$ is countably additive then p is countably additive. If $\bar{p} \ll \lambda$ then $p \ll \lambda$.*

Proof. Let (E_n) be a sequence of events such that $E_n \downarrow \emptyset$. Then $\bar{p}(E_n) = \sum_{p \in supp(\pi)} p(E_n)\, \pi(p)$. Because $\bar{p}(E_n) \to 0$, then $p(E_n) \to 0$ for every $p \in supp(\pi)$. Similarly, if E is an event such that $\lambda(E) = 0$ then $\bar{p}(E) = 0$ hence $p(E) = 0$ for all $p \in supp(\pi)$. □

B.2 Utility menus

Theorem 2 provides a utility function $u : X \to \mathbb{R}$. By composition, we may define, for each act $f \in \mathcal{F}$, a corresponding *utility act* $u \circ f : \Omega \to \mathbb{R}$; we may think of this as a map $u^* : \mathcal{F} \to B_0 \subseteq \mathbb{R}^\Omega$, where B_0 is the set of simple, Borel measurable functions from Ω to $\mathbb{R}$.

Conversely, Unboundedness implies that the image of u is$\mathbb{R}$. Therefore, given any simple Borel function $f^u : \Omega \to \mathbb{R}$, there exists an act $f : \Omega \to X$ such that $u \circ f = f^u$; think of this as a map $u_* : B_0 \to \mathcal{F}$. Notice that given f, the utility act $u \circ f$ is necessarily unique, but given f^u there may be many acts f satisfying $u \circ f = f^u$.

We may also define, for each menu F, the corresponding *utility menu*

$$u \circ F = \{u \circ f : f \in F\}.$$

It is a simple corollary of Claim 3 that

$$u \circ F = u \circ G \implies F \sim G.$$

This means that the preference relation $\succsim$ is isomorphic to a preference relation $\succsim^u$ over finite subsets of B_0, defined by

$$F \succsim G \iff u \circ F \succsim^u u \circ G.$$

The following lemmata show why this is a useful translation:

Lemma 2. *The following statements are equivalent:*

1. *$\succsim$ satisfies Singleton Independence*
2. *For every F^u, G^u finite subsets of B_0 and $h^u, h'^u \in B_0$, we have*

$$F^u + h^u \succsim^u G^u + h^u \iff F^u + h'^u \succsim^u G^u + h'^u.$$

Proof. Let F, G be menus and h, h' be acts such that

$$u \circ F = 2F^u \quad u \circ G = 2G^u \quad u \circ h = 2h^u.$$

Then

$$\begin{aligned} \frac{1}{2}F + \frac{1}{2}h \succsim \frac{1}{2}G + \frac{1}{2}h \quad &\iff \quad u \circ \left(\frac{1}{2}F + \frac{1}{2}h\right) \succsim^u u \circ \left(\frac{1}{2}G + \frac{1}{2}h\right) \\ &\iff \quad \frac{1}{2}u \circ F + \frac{1}{2}u \circ h \succsim \frac{1}{2}u \circ G + \frac{1}{2}u \circ h \\ &\iff \quad F^u + h^u \succsim^u G^u + h^u. \end{aligned}$$

The proof of the converse is analogous. $\square$

Lemma 3. *The following statements are equivalent:*

1. $\succsim$ *satisfies Separability in Orthogonal Decisions*
2. *Let* F^u, F'^u *be* $\mathcal{P}$*-measurable and* G^u, G'^u *be* $\mathcal{Q}$*-measurable, where* $\mathcal{P}$ *is orthogonal to* $\mathcal{Q}$*. Then*
$$F^u + G^u \succsim F'^u + G^u \implies F^u + G'^u \succsim F'^u + G'^u$$

Proof. The proof is analogous to that of Lemma 2.

$\square$

B.2.1 Utility implications

Given the utility function for menus V, we may define a utility function for utility menus V^u, by $V^u(u \circ F) = V(F)$. The following is a simple consequence of the niveloid properties of V.

Lemma 4. V^u *satisfies*

1. *If* F^u *is* $\mathcal{P}$*-measurable and* G^u *is* $\mathcal{Q}$*-measurable, where* $\mathcal{P}$ *and* $\mathcal{Q}$ *are orthogonal, then*
$$V^u(F^u + G^u) = V^u(F^u) + V^u(G^u)$$
2. *If* F^u *is a utility menu and* h^u *is a utility act, then*
$$V^u(F^u + h^u) = V^u(F^u) + \langle h^u, \overline{p} \rangle$$

The proof of this is a simple result of the niveloid properties of V. From now on, we will only consider utility acts and utility menus, and the superscripts u will be omitted.

B.3 Linear cost

Let $\mathcal{H}$ be the set of acts that are irrelevant to the singleton $\{0\}$, that is,

$$\mathcal{H} = \{h \in \mathcal{F} :\ 0 \sim \{0, h\}\}. \tag{B.1}$$

Proposition 3. $\mathcal{H}$ *satisfies the following properties:*

1. *For any menu F. we have $F \subseteq \mathcal{H}$ if and only if $0 \sim F \cup \{0\}$;*

2. *$\mathcal{H}$ is convex;*

3. *For every menu F, there exists an act g s.t. $F - g \subseteq \mathcal{H}$ and $0 \in F$;*

4. *$\mathcal{H}$ is symmetric: if σ is a rearrangement and $h \in \mathcal{H}$ then $h \circ \sigma \in \mathcal{H}$.*

Proof. (1) Let $F = \{f_1, f_2 \ldots, f_n\}$. Suppose that $F \subseteq \mathcal{H}$. The result can be proved by induction. If F has one element, it follows by the definition of $\mathcal{H}$. Now let $F_n = F_{n-1} \cup \{f_n\}$ be of size n. By the induction hypothesis, we have $F_{n-1} \cup \{0\} \sim 0$. Since $f_n \in F_n \subseteq \mathcal{H}$, it follows that $0 \sim \{0, f_n\}$. By IIA, we must have $F_n \cup \{0\} = F_{n-1} \cup \{0, f_n\} \sim 0$, as we wanted. To see the converse, note that $0 \sim F \cup \{0\} \succsim \{f_i, 0\} \succsim 0$ for every $f_i \in F$.

(2) Let $h, h' \in \mathcal{H}$. By (1), we know that $0 \sim \{0, h, h'\}$. Since $\{0, \alpha h + (1-\alpha) h'\} \subseteq co\{0, h, h'\}$, it follows from claim 4 that $0 \sim \{0, h, h'\} \succsim \{0, \alpha h + (1-\alpha) h'\}$, so $h + (1-\alpha) h' \in \mathcal{H}$.

(3) By linearity, for every F, there exists an act g such that $g \sim F \sim F \cup \{g\}$. Letting $F' = F - g$, we have $F' \sim 0 \sim F' \cup \{0\}$, so we must have $F \subseteq \mathcal{H}$.

(4) Since $0 \circ \sigma = 0$, we have, by Symmetry, that $\{0, h\} \sim \{0, h\} \circ \sigma = \{0, h \circ \sigma\}$. □

Now define

$$\psi(p) = \sup_{h \in \mathcal{H}} \int h \, dp.$$

Let $H = \{0, h_1, h_2, \ldots h_n\} \subseteq \mathcal{H}$.

Lemma 5. *The cost function c is given by*

$$c(\pi) = \sup_{\substack{F \subseteq \mathcal{H} \\ 0 \in F}} \langle \phi_F, \pi \rangle .$$

Proof. Let F be any menu. By property (3) in Proposition 3 , there exists an act g such that $F - g \subseteq \mathcal{H}$ and $0 \in F$. Since $V(F-g) = V(F) - \phi_g(\overline{p})$ and $\langle \phi_{F-g}, \pi \rangle = \langle \phi_F - \phi_g, \pi \rangle = \langle \phi_F, \pi \rangle - \phi_g(\overline{p})$, we have

$$\langle \phi_{F-g}, \pi \rangle - V(F-g) = \langle \phi_F, \pi \rangle - V(F).$$

Therefore

$$c(\pi) = \sup_{\substack{F \subseteq \mathcal{H} \\ 0 \in F}} \langle \phi_F, \pi \rangle - V(F) = \sup_{\substack{F \subseteq \mathcal{H} \\ 0 \in F}} \langle \phi_F, \pi \rangle,$$

since $F \subseteq \mathcal{H}$ and $0 \in F$ implies $F \sim 0$. $\square$

Finally, we can prove the linearity of the cost function.

Lemma 6. *We can write $c(\pi) = \langle \psi, \pi \rangle$, where $\psi : \Delta(\Omega) \to \mathbb{R}$ is given by*

$$\psi(p) = \sup_{h \in \mathcal{H}} \int h \, dp \tag{B.2}$$

where $\mathcal{H}$ is defined as in (B.1).

Proof. For any menu $F \subseteq \mathcal{H}$, we have $\phi_F \leqslant \psi$, so that

$$c(\pi) = \sup_{\substack{F \subseteq \mathcal{H} \\ 0 \in F}} \langle \phi_F, \pi \rangle \leqslant \langle \psi, \pi \rangle.$$

To show the converse inequality, fix $\epsilon > 0$ and $\pi \in \Pi(\overline{p})$ with support $p_1, \ldots, p_n$. From the definition of ψ, we can find $h_1, \ldots, h_n$ such that

$$\psi(p_i) < \langle h_i, p_i \rangle + \epsilon.$$

Letting $F = \{0, f_1, \ldots, f_n\}$ we have

$$c(\pi) \geqslant \langle \phi_F, \pi \rangle \geqslant \sum_i \langle h_i, p_i \rangle \pi_i > \langle \psi, \pi \rangle - \epsilon.$$

Since ϵ was chosen arbitrarily, this shows that $c(\pi) \geqslant \langle \psi, \pi \rangle$, which completes the proof. $\square$

B.3.1 Properties of ψ

The remainder of the proof consists of showing properties of ψ that will imply that it must be given by the entropy reduction, that is,

$$\psi(p) = \int_0^1 \log \frac{dp}{d\overline{p}} dp,$$

This section shows some of the more immediate properties of ψ.

Lemma 7. *The following properties hold:*

1. *ψ is convex*
2. *ψ is lower semi-continuous in the topology of $\Delta(\Omega)$, that is, if $\int \xi dp_\alpha \to \int \xi dp$ for every $\xi \in B(\Omega)$, then $\liminf \psi(p_\alpha) \geqslant \psi(p)$.*
3. *When restricted to densities, ψ is lower semi-continuous in the L^1 norm.*

Proof. (1) follows from the fact that $\psi$is the supremum of linear functions. To prove (2), fix $h \in \mathcal{H}$. Since every utility act is in $B(\Omega)$, we have $\int h\, dp_\alpha \to \int h\, dp$. Now, for any $\epsilon > 0$, there exists $h \in \mathcal{H}$ such that $\psi(p) < \int h\, dp + \epsilon$. Therefore,

$$\liminf \psi(p_\alpha) \geqslant \liminf \int h\, dp_\alpha = \int h\, dp > \psi(p) - \epsilon.$$

Since this holds for all $\epsilon > 0$, we have (2). The proof of (3) is analogous. Just note that since every act h is bounded, $\int h\, dp$ is a continuous function of p in the L^1 norm. $\square$

B.4 Convergence in $\mathcal{H}$

Lemma 8. *Let $(E_i)_{i\in\mathbb{N}}$ be a sequence of events of $[0,1]$ such that $\overline{p}(E_i) \to 0$. Then there exists a subsequence $(E_j)_{j\in\mathbb{N}}$, such that, if we define*

$$D_k = \bigcup_{j=k}^{\infty} E_j,$$

we still have $\overline{p}(D_k) \to 0$.

Proof. Just take E_j such that $\overline{p}(E_j)$ converges to zero fast enough so that $\sum_j \overline{p}(E_j)$ converges. $\square$

Now let $\|f\|_{\overline{p}} = \int |f|\ d\overline{p}$.

Lemma 9. *Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of measurable functions such that $\|f_n - f\|_{\overline{p}} \to 0$. Then there exists a subsequence f_k and a sequence of sets D_k such that $D_1 \supseteq D_2 \supseteq \ldots$ satisfying $\overline{p}(D_k) < 1/k$ and*

$$\sup_{\omega \notin D_k} |f_k(\omega) - f(\omega)| < \frac{1}{k}.$$

Proof. Let $E_n(t) = \{\omega : |f_n(\omega) - f(\omega)| > t\}$. Then

$$\|f_n - f\|_{\overline{p}} \geqslant t\overline{p}(E_n(t))$$

and

$$\sup_{\omega \notin E_n(t)} |f_n(\omega) - f(\omega)| \leqslant t.$$

For each $k \in \mathbb{N}$ take $t = 1/k$ and choose n_k large enough so that

$$\overline{p}\left(E_{n_k}\left(\frac{1}{k}\right)\right) \leqslant k \|f_{n_k} - f\|_{\overline{p}} < \frac{1}{k}.$$

We may then take another subsequence of n_k so that the same inequality holds for $D_k = \bigcup_{j=k}^{\infty} E_j$. $\square$

Proposition 4. *Let $\|f_n - f\|_{\overline{p}} \to 0$, where f_n and f are simple acts. If $f_n \in \mathcal{H}$ for all n and $f_n \geqslant y$ for some $y \in \mathbb{R}$, then $f \in \mathcal{H}$.*

Proof. Suppose that $f \notin \mathcal{H}$, that is, $\{0, f\} \succ 0$. By Lemma 9, we may assume that

$$\sup_{\omega \notin D_k} |f_k(\omega) - f(\omega)| < \frac{1}{k}$$

where $D_1 \supseteq D_2 \supseteq \ldots$ satisfies $\overline{p}(D_k) < 1/k$. Since the set $\bigcap_k D_k$ is null, there exists a k such that $\{0, fD_k y\} \succ 0$. Taking an even larger k, we can guarantee that $\left\{0, fD_k y - \frac{1}{k}\right\} \succ 0$. But since f_k dominates $fD_k y - 1/k$, we must have $\{0, f_k\} \succ 0$, contradicting the assumption that $f_k \in \mathcal{H}$. $\square$

B.5 Representing partitions

This section introduces an alternative way to represent partitions through functions.

Definition 9. An *indexed partition* is a measurable function $\theta : \Omega \to \{1, \ldots, I\}$, where $I \in \mathbb{N}$. θ is an *indexation* of a partition $\mathcal{P}$ if it satisfies $\theta_{\mathcal{P}}^{-1}(i) \in \mathcal{P}$ for $i = 1, \ldots, I$.

When θ is an indexation of $\mathcal{P}$, it assigns an index to each event in $\mathcal{P}$. When we write $\mathcal{P} = \{E_1, \ldots, E_I\}$, we are implicitly assuming a function associating each set in $\mathcal{P}$ to a label in $\{1, \ldots, I\}$, which defines uniquely an indexation of $\mathcal{P}$, given by $\theta^{-1}(i) = E_i$. It should be noted that, given a partition $\mathcal{P}$, there may be multiple functions θ that are indexations of $\mathcal{P}$. In fact, if τ is any permutation of the set $\{1,\ldots,I\}$ and θ is an indexation of $\mathcal{P}$, then $\tau\circ\theta$ is also an indexation of the same partition $\mathcal{P}$. It is easy to see that this exhausts the class of all indexations of $\mathcal{P}$: the indexation is unique *up to a permutation of the image.*

The notion of indexation is useful because, as we will see, we can think of the set $\{1,\ldots,I\}$ as a (finite) state space, and $\theta_{\mathcal{P}}$ provides an "embedding" of the finite state space $\{1,\ldots,I\}$ in Ω. This will allow us to use some results for finite state spaces when we restrict attention to $\mathcal{P}$-measurable menus.

B.5.1 Acts and menus

The notions of measurability naturally translate to indexed partition: If θ is an indexation of $\mathcal{P}$, then f is $\mathcal{P}$-measurable if and only if there exists a unique function $f_I:\{1,\ldots,I\}\to\mathbb{R}$ such that $f=f_I\circ\theta$. The function f_I will be called *the factorization of f through θ.*

Letting $\mathcal{F}_\theta\subseteq\mathcal{F}$ denote the set of acts that have a factorization through θ, we can define an invertible function $T_\theta:\mathcal{F}_\theta\to\mathbb{R}^I$ associating the factorization of f through θ with each act $f\in\mathcal{F}_\theta$. With some abuse of notation we will also denote by T_θ the function that takes a menu $F\subseteq\mathcal{F}_\theta$ to the finite subset of $\mathbb{R}^I$ of the factorizations of F.

The map T_θ associates the restriction of the preference $\succsim$ to the menus contained in $\mathcal{F}_\theta$ with a preference $\succsim_\theta$ over finite subsets of $\mathbb{R}^I$, defined by

$$F\succsim G \iff T_\theta F\succsim_\theta T_\theta G.$$

This defines a utility function V_θ over finite subsets of $\mathbb{R}^I$.

B.5.2 Induced Preferences

Suppose that $\mathcal{Q}$ is a rearrangement of $\mathcal{P}$, that is, there exists a $\sigma:\Omega\to\Omega$ such that, for every $D_j\in\mathcal{Q}$ there exists an $E_i\in\mathcal{P}$ such that $D_j=\sigma^{-1}(E_i)$; in short, $\mathcal{Q}=\sigma^{-1}\circ\mathcal{P}$. Symmetry guarantees that the preference $\succsim$ restricted to $\mathcal{P}$-measurable menus is isomorphic to the restriction to $\mathcal{Q}$-measurable menus: if F and G are $\mathcal{P}$-measurable, then

$$F\succsim G \iff F\circ\sigma\succsim G\circ\sigma,$$

and $F\circ\sigma$ and $G\circ\sigma$ are $\mathcal{Q}$-measurable. The correspondence $F\to F\circ\sigma$ is a bijection between $\mathcal{P}$-measurable and $\mathcal{Q}$-measurable menus.

Given $\theta : \Omega \to \{1, \dots I\}$ an indexation of $\mathcal{P}$, the function $\theta \circ \sigma$ is an indexation for $\mathcal{Q}$. Therefore, the rearrangement σ defines a map from the indexations of $\mathcal{P}$ to the indexations of $\mathcal{Q}$. By the argument above, the preference $\succsim_\theta$ induced by θ is the same as the preference induced by $\theta \circ \sigma$. If θ' is any other indexation of $\mathcal{Q}$, there exists a permutation τ of $\{1, \dots, I\}$such that $\theta' = \tau \circ \theta \circ \sigma$. Therefore, $\succsim_{\theta'}$ is equivalent to $\succsim_{\theta\circ\sigma}$ up to a permutation of $\{1, \dots, I\}$.

B.5.3 Interval partitions

Here we use the particular structure of the interval $[0, 1]$.

Definition 10 (Interval Partition)**.** A partition $\mathcal{P} = \{E_1, \dots, E_I\}$ is an interval partition if its events are intervals in $[0, 1]$.

The following lemma shows why it is without loss of generality (given the assumption of Symmetry) to restrict attention to interval partitions.

Lemma 10. *Let $\mathcal{P}$ be any partition. There exists an interval partition $\mathcal{Q}$ and a rearrangement $\sigma : [0, 1] \to [0, 1]$ such that $\mathcal{Q} = \sigma \circ \mathcal{P}$.*

Proof. Let $\theta_{\mathcal{P}} : [0, 1] \to \{0, \dots, I\}$ be an indexation of $\mathcal{P}$. The decreasing rearrangement $\theta^{\downarrow}_{\mathcal{P}}$ (see Appendix C.1 for the definition) is a decreasing function, so it represents an interval partition $\mathcal{Q}$. By Lemma 2 in Ryff [1965], there exists a rearrangement $\sigma : [0, 1] \to [0, 1]$ such that $\theta_{\mathcal{P}} = \theta^{\downarrow}_{\mathcal{P}} \circ \sigma$, that is, $\mathcal{Q} = \sigma \circ \mathcal{P}$. □

By Lemma 10, to characterize the preference $\succsim$ restricted to $\mathcal{P}$-measurable menus, we may as well look at the preference $\succsim$ restricted to the interval partition $\mathcal{Q} = \sigma^{-1} \circ \mathcal{P}$. In this sense, this restriction is without loss of generality.

B.5.4 Relations between indexations

Suppose that $\mathcal{Q} = \{D_1, \dots, D_J\}$ is finer than $\mathcal{P} = \{E_1, \dots, E_I\}$. In particular, $J \geqslant I$. If $\theta_{\mathcal{Q}}$ and $\theta_{\mathcal{P}}$ are indexations for them, there must exist a function $T : \{1, \dots, J\} \to \{1, \dots, I\}$ such that $\theta_{\mathcal{P}} = T \circ \theta_{\mathcal{Q}}$.

An interesting particular case is when $\mathcal{P}$ and $\mathcal{Q}$ are orthogonal equipartitions, and we consider their join $\mathcal{P} \vee \mathcal{Q}$. In that case, an indexation of $\mathcal{P} \vee \mathcal{Q}$ is given by the ordered pair $(\theta_{\mathcal{P}}, \theta_{\mathcal{Q}}) : \Omega \to I \times J$, where $\theta_{\mathcal{P}}$ is an indexation of $\mathcal{P}$ and $\theta_{\mathcal{Q}}$ is an indexation of $\mathcal{Q}$. The transformation T that solves $\theta_{\mathcal{P}} = T \circ (\theta_{\mathcal{P}}, \theta_{\mathcal{Q}})$ is then just the projection of the first coordinate of $I \times J$. We may refer to it simply as T_I and write $\theta_{\mathcal{P}} = T_I \circ \theta_{\mathcal{P} \vee \mathcal{Q}}$. Likewise, we write $\theta_{\mathcal{Q}} = T_J \circ \theta_{\mathcal{P} \vee \mathcal{Q}}$.

B.6 Symmetry

This section explores some further consequences of the Symmetry Axiom. The results here essentially translate the more general results in Appendix C.2. We start from the observation that the set of rearrangements Σ is a group, which acts linearly on the set of acts, menus, probabilities and channels.

B.6.1 Induced symmetries

We start by stating explicitly the nature of the actions of Σ over various sets, and the relationships between them.

Definition 11. Given a permutation σ and a probability distribution $p \in \Delta(\Omega)$, let $p \circ \sigma \in \Delta(\Omega)$ be defined by $p \circ \sigma(B) = p(\sigma(B))$, for every measurable set $B \subseteq \Omega$. For a channel π, we define $\sigma_* \pi$ by

$$\sigma_* \pi(P) = \pi\left(\sigma^{-1} P\right)$$

for every measurable set $P \subseteq \Delta(\Omega)$.

Under these definitions, the following change-of-variables formulae hold:

$$\int_0^1 f \circ \sigma d(p \circ \sigma) = \int_0^1 f dp \qquad \text{and} \qquad \int_{\Delta(\Omega)} \xi \circ \sigma d\pi = \int_{\Delta(\Omega)} \xi d(\sigma_* \pi).$$

Thus, if f is an act, the expected value of f under p is the same as the expected value of $f \circ \sigma$ under $p \circ \sigma$. Given a menu F, we have

$$\phi_{F \circ \sigma} \circ \sigma(p) \equiv \phi_{F \circ \sigma}(p \circ \sigma) = \max_{f \in F} \int_0^1 f \circ \sigma d(p \circ \sigma) = \max_{f \in F} \int_0^1 f dp = \phi_F(p).$$

In particular,

$$\langle \phi_{F \circ \sigma}, \sigma_* \pi \rangle \quad = \quad \langle \phi_{F \circ \sigma} \circ \sigma, \pi \rangle = \langle \phi_F, \pi \rangle.$$

Definition 12. A function $\xi : \Delta(\Omega) \to \mathbb{R}$ is *symmetric* if $\xi(\sigma_* p) = \xi(p)$ for every rearrangement σ.

B.6.2 Symmetric information cost

Lemma 11. *Under the DDMO axioms, the Linearity axioms, and Symmetry, ψ is symmetric.*

Proof. Let σ be measure-preserving. Note that, by symmetry of $\mathcal{H}$,

$$\psi(\sigma p) = \sup_{h \in \mathcal{H}} \langle h, \sigma p \rangle = \sup_{h \in \mathcal{H}} \langle \sigma h, \sigma p \rangle = \sup_{h \in \mathcal{H}} \langle h, p \rangle = \psi(p).$$

□

The analogous result can also be proven more generally for the cost function c, without using the Linearity axioms.

Corollary 2. *Under the same axioms, if $\pi \in \partial V(F)$, then $\sigma_* \pi \in \partial V(F \circ \sigma)$.*

Proof. Simply write

$$V(F \circ \sigma) = V(F) = \langle \phi_F - \psi, \pi \rangle = \left\langle \phi_{F \circ \sigma} - \psi \circ \sigma^{-1}, \sigma_* \pi \right\rangle = \langle \phi_{F \circ \sigma} - \psi, \sigma_* \pi \rangle.$$

□

B.6.3 Canonical partitions

It is convenient to assume that the prior $\overline{p}$ is the uniform distribution over $[0,1]$ (the Lebesgue measure). In fact, this can be done without loss of generality: by Theorem 3.4.23 in Srivastava [1998], there exists a Borel isomorphism $\tau : [0,1] \to [0,1]$ such that $\lambda(B) = \overline{p}(\tau^{-1}(B))$ for every Borel subset B of $[0,1]$. In what follows, we will assume that $\overline{p}$ is the Lebesgue measure.

We will restrict attention to acts and menus that are measurable with respect to a very particular class of partitions, the *canonical partitions.*

Definition 13. Given a number $I \in \mathbb{N}$,[6] the *canonical I-partition of* $[0,1]$, is the equipartition given by the events $E_i = \left(\frac{i-1}{I}, \frac{i}{I}\right]$ for $i = 2, \ldots, I$ and $E_1 = \left[0, \frac{1}{I}\right]$. An act (or a menu) that is measurable with respect to this partition will be said to be *I-measurable.* A *canonical act* (or menu) is one which is I-measurable for some $I \in \mathbb{N}$.

[6] Without creating confusion, we may let I also denote the set $\{1, \ldots, I\}$.

B.6.4 Rotations

The collection of all rearrangements of $[0,1]$ (with the uniform prior), denoted by Σ, is a group under composition. From it, we can define a group action on the set of acts, menus, measures and channels, by $f \circ \sigma$, $F \circ \sigma$, σp, and $\sigma_* \pi$, respectively. When working with interval partitions, two subgroups of this larger group of transformations will be used: rotations and finite permutations.

Definition 14. A rotation is a transformation $\sigma : [0,1] \to [0,1]$ such that $\sigma(\omega) = \omega + r \pmod 1$ for some $r \in [0,1]$, that is,

$$\sigma(\omega) = \begin{cases} \omega + r & \text{if } \omega + r \leqslant 1 \\ \omega + r - 1 & \text{if } \omega + r > 1 \end{cases}.$$

The group of rotations of $[0,1]$ is a compact topological group.[7] The Lebesgue measure λ over $[0,1]$ is the invariant measure under this group, see Definition 23.

Definition 15. Given a canonical and a canonical equipartition and an $r \in [0,1]$ we may define a "rotation" σ_i that only affects the points in the cell E_i:

$$\sigma_i(\omega) = \begin{cases} \omega & \text{if } \theta(\omega) \neq i \\ \omega + \frac{r}{I} & \text{if } \theta(\omega) = i = \theta\left(\omega + \frac{r}{I}\right) \\ \omega + \frac{r-1}{I} & \text{if } \theta(\omega) = i = \theta\left(\omega + \frac{r-1}{I}\right) \end{cases}.$$

The *group of I-rotations* is then simply the product of the groups of rotations for each cell of the I-partition; it is denoted by R_I. It is also a compact topological group and its Haar measure is the product measure: given sets $R(E_i) \subseteq [0,1]$ of rotations affecting only E_i, the Haar measure of $R(E_1) \times \cdots \times R(E_I)$ is the product of their Lebesgue measures.

This group acts on the set of acts, menus, probability measures and channels. Note that an act is I-measurable if and only if it is invariant with respect to R_I.

[7] This group is known as the *circle group*, and it is usually denoted by $\mathbb{R}/\mathbb{Z}$ or S^1. Its topology is the topology of the circle, which is the same as the topology of $[0,1]$ except at the endpoints, which may be identified.

B.6.5 Symmetrizations in $\mathcal{F}$

Given a function $f \in L^1[0,1]$ we may use the group R_I to define the *I-symmetrization* of f (see definition 25), denoted by $S_I f$. To understand this, first look at the group R_1—the group of rotations of the trivial partition. Then $f \circ r(0) = f(r)$, which means that $\int_{R_1} f \circ r(0)\, dr = \int_0^1 f(r)\, dr$. In fact, for any $\omega \in [0,1]$, we have $\int_{R_1} f \circ r(\omega)\, dr = \int_0^1 f(r)\, dr$: the symmetrization is the expected value of f in $[0,1]$.

For general I, the I-symmetrization of f is the act that gives, in each cell, the expected value of f in that cell: for $\omega \in E_i$,

$$S_I f(\omega) = \int_{R_I} f \circ r(\omega)\ dr = I \int_{R_I} f \circ r(\omega)\ dr = E[f|E_i].$$

Note that $S_I f$ is I-measurable.

Lemma 12. *If $h \in \mathcal{H}$, then $S_I h \in \mathcal{H}$.*

Proof. Since $\mathcal{H}$ is symmetric, we also have $h \circ \sigma \in \mathcal{H}$ for any rearrangement σ. The set

$$C = \{h \circ r :\ r \in R_I\}$$

is bounded in the supnorm, since h is a simple act, and $S_I h$ is the barycenter of the measure over C induced by the Haar measure of R_I. Considering C as a subset of L^1, Corollary 1.2.3 in Winkler [1985] implies that $S_I h$ belongs to the closed convex hull of C. Therefore, we can take a sequence of acts in the convex hull of C converging to $S_I h$, which is an act. Since $C \subseteq \mathcal{H}$ and $\mathcal{H}$ is convex, the convex hull of C is contained in $\mathcal{H}$. Lemma 4 implies that $S_I h \in \mathcal{H}$. □

B.6.6 Symmetrizations in $\Delta(\Omega)$

If a belief $p \in \Delta(\Omega)$ admits a density, the symmetrization of p is simply the measure resulting from the symmetrization of the density.

Definition 16. Say that p is uniform when conditioned on the I-partition if p has a density function which is measurable with respect to the partition.

For all $p \in \Delta(\Omega)$, the symmetrization $S_I p$ of p is uniform when conditioned on the I-partition. This follows from the fact that $S_I p$ must be invariant with respect to R_I.

Lemma 13. *For any $p \in \Delta(\Omega)$, we have $\psi(S_I p) \leqslant \psi(p)$.*

Proof. Given any $h \in \mathcal{H}$, we have $\langle h, S_I p\rangle = \langle S_I h, p\rangle$. By Lemma 12, $S_I \mathcal{H} \subseteq \mathcal{H}$, so that

$$\psi(p) = \sup_{h\in\mathcal{H}} \langle h, p\rangle \geqslant \sup_{h\in\mathcal{H}} \langle S_I h, p\rangle = \sup_{h\in\mathcal{H}} \langle h, S_I p\rangle = \psi(S_I p).$$

□

Remark 1. When p admits a density, this lemma can also follow from the Schur-convexity of ψ, as the density of p majorizes the density of $S_I p$ (see C.1).

Lemma 14. *Let p admit a density and consider the sequence $I_k = 2^k$ for $k \in \mathbb{N}$, then $\psi(S_{I_k} p) \to \psi(p)$ as $k \to \infty$.*

Proof. We may omit the subindex k in the proof. Let $\frac{dp}{d\lambda}$ denote the density of p. For any $\epsilon > 0$, let $h \in \mathcal{H}$ be such that $\psi(p) \le \langle h, p\rangle + \epsilon$. Recall that $S_I p$ is a measure whose density is given by $E\left[\frac{dp}{d\lambda}|\theta_I\right]$, the conditional expectation with respect to the algebra generated by the partition of cardinality I. The sequence $\left(E\left[\frac{dp}{d\lambda}|\theta_I\right]\right)_{I\in\mathbb{N}}$ is a martingale converging to $\frac{dp}{d\lambda}$ in L^1 (the convergence also follows from direct computation). Since h is bounded in the supnorm, $\langle h, S_I p\rangle \to \langle h, p\rangle$ as $I \to \infty$, so that $\psi(p) \le \langle h, S_I p\rangle + 2\epsilon$ for I sufficiently large. By Lemma 13, $\langle h, S_I p\rangle \le \psi(p) \le \langle h, S_I p\rangle + 2\epsilon$ for all I sufficiently large, which gives the result. □

B.6.7 Symmetrizations in $\Pi(\overline{p})$

Though $S_I f$ and $S_I p$ could be defined without using groups, the definition of $S_I \pi$ for $\pi \in \Pi(\overline{p})$ would be more difficult. Indeed, $S_I \pi$ has to be an invariant measure with respect to the action of R_I on π—it must put the same probability on an event P and the action of $\sigma \in R_I$ on that event σP. So while π is discrete, $S_I \pi$ need not be, since there are uncountably many $\sigma \in R_I$. On the other hand, we may define the *uniformization of* π, given by

$$U_I \pi(P) = \pi\left(S_I^{-1} P\right).$$

Notice that $U_I \pi(S_I \Delta) = \pi\left(S_I^{-1} \circ S_I \Delta\right) = 1$, so the support of $U_I \pi$ is contained in the image of S_I. In particular, every $p \in suppU_I\pi$ is uniform conditional on the I-partition. Moreover, if π is discrete so is $U_I \pi$.

Intuitively, when facing an I-measurable menu, the decision-maker should not have an incentive to acquire information about the relative likelihood of subevents of an event

E_i. In other words, the posteriors should be uniform when conditioned to E_i. It is perhaps surprising then that the Blackwell monotonicity of the information cost function is not enough to guarantee this. The next lemma shows that the intuition is correct when Symmetry holds.

Lemma 15. *Let F be I-measurable. Then there exists a $\pi \in \partial V(F)$ such that every $p \in supp(\pi)$ is uniform conditional on the I-partition.*

Proof. The cost function $c(\pi) = \langle \psi, \pi \rangle$ is Blackwell monotone. Since ψ is integrable we can extend c to all countably additive measures . By Proposition 7, we have that $S_I \pi \trianglerighteq U_I \pi$, so Blackwell-monotonicity implies that $c(\pi) = c(S_I \pi) \geqslant c(U_I \pi)$. On the other hand,

$$\langle \psi, S_I \pi \rangle = \int_{R_I} \langle \psi, r_* \pi \rangle \, dr = \int_{R_I} \langle \psi, r_* \pi \rangle \, dr = \int_{R_I} \langle \psi \circ r, \pi \rangle \, dr = \langle \psi, \pi \rangle \, .$$

This shows that $\langle \psi, \pi \rangle = \langle \psi, U_I \pi \rangle$.

Now we can show that $U_I \pi$ achieves the same utility as π. Indeed, since F is I-measurable, we have $S_I \varphi_F = \varphi_F$. Therefore,

$$\langle \varphi_F, \pi \rangle = \langle S_I \varphi_F, \pi \rangle = \langle \varphi_F, U_I \pi \rangle \, .$$

□

B.7 Moving to finite dimension

By the result of lemma 15, for a I-measurable menu F, we may restrict the set $\partial V(F)$ to only those π such that $U_I \pi = \pi$; doing so is without loss of generality, in the sense that the value in the maximization in the representation remains the same. Therefore, if we restrict attention only to I-measurable menus we may transform the problem into one with a finite state space. More precisely, we may define a bijection between the objects in the representation with the state space $[0,1]$ that are I-measurable and objects in a finite state space I:

1. To each I-measurable act $f : [0,1] \to \mathbb{R}$ we may associate its *I-factorization* $f_I : I \to \mathbb{R}$, defined by $f_I(i) = f(\omega)$ for $\omega \in E_i$;

2. To each menu F we associate the menu F_I of its I-factorizations;

3. To each posterior uniform within each E_i, we may associate a $p_I \in \Delta(I)$;

4. The prior $\overline{p}$ will be associated with the uniform distribution on $\Delta(I)$;

5. To each $\pi \in \Pi(\overline{p})$, we may associate a $\pi_I \in \Pi_I \subseteq \Delta(\Delta(I))$, with expectation given by the uniform $\overline{p}$;

6. We may define a new utility function V_I defined over finite subsets of $\mathbb{R}^I$, by $V_I(F_I) = V(F)$;

7. Given $p_I \in \Delta(I)$, let $p \in \Delta(\Omega)$ be uniform when conditioned on the I-partition satisfy $p_I = T_I p$. We define $\psi_I : \Delta(I) \to \mathbb{R}$ by $\psi_I(p_I) = \psi(p)$.

Moreover, these finite-dimensional problems are related. Since each act f which is I-measurable is also $I \times J$-measurable, we may consider f_I as an act $f_I : I \times J \to \mathbb{R}$.

Let $T_I : \Omega \to I$ be the embedding of the canonical I-partition and let

$$\mathcal{H}_I = \{h_I : I \to \mathbb{R} : \ V_I(\{0, h_I\}) = 0\}.$$

Lemma 16. $\mathcal{H}_I = T_I(S_I \mathcal{H})$

Proof. Let $h_I \in \mathcal{H}_I$. Then there exists an I-measurable act $h : \Omega \to \mathbb{R}$ such that $h_I = T_I h$. Now

$$V(\{0, h\}) = V_I(T_I\{0, h\}) = V_I(\{0, h_I\}) = 0,$$

so that $h \in \mathcal{H}$. On the other hand, if $h \in \mathcal{H}$ is I-measurable, then $S_I h = h$ (since S_I is idempotent), so that $S_I \mathcal{H} \subseteq \mathcal{H}$ is the set I-measurable acts in $\mathcal{H}$. □

Therefore, $\mathcal{H}_I$ is the set of I-factorizations of the I-measurable acts $h \in \mathcal{H}$. It is easy to see that, if p is uniform when conditioned on the I-partition, $p = S_I p$ and

$$\psi(p) = \sup_{h \in \mathcal{H}} \langle h, S_I p \rangle = \sup_{h \in \mathcal{H}} \langle S_I h, S_I p \rangle = \sup_{h_I \in \mathcal{H}_I} \sum_i h_I(i) \, p_I(i) = \psi_I(p_I)$$

Lemma 17. *$\mathcal{H}_I$ is a closed, convex subset of $\mathbb{R}^I$.*

Proof. Convexity follows from convexity of $\mathcal{H}$. Since $\mathcal{H}_I$ contains all the points dominated by any given point, it must have a non-empty interior. Now take any two utility acts h and h'. The set

$$\{\alpha : \ 0 \succsim \{0, \alpha h + (1-\alpha) h'\}\} = \{\alpha : \ \alpha h + (1-\alpha) h' \in \mathcal{H}_\theta\}$$

is closed by Mixture Continuity. If we take any point h in the boundary of $\mathcal{H}_I$ and h' in its interior, we will have that $\alpha h + (1-\alpha) h'$ belongs to $\mathcal{H}_I$ for all $\alpha \in [0,1)$; this implies that it also holds for $\alpha = 1$, so $h \in \mathcal{H}_I$. This proves that $\mathcal{H}_I$ is closed. □

B.7.1 Subgroups of permutations

From now on, we will consider the problem transformed to a finite state space. The set of permutations of states correspond to the set of all rearrangements preserving the partition; all the groups that we will consider will then be finite subgroups of this one.

Definition 17. Let Σ be a subgroup of the permutations of I. An I-menu F is Σ-symmetric if

$$f \in F \implies f \circ \sigma \in F$$

for all $\sigma \in \Sigma$. It is *symmetric* if this is satisfied for the group of all permutations $\mathcal{S}$.

In other words, F is symmetric if it contains all permutations of its acts. Given any I-menu $H \subseteq \mathcal{H}_I$ and permutation σ, we have that $H \circ \sigma \subseteq \mathcal{H}_I$ as well and thus, by IIA, so does the menu $M(H;\Sigma)$ defined by

$$M(H;\Sigma) = \bigcup_{\sigma \in \Sigma} H \circ \sigma \subseteq \mathcal{H}.$$

Definition 18. Σ is transitive if for any $i, i' \in I$, there exists a $\sigma \in \Sigma$ such that $\sigma(i) = i'$.

Lemma 18. *If Σ is transitive, the symmetrization $S(f;\Sigma)$ of any $f : I \to \mathbb{R}$ is given by*

$$S(f;\Sigma)(i) = \frac{1}{I} \sum_{i' \in I} f(i').$$

Definition 19. Given a function $f : I \to \mathbb{R}$, the *orbit of f under* Σ is the set

$$Orb_\Sigma(f) = \{f \circ \sigma : \sigma \in \Sigma\}.$$

Given a channel π, define $S(\pi;\Sigma)$ to be the symmetrization of π with respect to the group Σ, that is,

$$S(\pi;\Sigma)(p) = \frac{1}{|Orb_\Sigma(p)|} \pi(Orb(p)) = \frac{1}{|\Sigma|} \sum_{\sigma \in \Sigma} \sigma_* \pi. \tag{B.3}$$

That means that $S(\pi;\Sigma)$ puts the same probability on each permutation in Σ of any given p, while still putting the same probability on the whole orbit as π. When $\pi = \delta_p$ for a $p \in \Delta(\Omega)$, we write directly $S(p;\Sigma)$ instead of $S(\delta_p;\Sigma)$.

Lemma 19. *Let $H \subseteq \mathcal{H}$ be I-measurable and such that $0 \in H$. If $\pi \in \partial V(H)$, then $\pi^s \in \partial V(H^s)$ and $\langle \psi, \pi \rangle = \langle \psi, \pi^s \rangle$.*

Proof. We have, for any $\sigma \in \Sigma_\theta$,

$$V(H^s) = 0 = V(H) = \langle \phi_H - \psi, \pi \rangle = \langle \phi_{H\circ\sigma} - \psi, \sigma_* \pi \rangle \leqslant \langle \phi_{H^s} - \psi, \sigma_* \pi \rangle,$$

which implies that $\sigma_* \pi \in \partial V(H^s)$. Since $\partial V(H^s)$ is convex, we have $\pi^s \in \partial V(H^s)$.

That π and π^s have the same cost follows from the second equality in equation (B.3). $\square$

Given a I-measurable $p \in \Delta$, define the channel π_p to be the uniform distribution over $Orb(p)$. Since

$$\left(\frac{1}{|\Sigma_I|}\sum_{\sigma\in\Sigma_I}\sigma p\right)(i) = \frac{1}{|I|}\sum_i p(i) = \frac{1}{|I|},$$

π_p is a well-defined channel.

Proposition 5. *Let $h \in \mathcal{H}$ be I-measurable and satisfy*

1. *If $h' \in \mathcal{H}$ dominates h, we must have $h = h'$;*
2. *$M(h, \Sigma_I) \cup 0 \sim 0$*

Then there exists a $p \in \Delta$ such that $\pi_p \in \partial V(M(h,\Sigma_I))$ and for all $q \in \Delta$,

$$V(M(h,\Sigma_I)) = \langle h, p \rangle - \psi(p) \geqslant \max_{q\in\Delta} \langle h, q \rangle - \psi(q) \tag{B.4}$$

Proof. For convenience let $H = M(h,\Sigma_I)$. Take any $\pi \in \partial V(H)$ and take $p \in supp(\pi)$. We have

$$\begin{aligned} V(H) &= \langle \varphi_H - \psi, \pi \rangle \\ &= \langle \varphi_H - \psi, S(\pi, \Sigma_I) \rangle. \end{aligned}$$

But we can represent $S(\pi,\Sigma_I)$ by the formula

$$S(\pi,\Sigma_I) = \sum \pi(Orb(p))\, S(p,\Sigma_I),$$

where the sum ranges over a set containing a single representative from each orbit. Since the cost is linear, we must have

$$V(H) = \sum \pi(Orb(p)) \langle \varphi_H - \psi, S(p, \Sigma_I) \rangle,$$

but since $S(p, \Sigma_I)$ is a channel, we must have $V(H) \geqslant \langle \varphi_H - \psi, S(p, \Sigma_I) \rangle$ for each representative p, so that in fact all must hold with equality. In other words, for each $p \in supp\pi$, $S(p, \Sigma_I) \in \partial V(H)$. Choosing the representative p that maximizes $\langle h, p \rangle$, we also have

$$V(H) = \langle h, p \rangle - \psi(p) \geqslant \max_{q \in \Delta} \langle h, q \rangle - \psi(q)$$

$\square$

B.7.2 Product subgroups

We now consider the finite state space $I \times J$. One subgroup of permutations, denoted Σ_I, is given by the permutations that do not affect J, that is $\sigma \in \mathcal{S}_{I \times J}$ such that the J-coordinate of $\sigma(i, j)$ is j. We will denote by $\Sigma_I \times \Sigma_J$ the subgroup generated by Σ_I and Σ_J. This subgroup is transitive, see Definition 18.

B.8 Separability

Let $\theta : \Omega \to I \times J$ be an interval equipartition; we can write θ as resulting from two partitions: $\theta_I : \Omega \to I$ and $\theta_J : \Omega \to J$, so that $\theta = (\theta_I, \theta_J)$.

Definition 20. Let p_I, q_J factor $p, q \in \Delta$ through θ_I and θ_J, respectively. The belief $p \times q$ is the belief factoring through θ, with the factorization given by

$$p \times q(i, j) = p_I(i) q_J(j).$$

The purpose of this section is to prove the following lemma:

Proposition 6. *Let p be $I \times J$-measurrable and let p_I be the marginal of p on I and p_J be the marginal of p on J. Then ψ satisfies the following properties:*

1. (Subbaditivity) $\psi_{I \times J}(p) \geqslant \psi_I(p_I) + \psi_J(p_J)$;
2. (Additivity) $\psi_{I \times J}(p_I \times p_J) = \psi_I(p_I) + \psi_J(p_J)$.

B.8.1 An inequality

The first step is to prove the following inequality:

Lemma 20. *Let $h_I \in \mathcal{H}_I$ and $h_J \in \mathcal{H}_J$. If we let p_I and p_J be the solutions of equation (B.4) for h_I and h_J, respectively, then*

$$\psi_{I\times J}\left(p_I \times p_J\right) \geqslant \psi_I\left(p_I\right) + \psi_J\left(p_J\right).$$

Proof. We have

$$\begin{aligned} V\left(h_I^s + h_J^s\right) &\geqslant \left\langle h_I + h_J, p_I \times p_J\right\rangle - \psi\left(p_I \times p_J\right) \\ &= \left\langle h_I, p_I \times p_J\right\rangle + \left\langle h_J, p_I \times p_J\right\rangle - \psi\left(p_I \times p_J\right) \\ &= \left\langle h_I, p_I\right\rangle + \left\langle h_J, p_J\right\rangle - \psi\left(p_I \times p_J\right) \end{aligned}$$

The symmetrizations h_I^s (with respect to θ_I) and h_J^s (with respect to θ_J) satisfy the conditions of Separability, so

$$V\left(h_I^s + h_J^s\right) = V\left(h_I^s\right) + V\left(h_J^s\right).$$

But by our choice of p_I and p_J, we have

$$V\left(h_I^s\right) + V\left(h_J^s\right) = \left\langle h_I, p_I\right\rangle - \psi\left(p_I\right) + \left\langle h_J, p_J\right\rangle - \psi\left(p_J\right).$$

Combining these three, we obtain the result. □

B.8.2 Superadditivity

Lemma 21. *Let $g \in \mathcal{H}$ be $\mathcal{P}$-measurable and $h \in \mathcal{H}$ be $\mathcal{Q}$-measurable, where $\mathcal{P}$ and $\mathcal{Q}$ are orthogonal equipartitions. Then $g + h \in \mathcal{H}$.*

Proof. We know that $G = \{0, g\} \sim 0$ and $H = \{0, h\} \sim 0$ are orthogonal menus. By Lemma 3,

$$G + H \sim 0 + H = H \sim 0.$$

Since

$$0 \precsim \{0, g+h\} \sim \{0, g, h, g+h\} = G + H \sim 0,$$

we must have that $\{0, g+h\} \sim 0$, which proves the result. □

Let $g, h \in \mathcal{H}$. By Lemma 12, we know that $S_{I\times J}g$ and $S_{I\times J}h$ are in $\mathcal{H}$, as well as $S_I g$ and $S_J h$. Now let $G = M(S_I g, \Sigma_I)$ and $H = M(S_I h, \Sigma_J)$.

$$\begin{aligned}
\psi_{I\times J}(p) &\geqslant \sup_{g,h\in\mathcal{H}_{I\times J}} \langle S_I g + S_J h, p\rangle \\
&= \sup_{g\in\mathcal{H}_{I\times J}} \langle S_I g, p\rangle + \sup_{h\in\mathcal{H}_{I\times J}} \langle S_J h, p\rangle \\
&= \sup_{g\in\mathcal{H}} \langle S_I g, S_I p\rangle + \sup_{h\in\mathcal{H}} \langle S_J h, S_I p\rangle \\
&= \psi_I(p_I) + \psi_J(p_J)
\end{aligned}$$

B.8.3 Irrelevance of Orthogonal Flexibility

Let F factor through θ_I and G and H factor through θ_J. Recall that G^I factors through θ, with the factorization given by the acts

$$G^I \cong \{h : I \times J \to \mathbb{R} : \ h(i,j) = g_i(i,j), \ g_i \in G\}.$$

Now, since $G \subseteq G^I$ and $H \subseteq H^I$, we have that

$$F + G + H \subseteq F + G + H^I \subseteq F + G^I + H^I \sim F + G + H,$$

which implies indifference of all these menus, by preference for flexibility (which is implied by dominance). Therefore, we have

$$V\left(F + G + H^I\right) = V(F + G + H)$$

$$\max_{\rho\in\partial V(G)} \langle \phi_H, \rho\rangle = \max_{\pi\in\partial V(F+G)} \langle \phi_H, \pi\rangle = \max_{\pi\in\partial V(F+G)} \langle \phi_{H^I}, \pi\rangle$$

Now suppose $G = g^s$ has a unique optimal channel, which has to be the uniform distribution over the J-permutations of some p_J. The equation above shows that, for every menu H such that

B.8.4 Additivity

Let $f \in \mathcal{H}_I$ and suppose that there exists a unique solution $p \in \Delta(I)$ for the problem (B.4). Likewise, let $q \in \Delta(J)$ be the unique solution for the act $g \in \mathcal{H}_J$. Letting $F = S(f; \Sigma_I)$

and $G = S(g; \Sigma_J)$, we see that $\{\pi(p; \Sigma_I)\} = \partial V(F)$ and $\{\pi(q; \Sigma_J)\} = \partial V(G)$.

Now extend F and G to the product partition indexed by $I \times J$, so that F and G are measurable with respect to orthogonal partitions. Then $F + G$ is $\Sigma_I \times \Sigma_J$-symmetric and so is $\partial V(F + G)$. Therefore, given any r in the support of some $\rho \in \partial V(F + G)$, we must also have $\pi(r; \Sigma_I \times \Sigma_J) \in \partial V(F + G)$. We will now show that $\pi(p \times q; \Sigma_I \times \Sigma_J)$ is optimal for $F + G$.

To see this, suppose first that $\partial V(F + G)$ is a singleton $\rho = \pi(r; \Sigma_I \times \Sigma_J)$. Letting $H = S(h; \Sigma_J)$ be J-measurable, we have that $\langle \phi_H, \rho \rangle = \langle \phi_{H^I}, \rho \rangle$ for every menu H. By Lemma 29, this can only be the case if $r = p \times q$.

B.9 Entropic Cost

Aczel-Forte-Ng prove the following result:

Theorem 3. *[Aczel-Forte-Ng] If ψ_I satisfies Symmetry, Additivity and Superadditivity, then there exists a constant $a > 0$ and a function $A : \mathbb{N} \to \mathbb{R}$ such that*

$$\psi_I(p) = -aH_I(p) + A(I). \tag{B.5}$$

Proof. Follows by applying Lemma 5 in Aczél et al. [1974]to $-\psi$. □

Corollary 3. *If ψ satisfies Symmetry, Additivity, Superadditivity and*

$$\psi_I(\overline{p}_I) = \psi_I\left(\frac{1}{I}, \ldots, \frac{1}{I}\right) = 0$$

then there exists a constant $a > 0$ such that

$$\psi_I(p) = a\left(H_I(\overline{p}_I) - H_I(p_I)\right).$$

Proof. Evaluating equation (B.5) at the uniform prior, we obtain $A(I) = aH_I(\overline{p}_I)$. □

For p uniform with respect to the I-partition, we have that $\psi(p) = \psi_I(p_I)$ and $H(p) = H_I(p_I)$, so the results prove that the cost function is given by

$$c(\pi) = \langle \psi, \pi \rangle = a \int (H(\overline{p}) - H(p))\, \pi(dp),$$

where $a > 0$ is a constant. Since multiplication by a constant does not affect the representation, we may assume without loss of generality that $a = 1$. This concludes the proof of Theorem 1. By Lemma 14, ψ is given by this formula for any p that admits a density.

C Mathematical Appendix

C.1 Majorization

Given any function $f \in L^1[0,1]$, we may define a right-continuous and nonincreasing function $m : \mathbb{R} \to \mathbb{R}$ by

$$m(t) = \lambda \{\omega : f(\omega) > t\}.$$

This function can be inverted, using the formula

$$f_{\downarrow}(\omega) = \sup_{m(t) > \omega} t.$$

The decreasing function $f_{\downarrow}$ is called the *decreasing rearrangement* of f. This name is justified by the following lemma (for a proof, see Lemma 2 Ryff [1965])

Lemma 22. *To each $f \in L^1[0,1]$ there corresponds a measure-preserving $\sigma : [0,1] \to [0,1]$ such that $f = f_{\downarrow} \circ \sigma$.*

Definition 21. Let $f, g \in L^1[0,1]$. We say that f *majorizes* g, denoted $f \gg g$, if, for all $\overline{\omega} \in [0,1]$, we have

$$\int_0^{\overline{\omega}} f_{\downarrow}(\omega)\, d\omega \geqslant \int_0^{\overline{\omega}} g_{\downarrow}(\omega)\, d\omega$$

with equality for $\overline{\omega} = 1$.

The following is a standard result in the theory of majorization (see Marshall et al. [2011])

Lemma 23. *If $\xi : L^1[0,1] \to \mathbb{R}$ is convex and symmetric, then ξ is* Schur-convex*, that is,*

$$f \gg g \implies \xi(f) \geqslant \xi(g).$$

C.2 Topological groups

Definition 22. A *group* is a set G where we define a function $g \cdot h$ of two arguments $g, h \in G$, taking values in G and satisfying the following properties:

1. $g \cdot (h \cdot i) = (g \cdot h) \cdot i$

2. There exists an element $e \in G$ such that $e \cdot g = g \cdot e = g$

3. For every $g \in G$ there exists an element g^{-1} such that $g \cdot g^{-1} = g^{-1} \cdot g = e$.

A *topological group* is a group G together with a topology on G such that the functions $h \to g \cdot h$ (for $g \in G$) and $h \to h^{-1}$ are continuous functions from G to itself.

For the remainder of this subsection G will denote a topological group.

C.2.1 Invariant measures

Definition 23. A measure μ over the Borel subsets of G is said to be *left-invariant* if $\mu(gE) = \mu(E)$ for all $g \in G$ and E a Borel subset of G. Equivalently, for every $f : G \to \mathbb{R}$ integrable,

$$\int_G f(g)\,\mu(dg) = \int_G f(hg)\,\mu(dg).$$

μ is *right-invariant* if $\mu(Eg) = \mu(E)$ and it is *invariant* if it is both left and right-invariant.

The existence of an invariant probability measure is guaranteed if the topological group is compact Weil [1965], in which case it is unique. When G is a finite group, the uniform distribution is the invariant measure.

C.2.2 Group actions

Definition 24. Let X be a set and G a group. A *group action* is a set of functions $f : X \to X$, that is a group under composition. We say that G acts on X if there's a group action which is isomorphic to G. In this case, we can denote the functions $f(x)$ by gx, where $g \in G$.

If X is a vector space, we say that G *acts linearly on* X if

$$g(\alpha x + \beta y) = \alpha g x + \beta g y$$

for all $x, y \in X$, $\alpha, \beta \in \mathbb{R}$ and $g \in G$.

Example 1. Let X be a set and let G act on X. Then

1. We can define a linear action of G over the set of functions $f : X \to \mathbb{R}$, defined by $gf(x) = f(gx)$

2. We can define an action of G over subsets of X, defined by $gA = \{gx : x \in A\}$.

3. If $\langle X, X' \rangle$ is a dual pair and G acts linearly on X, we can define an action over X' by imposing, for all $x \in X$,
$$\langle x, x' \rangle = \langle gx, gx' \rangle .$$
In particular, if X is the space of bounded Borel-measurable functions over a set, this defines an action over the set of countably additive measures over that set, characterized by $g\mu(gB) = \mu(B)$ for all Borel sets B.

4. If $Y \subseteq X$ is such that $gY \subseteq Y$ for all $g \in G$, then G acts on Y by simply taking the restriction of the group actions to Y.

C.2.3 Symmetrization

Here, suppose that G is a compact group that acts linearly on the vector space X. Integrals over G are always going to be using the Haar measure of G.

Definition 25. Let $x \in X$. The *G-symmetrization of x* is defined by
$$S_G x = \int_G gx \, dg,$$
where the integral is the Bochner integral, see Aliprantis and Border [2006].We call the map $S_G : X \to X$ the *G-projection.*

Lemma 24. *For any $g \in G$, we have $gS_G(x) = S_G x$.*

Proof. Fix $x \in X$ and let $L : X \to \mathbb{R}$ be a linear function. Then we can define a function $f : G \to \mathbb{R}$ by $f(g) = L(gx)$, so that
$$\begin{aligned} L[gS_G(x)] &= L\left[\int_G ghxdh\right] \\ &= \int_G f(gh)\, dh \\ &= \int_G f(h)\, dh \\ &= L[S_G x]. \end{aligned}$$
where the third equality follows from the left-invariance of the Haar measure. $\square$

Lemma 25. *S_G is idempotent, that is, $S_G \circ S_G = S_G$.*

Proof. Using the result from the previous lemma, we have, for every $x \in X$,

$$\begin{aligned} S_G^2 x &= \int_G g\left(S_G x\right) dg \\ &= \int_G S_G x \, dg \\ &= S_G x. \end{aligned}$$

$\square$

C.2.4 Uniformization

Beyond the assumptions of the preceding section, suppose also that X has a topology.

Definition 26. Let G act linearly on X and define the action of G on the functions $f : X \to \mathbb{R}$ and the Borel measures $\mu \in ca\left(X\right)$ as in Example 1. The *G-uniformization* of f and μ, denoted $U_G f$ and $U_G \mu$ (abusing notation) are defined by

$$U_G f\left(x\right) = f\left(S_G x\right) \qquad U_G \mu\left(B\right) = \mu\left(S_G^{-1} B\right).$$

Thus, the following formula is satisfied:

$$\left\langle U_G f, \mu \right\rangle = \left\langle f, U_G \mu \right\rangle.$$

Lemma 26. *U_G is a projection.*

Proof. We have

$$U_G^2 f\left(x\right) = U_G f\left(S_G x\right) = f\left(S_G^2 x\right) = f\left(S_G x\right) = U_G f\left(x\right).$$

Likewise,

$$U_G^2 \mu\left(B\right) = \mu\left(\left\{x : \; S_G^2 x \in B\right\}\right) = \mu\left(\left\{x : \; S_G x \in B\right\}\right) = U_G \mu.$$

$\square$

C.2.5 Convex order

Let Φ denote the set of all continuous convex functions $\varphi : X \to \mathbb{R}$. The convex order over measures is defined by

$$\mu \geqslant_{cx} \eta \iff \int \varphi d\mu \geqslant \int \varphi d\eta, \ \forall \varphi \in \Phi.$$

Proposition 7. *Let $\pi \in ca\,(X)$. Then $S_G\pi \geqslant_{cx} U_G\pi$.*

Proof. Let $\varphi : X \to \mathbb{R}$ be convex. By Jensen's inequality,

$$U_G\varphi\,(x) = \varphi\left(\int_G gx\,dg\right) \leqslant \int_G \varphi\,(gx)\,dg = S_G\varphi\,(x)\,.$$

Now

$$\begin{aligned} \langle \varphi, S_G\pi \rangle &= \langle S_G\varphi, \pi \rangle \\ &\geqslant \langle U_G\varphi, \pi \rangle \\ &= \langle \varphi, U_G\pi \rangle\,. \end{aligned}$$

□

C.3 Other projections

Here I will consider a finite state-space of the form $I \times J$. M denotes the group of permutations over I alone. Define $T_M : \mathbb{R}^{I\times J}_{++} \to \mathbb{R}^{I\times J}_{++}$ by

$$(T_M x)\,(i,j) = \frac{1}{|I|}\sum_k x\,(k,j)\,,$$

that is, T_M is the symmetrization with respect to M. I also define

$$(T_k x)\,(i,j) = x\,(k,j) \qquad T_\Delta x = \frac{x}{\|x\|_1}.$$

Lemma 27. *For all $x \in \mathbb{R}^{I\times J}_{++}$, we have*

$$T_M x = \frac{1}{|I|}\sum_{k=1}^{I} T_k x$$

Proof. We have

$$\left(\sum_{k=1}^{I} T_k x\right)(i,j) \quad = \tfrac{1}{|I|} \quad \sum_{k=1}^{I} x(k,j) = (T_M x)(i,j)\,.$$

□

Lemma 28. *Let $x_k \in X$ and $\alpha_K \in \mathbb{R}_{++}$ and let $x = \sum_k \alpha_k x_k$. If for every convex, H1 function ϕ,*

$$\sum_k \alpha_k \phi(x_k) = \phi(x)\,,$$

then $T_\Delta x_1 = T_\Delta x_2 = \ldots = T_\Delta x$.

Proof. Suppose not; without loss of generality $T_\Delta x_1 \neq T_\Delta x$, that is

$$x_1 \notin R_x = \{tx : \ t \in \mathbb{R}_{++}\}\,.$$

Since R_x is convex, there exists a separating hyperplane y such that

$$\langle x_1, y\rangle > \alpha \geqslant \langle tx, y\rangle$$

for every $t \in \mathbb{R}_{++}$. Thus, we must have that $\langle x, y\rangle \leqslant 0$ and we may take $\alpha = 0$. Now let

$$\phi(x') = \max\{\langle x', y\rangle, 0\}\,.$$

Then ϕ is convex and H1, $\phi(x) = 0$ and

$$\sum_k \alpha_k \phi(x_k) \geqslant \alpha_1 \phi(x_1) > 0.$$

□

Lemma 29. *Let π be a positive discrete measure over $\mathbb{R}_{++}^{I\times J}$. Suppose that for every convex, H1 function $\phi : \mathbb{R}_{++}^{I\times J} \to \mathbb{R}_{++}^{I\times J}$, we have*

$$\int \left[\frac{1}{|I|}\sum_k \phi(T_k x) - \phi(T_M x)\right] \pi(dx) = 0.$$

Then for every x in the support of π, we have $T_\Delta T_k x = T_M x$.

Proof. For all $x \in X$, we have

$$T_M x = \frac{1}{|I|} \sum_{k=1}^{I} T_k x.$$

Since ϕ is convex, this means that the integrand is always nonnegative. Since π is a positive discrete measure, this means that, for every x in the support of π, we must have

$$\frac{1}{|I|} \sum_{k} \phi\left(T_k x\right) = \phi\left(T_M x\right).$$

By Lemma 28, we have $T_\Delta T_k x = T_\Delta T_M x = T_M x$. $\square$

References

János Aczél, B Forte, and CT Ng. Why the shannon and hartley entropies are'natural'. *Advances in Applied Probability*, pages 131–146, 1974.

C.D. Aliprantis and K.C. Border. *Infinite dimensional analysis: a hitchhiker's guide*. Springer Verlag, 2006.

F.J. Anscombe and R.J. Aumann. A definition of subjective probability. *Annals of mathematical statistics*, pages 199–205, 1963.

D. Blackwell. Comparison of experiments. In *Second Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 93–102, 1951.

D. Blackwell. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, pages 265–272, 1953.

Antonio Cabrales, Olivier Gossner, and Roberto Serrano. Entropy and the value of information for investors. *The American economic review*, 103(1):360–377, 2013.

Andrew Caplin and Mark Dean. Rational inattention, entropy, and choice: The posterior based approach. Technical report, Mimeo, Center for Experimental Social Science, New York University, 2013.

Henrique De Oliveira, Tommaso Denti, Maximilian Mihm, and M Kemal Ozbek. Rationally inattentive preferences. *Available at SSRN*, 2013.

E. Dekel, B.L. Lipman, and A. Rustichini. Representing preferences with a unique subjective state space. *Econometrica*, 69(4):891–934, 2001.

H. Ergin and T. Sarver. The unique minimal dual representation of a convex function. *Journal of Mathematical Analysis and Applications*, 370(2):600–606, 2010.

Bartosz Mackowiak and Mirko Wiederholt. Information processing and limited liability. *The American Economic Review*, 102(3):30–34, 2012.

Albert W Marshall, Ingram Olkin, and Barry Arnold. *Inequalities*. Springer, 2011.

Daniel Martin. Strategic pricing and rational inattention to quality. *Memeo, New York University*, 2013.

F. Matejka and A. McKay. Rational inattention to discrete choices: a new foundation for the multinomial logit model. 2013.

Filip Matejka and Alisdair McKay. Simple market equilibria with rationally inattentive consumers. *The American Economic Review*, 102(3):24–29, 2012.

Luigi Paciello and Mirko Wiederholt. Exogenous information, endogenous information and optimal monetary policy. 2011.

Andrzej Ruszczyński and Alexander Shapiro. Optimization of convex risk functions. *Mathematics of operations research*, 31(3):433–452, 2006.

John V Ryff. Orbits of L^1-functions under doubly stochastic transformation. *Transactions of the American Mathematical Society*, 117:92–100, 1965.

Christopher A. Sims. Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690, 2003.

Sashi Mohan Srivastava. *A course on Borel sets*, volume 180. Springer, 1998.

André Weil. *L'intégration dans les groupes topologiques et ses applications*, volume 1145. Hermann Paris, 1965.

Gerhard Winkler. *Choquet order and simplices: with applications in probabilistic models*, volume 1145. Springer-Verlag Berlin, New York, 1985.

M. Yang. Coordination with flexible information acquisition. 2011.