FOUNDATIONS FOR OPTIMAL INATTENTION

ANDREW ELLIS
(JOB MARKET PAPER)

Abstract. This paper models an agent who has a limited capacity to pay attention to information and thus conditions her actions on a coarsening of the available information. An optimally inattentive agent chooses both her coarsening and her actions by constrained maximization of an underlying subjective expected utility preference relation. The main result axiomatically characterizes the conditional choices of actions by an agent that are necessary and sufficient for her behavior to be seen *as if* it is the result of optimal inattention. Observing these choices permits unique identification of the agent's utility index, cognitive constraint and prior (the last under a suitable richness condition). An application considers a market in which strategic firms offer differentiated products. If the consumer's information concerns firms' quality, then equilibrium consumer surplus may be higher with an optimally inattentive consumer than with one who processes all available information.

1. Introduction

1.1. **Objectives and Outline.** Individuals often appear not to process all available information. This phenomenon, documented in both psychology and economics,[1] is usually attributed to agents' limited capacity to pay attention to information (Sims [2003]). When the available information exceeds their capacity, agents exhibit inattention, i.e. condition their choices on coarser information. This inattentiveness has significant economic consequences.[2]

This paper models agents who respond optimally to their limited attention. An *optimally inattentive* agent has a constraint that limits the information to which she can pay attention,

Key words and phrases. Inattention, optimal inattention, conditional choice.
Department of Economics, Boston University. Email: ellisa@bu.edu. October, 2012. I am especially grateful to Larry Epstein, Bart Lipman and Jawwad Noor for their advice, support and guidance. I would also like to thank Laurent Bouton, Sambuddha Ghosh, Sean Horan, Asen Kochov, Albert Ma, Andy Newman, Pietro Ortoleva, Dana Rotz, Todd Sarver, Jeremy Smith, Tomasz Strazlecki and conference participants at RUD 2012, NASM 2012 and NSF/NBER/CEME CGEME 2012 for helpful conversations and comments. Comments welcome.

[1]See Pashler [1998] for a book-length treatment of attention in psychology. Economists have also empirically documented that agents, e.g. restaurant patrons (Luca [2011]), stock traders (DellaVigna and Pollet [2009]) and professional forecasters (Coibion and Gorodnichenko [2011a,b]), fail to process all available information.

[2]For instance, it can imply delayed response to shocks (Sims [1998, 2003]), sticky prices (Mackowiak and Wiederholt [2009]), under-diversification (Van Nieuweburgh and Veldkamp [2010]), sticky investment (Woodford [2008]), coordination failure (Hellwig and Veldkamp [2009]), specialization (Dow [1991]), exploitation (Rubinstein [1993]) and extreme price swings (Gul et al. [2011]).

and she chooses both her coarsening and her actions (or acts) conditional on that coarsening by maximizing a subjective expected utility preference relation. I axiomatically characterize the conditional choices of actions (acts) by a decision maker (DM) that are necessary and sufficient for her behavior to be seen *as if* it is the result of optimal inattention. These axioms clarify the model's implications for choice behavior and provide a choice-theoretic justification for it.

The modeler observes an objective state space and a partition describing the objective information. In contrast, the DM's tastes, her prior beliefs, her capacity for attention, and the information to which she pays attention (which I call her *subjective information* to distinguish it from the finer, objective information) are taken to be *unobservables* that must be inferred from choices. I assume a rich set of choice data, namely the DM's choices from each feasible set of acts and conditional on each state of the world.[3]

The rationale for assuming that the indicated range of behavior is observable is easily understood. First, with a narrower range of behavior, the model cannot be characterized. Choice out of a single feasible set cannot reveal much about underlying behavior for the reasons familiar from standard choice theory. Furthermore when the state space is unobservable and choice is observed conditional on a single state, earlier work by Van Zandt [1996] shows that optimal inattention has no testable implications.[4,5] Second, my setting allows analysis directly in terms of the economic object of interest – namely, the agent's chosen action, such as setting a price, selecting a bundle of goods, or deciding from which firm to purchase. Finally, this range of behavior permits unique identification of the unobservables, even though the DM's choices violate many of the well-understood properties that permit identification in other models, including the Weak Axiom of Revealed Preference (WARP).

The remainder of the paper proceeds as follows. In the next subsection, I use an example to illustrate my setting, the behavior of interest and how I achieve identification. Section 2 presents the model in detail. In Section 3, I formally describe the behavior of interest through five axioms. Theorems 1 and 2 show that these axioms characterize an optimally

[3]This data is an extension of that considered by the papers cited in Footnote 2, which study all conditional choices from a single feasible set. It is easily obtainable in a laboratory environment, and it could, in principle, be gathered from a real-world setting where both the realized state and the information received by the agent are independently and identically distributed across time.

[4]Van Zandt [1996] studies hidden information acquisition, which is readily reinterpreted as optimal inattention. Specifically, he takes as given any choice correspondence on a finite collection of alternatives. He shows that one can construct a state space and an information acquisition problem so that for every choice problem, the alternative selected in a fixed state matches the choice correspondence if the DM chooses information optimally.

[5]The model's implications when the state space and choice conditional on only one state are observable an open question. A partial answer is given by the axioms Monotonicity and ACI (below): conditional choices in a fixed state satisfy WARP when the problem contains only state-independent acts. However, they violate WARP (as well as many weaker properties implied by it) in general, and although identification of the utility index is possible, the attention constraint cannot be identified.

inattentive DM's choices. I also characterize two special cases: a DM who processes all information and a DM who processes the same information, regardless of the menu faced. In Section 4, Theorem 3 shows that the utility index, the attention constraint and, in many circumstances, the prior are uniquely identified by the agent's conditional choices.

In Section 5, Theorem 4 gives an intuitive, behavioral comparison equivalent to one optimally inattentive DM having a higher capacity for attention than another. I then argue that an optimally inattentive DM values information differently than a Bayesian DM because the former may not process all available information. Even if one information partition is *objectively* more valuable than another (in the sense of Blackwell [1953]), it may not be *subjectively* more valuable. That is, because of information overload a DM may reject an objectively more valuable information partition in favor of a coarser one. After generalizing my setting to allow the objective information to vary, Theorem 5 characterizes the subjective value of information to an optimally inattentive DM in terms of her choices.

In Section 6, I analyze a market where firms compete over optimally inattentive consumers. Intuition suggests that firms can exploit these consumers, and previous work (cf. Rubinstein [1993]) focuses on that aspect. Fixing prices, consumer surplus increases as capacity for attention increases. However, if strategic effects are taken into account, then *lower* consumer capacity for attention may lead to *higher* equilibrium consumer surplus. In fact, firms may *benefit* from facing *more attentive* consumers. The key difference from earlier work is that consumers perceive the price perfectly, but they are inattentive to information about the quality of the products. Intuitively, if consumers allocate their attention optimally, then a firm attracts attention only if it offers the consumer more surplus. This induces competition among firms who would not otherwise compete, which lowers prices and increases consumer surplus.

Section 7 concludes by discussing the relationship with other models of inattention. Proofs are collected in appendices.

1.2. **Example.** Consider a benevolent doctor who treats patients suffering from a given disease. Glaxo, Merck, and Pfizer all produce pharmaceuticals that treat the disease, but the doctor knows that one of the three drugs will be strictly more effective than the other two. The one that works best for each patient is initially unknown, and the doctor can, in principle, determine it; for instance, by constructing a very detailed medical history. Uncertainty is modeled by the state space $\Omega = \{\gamma, \mu, \phi\}$, and the objective information by the partition

$$P = \{\{\gamma\}, \{\mu\}, \{\phi\}\}.$$

The state indicates whether the most effective drug is produced by Glaxo (γ), by Merck (μ) or by Pfizer (ϕ), and P indicates that the doctor can determine which state obtains.

TABLE 1. Conditional choices

	γ	μ	ϕ
$c(\{g,m,f\}\vert\cdot)$	$\{m\}$	$\{m\}$	$\{f\}$
$c(\{g,m\}\vert\cdot)$	$\{g\}$	$\{m\}$	$\{m\}$

Suppose there are two patients who are identical except that they have different insurance plans: one's covers all three drugs, and the other's does not cover Pfizer's drug. Each patient is a choice problem, in which prescribing a drug corresponds to choosing an act (g, m and f represent prescribing the drugs produced by Glaxo, Merck and Pfizer respectively). The drug prescribed to each patient conditional on each state of the world is given by a *conditional choice correspondence*, a family of choice correspondences indexed by the state of the world. Table 1 lists the conditional choices of a doctor when facing $\{g, m, f\}$ (the problem associated with unrestricted insurance) and $\{g, m\}$ (the problem associated with restricted insurance).

Under the assumption that the doctor's choices result from optimal inattention, what can be inferred from them? One complication is that the doctor's choices violate WARP in state γ: she chooses g but not m from $\{g, m\}$ and chooses m but not g from $\{g, m, f\}$. Although WARP violations prevent identification of preference through the usual methods, the doctor's tastes, subjective information, and attention constraint can nevertheless be inferred from her choices. To begin with, the above choices reveal that the doctor cannot pay attention to all the objective information. If she did, then her choice when Glaxo's drug is effective from the first menu would reveal that she strictly prefers to prescribe it rather than to prescribe Merck's drug. Therefore, if Glaxo's drug is available in the larger problem and it is the most effective, then she should not prescribe Merck's. But because she chooses to prescribe the latter when facing $\{g, m, f\}$ in state γ, she does not pay attention to the objective information.

Since the doctor does not pay attention to the objective information, I turn to inferring her subjective information, i.e. the information to which she does pay attention. When facing $\{g, m\}$, she chooses differently conditional on γ than she does conditional on either μ or ϕ, so her subjective information must be at least as fine as $\{\{\gamma\}, \{\mu, \phi\}\}$. Moreover, it cannot be strictly finer because it would then be the objective information. Consequently, her subjective information is exactly $\{\{\gamma\}, \{\mu, \phi\}\}$ when facing $\{g, m\}$. Similarly, her subjective information must be $\{\{\phi\}, \{\gamma, \mu\}\}$ when facing $\{g, m, f\}$. Therefore, the doctor chooses as if she knows the answer to the question "Is Glaxo's drug the most effective?" when facing $\{g, m\}$ and "Is Pfizer's drug the most effective?" when facing $\{g, m, f\}$. With her subjective information known, her choices reveal her conditional preferences, which can then be aggregated to reveal her underlying unconditional preferences.

How can one tell if the doctor's choices have an optimal inattention representation? Theorems 1 and 2 show that a set of properties characterizes a doctor whose choices can be seen as if they result from optimal inattention. The doctor's choices do not violate any of these properties, so they are compatible with optimal inattention. However, consider a second doctor who chooses according to $c'(\cdot)$, where $c'(\cdot)$ is the same same as $c(\cdot)$ except $c'(\{g, m, f\}|\phi) = \{m\}$. Although both doctors select the same prescription from each choice problem in state γ, when the other conditional choices are considered, $c'(\cdot)$ cannot have an optimal inattention representation. To see this, note that the second doctor chooses Merck's drug when the patient has good insurance, regardless of the state of the world. As above, the modeler infers that the doctor knows whether Glaxo's drug is the most effective when facing $\{g, m\}$, so her choice of g in state γ reveals that she strictly prefers prescribing it to prescribing Merck's in that state. However, this implies that her choices from the smaller problem yield a better outcome in every state of the world than those from the larger problem, an impossibility if her subjective information is optimal when facing both problems.

2. Setup and Model

2.1. **Setup.** I adopt the following version of the classic Anscombe-Aumann setting. Uncertainty is captured by a set of states Ω and a set of events Σ, a $\sigma-$algebra of subsets of Ω. Consequences are elements of a separable metric space, Z. Let the set X consist of all finite-support probability measures on Z, endowed with the weak* topology. Objects of choice are acts, $\Sigma-$measurable simple (finite-ranged) functions $f : \Omega \to X$. Let $\mathcal{F}$ be the set of all acts, endowed with the topology of uniform convergence. Since $\mathcal{F}$ is metrizable, let $d(\cdot)$ be a compatible metric.[6]

The DM must choose an act from a compact set, i.e. her choice problem is a non-empty, compact subset of $\mathcal{F}$. Let $K(\mathcal{F})$ be the set of all choice problems, endowed with the Hausdorff topology generated by the metric $d(\cdot)$. She has access to objective information, represented by P, a finite partition of Ω.[7] I require every element of the partition be an element of Σ. Knowing the objective information allows the modeler to distinguish imperfect information from limited capacity for attention.[8]

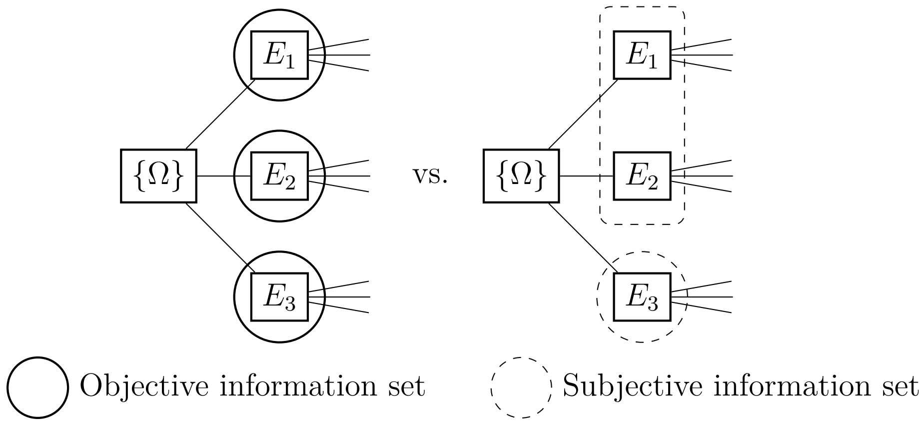
Choice results from a three-stage process. In stage 1, the state is realized but remains unknown to the DM. In stage 2, the DM chooses an act. Although in principle she observes the realized cell of P before making this choice, she instead acts as if she observes the realized cell of her subjective information. In stage 3, all uncertainty is resolved and the DM gets the

[6]X is metrizable by Theorem 15.12 of Aliprantis and Border [2006]. Let $\hat{d} : X \times X \to \mathbb{R}_+$ be a compatible metric. Since $\mathcal{F}$ has the topology of uniform convergence, $f_n \to f \iff \sup_\omega \hat{d}(f_n(\omega), f(\omega)) \to 0$. Therefore $d : \mathcal{F} \times \mathcal{F} \to \mathbb{R}_+$ given by $d(f, g) = \sup_\omega \hat{d}(f(\omega), g(\omega))$ is a compatible metric on $\mathcal{F}$; since $\mathcal{F}$ contains only simple acts, the supremum is attained and $d(f, g) < \infty$ for any $f, g \in \mathcal{F}$.

[7]With minor additional assumptions, P can be taken to be countable rather than finite.

[8]Because the modeler knows the objectively available information, inattention can be distinguished from imperfect information. For instance, if the modeler observes that the DM never distinguishes between states ω_1 and ω_2, then this is interpreted as inattention only if the objective information distinguishes ω_1 from ω_2.

Figure 1. Timeline



consequence specified by her chosen act and the realized state. The tree on the left in Figure 1 illustrates the timing. If the DM exhibits inattention, she acts as if facing a different tree than the objective one; for instance, the one on the right in Figure 1.

The modeler observes the DM's choices in stage 2 (or later) and the realization of the objective information, but does not observe the DM's subjective information. The choice data generates a conditional choice correspondence $c(\cdot)$, such that the DM is willing to choose the acts in $c(B|\omega)$ from the problem B when the state is ω. Formally, this is a set-valued, P-measurable function $c : K(\mathcal{F}) \times \Omega \to K(\mathcal{F})$ with $c(B|\omega) \subset B$ for all $B \in K(\mathcal{F})$ and all $\omega \in \Omega$.

I adopt the following notation throughout. Identify X with the subset of acts that do not depend on the state, i.e. $x \in X$ corresponds to the act $x \in \mathcal{F}$ such that $x(\omega) = x\ \forall \omega \in \Omega$, and let $K(X)$ be the set of compact, non-empty subsets of X, noting that $K(X) \subset K(\mathcal{F})$. For any partitions Q and Q', write $Q \gg Q'$ if Q is finer than Q'. For any acts $f, g \in \mathcal{F}$ and any event $E \in \Sigma$, define fEg to be the act that yields $f(\omega)$ if $\omega \in E$ and $g(\omega)$ if $\omega \notin E$. For any $\alpha \in [0, 1]$ and any two $f, g \in \mathcal{F}$ let $\alpha f + (1-\alpha)g \in \mathcal{F}$ be the state-wise mixture of f and g, or the act taking the value in state ω of $\alpha f(\omega) + (1-\alpha)g(\omega)$, defined by the usual mixture operation on lotteries. For any $A, B \in K(\mathcal{F})$ and $\alpha \in [0, 1]$, let $\alpha A + (1-\alpha)B \in K(\mathcal{F})$ be $\{\alpha a + (1-\alpha)b | a \in A, b \in B\}$. Any act in $\alpha A + (1-\alpha)B$ is an α mixture of an act in A with a $(1-\alpha)$ mixture of an act in B.

2.2. **Model.** *An optimally inattentive agent* is a tuple $(u(\cdot), \pi(\cdot), \mathbb{P}^*, \hat{P}(\cdot))$, where:

- $u : X \to \mathbb{R}$ is continuous and affine,
- $\pi : \Sigma \to [0, 1]$ is finitely-additive and $\pi(E) > 0$ for every $E \in P$,
- $\mathbb{P}^* \subset \{Q : P \gg Q\}$ has the property that if $Q \in \mathbb{P}^*$ and $Q \gg Q'$, then $Q' \in \mathbb{P}^*$, and
- $\hat{P} : K(\mathcal{F}) \to \mathbb{P}^*$.

The *utility index* $u(\cdot)$ and *prior* $\pi(\cdot)$ have familiar interpretations. Neither varies with the problem, so an optimally inattentive DM has stable tastes and beliefs. I focus on interpreting the two new objects, the *attention constraint* $\mathbb{P}^*$ and the *attention rule* $\hat{P}(\cdot)$. The former describes to what the DM *can* pay attention, while the latter describes to what she *does* pay attention. The attention constraint $\mathbb{P}^*$ is a set of partitions, all of which are coarser than the objective information. If $Q \in \mathbb{P}^*$, then the DM has the capacity to pay attention to Q. I assume that $\mathbb{P}^*$ satisfies free disposal of information, in the sense that whenever she can pay attention Q, she can also pay attention to any Q' that is coarser than Q. Depending on the problem, the DM may have different subjective information, given by the attention rule $\hat{P}(\cdot)$. That is, $\hat{P}(B)$ is her subjective information when facing the problem B.

Definition 1. A conditional choice correspondence $c(\cdot)$ has an *optimal inattention representation* if there exists an optimally inattentive agent so that for every problem B,

$$\hat{P}(B) \in \arg\max_{Q\in\mathbb{P}^*} \sum_{E\in Q} \pi(E)[\max_{f\in B} \int u \circ f d\pi(\cdot|E)], \tag{1}$$

and for every problem B and state ω,

$$c(B|\omega) = \arg\max_{f\in B} \int u \circ f d\pi(\cdot|\hat{P}(B)(\omega)). \tag{2}$$

The DM's choices have an optimal inattention representation if they satisfy two properties. Equation (1) requires that her subjective information gives at least as high expected utility as any other partition in $\mathbb{P}^*$, i.e. it is chosen optimally. Equation (2) requires that the DM's choice from B in state ω maximizes expected utility conditional on the realized cell of her subjective information.

An optimally inattentive DM considers all available acts. In contrast, Masatlioglu et al. [2012] studies an agent who does not pay attention to the entire set of available actions. Although both models are motivated by the same underlying mechanism, neither nests the other: there are choices compatible with optimal inattention but not inattention to alternatives and vice versa.[9] While DMs conforming to either model may violate WARP, the reason for such violations is different.[10] In fact, an optimally inattentive DM may exhibit inattention yet satisfy WARP (Corollary 2).

One special case of optimal inattention is a DM who processes all available information. I call such a DM Bayesian, and say that $c(\cdot)$ *has a Bayesian representation* if

$$c(B|\omega) = \arg\max_{f\in B} \int u \circ f d\pi(\cdot|P(\omega))$$

[9]See Appendix C.6 and C.7.

[10]In Masatlioglu et al. [2012], removing unchosen alternatives may affect the options considered by the DM; in this paper, removing alternatives not chosen in a given state may alter the information processed.

for every B and ω. In the model, this corresponds to an optimally inattentive agent with $\mathbb{P}^* = \{Q : Q \ll P\}$ and $\hat{P}(B) = P$ for every problem B.

Another special case is a DM who always pay attention to the same information (which may differ from the objective information), regardless of the problem faced. I say that such a DM has fixed attention, and that $c(\cdot)$ has *fixed attention representation* if there is a partition Q satisfying $P \gg Q$ so that

$$c(B|\omega) = \arg\max_{f\in B} \int u \circ f d\pi(\cdot|Q(\omega))$$

for every B and ω. In the model, this corresponds to an optimally inattentive agent with $\mathbb{P}^* = \{Q' : Q' \ll Q\}$ and $\hat{P}(B) = Q$ for every problem B.

In addition to the above special cases, the model admits many others considered in the literature. For instance, $\mathbb{P}^*$ could equal the set of all partitions that are both coarser than P and have at most $\kappa \geq 1$ elements (Gul et al. [2011]). Alternatively, $\mathbb{P}^*$ could equal the set of all partitions that have mutual information with respect to P less than κ (similar to Sims [2003]).[11]

3. Foundations

3.1. **Axioms.** I impose the following axioms. The quantifier "for all $f, g \in \mathcal{F}$, $A, B \in K(\mathcal{F})$, $\omega \in \Omega$ and $\alpha \in (0, 1]$" is suppressed throughout.

A DM satisfies *WARP*, sometimes referred to as Independence of Irrelevant Acts, if for any $A \subset B$, whenever $c(B|\omega) \cap A \neq \emptyset$ it follows that $c(A|\omega) = c(B|\omega) \cap A$. If an inattentive DM's choices from problems A and B are conditioned on the same subjective information, then her choice in each state maximizes the same conditional preference relation, so these choices do not violate WARP. Therefore, if she violates it, then her choices from A and B must be conditioned on different subjective information. The first axiom, *Independence of Never Relevant Acts* or *INRA*, gives one situation where the DM should not violate WARP.

Axiom 1. *(INRA) If $A \subset B$ and $c(B|\omega') \cap A \neq \emptyset$ for every state ω', then*

$$c(A|\omega) = c(B|\omega) \cap A.$$

Within the context of Section 1.2, INRA says that if two patients differ only in that one's plan drops the drug h but the doctor *never* prescribes h to the patient with better insurance, then she prescribes the same drug to both patients. To interpret the axiom, consider a problem B and a "never relevant" act f (i.e. $f \notin c(B|\omega')$ for all ω'), and let $A = B\backslash\{f\}$.[12] Suppose that her choices from B are conditioned on the subjective information Q. Because

[11]The mutual information is a measure of the information provided about the realization of one random variable by another. It corresponds to the reduction in entropy and is used by the rational inattention literature.

[12]Whenever B is finite, INRA is equivalent to "if $c(B|\omega') \neq \{f\}$ for all ω', then $c(B|\omega)\backslash\{f\} = c(B\backslash\{f\}|\omega)$."

she never chooses f from B, the benefit of paying attention to Q when facing A is the same as it is when facing B. If Q is *optimal* when facing B, then Q is *still optimal* when facing A. Therefore, the DM should have the same subjective information when facing B as when facing A, so her choices from A and B should not violate WARP. More generally, the statement $c(B|\omega') \cap A \neq \emptyset$ for every state ω' implies that the entire set of acts that are in B but not in A is "never relevant" and removing them would not decrease the benefit of her subjective information when facing B. As above, if her subjective information is optimal when facing B, then it is still optimal when facing A. Consequently, the DM's choices should not violate WARP.[13]

In the present context, a DM satisfies *Independence* if

$$f \in c(A|\omega) \, and \, g \in c(B|\omega) \iff \alpha f + (1-\alpha)g \in c(\alpha A + (1-\alpha)B|\omega).$$

That is, if the DM chooses f over each h in A and g over each h' in B, then she chooses $\alpha f + (1-\alpha)g$ over each $\alpha h + (1-\alpha)h'$ in $\alpha A + (1-\alpha)B$.[14] If an optimally inattentive DM pays attention to the same information when facing the problems A, B and $\alpha A + (1-\alpha)B$, then her choice in each state maximizes the same conditional preference relation. Because her conditional preferences are expected utility, her choices do not violate Independence. This implies that whenever the DM violates this property for A, B and $\alpha A + (1-\alpha)B$, she must not pay attention to the same information when facing all three problems. The second axiom, *Attention Constrained Independence* or *ACI*, gives one situation where the DM should not violate Independence.

Axiom 2. *(ACI) $f \in c(B|\omega)$ if and only if $\alpha f + (1-\alpha)g \in c(\alpha B + (1-\alpha)\{g\}|\omega)$.*

In my example, this says that if there is a state-independent chance that the patient will take some drug h regardless of what the doctor actually prescribes, then her choice of prescription is unaffected by both the identity of h and the magnitude of that chance. To interpret the axiom, fix problems B and $\{g\}$. Because $\{g\}$ is a singleton, the DM makes the same choice from it no matter what her subjective information is. Therefore, the relationship between the benefits of any two subjective information partitions is the same for the problem B as it is for the problem $\alpha B + (1-\alpha)\{g\}$.[15] If paying attention to Q is *optimal* when facing

[13]INRA can be illustrated by the choices in the introduction. Let $A = \{g, m\}$ and $B = \{g, m, f\}$. The doctor does not violate the axiom because she chooses only f when facing $\{g, m, f\}$ in state ϕ, i.e. $c(B|\phi) \cap A = \emptyset$. The second doctor, whose choices are represented by $c'(\cdot)$ and who cannot be represented as optimal inattention, violates the axiom because she never chooses f when facing $\{g, m, f\}$, regardless of the state of the world, i.e. $m \in A$ and $c(B|\omega) = \{m\}$ for any $\omega \in \{\gamma, \mu, \phi\}$ but $c(A|\gamma) \neq c(B|\gamma) \cap A$.

[14]This follows from the standard formulation of Independence for a binary relation: $f \succeq g \iff \alpha f + (1-\alpha)h \succeq \alpha g + (1-\alpha)h$.

[15]One can think of $\alpha B + (1-\alpha)\{g\}$ as flipping a (possibly biased) coin, choosing from B if the coin comes up heads and otherwise choosing from $\{g\}$, where the DM must choose her subjective information before observing the outcome of the coin-flip. Since information only has value if the coin comes up heads, a partition is optimal when facing $\alpha B + (1-\alpha)\{g\}$ if and only if it would be optimal when facing B for sure.

B, then paying attention to Q is *also optimal* when facing $\alpha B + (1-\alpha)\{g\}$. Consequently, an optimally inattentive DM conditions her choices on the same subjective information when facing $\alpha B+(1-\alpha)\{g\}$ as she does when facing B. Because her conditional preferences satisfy Independence, she chooses the mixture of her choices from B with g from $\alpha B + (1-\alpha)\{g\}$.

The next axiom adapts the standard Monotonicity axiom to the present setting. It also implies that tastes are state independent. For any lotteries x and y, say that x is *revealed* (resp. *strictly*) *preferred* to y if $x \in c(\{x, y\}|\omega)$ (resp. and $y \notin c(\{x, y\}|\omega)$) for some ω.[16]

Axiom 3. *(Monotonicity) (i) If $A \in K(X)$, then $c(A|\omega) = c(A|\omega')$ for all $\omega' \in \Omega$.*
(ii) If $f, g \in B$ and $f(\omega')$ is revealed preferred to $g(\omega')$ for every $\omega' \in \Omega$, then $g \in c(B|\omega) \implies f \in c(B|\omega)$. Moreover, if $f(\omega')$ is revealed strictly preferred to $g(\omega')$ for each $\omega' \in P(\omega)$, then $g \notin c(B|\omega)$.

In my example, this says that the doctor cares only about the realized consequence of her choice, and if one drug gives a better consequence in every state than another, then she never prescribes the inferior drug. For interpretation, consider $B = \{x, y\}$ where x and y are lotteries. If the DM's tastes are state independent and she chooses x over y in state ω, then she also chooses x over y in state ω'. This reveals that the DM prefers x to y, i.e. considers x to be a better consequence than y. Now, consider acts f and g so that f yields a better consequence than g in every state of the world. Even if the DM received information revealing that the state on which g gives the best consequence would occur for sure, she would still be willing to choose f over g. Consequently, she never chooses only g when f is available. In addition, if f yields a strictly better consequence than g in every state in $P(\omega)$, then the DM does not choose g. Thus, Monotonicity limits the scope of inattention; an inattentive DM will never pick a dominated act.

Another common property satisfied by most models of choice under uncertainty is *Consequentialism:* if $f, g \in B$ and $f(\omega') = g(\omega')$ for all $\omega' \in P(\omega)$, then

$$f \in c(B|\omega) \iff g \in c(B|\omega).$$

A DM who satisfies Consequentialism respects the objective information, in the sense that whenever two acts give the same outcome on every objectively possible state, then one of the acts is chosen if and only if the other is. A DM whose subjective information differs from the objective information will violate this property. The next axiom, Subjective Consequentialism, weakens Consequentialism to take this into account.

Axiom 4. *(Subjective Consequentialism) If $f, g \in B$ and $\forall\omega'[f(\omega') \neq g(\omega') \implies c(B|\omega') \neq c(B|\omega)]$, then $f \in c(B|\omega) \iff g \in c(B|\omega)$.*

[16] Recall that $K(X)$ is the set of problems that contain only lotteries.

Subjective Consequentialism implies that choice between any two acts is unaffected by their outcomes in states that the DM knows did not occur. To see this, fix B, f, and g as above, and suppose that the DM faces the problem B and that the realized state is ω. Whenever ω and ω' are in the same cell of her subjective information when facing B, the DM's choices in those states maximize the same conditional preference relation, so $c(B|\omega) = c(B|\omega')$. Consequently, if $c(B|\omega') \neq c(B|\omega)$, i.e. the DM makes different choices in states ω and ω', then these two states must be in different cells of her subjective information. By hypothesis, if f and g give a different consequence in state ω', then $c(B|\omega') \neq c(B|\omega)$, so the DM must know that ω' did not occur. Therefore, the DM knows that she receives the same consequenence from choosing either f or g, so she chooses f if and only if she chooses g.[17]

My final axiom is a technical condition ensuring the continuity of the underlying preference relation. Complicating its statement is that the DM's choices from different menus may be conditioned on different information. Consequently, her choices may appear discontinuous to the modeler, and the axiom must take into account that the underlying preference is revealed by choices that are not conditioned on the same subjective information. To state the axiom, I need two preliminary definitions. First, say that f *dominates* g if f is chosen from $\{f, g\}$ in every state of the world. If the DM has optimal inattention and f dominates g, then f must be (weakly) preferred to g conditional on every cell of every feasible subjective information partition. The second definition is:

Definition 2. *The acts in A are indirectly selected over the acts in B*, written $A\ IS\ B$, if there are problems $B_1, ..., B_n \in K(\mathcal{F})$ so that $B_1 = A$ and $B_n = B$ and for each $i \in \{1, ..., n-1\}$ and every ω, $c(B_{i+1}|\omega) \cap B_i \neq \emptyset$.

Suppose that the DM faces B and chooses an act in A regardless of the state of the world. Since her choices from B are available in A, her set of choices from A is selected over her choices from B. Moreover, if she chooses an act from B in every state of the world when facing C, then her set of choices from B is selected over any choices in C. Since the acts in A are selected over the acts in B that are in turn selected over the acts in C, the acts in A are indirectly selected over the acts in C. Write $A\ \overline{IS}\ B$ if there are sequences $(A_n)_{n=1}^{\infty}$ and $(B_n)_{n=1}^{\infty}$ so that $A_n \to A$, $B_n \to B$ and $A_n\ IS\ B_n$ for all n, i.e. $\overline{IS}$ is the sequential closure of IS.

The final axiom, Continuity, requires that each $c(\cdot|\omega)$ satisfies a weak continuity condition and that sequences of indirect selections do not contradict domination.

[17]Consequentialism implies Subjective Consequentialism: since $c(B|\cdot)$ is P-measurable, $P(\omega) \subset \{\omega'' : c(B|\omega'') = c(B|\omega)\}$ for every B and ω.

Axiom 5. *(Continuity) For any* $\{B_n\}_{n=1}^{\infty} \subset K(\mathcal{F})$*:*
(i) If $f_n \in c(B_n|\omega)$ *and*

$$\{\omega' : c(B|\omega') = c(B|\omega)\} = \{\omega' : c(B_n|\omega') = c(B_n|\omega)\}$$

for every $n \in \mathbb{N}$*, then* $B_n \to B$ *and* $f_n \to f$ *imply that* $f \in c(B|\omega)$*.*
(ii) If $\{f\}\ \overline{IS}\ \{g\}$ *and* g *dominates* f*, then* f *dominates* g *as well.*

The first condition of Continuity is a restriction of upper hemi-continuity.[18] It requires that this property holds only if the DM reveals that her choice is conditioned on the same information along the sequence and as it is at the limit. Both parts of Continuity are implied by combining WARP and upper hemi-continuity.

To interpret the second condition of the axiom, note that INRA suggests that the DM's set of choices from problem A is better than her set of choices from problem B whenever she chooses an act in A when facing B conditional on every state of the world. This direct ranking is incomplete but can be extended using finite sequences of choices to allow for indirect comparisons as well. These indirect comparisons are captured when the acts in A are indirectly selected over the acts in B. Because this ranks many more sets of choices, indirect selections are important for characterizing optimal inattention. Continuity insures a minimal consistency between these indirect comparisons and her direct comparisons. Specifically, suppose that f dominates g and g does not dominate f. Continuity implies that if B is sufficiently close to $\{g\}$ and A is sufficiently close to $\{f\}$, then the DM does not indirectly select the acts in B over the acts in A.

3.2. **Characterization Result.** I can now state the main result: if the DM's choices satisfy the five axioms above, then she acts as if she has optimal inattention.

Theorem 1. *If* $c(\cdot)$ *satisfies INRA, ACI, Monotonicity, Subjective Consequentialism and Continuity, then* $c(\cdot)$ *has an optimal inattention representation.*

Theorem 1 shows that the above axioms are sufficient for the DM to have optimal inattention. For a discussion of the key ideas in its proof, see Appendix A. Necessity is more complicated because I have not restricted attention to well-behaved tie-breaking rules. Consider two conditional choice correspondences, $c(\cdot)$ and $c'(\cdot)$, that both have optimal inattention representations with the same prior, utility index, and attention constraint. When their attention constraint is not a singleton, it is possible that the former has subjective information Q when facing the problem B, while the latter has subjective information Q' when facing B. This arises when a problem has multiple optimal subjective information partitions (i.e. the right hand side of (1) is not a singleton) and the DM must break ties between them. If

[18]This follows from Aliprantis and Border [2006, Cor 17.17].

she breaks these ties non-systematically, then the DM may violate ACI or INRA.[19] Though the axioms become necessary if I impose some conditions on tie-breaking when defining the model, Theorem 2 shows that the set of problems for which an optimally inattentive DM fails to satisfy either INRA or ACI is non-generic even without any such conditions.[20]

Theorem 2. *If $c(\cdot)$ has an optimal inattention representation, then $c(\cdot)$ satisfies Monotonicity, Subjective Consequentialism and Continuity. Moreover, there is a conditional choice correspondence $c'(\cdot)$ satisfying INRA, ACI, Monotonicity, Subjective Consequentialism and Continuity as well as an open, dense $K \subset K(\mathcal{F})$ so that*

(i) $c(\cdot)$ and $c'(\cdot)$ have optimal inattention representations parametrized by $(u(\cdot), \pi(\cdot), \hat{P}(\cdot), \mathbb{P}^*)$ and $(u(\cdot), \pi(\cdot), \hat{Q}(\cdot), \mathbb{P}^*)$, respectively, and

(ii) $c(B|\omega) = c'(B|\omega)$ for every $\omega \in \Omega$ and $B \in K$.

Theorem 2 implies that INRA and ACI are generically necessary. This is because the set of problems for which ties can occur is "small." Consequently, INRA and ACI capture the economic content of optimal inattention. Though not strictly necessary, for any given prior, utility index and attention constraint, there are always attention rules that satisfy INRA and ACI.

3.3. **Special cases.** To understand the role of the axioms in the characterization, I characterize the two special cases of optimal inattention mentioned at the start of this section, the Bayesian model and the fixed attention model.

Corollary 1. *$c(\cdot)$ satisfies Consequentialism in addition to INRA, ACI, Monotonicity, and Continuity if and only if $c(\cdot)$ has a Bayesian representation.*

Intuitively, Consequentialism requires that the DM respects the objective information structure. For an optimally inattentive DM, this implies that she processes all information and chooses the act that maximizes expected utility. Since Consequentialism implies Subjective Consequentialism, $c(\cdot)$ has an optimal inattention representation and must have a Bayesian representation.

[19] A similar issue exists for random expected utility (Gul and Pesendorfer [2006]) with a finite state space. If ties are broken using a "regular" random expected utility function, then choices satisfy linearity, but if ties are broken differently, then linearity may fail.

[20] Say that an optimal inattention representation is *regular* if for any $A, B \in K(\mathcal{F})$ and $g \in \mathcal{F}$, $\hat{P}(B) = \hat{P}(\alpha B + (1-\alpha)\{g\})$ and $\arg\max_{f \in A} \mathbb{E}_\pi[u \circ f | \hat{P}(A)(\omega)] = \arg\max_{f \in A} \mathbb{E}_\pi[u \circ f | \hat{P}(B)(\omega)]$ for all ω whenever $c(B|\omega) \cap A \neq \emptyset$ for all ω. Given Theorems 1 and 2, one can verify that $c(\cdot)$ has regular optimal inattention if and only if $c(\cdot)$ satisfies all six axioms.

Corollary 2. *The following are equivalent:*

(i) $c(\cdot)$ *satisfies Independence in addition to INRA, Monotonicity, Subjective Consequentialism and Continuity.*

(ii) $c(\cdot)$ *satisfies WARP in addition to ACI, Monotonicity, Subjective Consequentialism and Continuity.*

(iii) $c(\cdot)$ *has a fixed attention representation.*

It immediately follows that WARP and Independence are equivalent for an optimally inattentive DM. The intuition behind Corollary 2 is that an optimally inattentive DM's choices from A and B violate Independence or WARP only if her subjective information differs at A, B or $\alpha A+(1-\alpha)B$. If her subjective information never changes, then she never violates either condition. Consequently, she has fixed attention if she satisfies either WARP or Independence.

3.4. **Counter-examples.** To help understand the role of the axioms, I provide a series of counterexamples showing what may go wrong if one or more of the axioms are not satisfied. An alternative model of particular interest is the *inattention* model. An inattentive DM maximizes expected utility conditional on her subjective information, but her subjective information is not necessarily optimal. Although she has stable tastes and beliefs, the information to which she pays attention varies with the problem in a general manner. Formally, $c(\cdot)$ has an *inattention representation* if Equation (2) holds for all problems B and states ω but the source of $\hat{P}(\cdot)$ is left unspecified.

Proposition 1. *If* $c(\cdot)$ *has an inattention representation, then* $c(\cdot)$ *satisfies Monotonicity, Subjective Consequentialism and Continuity (i).*

In particular, an inattentive DM's choices may violate INRA, ACI or Continuity (ii). Consequently, these three axioms reflect the optimality of her subjective information. They capture her reaction to her attention constraint but not that she exhibits inattention in the first place.[21]

ACI reflects that the DM has an attention constraint. Consider the alternative model is *costly attention.* In this case, rather than being subject to a constraint on the information to which she can attend, the DM incurs a cost if she pays attention to a given partition. A function $\rho : \{Q : Q \ll P\} \to [0,\infty]$ is a *cost function* if $\rho(\{\Omega\}) = 0$ and $Q \gg Q'$ implies $\rho(Q) \geq \rho(Q')$. Formally, $c(\cdot)$ has a *costly attention representation* if there is an optimally inattentive agent and a cost function $\rho(\cdot)$ so that

$$\hat{P}(B) \in \arg\max_{Q} \sum_{E\in Q} \pi(E)[\max_{f\in B} \int u \circ f d\pi(\cdot|E)] - \rho(Q) \tag{3}$$

[21]A characterization of the inattention model is available as supplementary material.

for every problem B and Equation (2) holds for every problem B and state ω. Given appropriate tie-breaking, this model satisfies all of the axioms except ACI. In fact, it satisfies the following weaker version of ACI:

$$\alpha f + (1-\alpha)g \in c(\alpha B + (1-\alpha)\{g\}|\omega) \iff \alpha f + (1-\alpha)h \in c(\alpha B + (1-\alpha)\{h\}|\omega)$$

for a fixed $\alpha \in (0,1]$.[22]

INRA reflects that the DM's subjective information is optimal. If her subjective information cannot be represented as maximizing behavior, then the DM's choices violate INRA. For instance, suppose that Equation (1) holds when a minimum replaces each maxima and Equation (2) is satisfied. In this case, the DM's choices violate INRA but satisfy the remaining axioms.[23]

I now turn to Monotonicity, Subjective Consequentialism, and Continuity. If the utility index depends on the state, then the DM satisfies all axioms except Monotonicity. Fix a set of full support probability measures $\{\pi_\omega\}_{\omega\in\Omega}$ that containing at least two distinct measures. If

$$c(B|\omega) = \arg\max_{f\in B} \int u \circ f d\pi_\omega,$$

then the DM violates Subjective Consequentialism but satisfies the other axioms. The counter-example for Continuity involves lexicographic preferences. I defer it to Appendix C, which also contains details on the above counter-examples.

4. Identification

To interpret a model, it is important to understand how precisely the parameters are identified, i.e. what are the uniqueness properties of the representation. For instance, if certain parameters of the representation are not unique, then doing comparative statics is impossible. How much identification is possible within the current framework, given that the modeler does not directly observe ex ante preference, subjective information, or capacity for attention and that the DM's choices violate WARP? Of the four components that characterize an optimally inattentive agent, Theorem 3 shows that three are suitably unique, and in many cases of interest, all four are unique.

Before stating Theorem 3, one issue deserves elaboration. In general, many attention rules represent the same choice correspondence; for instance, if B contains only constant acts, then $\hat{P}(B)$ could be any partition. However, there is a unique *canonical* attention rule, given by the coarsest attention rule that represents choice. That is, $\hat{P}(\cdot)$ is canonical if $(u(\cdot), \pi(\cdot), \mathbb{P}^*, \hat{P}(\cdot))$ represents $c(\cdot)$ and for any $(u'(\cdot), \pi'(\cdot), \mathbb{P}'^*, \hat{Q}(\cdot))$ that also represents

[22] A full characterization of this model is work in progress.

[23] It is possible that Continuity (ii) is also violated. This is not surprising the interpretation of Continuity (ii) relies on INRA.

$c(\cdot)$, $\hat{Q}(B) \gg \hat{P}(B)$ for every B. The canonical attention rule is the partition

$$\hat{\hat{P}}(B) = \{\{\omega' : c(B|\omega') = c(B|\omega)\} : \omega \in \Omega\}. \tag{4}$$

To interpret this normalization, if paying attention to a finer partition has an arbitrarily small but positive cost, then the DM would always choose a canonical attention rule – she can make the same conditional choice in every state but avoid paying the cost. On the one hand, her subjective information may be finer than that given by her canonical attention rule. That is, she may pay attention to a partition strictly finer than it but make the same conditional choices on at least two cells. On the other hand, her subjective information must be at least as fine as it because otherwise, she does not pay attention to information that distinguishes two state on which she makes different conditional choices.

Theorem 3. *If $c(\cdot)$ is non-degenerate and represented by the optimally inattentive agents $(u_1(\cdot), \pi_1(\cdot), \mathbb{P}^*_1, \hat{P}_1(\cdot))$ and $(u_2(\cdot), \pi_2(\cdot), \mathbb{P}^*_2, \hat{P}_2(\cdot))$, then:*
(i) $u_1(\cdot)$ is a positive affine transformation of $u_2(\cdot)$,
(ii) there is a partition $\mathbb{Q} \ll P$ so that $\pi_1(\cdot|E) = \pi_2(\cdot|E)$ for any $E \in \mathbb{Q}$,
(iii) $\mathbb{P}^*_1 = \mathbb{P}^*_2$, and
(iv) $\hat{P}_1(\cdot) = \hat{P}_2(\cdot)$ whenever $\hat{P}_1(\cdot)$ and $\hat{P}_2(\cdot)$ are both canonical.

Theorem 3 establishes that an optimally inattentive DM's utility index, attention rule and attention constraint are unique. However, her prior probability measure may not be. Because ex ante preference is unobserved, the DM's choices only reveal the likelihood of events that are relevant for choosing either her act or her subjective information. For this reason, the modeler can uniquely identify the DM's prior only up to conditioning on a partition $\mathbb{Q}$, which will be characterized below. Note that a coarser $\mathbb{Q}$ implies more precise identification, in the sense that fewer probability measures represent choices for given "true" prior beliefs. Since $P \gg \mathbb{Q}$, the prior of an optimally inattentive DM is identified at least as precisely as that of a Bayesian DM.

In many cases, $\mathbb{Q} = \{\Omega\}$, and the prior is uniquely identified. For instance, $\mathbb{Q} = \{\Omega\}$ whenever $\mathbb{P}^*$ is all partitions coarser than P with at most κ elements and κ is less than the number of cells in P. One notable case where uniqueness does not obtain is when the DM is Bayesian; in this case, the coarsest $\mathbb{Q}$ is equal to P.

I now turn to characterizing the coarsest $\mathbb{Q}$. This partition is the set of "minimal isolatable events." Intuitively, E is an isolatable event if any choice problem can be partitioned into two distinct problems – one that depends on E and one that depends on E^c – so that either of the two can be varied without changing the DM's conditional choices of acts. The relative likelihood of events contained in different isolatable events is not relevant for her choices.

To define an isolatable event formally, first let

$$B_{E,x}B' = \{fEx : f \in B\} \cup \{gE^c x : g \in B'\}$$

for any problems B, B' and any consequence x. The problem $B_{E,x}B'$ is formed by combining the two problems B and B' into a single problem containing modifications of the acts in B so they differ from each other only on E and of the acts in B' so they differ from each other only on E^c.

Definition 3. E is an *isolatable event* if for any B so that the right hand side of Equation (1) is a singleton and any B', whenever $z \notin c(\{g(\omega'), z\}|\omega')$ for any $g \in B \cup B'$ and ω', both

$$f \in c(B|\omega) \implies fEz \in c(B_{E,z}B'|\omega)$$

for all $\omega \in E$ and

$$f \in c(B|\omega) \implies fE^c z \in c(B_{E^c,z}B'|\omega)$$

for all $\omega \in E^c$ hold.[24]

That is, whenever z is a bad enough outcome, the modifications of the DM's choices from B are still chosen from the problems $B_{E,z}B'$ and $B_{E^c,z}B'$, regardless of the contents of B'. Say that an isolatable event is *minimal* if it does not contain any other non-empty isolatable events. Note that Ω is always an isolatable event, but may not be minimal.

Lemma 1. *If $\mathbb{Q}$ is the coarsest partition satisfying (ii), then E is a minimal isolatable event if and only if $E \in \mathbb{Q}$.*

If the DM is Bayesian, then each $E \in P$ is a minimal isolatable event. However in the introductory example, the only isolatable event is Ω. When the DM is Bayesian, each element of the objective information is a minimal isolatable event, so it follows that her prior is only identified up to its conditional probabilities on every element of P.

5. Comparative Attention and the Value of Information

There are two comparatives of interest. The first is to compare two distinct DMs that have the same information. The second is to compare a single DM with two different objective information partitions.

Consider DM1 and DM2 with conditional choice correspondences given by $c(\cdot)$ and $c'(\cdot)$, respectively. Denote by $\hat{\hat{P}}_c(B)$ the canonical subjective information of $c(\cdot)$ when facing B and by $\hat{\hat{P}}_{c'}(B)$ the canonical subjective information of $c'(\cdot)$ when facing B using Equation (4). Note that these are defined from choices alone.

[24]One can define this condition from choices without referring to $\mathbb{P}^*$, but it is simpler to define it this way.

Definition 4. $c(\cdot)$ *is more attentive than* $c'(\cdot)$ if for any B, there exists a B' so that

$$\hat{\hat{P}}_c(B') = \hat{\hat{P}}_{c'}(B).$$

To understand this comparison, suppose that the modeler observes that $c'(\cdot)$ pays attention to Q when facing B, i.e. her canonical subjective information is Q. If $c(\cdot)$ is more attentive than $c'(\cdot)$, then there is a B' so that the modeler observes that $c(\cdot)$ pays attention to Q when facing B'. That is, whenever the modeler observes DM2 using information Q, the modeler also observes DM1 using Q, though possibly when facing a different choice problem. Theorem 4 shows that this comparison is equivalent to comparing their attention constraints when both DMs have optimal inattention.

Theorem 4. *If* $c(\cdot)$ *and* $c'(\cdot)$ *are non-degenerate and have optimal inattention representations, parametrized by* $(u_c, \pi_c, \mathbb{P}^*_c, \hat{P}_c)$ *and* $(u_{c'}, \pi_{c'}, \mathbb{P}^*_{c'}, \hat{P}_{c'})$ *respectively, then:*

$$c(\cdot)\, is\, more\, attentive\, than\, c'(\cdot) \iff \mathbb{P}^*_{c'} \subset \mathbb{P}^*_c.$$

That is, $c(\cdot)$ is more attentive than $c'(\cdot)$ if and only if her attention constraint is larger. Note that their representations may have different priors and tastes. Therefore, $\mathbb{P}^*$ reflects the DM's capacity for attention: whenever $\mathbb{P}^*_{c'} \subset \mathbb{P}^*_c$, DM1 has a higher capacity for attention than DM2.

Another interesting comparative is how a DM reacts to changes in the available information. Up to now, I have considered a fixed information structure. I modify the primitives in order to allow the objective information to vary. This generalization allows me to consider the value of information to an optimally inattentive DM. Consider a set of finite partitions, $\mathcal{P}$, that represent the possible objective information. Suppose the DM's choices given objective information P are represented by a conditional choice correspondence indexed by $P \in \mathcal{P}$, i.e. $c_P : K(\mathcal{F}) \times \Omega \to K(\mathcal{F})$ where $c_P(B|\omega) \subset B$ and $c_P(B|\cdot)$ is P-measurable. I assume throughout that each $c_P(\cdot)$ has an optimal inattention representation parametrized by $(u, \pi, \mathbb{P}^*_P, \hat{P}(\cdot))$. In Appendix B.3, I provide a sufficient condition for this specification.

I now consider the value of inattention to an optimally inattentive DM. The typical formulation (for instance, Blackwell [1953]) says that Q_1 is objectively more valuable than Q_2 if the expected utility that a DM can obtain by choosing from any problem B is higher when she conditions her choices on Q_1 than when she conditions her choices on Q_2, regardless of her utility index and prior. To state this formally, define

$$V(Q, B, u, \pi) = \sum_{E \in Q} \pi(E) \max_{f \in B} \int u \circ f d\pi(\cdot|E)$$

for every $Q \in \mathcal{P}$, problem B, utility index u and probability measure π. Say that Q_1 *is objectively more valuable than* Q_2 if and only if

$$V(Q_1, B, u, \pi) \geq V(Q_2, B, u, \pi)$$

for every problem B, utility index u and prior π.

This definition only makes sense if the DM in the absence of inattention because an inattentive DM may not be able to condition her choices on the objective information. Instead, I propose the following alternative: Q_1 *is subjectively more valuable than* Q_2 if and only if

$$\max_{Q' \in \mathbb{P}^*_{Q_1}} V(Q', B, u, \pi) \geq \max_{Q' \in \mathbb{P}^*_{Q_2}} V(Q', B, u, \pi)$$

for every problem B, utility index u and prior π. A key difference between the two notions is that whether Q_1 is subjectively more valuable than Q_2 depends on the DM under consideration – DM1 may regard Q_1 as subjectively more valuable than Q_2 while DM2 reverses the ranking.

Theorem 5 relates this comparison to comparative capacity for attention. Note that "more attentive than" is defined for a fixed information structure, but can be easily adapted to our present context where the information structure varies. I omit a formal restatement.

Theorem 5. *For any $Q_1, Q_2 \in \mathcal{P}$, Q_1 is subjectively more valuable than Q_2 if and only if $c_{Q_1}(\cdot)$ is more attentive than $c_{Q_2}(\cdot)$.*

One case where objectively and subjectively more valuable agree is if the DM is Bayesian. Another case where the equivalence between (i) and (iii) holds is if

$$\mathbb{P}^*_P = \{Q' \ll P : Q' \, has \, at \, most \, \kappa \, elements\}$$

for every $P \in \mathcal{P}$. In general, however, the equivalence fails.[25] This accords with some real-life evidence on information overload. For instance, when choosing between health care plans, DMs may become overwhelmed by the sheer amount of information available and make decisions based on less information as more is provided. The contrast between objective and subjective valuation of information is one step towards analyzing information overload.[26]

6. Application: Markets with Optimally Inattentive Consumers

This section argues that inattention may increase competition among firms and benefit consumers. I use a simple model to show that, in equilibrium, more attentive consumers may have less expected consumer surplus. The key idea is that in order for firms that produce differentiated products to exploit their market power, consumers must pay attention to the differences between the products. If consumers are optimally inattentive, then firms must compete with each other for attention, even if their products would not compete given that

[25] Suppose that $\Omega = \{a, b, c, d\}$, $P = \{\{a, b\}, \{c, d\}\}$, $Q = \{\{a\}, \{b\}, \{c\}, \{d\}\}$ and $\mathbb{P}^*_{Q'} = \{\{E, E^c\} : E \in Q\}$ for $Q' \in \{P, Q\}$. Then for $B = \{(100, 100, -100, -100), (-100, -100, 100, 100)\}$, facing P yields a higher ex ante expected utility than facing Q.

[26] Further work along this line is in progress.

consumers were Bayesian. This result contrasts with several papers demonstrating that firms benefit when consumers exhibit inattention; for instance, Rubinstein [1993].

To illustrate the model, I return to the example in Section 1.2, but suppose now that the drugs are non-prescription, the patient must pay out of pocket, and the patient has access to the same information that the doctor did. The patient purchases at most one of the drugs and observes the price of all three drugs before deciding which to purchase. If she processes all available information and the three firms compete by setting prices, then each firm picks a price that extracts her entire surplus whenever its drug is most effective.[27] In contrast, if the patient cannot process all of her information and is optimally inattentive, then consumer surplus must be positive in equilibrium. To see this, suppose that one firm sets a price that would extract all surplus. The patient has no incentive to pay attention to information revealing if that firm's drug is effective. Consequently, to induce the patient to pay attention to information about its drug's effectiveness, each firm must set a price that gives positive consumer surplus.

6.1. **Model.** There are n risk-neutral firms. Each costlessly produce one of $m \geq 3$ distinct, non-divisible goods. A market ϕ is an element of $\{1, ..., m\}^n$ with the interpretation that firm i produces product ϕ_i. Let $n_\mu(\phi)$ be the number of firms who have a monopoly on producing a good of a given type, and $n_c(\phi)$ be the number of goods produced competitively, i.e. by at least two firms. All consumers and firms know the type of product that each firm produces.

A risk-neutral consumer purchases at most one unit of the good. The state space is $\Omega = \{1, ..., m\}$, and the consumer values a good at 1 if its type matches the state and otherwise values it at 0. She initially assigns equal probability to each state and has access to information that reveals the state of the world perfectly, i.e. her objective information is $P = \{\{\omega\} : \omega \in \Omega\}$. She has optimal inattention with an attention constraint parametrized by κ where

$$\mathbb{P}^* = \{Q \ll P : \#Q \leq \kappa\}.$$

The timing of the game is as follows. First, the state of the world is determined. Then, firms simultaneously choose a price without observing the state. Next, each consumer observes the price. Finally, each consumer chooses her subjective information, observes its realization, and purchases from one of the firms.

Let $\phi_i(q)$ be the act of buying from firm i in market ϕ at price q, so that

$$u(\phi_i(q))(\omega) = \begin{cases} 1 - q & if\, \omega = \phi_i \\ -q & otherwise \end{cases}.$$

[27] Similarly, if the consumer has fixed attention, then any firm whose information she distinguishes can extract all surplus.

A pair (ϕ, p), where ϕ is a market and $p \in \mathbb{R}^n_+$ is a price vector, corresponds to the problem $\{\phi_i(p_i) : i \leq n\} \cup \{\emptyset\}$, where $\emptyset$ is not buying from any firm. An equilibrium for a given market ϕ is a price vector $p \in \mathbb{R}^n_+$ so that each firm j in ϕ maximizes expected profit given p_{-j} and the choices of the consumer.

6.2. **Equilibrium.** In any equilibrium where there are more products available than the consumer has the capacity to differentiate between, i.e. attention is scarce, then the effective equilibrium price is zero.

Proposition 2. *For any market ϕ, if p is an equilibrium for ϕ and the consumer purchases from firm j, then $n_c(\phi) + n_\mu(\phi) > \kappa$ implies that $p_j = 0$, and $n_c(\phi) + n_\mu(\phi) \leq \kappa$ implies that for any j, either $p_j = 1$ or $p_j = 0$, where $p_j = 1$ if and only if j has a monopoly.*

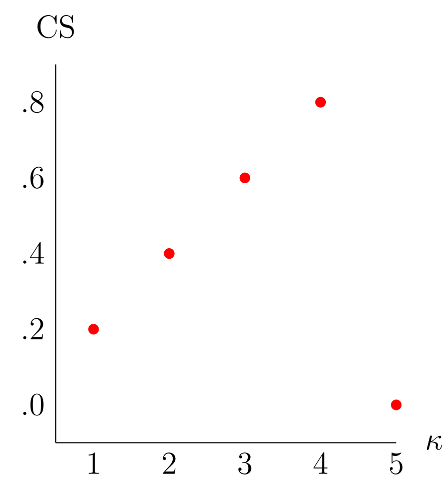
To illustrate, consider first the case where there are $n = 6$ firms, $m = 3$ products and the market is $\phi = (1, 2, 3, 1, 2, 3)$. Two firms produce each type of product. If $\kappa = 3$, then the consumer is Bayesian and a firm makes a sale only if the product it produces matches the state of the world. Since consumers know the state of the world, in every state j, the two firms of type j play the Bertrand duopoly game with marginal cost equal to zero. Consequently, the only equilibrium is $p = (0, 0, 0, 0, 0, 0)$.

Consider the same market where $\kappa = 2$. The same price vector is an equilibrium, but the consumer behaves differently. The consumer pays attention to $\{\{j\}, \{j\}^c\}$ for some $j \in \{1, 2, 3\}$, and in state j, she is indifferent between purchasing from either of the two firms with type j; in any other state, she is indifferent between purchasing from any of the four remaining firms. WLOG, assume that the consumer's subjective information is $\{\{1\}, \{2, 3\}\}$.

Now, suppose firms 5 and 6 exit the market, so there are $n' = 4$ firms and the market is $\phi' = (1, 2, 3, 1)$. If $\kappa = 3$, then firms 2 and 3 have monopolies on producing goods of type 2 and 3, respectively, so both these firms charge 1. In contrast, firms 1 and 4 both produce good 1, so they compete as a Bertrand duopoly. The unique equilibrium price vector is $(0, 1, 1, 0)$, and consumers get expected consumer surplus equal to $\frac{1}{3}$.

But if $\kappa = 2$, then an equilibrium price vector is $(0, 0, 0, 0)$. Firms 1 and 4 compete as a duopoly; if either charged a positive price and made a sale, the other could undercut the price and make a larger profit. Since firm 1 sets a price equal to zero, the consumer must decide whether to pay attention to information that distinguishes either state 2 or state 3. To attract the customer, firms 2 and 3 must offer the consumer surplus conditional on paying attention to the information that reveals whether their product is optimal. Again, if either charged a positive price and made a sale, the other makes zero profit. The firm without

Figure 2. Expected Consumer Surplus with $n = m = 5$ and $n_c(\phi) = 0$



a sale could undercut the other's price, causing the consumer to pay attention to different information and make a sale.[28]

Though the above equilibrium is not unique, in *any* equilibrium to the game, no firm that charges a positive price makes a sale with positive probability in equilibrium. Intuitively, if two firms share a type, and the first charges a positive price and makes a sale with positive probability, then the second can undercut its price to capture the whole market. Competition between the two firms drives the price to zero. If no other firm shares a type with a firm that charges a positive price, then the consumer does not pay attention to information about that firm's product. Consequently, if any firm charges a positive price, then no consumer purchases from it.

For a given market ϕ and price vector p, expected consumer surplus weakly increases with κ. However, *equilibrium* consumer surplus is non-monotonic in κ for the above market – it is maximized at $\kappa = 2$. Proposition 3 characterizes equilibrium consumer surplus for any market.

[28]This can also be seen using Monotonicity and INRA. Define $B(p_2, p_3) = (\phi, (0, p_2, p_3, 0, 0, 0))$ and $B'(p_2, p_3) = (\phi', (0, p_2, p_3, 0))$ for $p_2, p_3 \geq 0$. Identify $B'(p_2, p_3)$ as the natural subset of $B(p_2, p_3)$, i.e. $\phi'_i(p)$ is $\phi_i(p)$ for $i \leq 4$. Suppose $p_3 = 0$ and that $p_2 > 0$. Because $\phi_5(0)$ dominates $\phi_2(p_2)$,

$$c(B(p_2, p_3)|\omega) = c(B(0, p_3)|\omega) \backslash \{\phi_2(p_2)\}$$

for every $\omega \in \Omega$ by Monotonicity and INRA. Since

$$c(B(p_2, p_3)|\omega) \cap B'(p_2, p_3) \neq \emptyset$$

for every $\omega \in \Omega$, INRA implies that

$$\phi_2(p_2) \notin c(B'(p_2, p_3)|\omega)$$

for every ω. Consequently, firm 2 is indifferent between charging 0 and any other price. Repeating the above arguments but swapping p_3 with p_2 and firm 2 with firm 3 shows that the same holds for firm 3, so $(0, 0, 0, 0)$ is an equilibrium. In this equilibrium, the expected consumer surplus is $\frac{2}{3}$, larger than with $\kappa = 3$.

Proposition 3. *In any equilibrium for ϕ, expected total surplus is equal to $\frac{1}{m}\min(n_\mu(\phi) + n_c(\phi), \kappa)$, and expected consumer surplus equals $\frac{1}{m}\kappa$ if $\kappa < n_\mu(\phi) + n_c(\phi)$ and equals $\frac{1}{m}n_c(\phi)$ if $\kappa \geq n_\mu(\phi) + n_c(\phi)$.*

Figure 2 illustrates Proposition 3 graphically. In a market with five firms that each produce different products, expected consumer surplus is maximized at $\kappa = 4$ and minimized at $\kappa = 5$. Similarly, expected profit is maximized when $\kappa = 5$ and equals 0 for any other value of κ. Proposition 3 shows that this non-monotonicity occurs whenever $n_c(\phi) < \kappa$.

7. Conclusion

In this paper, I have axiomatically characterized the properties of conditional choices that are necessary and sufficient for the DM to act as if she has optimal inattention. These axioms provide a choice-theoretic justification for the theory that agents respond to their limited attention optimally. The optimal inattention model is a versatile model with interesting implications: Dow [1991], Rubinstein [1993], and Gul et al. [2011] all consider consumers who conform exactly to the optimal inattention model.

Related papers by de Olivera [2012] and Mihm and Ozbek [2012] study rational inattention as revealed by a DM's ex ante preference over menus of acts.[29] Their representation of preference is similar to my own, but the primitives are very different. The DM chooses a menu in the anticipation that she will receive information and can choose what information to process at some cost.

Caplin and Martin [2012] study a related model, optimal framing. If frames are interpreted as states, then their analysis can be interpreted similarly to mine. Our papers are complementary, as their framework is designed for testing in the laboratory but does not achieve as precise identification. Moreover, Caplin and Martin relate choices to one another only through the existence of a utility function that solves a system of inequalities. In contrast to this paper, their primitive is stochastic choice, and their DM's prior is known to the modeler.

By way of conclusion, I compare the optimal inattention model with some other models of inattention that have been considered by the literature. The most prominent example is the rational inattention model, due to Sims [1998, 2003]. In this model, the constraint on attention takes the form of restricting the mutual information, i.e. the reduction in entropy, between actions and the state of the world. This constraint implies that conditional choices are stochastic.[30] One interpretation is that the agent has access to arbitrarily precise, and

[29] Ergin and Sarver [2010] can also be interpreted in this way, but it is not their focus.

[30] Recently, Matejka and McKay [2012] have studied this model's implications in the context of discrete choices. Their focus is on solving the model in a discrete setting, and in the course of analysis, they provide testable implications in terms of choices from a suitably rich feasible set of actions. A full behavioral characterization of the model, even in this setting, remains an open question.

arbitrarily imprecise, signals about the state of the world, but the modeler does not observe the realization of this information. Another, offered by Woodford [2012], is that the agent's perception of information is stochastic. Both these interpretations are outside the scope of my model: the objective information is known, and the agent's perceptions are deterministic.

Mankiw and Reis [2002] introduce the sticky information model. It postulates that agents update their information infrequently, and when they update, they obtain perfect information. The key difference between this model in a static setting and optimal inattention is that agents do not choose the information to which they pay attention.[31]

References

Charalambos Aliprantis and Kim Border. *Infinite Dimensional Analysis*. Springer, Berlin, Germany, 3rd edition, 2006.

Robert J. Aumann. Utility theory without the completeness axiom. *Econometrica*, 30(3): 445–462, 1962.

David Blackwell. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, 24(2):265–272, 1953.

Andrew Caplin and Daniel Martin. Framing effects and optimization. *Mimeo*, 2012.

Olivier Coibion and Yuriy Gorodnichenko. What can survey forecasts tell us about informational rigidities? *Journal of Political Economy*, Forthcoming, 2011a.

Olivier Coibion and Yuriy Gorodnichenko. Information rigidity and the expectations formation process: A simple framework and new facts. *mimeo*, 2011b.

Henrique de Olivera. Axiomatic foundations for rational inattention. *Mimeo*, 2012.

Stefano DellaVigna and Joshua Pollet. Investor inattention and friday earnings announcements. *Journal of Finance*, 64, 2009.

David Dillenberger, Juan Sebastian Lleras, Philipp Sadowski, and Norio Takeoka. A theory of subjective learning. *Mimeo*, 2012.

James Dow. Search decisions with limited memory. *Review of Economic Studies*, 58(1):1–14, 1991.

Haluk Ergin and Todd Sarver. A unique costly contemplation representation. *Econometrica*, 78(4), 2010.

Itzhak Gilboa, Fabio Maccheroni, Massimo Marinacci, and David Schmeidler. Objective and subjective rationality in a multiple prior model. *Econometrica*, 78(2), 2010.

Jean-Michel Grandmont. Continuity properties of von Neumann-Morgenstern utility. *Journal of Economic Theory*, 4(1):45–57, 1972.

[31]The subjective learning literature, e.g. Dillenberger et al. [2012] and Natenzon [2012], studies an agent who has or anticipates receiving information (which is unobserved to the modeler) before making her choice. Though they do not focus on interpreting this behavior as inattention, these models have the same relationship to mine as does Mankiw and Reis [2002].

Faruk Gul and Wolfgang Pesendorfer. Random expected utility. *Econometrica*, 74(1):pp. 121–146, 2006.

Faruk Gul, Wolfgang Pesendorfer, and Tomasz Strzalecki. Behavioral competitive equilibrium and extreme prices. *mimeo*, 2011.

Christian Hellwig and Laura Veldkamp. Knowing what others know: Coordination motives in information acquisition. *Review of Economic Studies*, 76(1):223–251, 2009.

Michael Luca. Reviews, reputation, and revenue: The case of yelp.com. *mimeo*, 2011.

Bartosz Mackowiak and Mirko Wiederholt. Optimal sticky prices under rational inattention. *American Economic Review*, 99:769–803, 2009.

N. Gregory Mankiw and Ricardo Reis. Sticky information versus sticky prices: A proposal to replace the new keynesian phillips curve. *The Quarterly Journal of Economics*, 117(4): 1295–1328, 2002.

Yusufcan Masatlioglu, Daisuke Nakajima, and Erkut Y. Ozbay. Revealed attention. *American Economic Review*, 105:2183–2205, 2012.

Filip Matejka and Alisdair McKay. Rational inattention to discrete choices: A new foundation for the multinomial logit model. *mimeo*, 2012.

Maximilian Mihm and M. Kemal Ozbek. Decision making with rational inattention. *Mimeo*, 2012.

Paulo Natenzon. Random choice and learning. *Mimeo*, 2012.

Harold E. Pashler. *The Psychology of Attention*. MIT Press, Cambridge, MA, USA, 1998.

Ariel Rubinstein. On price recognition and computational complexity in a monopolistic model. *Journal of Political Economy*, 101(3):473–484, 1993.

Lloyd S. Shapley and Manel Baucells. A theory of multiperson utility. *mimeo*, 1998.

Christopher A. Sims. Stickiness. In *Carnegie-Rochester Conference Series on Public Policy*, volume 49, pages 317–356. Elsevier, 1998.

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

Stijn Van Nieuweburgh and Laura Veldkamp. Information acquisition and underdiversification. *Review of Economic Studies*, 77(2):779–805, 2010.

Timothy Van Zandt. Hidden information acquisition and static choice. *Theory and Decision*, 40:235–247, 1996.

Michael Woodford. Inattention as a source of randomized discrete adjustment. *Mimeo*, 2008.

Michael Woodford. Inattentive valuation and reference-dependent choice. *Mimeo*, 2012.

Appendix A. Preview of the Proof of Theorem 1

This Section discusses the key idea behind the proof of Theorem 1. The key idea of the proof is to map choices of acts onto a larger domain where it is suitably well-behaved. In particular, I consider the space of "plans." A plan is a mapping from each state to an act. Each set of conditional choices from a given menu defines one plan. In the example from Section 1.2, the modeler observes the doctor's conditional choices of acts. Instead of looking at her each of her conditional choices in isolation, one can think of them as choosing one plan. For instance, the doctor chooses the plan "pick g in state γ, otherwise pick m" from $\{g, m\}$ and chooses the plan "pick f in state ϕ, otherwise pick m" from $\{g, m, f\}$.

Although choice in a given state may violate WARP, INRA guarantees that her choice over plan maximizes a preference relation (whose domain is plans rather than acts). However, this preference relation may be discontinuous, incomplete and intransitive. Given the other axioms, one can extend it to a well-behaved preference relation so that the DM's choices are a maximal element of this preference relation.[32] I then show that this preference over plans can be represented as expected utility over a subset of "feasible" plans. I identify a candidate for $\mathbb{P}^*$ and show that any plan measurable with respect to some $Q \in \mathbb{P}^*$ is feasible. The coarsest partition that measures her chosen plan is identified as the DM's subjective information. This subjective information is optimal, in the sense that it maximizes expected utility according to the utility index and prior representing the preference over plans. The final step shows that her conditional choices can also be represented as maximizing this preference relation.

This suggests an alternative domain on which optimal inattention admits foundations: preference over plans. In supplementary material, I show that one can use preference over plans to derive both optimal inattention and costly attention representations. Observing choice of plan is more convenient because it requires observing a single ex ante choice rather than choices in each state of the world. It has also been used in applications (for instance, Gul et al. [2011]). However, this has some significant drawbacks. First, choice of plan is difficult to observe outside of a laboratory. Second, choice of plan typically reflects both constraints and true preference. Third, what a DM plans to choose may differ from what she actually chooses. Fourth, economic objects of interest are conditional choices, not ex ante choices. Therefore, I focus in the main paper on choice of acts. This data is closer to what economists typically work with, and reflects the DM's response to whatever constraints she faces rather than those she thinks that she will face. Moreover, if the DM follows through with her choice of plan, then her final conditional choices of acts satisfy the axioms from Section 3.

Appendix B. Proofs

B.1. **Proofs from Section 3.**

[32] Although this is an extension, it is typically not a "compatible extension" in the sense that it may not preserve strict preference.

B.1.1. *Proof of Theorem 1:*

Proof. If $c(B|\omega) = B$ for all $B \in K(X)$, then by Monotonicity, it follows that $c(B|\omega) = B \forall B \in K(\mathcal{F})$. Taking $\mathbb{P}^* = \{\Omega\}$ and $u(x) = 0 \forall x$ establishes the desired result. Therefore, assume that there are $x^*, x_* \in X$ and ω so that $x_* \notin c(\{x^*, x_*\}|\omega)$.

Lemma 2. *There exists an affine, continuous $u : X \to \mathbb{R}$ so that for any $B \in K(X)$, $x \in c(B|\omega) \iff u(x) \geq u(y)$ for all $y \in B$.*

Proof. Fix any $A \subset B \in K(X)$. By Monotonicity, if $x \in c(B|\omega)$, then $x \in c(B|\omega')$ for any ω'. By INRA, $c(B|\omega) \cap A \neq \emptyset$ implies that $c(B|\omega) \cap A = c(A|\omega)$, i.e. $c(\cdot|\omega)$ satisfies WARP when restricted to problems of lotteries. It is routine to verify that the resulting revealed preference relation satisfies the hypothesis of Grandmont [1972, Thm 2] and therefore an affine, continuous $u : X \to \mathbb{R}$ exists. □

Let $\mathcal{F}^\Omega$ be the set of functions from Ω to $\mathcal{F}$ that are $\sigma(P)$ measurable. I refer to elements of $\mathcal{F}^\Omega$ as "plans" with the interpretation that the DM chooses $F(\omega)$ in state ω. Since P is finite, any $F \in \mathcal{F}^\Omega$ is simple. I denote elements of X by $x, y, z, \ldots$, elements of $\mathcal{F}$ by $f, g, h, \ldots$ and elements of $\mathcal{F}^\Omega$ by $F, G, H, \ldots$. Identify X with the subset of $\mathcal{F}$ that does not vary with the state and $\mathcal{F}$ with the subset of $\mathcal{F}^\Omega$ that does not vary with the state, so $X \subset \mathcal{F} \subset \mathcal{F}^\Omega$.

Denote by $\mathcal{A}(\cdot)$ the partition defined as $\hat{\hat{P}}(\cdot)$ in Equation (4) and define $\hat{c} : K(\mathcal{F}) \twoheadrightarrow \mathcal{F}^\Omega$ by $F \in \hat{c}(B) \iff F(\omega) \in c(B|\omega)$ for every ω and $\sigma(F) \subset \sigma(\mathcal{A}(B))$. Since $\sigma(\mathcal{A}(B)) \subset \sigma(P)$, any $\sigma(\mathcal{A}(B))$-measurable selection from $c(B|\cdot)$ is in $\mathcal{F}^\Omega$. For any $F \in \mathcal{F}^\Omega$, define $\{F\} \in K(\mathcal{F})$ by $\{F\} = \{F(\omega) : \omega \in \Omega\}$. Since F is simple, $\{F\}$ is finite and compact. By INRA, if $F \in \hat{c}(B)$ then $F \in \hat{c}(\{F\})$.

Define $\mathcal{C} \subset \mathcal{F}^\Omega$ by $F \in \mathcal{C} \iff F \in \hat{c}(\{F\})$; $\mathcal{C}$ is the set of plans that the DM chose from some problem. Define a binary relation $\hat{\succsim}$ on $\mathcal{C}$ by

$$F \hat{\succsim} G \iff \{F\} \, \overline{IS} \, \{G\}.$$

Note that $F \in \hat{c}(B) \implies F \hat{\succsim} G$ for every $G \in \mathcal{C}$ so that $\{G\} \subset B$. For any $F \in \mathcal{F}^\Omega$, define $F^* \in \mathcal{F}$ to be $F^*(\omega) = F(\omega)(\omega)$.

Lemma 3. *If $F \in \hat{c}(\{F\})$, then $F^* \hat{\sim} F$.*

Proof. Assume that $F = \hat{c}(\{F\})$ and define $\hat{F}$ so that

$$\hat{F}(\omega) = F(\omega)\mathcal{A}(\{F\})(\omega)\underline{x}$$

for some $\underline{x} \in X$ so that $u(\underline{x}) \leq u(F(\omega)) \forall \omega$. It holds that $\hat{F} \in \hat{c}(\{F\} \cup \{\hat{F}\})$. To see this, note that

$$u \circ \hat{F}(\omega) \leq u \circ F(\omega)$$

for all ω so it must be that $F \in \hat{c}(\{F\} \cup \{\hat{F}\})$ by monotonicity and INRA. Therefore, $\mathcal{A}(\{F\} \cup \{\hat{F}\}) \gg \mathcal{A}(\{F\})$. For any ω, so that $\mathcal{A}(\{F\} \cup \{\hat{F}\})(\omega) \subset \mathcal{A}(\{F\})(\omega)$ so since

$$F(\omega)(\omega') = \hat{F}(\omega)(\omega') \forall \omega' \in \mathcal{A}(\{F\})(\omega),$$

it follows from Subjective Consequentialism and $F(\omega) \in c(\{F\} \cup \{\hat{F}\}|\omega)$ that $\hat{F}(\omega) \in c(\{F\} \cup \{\hat{F}\}|\omega)$. By INRA, $\hat{F} = \hat{c}(\{\hat{F}\})$. Further, if $F^* \hat{\sim} \hat{F}$, it follows that $F^* \hat{\sim} F$.

By construction, $\mathcal{A}(\{\hat{F}\})$ is finer than $\mathcal{A}(\{F^*\})$. By ACI,

$$\alpha \hat{F} + (1-\alpha) F^* \in \hat{c}(\alpha \{\hat{F}\} + (1-\alpha)\{F^*\})$$

for all $\alpha \in [0,1]$. Further

$$u \circ \alpha \hat{F}(\omega) + (1-\alpha) F^* \geq u \circ \hat{F}(\omega)$$

by construction. Therefore, when $B_\alpha = (\alpha\{F^*\} + (1-\alpha)\{\hat{F}\}) \cup \{\hat{F}\}$

$$\alpha \hat{F} + (1-\alpha) F^* \in \hat{c}(B_\alpha)$$

by INRA. Since $\mathcal{A}(B_\alpha) \gg \mathcal{A}(\{\hat{F}\})$ for any $\alpha \in [0,1)$ and for any ω,

$$[\alpha F^* + (1-\alpha)\hat{F}(\omega)](\omega') = \hat{F}(\omega)(\omega')$$

for all $\omega' \in \mathcal{A}(\{\hat{F}\})(\omega)$, by Subjective Consequentialism

$$\hat{F} \in \hat{c}(B_\alpha).$$

Therefore, for $n \in \{1, 2, ...\}$, $\{F\}$ IS $\frac{n-1}{n}\{F^*\} + \frac{1}{n}\{\hat{F}\}$, which goes to $\{F^*\}$, and by definition $F \hat{\succeq} F^*$. Since $F^* \in \hat{c}(\{F^*\} \cup \{\hat{F}\})$, it follows that $F^* \hat{\succeq} F$; combining yields $F \hat{\sim} F^*$. □

Lemma 4. *For any $h \in \mathcal{F}$, $\alpha \in [0,1]$ and $A, B \in K(\mathcal{F})$, if $A \, \overline{IS} \, B$, then $\alpha A + (1-\alpha)\{h\} \, \overline{IS} \, \alpha B + (1-\alpha)\{h\}$.*

Proof. There are sequences $(A_n)_{n=1}^\infty$ and $(C_n)_{n=1}^\infty$ that converge to A and B respectively where A_n IS C_n.

Consider an arbitrary n and the finite sequence $B_1, ..., B_m$ so that $A_n = B_1$ and $B_m = C_n$ and $c(B_i|\omega) \cap B_{i-1} \neq \emptyset$ for every ω. Since $\mathcal{A}(B_i) \gg \mathcal{A}(\{h\})$, $c(\alpha B_i + (1-\alpha)\{h\}|\omega) \cap [\alpha B_{i-1} + (1-\alpha)\{h\}]$ for every ω by ACI. Since $\alpha A_n + (1-\alpha)\{h\} = \alpha B_1 + (1-\alpha)\{h\}$ and $\alpha B_m + (1-\alpha)\{h\} = \alpha C_n + (1-\alpha)\{h\}$, $\alpha A_n + (1-\alpha)\{h\}$ IS $\alpha C_n + (1-\alpha)\{h\}$.

Since n was arbitrary, we can do this for all n. Note that $\alpha A_n + (1-\alpha)\{h\} \to \alpha A + (1-\alpha)\{h\}$ and $\alpha C_n + (1-\alpha)\{h\} \to \alpha B + (1-\alpha)\{h\}$, it follows that $\alpha A + (1-\alpha)\{h\} \, \overline{IS} \, \alpha B + (1-\alpha)\{h\}$. □

Lemma 5. $\hat{\succeq}$ *is transitive.*

Proof. Suppose $F \hat{\succeq} G$ and $G \hat{\succeq} H$. Set $A = \{F\}$, $B = \{G\}$ and $C = \{H\}$. Then $F, G, H \in \mathcal{C}$, $A \, \overline{IS} \, B$ and $B \, \overline{IS} \, C$.

Then there are sequence A_n, B_n, B'_n, C_n that converge to A, B, B, C respectively so that $A_n\, IS\, B_n$ and $B'_n\, IS\, C_n$. Pick $G_n \in \hat{c}(B'_n)$, noting that $\{G_n\}\, IS\, C_n$ by INRA. Let y be the worst outcome of any act in B. Let $z \in X$ be so that $u(y) - u(z) = k > 0$ (if such an outcome does not exist, replace each problem by mixing it with x^*). Pick $F_n \in \hat{c}(B'_n)$ for every n. Note that $\{F_n\}$ IS C_n for every n using INRA.

Because $u(\cdot)$ is continuous, for any $\epsilon > 0$ there is a $\delta(\epsilon)$ so that $d(x, x') < \delta(\epsilon)$ implies that $|u(x) - u(x')| < \epsilon$. Therefore, for any ϵ there is an $n(\epsilon)$ so that for any every $n > n(\epsilon)$ and any act $f' \in \{G_n\}$ there is an act $f \in B_n$ so that $u(f(\omega)) - u(f'(\omega)) < \epsilon$ (for every ω) and $u(f'(\omega)) > u(y) - \epsilon$.

Take a sub-sequence of B'_n and B_n so that $B_{n_i} = B_{n(\frac{1}{i})+1}$ and $B'_{n_i} = B'_{n(\frac{1}{i})+1}$. Pick $\bar{i}$ so that $\frac{1}{i} < k$. Set $\alpha_i = \frac{\frac{1}{n_i}}{(u(y) - \frac{1}{n_i} - u(z))}$ for every $i > \bar{i}$. Consider $f = F_{n_i}(\omega)$ for an arbitrary ω and $i > \bar{i}$. Pick $f' \in \{G_{n_i}\}$ so that $u(f(\omega')) - u(f'(\omega')) < \epsilon$ for every ω'. Note that for every ω'

$$
\begin{aligned}
(1-\alpha)u(f(\omega')) + \alpha u(z) - u(f'(\omega')) &= u(f(\omega')) - u(f'(\omega')) - \alpha(u(f(\omega')) - u(z)) \\
&< \frac{1}{n_i} - \alpha(u(f(\omega')) - u(z)) \\
&< \frac{1}{n_i} - \alpha(u(y) + \frac{1}{n_i} - u(z)).
\end{aligned}
$$

Therefore, for every ω,

$$u \circ ((1-\alpha_i)F_{n_i}(\omega) + \alpha_i z) \leq u \circ f'$$

for some $f' \in \{G_{n_i}\}$.

By Monotonicity and INRA, $\hat{c}(B_{n_i}) \subset \hat{c}(B_{n_i} \cup ((1-\alpha_i)\{G_{n_i}\} + \alpha_i\{z\}))$ for $i < \bar{i}$. By Lemma 4, $\{G_{n_i}\}$ IS C_{n_i} implies that $(1-\alpha_i)\{G_{n_i}\} + \alpha_i\{z\}$ IS $(1-\alpha_i)C_{n_i} + \alpha_i\{z\}$. But then for $i > \bar{i}_k$ A_{n_i} IS $(1-\alpha_i)C_{n_i} + \alpha_i\{z\}$ and since $\alpha_i \to 0$, $A_{n_i} \to A$ and $C_{n_i} \to C$, we have $A\, \overline{IS}\, C$. Suppose we needed to mix all problems with x^* first. The amount of this mixture can be arbitrarily small. So using the same logic, we can find a sub-sequence $(n_k)_{k=1}^{\infty}$and a sequence $(\alpha_k)_{k=1}^{\infty}$ where $\alpha_k \to 0$ so that $(\frac{k-1}{k}A_{n_k} + \frac{1}{k}\{x^*\})$ IS $(1-\alpha_k)(\frac{k-1}{k}C_{n_k} + \frac{1}{k}\{x^*\}) + \alpha_k\{z\}$ for every k. Again, this gives $A\, \bar{IS}\, C$. Conclude that $F \hat{\succeq} H$. □

Define $\mathbb{P}^* = \{\mathcal{A}(B) : B \in K(\mathcal{F})\}$ and $\mathbb{P}^{**} = \{Q \in \mathbb{P}^* : \nexists Q' \in \mathbb{P}^*\, s.t.\, Q' \gg Q\}$. Identify $\mathcal{F}_Q^{\Omega}$ with the subset of $\mathcal{F}^{\Omega}$ that is $\sigma(Q)-$measurable.

Define the binary relation $\succeq$ on $\mathcal{F}^{\Omega}$ by

$$F \succeq G \iff \text{either } F \in \mathcal{H} \text{ and } F^* \hat{\succeq} G^* \text{ or } G \notin \mathcal{H}$$

$\succeq$ is a consistent extension of $\hat{\succeq}$.

Lemma 6. *$\succeq$ extends $\hat{\succeq}$ consistently.*

Proof. Suppose first that $F, G \in \mathcal{H}$ and that $G \in \mathcal{F}_Q^{\Omega}$.

[$F\hat{\succsim}G$ implies $F \succeq G$] Suppose $F\hat{\succsim}G$. Then $F, G \in \mathcal{C}$. By Lemma 3, $F^*\hat{\sim}F$ and $G^*\hat{\sim}G$. Since $\hat{\succsim}$ is transitive, $F^*\hat{\sim}F\hat{\succsim}G \implies F^*\hat{\succsim}G$. Since $G\hat{\sim}G^*$, we have that $F^*\hat{\succsim}G^*$ using transitivity of $\hat{\succsim}$.

[$F\hat{\succ}G$ implies $F \succ G$] There are two cases: $G \in \mathcal{H}$ and $G \notin \mathcal{H}$. In the latter case, it is impossible that $F\hat{\succ}G$. Therefore, it must be that both $F, G \in \mathcal{C}$. For contradiction, assume that $F\hat{\succ}G$ but $F \not\succ G$. Since $F\hat{\succsim}G$ implies $F \succeq G$, it must be that $G \succeq F$ (otherwise $F \succ G$) so $F^*\hat{\sim}G^*$. Since $F\hat{\sim}F^*$ and $G^*\hat{\sim}G$ and $\hat{\succsim}$ is transitive, it follows immediately that $G\hat{\sim}F$, a contradiction with $F\hat{\succ}G$. □

For any act $f \in \mathcal{F}$ and partition $Q \in \mathbb{P}^*$, define $\bar{f}_Q \in \mathcal{F}^\Omega$ as follows. If $\min_{x\in X} u(x)$ does not exist, then $\bar{f}_Q(\omega) = fQ(\omega)x$ where x is so that $u(x) = \min_{x\in\cup_{\omega\in\Omega} supp(G(\omega))} u(x) - 1$. If $\min_{x\in X} u(x)$ exists, then by $\bar{f}_Q(\omega) = fQ(\omega)x$ where x is so that $u(x) = \min_{x\in X} u(x)$.

Lemma 7. *$\succeq$ is a preorder.*

Proof. [Transitive] Suppose that $F \succeq G$ and $G \succeq H$. If $F \notin \mathcal{H}$ then $G \notin \mathcal{H}$ so $H \notin \mathcal{H}$ so $F \succeq H$. If $G \notin \mathcal{H}$ then $H \notin \mathcal{H}$ so $F \succeq H$. If $H \notin \mathcal{H}$, then $F \succeq H$. All that remains is the case where $F, G, H \in \mathcal{H}$ which would imply that

$$F^*\hat{\succsim}G^*\hat{\succsim}H^*$$

Since $\hat{\succsim}$ is transitive, $F^*\hat{\succsim}H^*$ and $F \succeq H$.

[Reflexive] Fix arbitrary F. If $F \notin \mathcal{H}$, then $F \succeq F$ by definition. If $F \in \mathcal{H}$, then noting that $F^*\hat{\sim}F^*$ (since $F^* \in \hat{c}(\{F^*\})$) immediately gives $F^*\hat{\succsim}F^*$ and $F \succeq F$. □

Lemma 8. *If $x, y \in X$ then either $x \succeq y$ or $y \succeq x$.*

Proof. This follows immediately from Lemma 2. □

Lemma 9. *For all $e, f, g, h \in \mathcal{F}$, the set $U = \{\lambda \in [0,1] : \lambda f + (1-\lambda)g \succeq \lambda h + (1-\lambda)e\}$ is closed in* $[0,1]$.

Proof. Suppose $\lambda_n \to \lambda$ and $\lambda_n \in U$ for all n. Then $\{\lambda_n f + (1-\lambda_n)g\} = B_n$, $\{\lambda_n h + (1-\lambda_n)e\} = C_n$ and $B_n\, \overline{IS}\, C_n$ by definition of $\succeq$. Therefore, for every n, there are sequences $(B^n_m)_{m=1}^\infty$ and $(C^n_m)_{m=1}^\infty$ so that B^n_m IS C^n_m and $d(B^n_m, B_n) + d(C^n_m, C_n) \to 0$. For every ϵ, there is an M^n_ϵ so that $m > M^n_\epsilon$ implies that $d(B^n_m, B_n) + d(C^n_m, C_n) < \epsilon$. Since $B_n \to \{\lambda f + (1-\lambda)g\} = B$ and $C_n \to \{\lambda h + (1-\lambda)e\} = C$, for every ϵ, there is an N_ϵ so that $n > N_\epsilon$ implies that $d(B_n, B) + d(C_n, C) < \epsilon$.

For every $n \in \{1, 2, ...\}$ define B'_n and C'_n by

$$B'_n = B^{M_{\frac{1}{n}}+1}_{N_{\frac{1}{n}}+1}$$

and

$$C'_n = C^{M_{\frac{1}{n}}+1}_{N_{\frac{1}{n}}+1}.$$

By the triangle inequality,

$$d(B'_n, B) \leq d(B, B_{N_{\frac{1}{n}}+1}) + d(B_{N_{\frac{1}{n}}+1}, B_{N_{\frac{1}{n}}+1}^{M_{\frac{1}{n}}+1})$$

and

$$d(C'_n, C) \leq d(C, C_{N_{\frac{1}{n}}+1}) + d(C_{N_{\frac{1}{n}}+1}, C_{N_{\frac{1}{n}}+1}^{M_{\frac{1}{n}}+1}).$$

Since

$$d(B, B_{N_{\frac{1}{n}}+1}) + d(B_{N_{\frac{1}{n}}+1}, B'_n) + d(C, C_{N_{\frac{1}{n}}+1}) + d(C_{N_{\frac{1}{n}}+1}, C'_n) \leq \frac{4}{n}$$

which goes to zero, $B'_n \to B$ and $C'_n \to C$; since B'_n IS C'_n for every n, $B \, \overline{IS} \, C$. It immediately follows from the definition of $\succeq$ and U that $\lambda f + (1-\lambda)g \succeq \lambda h + (1-\lambda)e$ so $\lambda \in U$. □

Lemma 10. *For any $f, g \in \mathcal{F}$, if $f(\omega) \succeq g(\omega)$ for all ω, then $f \succeq g$.*

Proof. For any $f, g \in \mathcal{F}$ so that $f(\omega) \succeq g(\omega)$ for every ω, it follows that $f(\omega) \in c(\{f(\omega), g(\omega)\}|\omega)$ from Lemma 2. From monotonicity, $f \in c(\{f, g\}|\omega) \forall \omega$ so $f \in \hat{c}(\{f, g\})$ and $f \succeq g$. □

Lemma 11. *For any $f, g, h \in \mathcal{F}$ and $\alpha \in (0, 1]$, $f \succeq g$ if and only if $\alpha f + (1-\alpha)h \succeq \alpha g + (1-\alpha)h$.*

Proof. Fix $f, g, h \in \mathcal{F}$ so that $f \succeq g$. Let $\alpha \in (0, 1]$ be arbitrary. Since $f \succeq g$, $\{f\} \, \bar{IS} \, \{g\}$. Apply Lemma 4 to get $\{\alpha f + (1-\alpha)h\} \, \bar{IS} \, \{\alpha g + (1-\alpha)h\}$, implying that $\alpha f + (1-\alpha)g \hat{\succeq} \alpha f + (1-\alpha)g$ so $\alpha f + (1-\alpha)g \succeq \alpha f + (1-\alpha)g$. □

Given the above and Lemma 9, Lemma 1.2 of Shapley and Baucells [1998] gives that $\alpha f + (1-\alpha)h \succeq \alpha g + (1-\alpha)h$ for $\alpha > 0$ implies $f \succeq g$.

Lemma 12. *There are $x, y \in X$ so that $x \succ y$.*

Proof. Recall that $x_* \notin \hat{c}(\{x^*, x_*\})$. Since $\mathcal{A}(\{x^*\}) = \mathcal{A}(\{x_*\}) = \{\Omega\}$ and $x^* \in \hat{c}(\{x_*, x^*\})$ by Lemma 2, x^* weakly dominates x_*. It follows that $x^* \succeq x_*$. Further, by Continuity, it is not the case that $\{x_*\} \, \bar{IS} \, \{x^*\}$ so $x^* \succ x_*$. □

Lemma 13. *If there is an $x \in X$ so that $u(f(\omega)) > u(x)$ for every ω, then $\bar{f}_Q \in \hat{c}(\{\bar{f}_Q\})$.*

Proof. Fix any such f and relabel $\bar{f}_Q = \bar{F}$. Let $y \in \arg\min_{x' \in \cup_\omega supp(\bar{F}(\omega))} u(x')$. Suppose not: $\bar{F} \notin \hat{c}(\{\bar{F}\})$ and $H \in \hat{c}(\{\bar{F}\})$.

It must be that $\mathcal{A}(\{\bar{F}\}) \not\gg Q$. If $\mathcal{A}(\{\bar{F}\}) \gg Q$, pick any ω so that $\bar{F}(\omega) \notin c(\{\bar{F}\}|\omega)$. Take $h = H(\omega)\mathcal{A}(\{\bar{F}\})x$. Since $u \circ H(\omega) \geq u \circ h$, $H \in \hat{c}(\{\bar{F}\} \cup \{h\})$ and $\mathcal{A}(\{\bar{F}\} \cup \{h\}) \gg \mathcal{A}(\{\bar{F}\})$. Further, by Subjective Consequentialism, $h \in c(\{\bar{F}\} \cup \{h\}|\omega)$. However, $u \circ \bar{F}(\omega) \geq u \circ h$ so monotonicity implies that $\bar{F}(\omega) \in c(\{\bar{F}\} \cup \{h\}|\omega)$. Since h is never strictly relevant, INRA implies that $\bar{F}(\omega) \in c(\{\bar{F}\}|\omega)$, a contradiction.

Since $\mathcal{A}(\{\bar{F}\}) \not\gg Q$, there is some $E \in P$ so that $u(H^*(\omega)) = u(y) \forall \omega \in E$. For every $\omega \notin E$, $u(F^*(\omega)) \geq u(H^*(\omega))$. Therefore, F^* weakly dominates H^* by monotonicity.

Define $\bar{H} \equiv \bar{H}^*{}_Q$ where $Q = \mathcal{A}(\{\bar{F}\})$. By definition, there is some J so that $J \in \hat{c}(\{J\})$ and $\mathcal{A}(\{J\}) = Q$. Since $\mathcal{A}(\{F^*\}) = \{\Omega\}$, $\alpha F^* + (1-\alpha)J \in \hat{c}(\alpha\{F^*\} + (1-\alpha)J)$

By INRA, monotonicity and Subjective Consequentialism, $\bar{H} \in \hat{c}(\{\bar{H}\} \cup \{\bar{F}\})$ and $\bar{H}^* \in \hat{c}(\{\bar{H}\} \cup \{\bar{H}^*\})$. Let $B_0 = \alpha\{\bar{H}^*\} + (1-\alpha)\{J^*\}$, $B_1 = \alpha\{\bar{H}^*\} \cup \{\bar{H}\} + (1-\alpha)\{J^*\}$ and $B_2 = \alpha\{\bar{H}\} \cup \{\bar{F}\} + (1-\alpha)\{J^*\}$. By ACI,

$$\alpha\bar{H}^* + (1-\alpha)J^* \in \hat{c}(B_1) \cap \hat{c}(B_0)$$

and

$$\alpha\bar{H} + (1-\alpha)J^* \in \hat{c}(B_2).$$

This implies that

$$\alpha\bar{H}^* + (1-\alpha)J^* \hat{\succeq} \alpha\bar{H} + (1-\alpha)J^* \hat{\succeq} \alpha\bar{F} + (1-\alpha)J^*. \tag{5}$$

Set $B_4 = \alpha\{F^*\} + (1-\alpha)\{J\}$. By ACI, $\alpha F^* + (1-\alpha)J(\omega) \in c(B_4|\omega)$. Let $B_5 = B_4 \cup \{\alpha\bar{F} + (1-\alpha)J\}$. By INRA, monotonicity and Subjective Consequentialism, $\alpha F^* + (1-\alpha)J \in \hat{c}(B_5)$. By Subjective Consequentialism, $\alpha\bar{F}(\omega) + (1-\alpha)J(\omega) \in c(B_5|\omega)$ for all ω so by INRA, $\alpha\bar{F} + (1-\alpha)J \in \mathcal{C}$. Set $B_3 = B_2 \cup \{\alpha\bar{F} + (1-\alpha)J\}$. By Monotonicity and INRA,

$$\alpha\bar{H} + (1-\alpha)J^* \in \hat{c}(B_3)$$

so $\alpha\bar{H}^* + (1-\alpha)J^* \hat{\succeq} \alpha\bar{F} + (1-\alpha)J$.

By Lemma 3, $\alpha\bar{F} + (1-\alpha)J \hat{\sim} \alpha F^* + (1-\alpha)J^*$. By Lemma 5,

$$\alpha\bar{H}^* + (1-\alpha)J^* \hat{\succeq} \alpha F^* + (1-\alpha)J^*$$

which, by definition, is equivalent

$$\{\alpha\bar{H}^* + (1-\alpha)J^*\} \, \overline{IS} \, \{\alpha F^* + (1-\alpha)J^*\}.$$

Since F^* dominates $\bar{H}^*$ and $\bar{H}^*$ does not dominate F^* by Monotonicity, $\alpha F^* + (1-\alpha)J^*$ dominates $\alpha\bar{H}^* + (1-\alpha)J^*$ and $\alpha\bar{H}^* + (1-\alpha)J^*$ does not dominate $\alpha F^* + (1-\alpha)J^*$. This contradicts Continuity, so $\bar{f}_Q \in \hat{c}(\{\bar{f}_Q\})$. □

If $Q \in \mathbb{P}^*$ and $Q \gg Q'$, then Lemma 13 implies that $Q' \in \mathbb{P}^*$.

Lemma 14. *Suppose $F \in \hat{c}(B)$. If $\{G\} \subset B$, then $F \succeq G$.*

Proof. If $G \notin \mathcal{H}$, then $F \succeq G$. If $G \in \mathcal{H}$, then pick $Q \in \mathbb{P}^*$ so that $G \in \mathcal{F}_Q^\Omega$. There are two cases.

First, suppose $u(G^*(\omega)) > u(x)$ for some $x \in X$ and every ω. Set $\bar{G} = (\bar{G^*})_Q$. By monotonicity $F \in \hat{c}(B \cup \{\bar{G}\})$. By Lemma 13, $\bar{G} \in \hat{c}(\{\bar{G}\})$, which implies that $\{F\} \, \overline{IS} \, \{\bar{G}\}$. Since $G^* = \bar{G}^*$ by construction, it follows that $F \succeq G$.

Now, suppose $u(G^*(\omega')) = \min_{x\in X} u(x)$ for at least one ω'. Consider $F' = \frac{1}{2}F + \frac{1}{2}x^*$ and $G' = \frac{1}{2}G + \frac{1}{2}x^*$ and $\bar{G}' = (\bar{G'^*})_Q$. Now, $u(G'^*(\omega)) > \min_{x\in X} u(x)$. Apply the above argument to get $\{F'\}\,\overline{IS}\,\{G'\}$ and $(\frac{1}{2}F + \frac{1}{2}x^*)^* \hat{\succeq} (\frac{1}{2}G + \frac{1}{2}x^*)^*$. Lemma 11 gives that $F^* \succeq G^*$, so $F \succeq G$. □

Lemma 15. *There is a finitely additive probability measure on Σ, $\pi(\cdot)$, that assigns positive probability to every $E \in P$ so that for any $f, g \in \mathcal{F}$, $f \succ g$ implies $\int u \circ f d\pi > \int u \circ g d\pi$ and $f \sim g$ implies $\int u \circ f d\pi = \int u \circ g d\pi$.*

Proof. Let $\mathcal{F}' \subset \mathcal{F}$ be the acts that are $\sigma(P)$ measurable.

Claim 1. For any $f \in \mathcal{F}$, there is an $f' \in \mathcal{F}'$ so that $f' \sim f$.

Proof. First, I show that for any act f and any $E \in P$, there is an act g so that $g \sim f$ and g is constant on E and agrees with f on E^c. Pick any $f \in \mathcal{F}$ and any $E \in P$. Let $\bar{x} = \arg\max_{\omega\in E} u(f(\omega))$, $\underline{x} = \arg\min_{\omega\in E} u(f(\omega))$, $\bar{g} = \bar{x}Ef$ and $\underline{g} = \underline{x}Ef$. For every $\alpha \in [0,1]$, define

$$B_\alpha = \{f, \alpha\bar{g} + (1-\alpha)\underline{g}\}.$$

By Subjective Consequentialism and because $c(B|\cdot)$ must be P measurable, there is at least one h in every B_α so that $h \in c(B_\alpha|\omega)$ for all $\omega \in \Omega$.

Fix $\omega \in E$. By Monotonicity and INRA,

$$\alpha\bar{g} + (1-\alpha)\underline{g} \in c(B_\alpha|\omega) \,\&\, \beta > \alpha \implies \beta\bar{g} + (1-\beta)\underline{g} \in c(B_\beta|\omega)$$

and conversely

$$\alpha\bar{g} + (1-\alpha)\underline{g} \notin c(B_\alpha|\omega) \,\&\, \beta < \alpha \implies \beta\bar{g} + (1-\beta)\underline{g} \notin c(B_\beta|\omega).$$

Using the above and that $\bar{g} \in c(B_1|\omega)$, there is an $\bar{\alpha}$ so that $\alpha > \bar{\alpha}$ implies that $\{\alpha\bar{g} + (1-\alpha)\underline{g}\}\,IS\,\{f\}$ and $\alpha < \bar{\alpha}$ implies that $\{f\}\,IS\,\{\alpha\bar{g} + (1-\alpha)\underline{g}\}\,IS\,\{f\}$. Conclude that $\bar{\alpha}\bar{g} + (1-\bar{\alpha})\underline{g} \sim f$. Since f and E were arbitrary, this establishes the first step.

Now, label $P = \{E_1, ..., E_n\}$. Fix f. By the above, there is an f_1 so that $f \sim f_1$ and $f_1$is constant on E_1 and agrees with f on E_1^c. For $i = 2, ..., n$, the above shows that there is f_{i+1} so that $f_i \sim f_{i+1}$ and $f_{i+1}$is constant on E_{i+1} and agrees with f_i on E_{i+1}^c. By construction, f_n is $\sigma(P)$-measurable, and $f \sim f_1 \sim f_2 \sim ... \sim f_n \implies f \sim f_n$ by Lemma 7, so $f_n \in \mathcal{F}'$ and $f_n \sim f$, establishing the claim. □

Moreover, $\mathcal{F}'$ is finite dimensional. Restricted to $\mathcal{F}'$, $\succeq$ satisfies reflexivity, transitivity and independence by Lemmas 7 and 4. Lemma 9 implies that if $\lambda f + (1-\lambda)g \succeq g$ for every $\lambda \in (0,1)$, then it is not the case that $g \succ f$. Applying Aumann [1962, Thm. A] yields the existence of a mixture linear $U(\cdot)$ so that $f \succ g$ implies $U(f) > U(g)$ and $f \sim g$ implies $U(f) = U(g)$. By Monotonicity using choice from problems in the set $\{\{xEy, y\} : E \in P\}$

where $u(x) > u(y)$ and Lemma 2, there is a $\pi(\cdot)$ with the desired properties, an $\alpha > 0$ and a $\beta \in \mathbb{R}$ so that $U(\cdot) = \int \alpha u \circ f d\pi(\cdot) + \beta$, the desired result. WLOG, take $\alpha = 1$ and $\beta = 0$. $\square$

Lemma 16. $c(B|\omega) = \arg\max_{f\in B} \int u \circ f d\pi(\cdot|\mathcal{A}(B)(\omega))$.

Proof. Fix $B \in K(B)$ and set $E = \mathcal{A}(B)(\omega)$. First, suppose $f \in c(B|\omega)$ and set $F \in \hat{c}(B)$ so that $F(\omega) = f$ for all $\omega \in E$. Specifically, $F \succeq G$ whenever there is a $g \in B$ so that $G(\omega) = F(\omega)$ for every $\omega \notin E$ and $G(\omega) = g$ for every $\omega \in E$, so $f \in \arg\max_{g\in B} \int u \circ g d\pi(\cdot|E)$, implying that $c(B|\omega) \subset \arg\max_{f\in B} \int u \circ f d\pi(\cdot|E)$.

Now, suppose that $\int u \circ g d\pi(\cdot|E) \in \arg\max_{g\in B} \int u \circ g d\pi(\cdot|E)$. Set $x \in X$ so that $u(x) < \min_{\{f(\omega): f\in B \, and \, \omega\in\Omega\}} u(f(\omega))$. Define $\hat{F}(\omega) = F(\omega)\mathcal{A}(B)(\omega)x$ for all ω and $\hat{G}(\omega) = F(\omega)\mathcal{A}(B)(\omega)x$ for all $\omega \notin E$ and $\hat{G}(\omega) = gEx$ for $\omega \in E$ and $B' = B \cup \{\hat{F}\} \cup \{\hat{G}\}$ and note that Subjective Consequentialism and Monotonicity imply $\hat{F} \in \hat{c}(B')$. Then take $B'' = \{\hat{F}\} \cup \{\hat{G}\}$ and by INRA, $\hat{F} \in \hat{c}(B'')$. Now, take $y \in X$ so that $u(y) > \int u \circ f d\pi(\cdot|E)$ define $B_n = (B''\backslash\{gEx\}) \cup \{\frac{1}{n}y + \frac{n-1}{n}gEx\}$. By Lemma 15, $G_n \succ F'$ for all F' so that $\{F'\} \subset B$ where $G_n(\omega) = \hat{F}(\omega)\mathcal{A}(B)(\omega)x$ for all $\omega \notin E$ and $G_n(\omega) = \frac{1}{n}y + \frac{n-1}{n}gEx$ for all $\omega \in E$. By Lemma 14, $G_n \in \hat{c}(B_n)$.[33]

By construction, $\mathcal{A}(B_n) = \mathcal{A}(B'')$ for all n. Since $\frac{1}{n}y + \frac{n-1}{n}g \in c(B_n|\omega)$ and $(\frac{1}{n}y + \frac{n-1}{n}g)Ex \to gEx$, it follows from Continuity that $gEx \in c(B''|\omega)$. By INRA, $c(B''|\omega) = c(B'|\omega) \cap B''$. Since $u \circ g \geq u \circ gEx$, $gEx \in c(B'|\omega) \implies g \in c(B'|\omega)$ by Monotonicity. By INRA, $c(B'|\omega) = c(B'|\omega) \cap B$, so $g \in c(B|\omega)$, completing the proof. $\square$

Set $\hat{P}(B) = \mathcal{A}(B)$. Lemma 15 and 14 give that

$$\hat{P}(B) \in \arg\max_{Q\in\mathbb{P}^*} \sum_{E'\in Q} \pi(E') \max_{f\in B} \int u \circ f d\pi(\cdot|E')$$

because if $F \in \hat{c}(B)$ then $F \succeq G$ for all $G \in \mathcal{H}$ and $\{G\} \subset B$ implies that $\int u \circ F^* d\pi \geq \int u \circ G^* d\pi$, implying that

$$\sum_{E'\in\hat{P}(B)} \pi(E') \int u(F(\omega)(\omega)) d\pi(\cdot|E') \geq \sum_{E'\in Q} \pi(E') \max_{g\in B} \int u \circ g d\pi(\cdot|E')$$

for any $Q \in \mathbb{P}^*$. Using that $\hat{P}(B) = \mathcal{A}(B)$, Lemma 16 gives that $c(B|\omega) = \arg\max_{f\in B} \int u \circ f d\pi(\cdot|\hat{P}(B)(\omega))$, completing the proof. $\square$

B.1.2. *Proof of Theorem 2.*

Proof. Suppose $c(\cdot)$ is represented by $(u(\cdot), \pi(\cdot), \mathbb{P}^*, \hat{P}(\cdot))$. First, I show that $c(\cdot)$ satisfies Continuity.

[33] These inequalities for $u(x)$ and $u(y)$ can be taken to be strict even if $u(\cdot)$ is bounded because the remainder relies only on properties of $\tilde{R}$. Mixing with B with a constant $z \in int(u(X))$ ensures that there exists such $x, y \in X$.

Lemma 17. *$c(\cdot)$ satisfies Continuity.*

Proof. Suppose that both $\{f\}\,\overline{IS}\,\{g\}$ and g weakly dominates f. Let $B_n \to \{f\}$ and $C_n \to \{g\}$ so that $B_n\,IS\,C_n$. If $B_n\,IS\,C_n$, $F_n \in c(B_n)$ and $G_n \in c(C_n)$, then $\int u\circ F_n^* d\pi \geq \int u\circ G_n^* d\pi$. To see this, let $B_1, ..., B_n$ be the sequence from the definition of indirectly selected. Then $F_i \in \hat{c}(B_i)$ and $\{F_{i+1}\} \subset B_i$ implies that $\int u \circ F_i^* d\pi \geq \int u \circ F_{i+1}^* d\pi$ by construction of $c(\cdot)$ for all i. Therefore, $\int u \circ F_n^* d\pi \geq \int u \circ G_n^* d\pi$.

Now, note that $F_n^* \to f$ since all components of $B_n \to f$. Similarly, $G_n^* \to g$. Therefore, since $u(\cdot)$ is continuous, $\int u \circ f d\pi \geq \int u \circ g d\pi$. Since g weakly dominates f, $\int u \circ g d\pi(\cdot|\hat{P}(\{f,g\})(\omega)) \geq \int u \circ f d\pi(\cdot|\hat{P}(\{f,g\})(\omega))$, without equality for some ω'. This implies that $\int u \circ g d\pi(\cdot) > \int u \circ f d\pi$, a contradiction. $\square$

For the second part, begin by defining a function $V : \{Q : P \gg Q\} \times K(\mathcal{F}) \to \mathbb{R}$ by

$$V(Q, B) = \sum_{E \in Q} \pi(E) \max_{f \in B} \int u \circ f d\pi(\cdot|E).$$

With this formulation, $\hat{P}(B) \in \arg\max_{Q \in \mathbb{P}^*} V(Q, B)$ for all B. By the maximum theorem, $V(Q, \cdot)$ is continuous and $\arg\max V(\cdot, B)$ is upper-hemi continuous.

If $u(\cdot)$ is constant, then set $K = K(\mathcal{F})$. Both INRA and ACI are satisfied because $c(B|\omega) = B$ for every B and ω. Clearly, K is open and dense in $K(\mathcal{F})$.

If not, then define K by

$$K = \{B \in K(\mathcal{F}) : \exists Q \in \mathbb{P}^* \, s.t. \, V(Q, B) > V(Q', B) \, \forall Q' \in \mathbb{P}^* \backslash \{Q\}\}.$$

I proceed by showing that $cl(K) = K(\mathcal{F})$ and then that K is open.

Lemma 18. $cl(K) = K(\mathcal{F})$

Proof. Pick any $B \in K(\mathcal{F})$ and any $\epsilon > 0$.

Fix $x \in X$ so that $u(x) \in int(u(X))$. Define $B' \in K(\mathcal{F})$ by $\alpha B + (1-\alpha)\{x\}$ for α close enough to 1 so that $d(B', B) < \frac{\epsilon}{3}$.

Pick a $Q \in \mathbb{P}^*$ so that $Q \gg \mathcal{A}(B')$ and $Q' \gg Q$ for $Q' \in \mathbb{P}^*$ implies that $Q = Q'$. Label $Q = \{E_1, ..., E_n\}$ and pick $f_1, ..., f_n$ so that $f_i \in c(B'|\omega)$ for some $\omega \in E_i$. Define f^* so that

$$f^*(\omega) = f_i(\omega)$$

whenever $\omega \in E_i$ and f^{**} so that $u(f^{**}) = u(f^*) + k$ for some $k > 0$. Since $u \circ f^* \in int(u(X)^\Omega)$ by construction of B', such a k exists.

Now, define f_α^i for every $\alpha \in [0,1]$ by $f_\alpha^i = (\alpha f_i + (1-\alpha) f^{**}) E_i f_i$ for every $i \in \{1, ..., n\}$. For α close enough to 1, $d(f_\alpha^i, f_i) < \frac{\epsilon}{3}$. Therefore, for α^* sufficiently high, note that $d(B'', B') < \frac{\epsilon}{3}$ where

$$B'' = B' \cup \{f_{\alpha^*}^i\}_{i=1}^n.$$

Conclude that $d(B'', B) \leq d(B'', B') + d(B', B) < \frac{2\epsilon}{3} < \epsilon$. Further, $V(Q, B'') > V(Q', B'')$ for all $Q' \in \mathbb{P}^*$ so that $Q' \neq Q$. Therefore, $B'' \in K$. Since B and ϵ are arbitrary, there is a $B'' \in K$ arbitrarily close to any $B \in K(\mathcal{F})$. Therefore, $cl(K) = K(\mathcal{F})$. □

Lemma 19. *K is open.*

Proof. Let $K^c = K(\mathcal{F}) \backslash K$. K is open if and only if K^c is closed. Because $K(\mathcal{F})$ is a metric space and thus first countable, it is sufficient to only show sequentially closed.

Pick $(B_n)_{n=1}^{\infty} \subset K^c$ and suppose that $B_n \to B$. Because $\mathbb{P}^*$ is finite, there are $Q \neq Q' \in \mathbb{P}^*$ and a sub-sequence $(B_{n_k})_{k=1}^{\infty}$ so that $Q, Q' \in \arg\max_{Q \in \mathbb{P}^*} V(Q, B_{n_k})$ for all $Q'' \in \mathbb{P}^*$. Because $\arg\max_{Q \in \mathbb{P}^*} V(Q, \cdot)$ is upper hemi-continuous and $B_{n_k} \to B$, $Q, Q' \in \arg\max_{Q \in \mathbb{P}^*} V(Q, B)$. Conclude that $B \in K^c$, so K^c is closed and K is open. □

Let $>$ be a linear order on $\mathbb{P}^{**}$ and set

$$\hat{Q}(B) = \max_{>} \arg\max_{Q \in \mathbb{P}^{**}} V(Q, B).$$

Define the conditional choice correspondence $c'(\cdot)$ by

$$c'(B|\omega) = \arg\max_{f \in B} \int u \circ f d\pi(\cdot|\hat{Q}(B)(\omega))$$

for every $B \in K(\mathcal{F})$. Clearly $c'(\cdot)$ has an optimal inattention representation and for every $B \in K$, $c'(B|\omega) = c(B|\omega)$ for every $\omega \in \Omega$.

Lemma 20. *$c'(\cdot)$ satisfies ACI.*

Proof. If $\mathcal{A}(B) \gg \mathcal{A}(C)$, then $\hat{Q}(B) \in \arg\max_{Q \in \mathbb{P}^{**}} V(Q, C)$. Therefore, $\hat{Q}(B) \in \arg\max_{Q \in \mathbb{P}^{**}} V(Q, \alpha B + (1-\alpha)C)$. Further, if $Q' \in \arg\max_{Q \in \mathbb{P}^{**}} V(Q, \alpha B + (1-\alpha)C)$, then $Q' \in \arg\max_{Q \in \mathbb{P}^{**}} V(Q, B) \cap \arg\max_{Q \in \mathbb{P}^{**}} V(Q, C)$. Since $\hat{Q}(B) > Q'$ for every $Q \in \arg\max_{Q \in \mathbb{P}^{**}} V(Q, B)$, it follows immediately that $\hat{Q}(\alpha B + (1-\alpha)C) = \hat{Q}(B)$. The conclusion follows immediately. □

Lemma 21. *$c'(\cdot)$ satisfies INRA.*

Proof. Suppose that $A \subset B$ and $c'(B|\omega) \cap A \neq \emptyset$ for all ω. Note that

$$\arg\max_{Q \in \mathbb{P}^{**}} V(Q, A) \subset \arg\max_{Q \in \mathbb{P}^{**}} V(Q, B)$$

and since $\hat{Q}(B) > Q'$ for all $Q' \in \arg\max_{Q\in\mathbb{P}^{**}} V(Q,B)$, $\hat{Q}(B) > Q'$ for all $Q \in \arg\max_{Q\in\mathbb{P}^{**}} V(Q,A)$ so $\hat{Q}(A) = \hat{Q}(B)$. Since

$$\begin{aligned}
c'(B|\omega) \cap A &= [\arg\max_{f\in B} \int u \circ f d\pi(\cdot|\hat{Q}(B)(\omega))] \cap A \neq \emptyset \\
&= \arg\max_{f\in A} \int u \circ f d\pi(\cdot|\hat{Q}(B)(\omega)) \\
&= \arg\max_{f\in A} \int u \circ f d\pi(\cdot|\hat{Q}(A)(\omega)) \\
&= c'(A|\omega)
\end{aligned}$$

it follows that $c'(B|\omega) \cap A = c'(A|\omega)$. □

Since K is open and $cl(K) = K(\mathcal{F})$, the Theorem follows immediately. □

B.1.3. *Proof of Corollary 2:*

Proof. (iii) clearly implies either (i) or (ii).

[(i) implies (iii)] Suppose $c(\cdot)$ satisfies Independence and has optimal inattention, with the canonical representation $\hat{P}(B) = \mathcal{A}(B)$. Let Q be coarsest common refinement of $\{\hat{P}(B)\}_{B\in K(\mathcal{F})}$. Claim that $c(B|\omega) = \arg\max_{f\in B} \int u \circ f d\pi(\cdot|Q(\omega))$ for every B. If not, there is a B' and an ω so that $c(B'|\omega) \neq \arg\max_{f\in B} \int u \circ f d\pi(\cdot|Q(\omega))$. Since $c(\cdot)$ has inattention, $c(B'|\omega) = \arg\max_{f\in B'} \int u \circ f d\pi(\cdot|\hat{P}(B')(\omega))$. There is a finite collection $\{B_1, ..., B_n\} \subset K(\mathcal{F})$ so that $[\cap_{i=1}^n \hat{P}(B_i)(\omega)] \cap \hat{P}(B')(\omega) = Q(\omega)$ and $c(B_i|\omega) \neq c(B_j|\omega)$ for all $i \neq j$ (perhaps after mixing B_i with a singleton). Set $B^* = \prod_{i=1}^n \frac{1}{n} B_i$ and note that $\mathcal{A}(B^*)(\omega) = Q(\omega)$. Since we can take $\hat{P}(B^*) = \mathcal{A}(B^*)$, it follows that

$$c(B^*|\omega) = \arg\max_{f\in B^*} \int u \circ f d\pi(\cdot|Q(\omega)).$$

Now, since $c(\frac{1}{2}B^* + \frac{1}{2}B'|\omega) = \frac{1}{2}c(B^*|\omega) + \frac{1}{2}c(B'|\omega)$, $\mathcal{A}(\frac{1}{2}B^* + \frac{1}{2}B') \gg \mathcal{A}(B^*)$. By construction, $\mathcal{A}(\frac{1}{2}B^* + \frac{1}{2}B')(\omega) = Q(\omega)$, so

$$\begin{aligned}
c(\frac{1}{2}B^* + \frac{1}{2}B'|\omega) &= \arg\max_{f\in \frac{1}{2}B^* + \frac{1}{2}B'} \int u \circ f d\pi(\cdot|Q(\omega)) \\
&= \frac{1}{2} \arg\max_{f\in B^*} \int u \circ f d\pi(\cdot|Q(\omega)) \\
&\quad + \frac{1}{2} \arg\max_{f\in B'} \int u \circ f d\pi(\cdot|Q(\omega)) \\
&\neq \frac{1}{2} \arg\max_{f\in B^*} \int u \circ f d\pi(\cdot|Q(\omega)) \\
&\quad + \frac{1}{2} \arg\max_{f\in B'} \int u \circ f d\pi(\cdot|\hat{P}(B')(\omega)) \\
&= \frac{1}{2}c(B^*|\omega) + \frac{1}{2}c(B'|\omega),
\end{aligned}$$

which contradicts Independence.

[(ii) implies (iii)] Suppose that $c(\cdot|\omega)$ satisfies WARP and has optimal inattention. Since $K(\mathcal{F})$ includes all two and three element subsets, there is a complete and transitive binary relation $\succeq_\omega$ so that This binary relation is equal to the revealed preference relation. Let Q be any maximal element of $\mathbb{P}^*$ according to $\gg$. I show that $f \sim_\omega fQ(\omega)y$ for any f and an arbitrarily bad y. Therefore, $\hat{P}(B)(\omega) \subset Q(\omega)$ for every B and ω and consequently $\hat{P}(B)(\omega) = Q(\omega)$ represents choices.

Fix $f \in \mathcal{F}$, $\omega^* \in \Omega$ and $x, y \in X$ so that $u(x) > u(f(\omega)) > u(y)$ for every $\omega \in \Omega$. Define g_ω by $g_\omega(\omega') = \frac{1}{2}x + f(\omega)$ if $\omega' \in Q(\omega)$ and $g_\omega(\omega') = y$ otherwise. Consider the problem $B = \{g_\omega : \omega \notin Q(\omega^*)\} \cup \{f, fQ(\omega^*)y\}$. Clearly, $\hat{P}(B) = Q$ (otherwise, this is not optimal) and also $f, fQ(\omega^*)y \in c(B|\omega^*)$. Conclude that $f \sim_\omega fQ(\omega^*)y$. □

B.1.4. *Proof of Corollary 1:*

Proof. That (ii) implies (i) is trivial, so suppose $c(\cdot)$ has optimal inattention and satisfies Consequentialism.

Set $y, x \in X$ so that $u(x) > u(y)$ and consider $B = \{xEy : E \in P\} \cup \{x\}$. Clearly $x \in c(B|\omega)\forall\omega$. For any ω, note that $xP(\omega)y \in B$ and $xP(\omega)y(\omega') = x(\omega')$ for every $\omega' \in P(\omega)$. By Consequentialism, $xP(\omega)y \in c(B|\omega)$. However, if $\omega' \notin P(\omega)$, then monotonicity implies that $xP(\omega)y \notin c(B|\omega')$. Therefore, $\mathcal{A}(B) = P$, which implies that $P \in \mathbb{P}^*$. Since there is no $Q \in \mathbb{P}^* \backslash \{P\}$ finer than P or coarser than P, $\{P\} = \mathbb{P}^*$, implying that $c(\cdot)$ is Bayesian. □

B.2. **Proof from Section 4.**

B.2.1. *Proof of Theorem 3.*

Proof. [i,iv]Affine-uniqueness of $u(\cdot)$ is standard, and canonical uniqueness of $\hat{P}(\cdot)$ is trivial.

[iii] Suppose $\mathbb{P}^*_1, \mathbb{P}^*_2$ both represent $c(\cdot)$. Since $c(\cdot)$ is non-degenerate, there are $x, y \in X$ so that $u(x) > u(y)$. For any $Q \in \mathbb{P}^*_1$, define $B_Q = \{xEy : E \in Q\}$. Clearly, $xQ(\omega)y \in c(B_Q|\omega)$ for every ω, so $\hat{P}(B_Q) \gg Q$. Since $\mathbb{P}^*_2$ represents $c(\cdot)$, $Q \in \mathbb{P}^*_2$. Reversing the role of $\mathbb{P}^*_1$ and $\mathbb{P}^*_2$ give the converse, so they must be equal. T

[ii] Suppose that both π_1 and π_2 represent $c(\cdot)$. By (iii), let $\mathbb{P}^* = \mathbb{P}^*_1 = \mathbb{P}^*_2$ and $\mathbb{P}^{**} = \{Q \in \mathbb{P}^* : Q' \gg Q \,\&\, Q' \in \mathbb{P}^* \implies Q = Q'\}$ be the set of the finest subjective information partitions in $\mathbb{P}^*$. Write $V(B) = \arg\max_{Q \in \mathbb{P}^*} \sum_{E \in Q} \pi(E)[\max_{f \in B} \int u \circ f d\pi(\cdot|Q(\omega))]$ for any B. Let $\mathbb{Q}$ be the set of minimal isolatable events for $\mathbb{P}^*$.

Lemma 22. *E is a isolatable event for $\mathbb{P}^*$ if and only if any $Q_1, Q_2 \in \mathbb{P}^{**}$ are such that $Q_1 \gg \{E, E^c\}$, $Q_2 \gg \{E, E^c\}$, and there is a $Q_3 \in \mathbb{P}^{**}$ so that*

$$Q_3 \gg \{E' \in Q_1 : E' \subset E\} \cup \{E' \in Q_2 : E' \subset E^c\}.$$

Proof. [$\Longrightarrow$] Suppose that E is an isolatable event for $\mathbb{P}^*$. Then pick any $Q_1, Q_2 \in \mathbb{P}^{**}$ and $x, y, z \in X$ so that $u(x) > u(y) > u(z)$; let $B_i = \{xE'y : E' \in Q_i\}$ for $i = 1, 2$.

Suppose $Q_i \not\gg \{E, E^c\}$. Then $\exists F \in Q_1$ s.t. $F \cap E \neq \emptyset$ and $F \cap E^c \neq \emptyset$. Fix $z \in X$ so that $u(x') = u(x) + \epsilon$ where $\epsilon > 0$ and $u(y) > u(z)$. Consider $B = B_i$ and $B' = B_i$, noting that $B_{E,z}B' = B_{E^c,z}B'$ by construction, and that Q_i is the only element of $V(B)$. If $\hat{P}(B_{E,z}B') \neq Q_i$, then E is not an isolatable event since there is no partition in $\mathbb{P}^*$ finer than Q_i except Q_i. However, if $\hat{P}(B_E B') = Q_i$, then either for any $\omega \in F \cap E$, $[xFy]Ex \notin c(B_{E,z}B'|\omega)$ or $[xFy]E^c z \notin c(B_{E^c,z}B'|\omega)$, implying that $xFy \notin c(B_E B'|\omega)$.[34] This contradicts that E is an isolatable event for $\mathbb{P}^*$.

Now, consider $B = B_1$ and $B' = B_2$. By construction, $B_{E,z}B' = B'_{E^c,z}B = B''$. Consequently, $[xQ_1(\omega)y]Ez \in c(B''|\omega)$ for any $\omega \in E$ and $[xQ_2(\omega)y]E^c z \in c(B''|\omega)$ for any $\omega \in E^c$, implying that $\hat{P}(B'') \gg \{E' \in Q_1 : E' \subset E\} \cup \{E' \in Q_2 : E' \subset E^c\}$. Therefore, there is some $Q_3 \in \mathbb{P}^{**}$ satisfying the desired property.

[$\Longleftarrow$] Suppose that any $Q_1, Q_2 \in \mathbb{P}^{**}$ are such that $Q_1 \gg \{E, E^c\}$, $Q_2 \gg \{E, E^c\}$, and there is a $Q_3 \in \mathbb{P}^{**}$ so that

$$Q_3 \gg \{E' \in Q_1 : E' \subset E\} \cup \{E' \in Q_2 : E' \subset E^c\}.$$

Consider any B so that $V(B)$ is singleton, and pick any B', labeling $\hat{P}(B) = Q_1$ and $\hat{P}(B') = Q$. If there is no z so that $z \notin c(\{f(\omega), z\}|\omega)$ for any $f \in B \cup B'$, then the condition is arbitrarily satisfied so suppose such a z exists and consider $B_{E,z}B'$.

I claim that

$$\{Q_1\} = \arg\max_{Q \in \mathbb{P}^*} \sum_{E' \in Q \& E' \subset E} \pi(E') \max_{f \in B} \int u \circ f d\pi(\cdot|E').$$

Suppose not, so $Q' \neq Q_1$ is in the argmax above. $Q'' = \{E' \cap E : E' \in Q'\} \cup \{E' \cap E^c : E' \in Q_1\}$ gives at least as high utility as Q_1 when facing, and by assumption, there is a $Q^* \in \mathbb{P}^{**}$ so that $Q^* \gg Q''$, contradicting that $V(B)$ is a singleton.

Consider now $B_{E,z}B'$. For the sake of contradiction, suppose that $\hat{P}(B_{E,z}B')$ cannot be written as $\{E' \in Q_1 : E' \subset E\} \cup \{E' \in Q_2 : E' \subset E^c\}$ for some $Q_2 \in \mathbb{P}^*$. Let $\hat{P}(B_{E,z}B') = Q'$ and take any $Q \gg Q'$ so that $Q \in \mathbb{P}^{**}$. There is a $Q_3 \gg \{E' \in Q_1 : E' \subset E\} \cup \{E' \in Q : E' \subset E^c\}$ that is in $\mathbb{P}^{**}$. Further,

$$\sum_{E' \in Q_3 \& E' \subset E^c} \pi(E') \max_{f \in B_{E,z}B'} \int u \circ f d\pi(\cdot|E') \geq \sum_{E' \in Q' \& E' \subset E^c} \pi(E') \max_{f \in B_{E,z}B'} \int u \circ f d\pi(\cdot|E')$$

because Q_3 is finer than Q' when restricted to E^c, and

$$\sum_{E' \in Q_3 \& E' \subset E} \pi(E') \max_{f \in B_{E,z}B'} \int u \circ f d\pi(\cdot|E') \geq \sum_{E' \in Q' \& E' \subset E} \pi(E') \max_{f \in B_{E,z}B'} \int u \circ f d\pi(\cdot|E')$$

[34] This procedure must be modified slightly when $\pi_1(F \cap E) = \pi_1(F \cap E^c)$ so that the bet on F gives slightly higher utility on $F \cap E^c$, but arguments otherwise extend.

by the above claim because for any $E' \subset E$,

$$\max_{f \in B_{E,z}B'} \int u \circ f d\pi(\cdot|E') \leq \max_{f \in B} \int u \circ f d\pi(\cdot|E'),$$

where equality holds whenever $E' \in Q_1$. But this contradicts the assumption that $V(B)$ is a singleton: since $\{E' \in Q_3 : E' \subset E\} \neq \{E' \in Q_1 : E' \subset E\}$, there is a $Q_4 \gg \{E' \in Q_3 : E' \subset E\} \cup \{E' \in Q_1 : E' \subset E^c\}$, $Q_4 \neq Q_1$ and $Q_4 \in V(B)$ by construction. Conclude that $\hat{P}(B_{E,z}B') = \{E' \in Q_1 : E' \subset E\} \cup \{E' \in Q : E' \subset E^c\}$ for some $Q \in \mathbb{P}^*$.

Now, fix any $\omega \in E$ and suppose $f \in c(B|\omega)$. Since $c(B|\omega) = \arg\max_{g \in B} \int u \circ g d\pi(\cdot|Q_1(\omega))$ and $c(B_{E,z}B'|\omega) = \arg\max_{f \in B_{E,z}B'} \int u \circ g d\pi(\cdot|Q_1(\omega))$ and $\int u \circ f d\pi(\cdot|Q_1(\omega)) = \int u \circ fEz d\pi(\cdot|Q_1(\omega))$, it follows that $fEz \in c(B|\omega)$. Similar arguments show that the same property holds for E^c. Conclude that E is an isolatable event for $\mathbb{P}^*$. $\square$

This implies that $\mathbb{Q} \ll P$.

Lemma 23. *If E is an isolatable event for $\mathbb{P}^*$ and F is a isolatable event $\mathbb{P}^*$, then $E \cap F$ is an isolatable event $\mathbb{P}^*$.*

Proof. Suppose E and F and isolatable events. Fix any $Q_1, Q_2 \in \mathbb{P}^{**}$. Since E is an isolatable event, there is a $Q_3 \gg \{E' \in Q_1 : E' \subset E\} \cup \{E' \in Q_2 : E' \subset E^c\} \in \mathbb{P}^{**}$. Since $Q_2, Q_3 \in \mathbb{P}^{**}$ and F is an isolatable event, there is a $Q_4 \gg \{E' \in Q_3 : E' \subset F\} \cup \{E' \in Q_2 : E' \subset F^c\} \in \mathbb{P}^{**}$. By construction, $Q_4 \gg \{Q_1 : E' \subset E \cap F\} \cup \{E' \in Q_2 : E' \subset (E \cap F)^c\}$. Since Q_1, Q_2 were arbitrary, this holds for any Q_1, Q_2. By Lemma 22, $E \cap F$ is a sub-problem. $\square$

Lemma 24. *If $Q \in \mathbb{P}^{**}$ and $E \in Q$, then $\pi_1(\cdot|E) = \pi_2(\cdot|E)$.*

Proof. Take any $Q \in \mathbb{P}^{**}$ and any $E \in Q$. Set $x, y \in int(u(X))$ so that $u(x) > u(y)$. WLOG, identify $x = u(x)$ and $y = u(y)$. Set ϵ so that

$$\pi_1(E)2\epsilon + (y - x)\pi_1(E') < 0$$

for all $E' \in P$. Take any simple $f, g \in \mathbb{R}^E$ so that $\int f d\pi_1(\cdot|E) \geq \int g d\pi_1(\cdot|E)$. There is an $\alpha \in (0, 1]$ so that $\alpha f + (1 - \alpha)x, \alpha g + (1 - \alpha)x \in [x - \epsilon, x + \epsilon]$. There are acts $f', g' \in \mathcal{F}$ so that $u(f')(\omega) = \alpha f(\omega) + (1 - \alpha)x$ for all $\omega \in E$ and $u(f')(\omega) = x$ otherwise and $u(g')(\omega) = \alpha g(\omega) + (1 - \alpha)x$ for all $\omega \in E$ and $u(g')(\omega) = x$ otherwise. Define $B = \{f'Q(\omega)y : \omega \in \Omega\} \cup \{g'Q(\omega)y : \omega \in \Omega\}$.

Claim that $\hat{P}(B) = Q$. If not, then there is a Q' so that $\hat{P}(B) = Q'$, so let $H \in \hat{c}(B)$ and consider H^*. It must be that $\int H^* d\pi_1 \geq \int f d\pi_1$. Since Q' is not finer than Q, there must be some $E' \in P$ so that $H^*(\omega) = y$ for all $\omega \in E'$. Let $E'' = \{\omega \in \Omega : u(H^*(\omega)) \geq u(f(\omega))\}$. By construction, $E'' \subset E$. Further, $2\epsilon \geq u(H^*(\omega)) - u(f(\omega))$. Therefore $\pi_1(E)2\epsilon + (y - x)\pi_1(E') + \int u \circ f' d\pi_1 \geq \int H^* d\pi_1$. However, $\int u \circ f' d\pi_1 > \pi_1(E)2\epsilon + (y - x)\pi_1(E') + \int u \circ f' d\pi_1$, contradicting that $\hat{P}(B) = Q'$.

Since $\int f d\pi_1(\cdot|E) \geq \int g d\pi_1(\cdot|E)$, $f'Q(\omega)x \in c(B|\omega)$. Since $\pi_2(\cdot)$ also represents $c(\cdot)$, $\int f d\pi_2(\cdot|E) \geq \int g d\pi_2(\cdot|E)$. Since f and g are arbitrary, $\int f d\pi_1(\cdot|E) \geq \int g d\pi_1(\cdot|E) \iff \int f d\pi_2(\cdot|E) \geq \int g d\pi_2(\cdot|E)$ for any f and g. Therefore, $\pi_1(\cdot|E) = \pi_2(\cdot|E)$ for every E. □

Lemma 25. *If $\pi_1(\cdot|E) = \pi_2(\cdot|E)$, $\pi_1(\cdot|E') = \pi_2(\cdot|E')$ and $E \cap E' \neq \emptyset$ then $\pi_1(\cdot|E \cup E') = \pi_2(\cdot|E \cup E')$.*

Proof. Suppose that $\pi_1(\cdot|E) = \pi_2(\cdot|E)$, $\pi_1(\cdot|E') = \pi_2(\cdot|E')$ and $E \cap E' \neq \emptyset$. Note that $\pi_1(F|E) = \pi_2(F|E)$ and $\pi_1(F|E') = \pi_2(F|E')$. Using Bayes' rule on the event $E \cap E'$, it follows that

$$\frac{\pi_1(E|E \cup E')}{\pi_1(E'|E \cup E')} = \frac{\pi_2(E|E \cup E')}{\pi_2(E'|E \cup E')}.$$

For any $F \in \Sigma$, it holds that

$$\pi_1(F|E \cup E') = \pi_1(E|E \cup E')(\pi_1(F|E) - \pi_1(F \cap E'|E)) + \pi_1(E'|E \cup E')\pi_1(F|E')$$

and that

$$\pi_2(F|E \cup E') = \pi_2(E|E \cup E')(\pi_2(F|E) - \pi_2(F \cap E'|E)) + \pi_2(E'|E \cup E')\pi_2(F|E').$$

Because $\pi_1(E \cup E'|E \cup E') = \pi_2(E \cup E'|E \cup E') = 1$, conclude that $\pi_2(\cdot|E \cup E') = \pi_1(\cdot|E \cup E')$. □

Let Q^* be the finest common coarsening of $\mathbb{P}^{**}$. By successive applications of Lemma 25, we have that $\pi_1(\cdot|E') = \pi_2(\cdot|E')$ for all $E' \in Q^*$.

Lemma 26. *For any $E \in \mathbb{Q}$, $\pi_1(\cdot|E) = \pi_2(\cdot|E)$.*

Proof. To save on notation, write $\pi_1 = \pi_1(\cdot|E)$ and $\pi_2 = \pi_2(\cdot|E)$ and assume it is understood that each event E' is contained in E. Label the events in Q^* that are contained in E as $E_1, E_2, ..., E_n$. If $n = 1$, then we are done by the above, so assume $n \geq 2$.

Consider E_1 and E_2. By construction, there must be $Q_1, Q_2 \in \mathbb{P}^{**}$ so that $E' \subset E_1$ is in Q_1 but not Q_2, $E'' \subset E_2$ is in Q_2 but not in Q_1, and there is no $Q_3 \gg \{E' \in Q_1 : E' \subset E_1\} \cup \{E' \in Q_2 : E' \subset E_1^c\}$. Fix $x, y \in X$ so that $u(x) > u(y)$ and $u(x), u(y) \in int(u(X))$. Define $B_1 = \{xFy : F \in Q_1\}$ and $B_2 = \{xFy : F \in Q_2\}$. Let $B_1^\epsilon = (B_1 \cup \{x'E'y\}) \backslash \{xE'y\}$ and $B_2^\epsilon = (B_1 \cup \{x'E''y\}) \backslash \{xE''y\}$ where $u(x') = u(x) + \epsilon$. For ϵ, ϵ' small enough but positive,

$$\hat{P}(B_1^\epsilon \cup B_2^{\epsilon'}) = \begin{cases} Q_1 & if\ \epsilon\pi_1(E') > \epsilon'\pi_1(E'') \\ Q_2 & if\ \epsilon\pi_1(E') < \epsilon'\pi_1(E''). \end{cases}$$

Therefore, there exists a $k = \frac{\pi_1(E'')}{\pi_1(E')}$ so that $\frac{\epsilon}{\epsilon'} > k$ implies $\hat{P}(B_1^\epsilon \cup B_2^{\epsilon'}) = Q_1$ and $\frac{\epsilon}{\epsilon'} < k$ implies $\hat{P}(B_1^\epsilon \cup B_2^{\epsilon'}) = Q_1$. Since π_2 also represents $c(\cdot)$, the same cutoff must hold for π_2.

Therefore,

$$\frac{\pi_2(E'')}{\pi_2(E')} = \frac{\pi_1(E'')}{\pi_1(E')}.$$

By Lemma 25 and Bayes rule,

$$\frac{\pi_2(E_1)}{\pi_2(E_2)} = \frac{\pi_1(E_1)}{\pi_1(E_2)}.$$

By replacing E_1 with E_i and E_2 with E_{i+1}, we must have $\frac{\pi_2(E_i)}{\pi_2(E_{i+1})} = \frac{\pi_1(E_i)}{\pi_1(E_{i+1})}$. Since $\sum_{i=1}^n \pi_1(E_i) = \sum_{i=1}^n \pi_2(E_i) = 1$, we must have that $\pi_1(E_i) = \pi_2(E_i)$ for all i. $\square$

Conclude that $\pi_1(\cdot|E) = \pi_2(\cdot|E)$ whenever E is a minimal isolatable event for $\mathbb{P}_1^*$, establishing the result. $\square$

Proof of Lemma 1:

Proof. Fix a π_1 that represents $c(\cdot)$. It suffices to show that any π so that $\pi(\cdot|E) = \pi_1(\cdot|E)$ for every $E \in \mathbb{Q}$ also represents $c(\cdot)$. Fix any such π. It's clear that $\pi(\cdot|\hat{P}(B)(\omega)) = \pi_1(\cdot|\hat{P}(B)(\omega))$ for all ω because $\hat{P}(B) \gg \mathbb{Q}$. By Lemma 22,

$$\arg\max_{Q\in\mathbb{P}^*} \sum_{E\in Q} \pi(E)[\max_{f\in B} \int u \circ f d\pi(\cdot|Q(\omega))] = \arg\max_{Q\in\mathbb{P}^*} \sum_{E\in Q} \pi_1(E)[\max_{f\in B} \int u \circ f d\pi_1(\cdot|Q(\omega))]$$

for every B. Conclude that $\pi(\cdot)$ also represents $c(\cdot)$. $\square$

B.3. **Proofs from Section 5.** Assume $\mathcal{P}$ is finite and label $\mathcal{P} = \{P_1, ..., P_n\}$. Write $A\,\overline{IS}_P\,B$ if $A\,\overline{IS}\,B$ for $c_P(\cdot)$.

Axiom. *(Agreement) If $\{f\}\ \overline{IS}_P\ \{g\}$ and not $\{g\}\ \overline{IS}_P\ \{f\}$, then for any $\alpha_1, ..., \alpha_n \in [0,1]$ and acts $f_1, ..., f_n, g_1, ..., g_n$ so that $\sum \alpha_i = 1$ and $\{g_i\}\ \overline{IS}_{P_i}\ \{f_i\}$, either $\sum \alpha_i g_i \neq g$ or $\sum \alpha_i f_i \neq f$.*

Theorem 6. *If $\mathcal{P}$ is finite, each $c \in \{c_P(\cdot)\}_{P\in\mathcal{P}}$ has an optimal inattention representation and $\{c_P(\cdot)\}_{P\in\mathcal{P}}$ satisfies Agreement, then there is a $\pi(\cdot)$ and a $u(\cdot)$ so that each $c_Q(\cdot)$ is represented by $(u, \pi, \hat{Q}(\cdot), \mathbb{P}_Q^*)$.*

Proof. Define $\succeq_P$ on $\mathcal{F}$ by $f \succeq_P g$ if and only if $\{f\}\ \overline{IS}_P\ \{g\}$. $\succeq_P$ is a preorder that satisfies Gilboa et al. [2010]'s c-completeness, monotonicity and independence by Lemmas 7, 8, 10, and 11 respectively. Let B_0 be the set of simple, Σ-measurable, real functions. There exists a $u_P : X \to \mathbb{R}$ by Lemma 2. By agreement, it is WLOG to take the same $u(\cdot)$ for $u_P(\cdot)$ and assume that $0 \in int(u(X))$. For every $\succeq_P$, define a cone $K_P \subset B_0$ by

$$K_P = \{\lambda(u \circ f - u \circ g) : f \succeq_P g\, and\, \lambda \in \mathbb{R}_+\}$$

and let $K = co(\cup_{P\in\mathcal{P}} K_P)$. K is a cone: suppose $v \in K$. Then there are γ_i and $v_i \in K_i$ so that $\sum \gamma_i v_i = v$. But then $\lambda v_i \in K_i$ and consequently $\sum \gamma_i \lambda v_i = \lambda v$ and $\lambda v \in K$.

Note that Agreement implies that if there exist $f_1, ..., f_n, g_1, ..., g_n \in \mathcal{F}$ and $\alpha_1, ..., \alpha_n \in [0,1]$ so that $\sum \alpha_i = 1$, $f_i \succeq_{P_i} g_i$, $f = \sum \alpha_i f_i$ and $g = \sum \alpha_i g_i$, then $g \not\succ_P f$ for any $P \in \mathcal{P}$.

Identify $\hat{f}$ with $u \circ f \in B_0$. Define $\succeq$ by $f \succeq g$ if and only if $\hat{f} - \hat{g} \in K$. Claim that $\succeq$ extends each $\succeq_P$ compatibly. First, note that if $f \succeq_P g$, then $\hat{f} - \hat{g} \in K_P \subset K$, so $f \succeq g$.

Suppose $g \succ_{P_i} f$ but $f \succeq g$, so that $\hat{f} - \hat{g} = v \in K$. Then there are $v_1, ..., v_n$ so that $v_i \in K_{P_i}$ and $\gamma_1, ...\gamma_n \in \mathbb{R}_+$ so that

$$\sum \gamma_i v_i = v.$$

Consequently, there is a λ_i and two acts f_i, g_i so that $f_i \succeq_{P_i} g_i$ and $\lambda_i(\hat{f}_i - \hat{g}_i) = v_i$. Rewriting,

$$\hat{f} - \hat{g} = \sum \lambda_i \gamma_i \hat{f}_i - \sum \lambda_i \gamma_i \hat{g}_i$$

so defining $h(\omega) = \sum \lambda_i \gamma_i \hat{f}_i(\omega) - \hat{f}(\omega) = \sum \lambda_i \gamma_i \hat{g}_i(\omega) - \hat{g}(\omega)$ gives that

$$\hat{f} + h = \sum \lambda_i \gamma_i \hat{f}_i$$

and

$$\hat{g} + h = \sum \lambda_i \gamma_i \hat{g}_i.$$

Moreover, by mixing f, g, f_i, g_i with an act 0 so that $u(0(\omega)) = 0$ in every state at a given probability, we can take $h \in u(X)^\Omega$. Now, we have that $\frac{1}{2}f + \frac{1}{2}h = \sum \lambda_i \gamma_i (\frac{1}{2}f_i + \frac{1}{2}0)$ and $\frac{1}{2}g + \frac{1}{2}h = \sum \lambda_i \gamma_i (\frac{1}{2}g_i + \frac{1}{2}0)$. Conclude that there are acts $f_i' = \frac{1}{2}f_i + \frac{1}{2}0$ and $g_i' = \frac{1}{2}g_i + \frac{1}{2}0$ and $\alpha_i \in [0,1]$ so that $f_i' \succeq_{P_i} g_i'$ for every i and that $\frac{1}{2}f + \frac{1}{2}h = \sum \alpha_i f_i'$ and $\frac{1}{2}g + \frac{1}{2}h = \sum \alpha_i g_i'$. By Agreement, it is not the case that $\frac{1}{2}g + \frac{1}{2}h \succ_{P_i} \frac{1}{2}f + \frac{1}{2}h$. However $g \succ_{P_i} f \iff \frac{1}{2}g + \frac{1}{2}h \succ_{P_i} \frac{1}{2}f + \frac{1}{2}h$ because $\succeq_{P_i}$ satisfies Independence, a contradiction.

Now, $\succeq$ has an Aumann utility for the same reasons as above. Further, it also satisfies Independence, monotonicity, reflexivity, transitivity and continuity from Gilboa et al. [2010]. Conclude from their Theorem 1 that $\succeq$ has a unique set of priors Π so that $f \succeq g \iff \int u \circ f d\pi \geq \int u \circ g d\pi$ for all $\pi \in \Pi$. The prior from the Aumann utility must be in Π by routine arguments. □

B.3.1. *Proof of Theorem 4:*

Proof. First, suppose that $c(\cdot)$ is more attentive than $c'(\cdot)$. Fix an arbitrary $Q \in \mathbb{P}^*_{c'}$ and $x, y \in X$ so that $u_{c'}(x) > u_{c'}(y)$. Define the problem B by $\{xEy : E \in Q\}$. By construction, $\hat{P}_{c'}(B) \gg Q$ so $\hat{\hat{P}}_{c'}(B) = Q$. It follows from $c(\cdot)$ more attentive than $c'(\cdot)$ that there exists a B' so that $\hat{\hat{P}}_c(B') = Q$. Consequently, $\hat{P}_c(B') \in \mathbb{P}^*_c$ and $\hat{P}_c(B') \gg \hat{\hat{P}}_c(B') = Q$ so $Q \in \mathbb{P}^*_c$.

Now, suppose that $\mathbb{P}^*_{c'} \subset \mathbb{P}^*_c$. Fix an arbitrary B and suppose that $\hat{\hat{P}}_{c'}(B) = Q$. It follows immediately that $Q \in \mathbb{P}^*_{c'}$. Fix $x, y \in X$ so that $u_c(x) > u_c(y)$. Define the problem B' by $\{xEy : E \in Q\}$. By construction, $\hat{P}_c(B') \gg Q$, implying that $\hat{\hat{P}}_c(B') = Q$. Since B is arbitrary, there exists such a B' for every B. It follows that $c(\cdot)$ is more attentive than $c'(\cdot)$. □

B.3.2. *Proof of Theorem 5:*

Proof. It is clear that $\mathbb{P}^*_Q \subset \mathbb{P}^*_P$ implies that that P is more valuable than Q. From Theorem 4, this follows from $c_P(\cdot)$ has a higher capacity for attention than $c_Q(\cdot)$.

Suppose, for the sake of contradiction, that P is not more valuable than Q but that $c_P(\cdot)$ does not have a higher capacity for attention than $c_Q(\cdot)$. Then there must be some u, π, $B \in K(\mathcal{F})$ so that

$$\max_{Q' \in \mathbb{P}^*_Q} V(u, \pi, B, Q') > \max_{Q'' \in \mathbb{P}^*_P} V(u, \pi, B, Q'').$$

Let $Q^* \in \arg\max_{Q' \in \mathbb{P}^*_Q} V(u, \pi, B, Q')$. From Theorem 4, $\mathbb{P}^*_Q \subset \mathbb{P}^*_P$, so $Q^* \in \mathbb{P}^*_P$. Therefore,

$$\max_{Q'' \in \mathbb{P}^*_P} V(u, \pi, B, Q'') \geq V(u, \pi, B, Q^*) = \max_{Q' \in \mathbb{P}^*_Q} V(u, \pi, B, Q'),$$

a contradiction. □

B.4. **Proofs from Section 6.**

B.4.1. *Proof of Proposition 2.*

Proof. Let p be an equilibrium to the market ϕ where there is at least one firm of each type. For $i = 1, ..., m$, let $p_i = \min\{p_j : \phi_j = i\}$ where $\min(\emptyset) = \infty$. Relabel so that $p_1 \leq p_2 \leq ...$ and suppose that $p_i \leq 1$ for $i \leq \kappa$ (this is WLOG because no firm of type i sells anything if $p_i > 0$). I claim that $p_{\kappa-1} = 0$. By assumption, there is at least one firm, j^*, so that $\phi_{j^*} = \kappa + 1$.

Suppose $p_{\kappa-1} > 0$. If the consumer purchases a good of type i, she purchases from the firm that charges p_i by monotonicity. If she pays attention to the partition

$$\{\{1\}, \{2\}, ..., \{\kappa, \kappa + 1, ..., m\}\}, \tag{6}$$

then her expected utility is

$$\sum_{i=1}^{\kappa-1} \frac{1}{m}(1 - p_i)$$

which is a maximum, unless $\frac{1}{m}(1 - p_\kappa) \geq p_1(m - \kappa)$, in which case

$$\{\{1\}, \{2\}, \{3\}, ..., \{\kappa\}, \{j, \kappa + 1, ...m\}\} \backslash \{j\} \tag{7}$$

for some $1 \leq j \leq \kappa$ so that $p_j = p_1$ attains an expected utility of

$$\sum_{i=1}^{\kappa} \frac{1}{m}(1 - p_i) - (m - \kappa)p_1.$$

In either case, firm j^* makes zero profit.

Suppose j^* deviates to charging a price $\frac{p_{\kappa-1}}{2} > 0$. Now, the optimal partition is either

$$\{\{1\}, \{2\}, ..., \{\kappa + 1\}, \{\kappa - 1, \kappa, \kappa + 2, ..., m\}\}$$

if (6) was maximal in the first problem or

$$\{\{1\},\{2\},\{3\},...,\{\kappa-1\},\{\kappa+1\},\{j,\kappa,\kappa+2,...,m\}\}\backslash\{j\}$$

for j as above if (7) was maximal in the first problem. In state $\kappa+1$, the consumer purchases from j^* so it attains an expected profit of

$$\frac{p_{\kappa-1}}{2m} > 0$$

which is a profitable deviation and contradicts that p is an equilibrium where $p_{\kappa-1} > 0$. Therefore, $p_{\kappa-1} = 0$ in any equilibrium.

Since $p_1 = 0$, the second partition is always optimal. Suppose $1 \geq p_\kappa > 0$, so the consumer purchases from a firm of type κ and pays a positive price in state κ in the equilibrium. Firm j^*can charge $\frac{p_\kappa}{2}$ and attract customers to make positive profit. Therefore, $p_\kappa = 0$ so κ firms of different types charge price 0 in equilibrium and the consumer purchases from one of these firms no matter what the state is. □

B.4.2. *Proof of Proposition 3.*

Proof. Expected total surplus is equal to the probability that the consumer purchases the good from a firm whose type matches the state. Clearly, this probability can be no larger than $\frac{n_m(\phi)+n_c(\phi)}{m}$. Further, the consumer makes at most κ different purchases, so it can also be no larger than $\frac{\kappa}{m}$. If $k \geq n_m(\phi)+n_c(\phi)$ but the consumer makes less than $n_m(\phi)+n_c(\phi)$ different purchases or $n_m(\phi)+n_c(\phi) > \kappa$ and the consumer makes less than κ, then a firm from which the consumer does not purchase can make positive profit by charging a price equal to $\epsilon > 0$ for ϵ small enough. Consequently, in equilibrium, the consumer purchases the good from a firm whose type matches the state with probability equal to $\max\{\frac{n_m(\phi)+n_c(\phi)}{m}, \frac{\kappa}{m}\}$.

For expected consumer surplus, if $\kappa < n_m(\phi)+n_c(\phi)$, the result follows from Proposition 2 and the above discussion. If $\kappa \geq n_m(\phi)+n_c(\phi)$, then it is easy to verify that in equilibrium, the monopolistic firms charge price 1 and at least one competitive firm of each type charges price 0. Consequently, expected consumer surplus equals $\frac{n_c(\phi)}{m}$. □

Appendix C. Counter-Examples

For the following, set $\Omega = \{\omega_1,\omega_2,\omega_3\}$, $P = \{\{\omega_1\},\{\omega_2\},\{\omega_3\}\}$ and $X = \Delta\mathbb{R}$. To economize on space, I write (a,b,c) for an act that gives a for sure in state ω_1, b for sure in state ω_2 and c for sure in state ω_3 and $c(B|\cdot) = (c(B|\omega_1), c(B|\omega_2), c(B|\omega_3))$.

C.1. **All but ACI.** Suppose $\mathbb{P}^* = \{Q \ll P : \#Q \leq 2\}$, $u(x) = x$ and $\pi(\omega) = \frac{1}{3}$ for every ω. Define a cost function ρ so that $\rho(Q) = \#Q - 1$ if $Q \in \mathbb{P}^*$ and $\rho(Q) = \infty$ otherwise. If

$$\hat{P}(B) = \max_{>} \arg\max_{Q\in\mathbb{P}^*} \sum_{E\in Q} \pi(E)[\max_{f\in B} \int u \circ f d\pi(\cdot|Q(\omega))] - \rho(Q)$$

where $\{\{\omega_1\},\{\omega_2,\omega_3\}\} > \{\{\omega_2\},\{\omega_1,\omega_3\}\} > \{\{\omega_3\},\{\omega_1,\omega_2\}\} > \{\Omega\}$ and Equation (2) holds, then $c(\cdot)$ violates ACI but satisfies the other 4 axioms. Define f, g, h by $(2,0,0)$, $(0,2,2)$ and $(0,0,0)$, respectively, and let $B_\alpha = \alpha\{f,g\} + (1-\alpha)\{h\}$. Note that $c(B_1|\cdot) = (\{f\},\{g\},\{g\})$. For $\alpha \geq \frac{1}{2}$, $\hat{P}(B_\alpha) = \{\{\omega_1\},\{\omega_2,\omega_3\}\}$, but for $\alpha < \frac{1}{2}$ $\hat{P}(B_\alpha) = \{\Omega\}$. Therefore,

$$c(B_{\frac{1}{4}}|\omega_1) = \{\frac{1}{4}g + \frac{3}{4}h\} \neq \{\frac{1}{4}f + \frac{3}{4}h\} = \frac{1}{4}c(B_1|\omega_1) + \frac{3}{4}\{h\},$$

contradicting ACI.[35] Monotonicity and Subjective Consequentialism are clearly satisfied. To see that INRA is satisfied, note that by Equation (2), if $A \subset B$ and $A \cap c(B|\omega) \neq \emptyset$ for all ω, then $\hat{P}(A) = \hat{P}(B)$. Again using equation (2), we have that $c(A|\omega) = c(B|\omega) \cap A$.

C.2. **All but INRA.** Keeping $\mathbb{P}^*$, $u(\cdot)$, $\pi(\cdot)$ and $>$ as above, suppose that

$$\hat{P}(B) = \max_{>} \arg\min_{Q \in \mathbb{P}^*} \sum_{E \in Q} \pi(E)[\min_{f \in B} \int u \circ f d\pi(\cdot|Q(\omega))]$$

for every B and that Equation (2) holds for each B and ω. This $c(\cdot)$ violates INRA. Take f, g, h, j so that $f = (3,1,2)$, $g = (1,3,1)$, $h = (1,0,0)$ and $j = (0,1,1)$. If $B = \{f,g,h,j\}$ and $A = \{f,g\}$, then $\hat{P}(B) = \{\{\omega_1\},\{\omega_2,\omega_3\}\}$ while $\hat{P}(A) = \{\{\omega_2\},\{\omega_1,\omega_3\}\}$. Consequently, $c(B|\cdot) = (\{f\},\{g\},\{g\})$ and $c(A|\cdot) = (\{f\},\{g\},\{f\})$, contradicting INRA.[36] Equation (2) implies that Subjective Consequentialism and Monotonicity hold. To see why ACI holds, note that

$$\min_{f \in \alpha B + (1-\alpha)\{g\}} \int u \circ f d\pi(\cdot|E) = \alpha \min_{f \in B} \int u \circ f d\pi(\cdot|E) + (1-\alpha) \int u \circ g d\pi(\cdot|E)$$

for any B, g and E. This implies that $\hat{P}(\alpha B + (1-\alpha)\{g\}) = \hat{P}(B)$, and Equation (2) gives that ACI holds.

C.3. **All but Monotonicity.** Let $v(x,\omega_1) = x$ and $v(x,\omega_2) = v(x,\omega_3) = -x$. Define

$$c(B|\omega) = \arg\max_{f \in B} \sum_{\omega \in \Omega} v(f(\omega),\omega)$$

and note that $0 \in c(\{0,1\}|\omega)$ for all ω. Set $f = (1,0,0)$ and $B = \{f, 0\}$. Since $\sum_{\omega \in \Omega} v(f(\omega),\omega) = 1$ and $\sum_{\omega \in \Omega} v(0,\omega) = 0$, $\{f\} = c(B|\omega)$. However, $0 \in c(\{0, f(\omega)\}|\omega)$ for all ω, so Monotonicity is contradicted. It is trivial to verify that the other axioms are satisfied.

C.4. **All but Subjective Consequentialism.** Return to the setup from the first two counter-examples. Set $\pi_1(\omega_1) = \pi_2(\omega_3) = \frac{1}{2}$, $\pi_1(\omega_2) = \pi_1(\omega_3) = \pi_2(\omega_1) = \pi_2(\omega_2) = \frac{1}{4}$, and $\pi_3 = \pi_2$. Suppose that

$$c(B|\omega_i) = \arg\max_{f \in B} \int u \circ f d\pi_i$$

[35] Similar choices occur for any B' with $d(B', B) < \epsilon$ for ϵ suitably small.

[36] Similar choices occur for any B' with $d(B', B) < \epsilon$ for ϵ suitably small.

and consider $f=(4,2,2)$, $g=(4,2,0)$, $h=(0,4,5)$ and $B=\{f,g,h\}$. By construction $c(B|\cdot)=(\{f\},\{h\},\{h\})$. Note that $\{\omega_1\}=\{\omega'':c(B|\omega'')=c(B|\omega_1)\}$ and that $f(\omega_1)=g(\omega_1)$, a contradiction of subjective consequentialism. The other properties are trivial to verify.

C.5. **All but Continuity.** Take $\mathbb{P}^*$, $u(\cdot)$ and $\pi(\cdot)$ as in the first example. Write $P_i=\{\{\omega_i\},\{\omega_i\}^c\}$. For every problem B, define an ordering $>_B$ by $P_i>_B P_j$ if and only if

$$
\begin{aligned}
&\max_{f\in B} u(f(\omega_i))) > \max_{f\in B} u(f(\omega_j))\, OR\\
&\quad [\max_{f\in B} u(f(\omega_i))) = \max_{f\in B} u(f(\omega_j))\, AND \max_{f\in B}\int u\circ f d\pi(\cdot|\{\omega_i\}^c) > \max_{f\in B}\int u\circ f d\pi(\cdot|\{\omega_j\}^c)]\\
&\quad OR\,[\max_{f\in B} u(f(\omega_i))) = \max_{f\in B} u(f(\omega_j))\, AND \max_{f\in B}\int u\circ f d\pi(\cdot|\{\omega_i\}^c) = \max_{f\in B}\int u\circ f d\pi(\cdot|\{\omega_j\}^c)\\
&\qquad AND\, i<j]
\end{aligned}
$$

Also, set every $P_i>_B\{\Omega\}$. For every problem B, take $\hat{P}(B)=\max_{>_B}\mathbb{P}^*$ and suppose Equation (2) holds. I will show that Continuity fails. Set $f=(1,.9.,1)$. and $g=(1,.8,1)$, noting that f dominates g but g does not dominate f. Set $h=(1,0,1)$, $j=(0,.9,0)$ and $k=(0,0,0)$, and for every $n>2$ define $B_{n,1}=\{g\}$, $B_{n,2}=\{g,\frac{n-1}{n}h+\frac{1}{n}k,\frac{n-1}{n}j+\frac{1}{n}k\}$, $B_{n,3}=\{\frac{n-1}{n}h+\frac{1}{n}k,\frac{n-1}{n}j+\frac{1}{n}k,\frac{n-2}{n}f+\frac{2}{n}k\}$, $B_{n,4}\{,\frac{n-2}{n}f+\frac{2}{n}k\}$. Note that

$$(\{g\},\{g\},\{g\})=c(B_{n,2}|\cdot),$$

that

$$(\{\frac{n-1}{n}h+\frac{1}{n}k\},\{\frac{n-1}{n}j+\frac{1}{n}k\},\{\frac{n-1}{n}h+\frac{1}{n}k\})=c(B_{n,3}|\cdot),$$

and that

$$(\{\frac{n-2}{n}f+\frac{2}{n}k\},\{\frac{n-2}{n}f+\frac{2}{n}k\},\{\frac{n-2}{n}f+\frac{2}{n}k\})=c(\{\frac{n-2}{n}f+\frac{2}{n}k\}|\cdot).$$

Therefore $\{g\}\,IS\,\{\frac{n-2}{n}f+\frac{2}{n}k\}$ for every $n>2$, and as $n\to\infty$, $\frac{n-2}{n}f+\frac{2}{n}k\to f$. Conclude that $\{g\}\,\overline{IS}\,\{f\}$ and f dominates g but g does not dominate f, contradicting Continuity. One can verify easily that the other four axioms are satisfied.

C.6. **Behavior compatible with optimal inattention but not inattention to alternatives.** Suppose $\Omega=\{a,b,c,d\}$ and $P=\{\{a\},\{b\},\{c\},\{d\}\}$. Consider π so that $\pi(\omega)=\frac{1}{4}$ for every ω and $\mathbb{P}^*=\{Q:Q\ll Q_1\}\cup\{Q:Q\ll Q_2\}$ where $Q_1=\{\{a\},\{b,c\},\{d\}\}$ and $Q_2=\{\{a,d\},\{b\},\{c\}\}$. Define acts x,y,z,w that give the utility values in the following table:

	a	b	c	d
$u \circ w$	6	6	6	4
$u \circ x$	8	9	0	0
$u \circ y$	0	0	0	16
$u \circ z$	2	0	9	0

One can verify that $\hat{P}(\{x, y, z, w\}) = Q_1$, $\hat{P}(\{x, z, w\}) = Q_2$, $\hat{P}(\{x, y, z\}) = Q_2$, and $\hat{P}(\{y, z\}) = Q_1$, so $c(\{x, y, z, w\}|a) = \{x\}$, $c(\{x, z, w\}|a) = \{w\}$, $c(\{x, y, z\}|a) = \{y\}$, and $c(\{y, z\}) = \{z\}$. But then by Lemma 1 and Theorem 3 of Masatlioglu et al. [2012], xPy and yPx so $c(\cdot|a)$ cannot be a choice with limited attention.

C.7. **Behavior compatible with inattention to alternatives but not optimal inattention.** Use the same setup as in C.6. Fix $x, y, z \in X$, i.e. all three are lotteries. Suppose $c(\cdot|a)$ is a choice with limited attention where $\Gamma(\{x, y, z\}) = \{y, z\}$, $\Gamma(\{x, y\}) = \{x, y\}$ and $x \succ y \succ z$. Then $c(\{x, y, z\}|a) = \{y\}$ and $c(\{x, y\}|a) = \{x\}$. If $c(\cdot)$ has an optimal inattention representation, then $c(\{x, y, z\}|a) = \{y\}$ implies $u(y) > u(x)$ but $c(\{x, y\}|a) = \{x\}$ implies $u(x) > u(y)$, a contradiction.