FOUNDATIONS FOR OPTIMAL INATTENTION

ANDREW ELLIS

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 maximization of an underlying subjective expected utility preference relation, net of a cognitive cost of attention. 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, the information to which she pays attention, her attention cost and her prior whenever information is costly.

1. Introduction

Individuals often appear not to pay attention to all available information. As argued by Simon (1971), "a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently."[1] This has motivated the application of models incorporating limited attention to economic settings, where it is found that inattention has significant consequences.[2]

This paper develops an axiomatic model of an agent who responds optimally to her limited attention, with the aim of clarifying its implications for observable choice behavior and providing a choice-theoretic justification for it. An *optimally inattentive* agent associates a

Date: September 2017.

Key words and phrases. Inattention, optimal inattention, conditional choice.

Department of Economics, London School of Economics. Email: a.ellis@lse.ac.uk. I am especially grateful to Larry Epstein, Bart Lipman and Jawwad Noor for their advice, support and guidance. I would also like to thank the editor, the referees, Laurent Bouton, Andrew Caplin, Sambuddha Ghosh, Faruk Gul, Sean Horan, Asen Kochov, Gilat Levy, Albert Ma, Francesco Nava, Andy Newman, Efe Ok, Pietro Ortoleva, Collin Raymond, Ronny Razin, Dana Rotz, Todd Sarver, Jeremy Smith, Tomasz Strazlecki, Balazs Szentes, and conference/seminar participants at RUD 2012, NASM 2012, NSF/NBER/CEME CGEME 2012, BRIC 2013, NYU, LSE, Michigan, Kellogg MEDS, Penn State, Duke, Johns Hopkins, Warwick, Oxford, and Rochester for helpful conversations and comments.

[1]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, 2011, 2015), fail to process all available information. Psychologists have also documented inattention, e.g. Pashler (1998)'s book-length treatment.

[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), self-reinforcing career dynamics (Meyer, 1991), exploitation (Rubinstein, 1993), and extreme price swings (Gul et al., 2017).

cost of paying attention to each information partition, and she chooses both her partition and her actions conditional on it by maximizing expected utility net of this cost. I axiomatically characterize the conditional choices of actions by a decision maker (DM) that are necessary and sufficient for her behavior to be seen *as if* it results from optimal inattention.

I take as primitive an objective state space and 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] I propose six natural properties of these choices, each of which weakens or is equivalent to one of the axioms that characterize a fully attentive, subjective expected utility DM. The key axiom, *Independence of Never Relevant Acts (INRA)*, requires that if two choice problems differ only because the second lacks an act that the DM *never* chooses when she faces the first, then she makes the same choices from each of the problems. The main theoretical results (Theorems 1 and 2) show that these axioms characterize a DM who pays attention to a partition Q that maximizes

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

when she faces the choice problem B, where u is a utility index, π is a probability measure, and γ is an attention cost function.

The range of behavior studied permits identification of the DM's tastes, beliefs, and attention costs, as well as the information to which she pays attention (which I call her *subjective information*). The challenge for identification stems from the modeler's inability to observe subjective information directly. This causes the DM to violate many of the properties that permit identification in other models, including the Weak Axiom of Revealed Preference (WARP). Nevertheless, I show that the DM's subjective information can be inferred from her choice behavior. Building on this insight, Theorem 3 shows that u, π, γ, and the agent's subjective information are all uniquely identified from choice data alone, whenever the likelihoods of all events are decision-relevant.

My analysis also distinguishes two special cases that have been studied in applied settings but are observationally equivalent within any fixed choice problem. A DM has a *fixed attention representation* if she always pays attention to the same information, regardless of the problem faced. A DM has a *constrained attention representation* if she has a constraint that limits the information to which she can pay attention, and she chooses her coarsening optimally within this constraint. Corollaries 1 and 2 characterize the three models in terms of the permitted violations of the Independence Axiom. Specifically, the fixed attention model never violates the Independence Axiom, and the constrained attention model never violates the Independence Axiom only when one of two mixed choice problems is a singleton.

[3]This data is typically used in dynamic decision-theoretic models and extends that considered by the papers cited in Footnote 2; Sections 2.3 and 6 elaborate further.

Table 1. Conditional choices

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

A third, particularly important, special case arises when the information available to the DM is known. This is the standard, fully attentive, subjective expected utility model (henceforth the *full attention* model). As a corollary of my results, I provide a novel characterization of the full attention model. Corollary 3 shows that the key behavioral distinction between a fully attentive DM and an optimally inattentive DM is that the former satisfies a standard property of models of choice under uncertainty known as Consequentialism while the latter need not.

I now offer an example to illustrate my setting, how I achieve identification, and what behavior rules out optimal inattention. 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, but the doctor can, in principle, determine it; for instance, by constructing a very detailed medical history. Uncertainty is modeled by the state space $\Omega = \{\gamma, \mu, \phi\}$.[4] The state indicates whether the most effective drug is produced by Glaxo (γ), by Merck (μ), or by Pfizer (ϕ), and the doctor has access to information that distinguishes which state obtains.

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).

Although 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 $\{g, m, f\}$ – the choices allow inference about her tastes, beliefs, and information. Her WARP violation reveals that the doctor is unable or unwilling to distinguish between all states. If she did, then her choice in state γ from $\{g, m\}$ would reveal that she strictly prefers to prescribe Glaxo's drug rather than to prescribe Merck's drug. Therefore, if Glaxo's drug is available in the larger problem and it is the most

[4]That the state space allows only perfect determination of which drug is appropriate is for illustrative purposes only. A more realistic problem would contain many more states but would fit within the scope of my formal analysis.

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 must not pay full attention. Additionally, the doctor's choices reveal the information to which she pays 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 $\{\{\gamma\}, \{\mu, \phi\}\}$. Similarly, her subjective information must be at least as fine as $\{\{\phi\}, \{\gamma, \mu\}\}$ 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 $\{g, m, f\}$.[5] With her subjective information known, her choices reveal her conditional preferences, which can then be aggregated to reveal her underlying unconditional preferences.

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. INRA, the key axiom in both results, requires that if a drug never prescribed to a patient is dropped from the insurance of a second, then the doctor prescribes the same drug to both patients in each state. The only drug never prescribed to the patient with good insurance is g, so the doctor's choices do not violate INRA, nor any of the other axioms. Thus, they do not rule out optimal inattention. Indeed, these choices are consistent with the following story. The doctor, perhaps in a rush or constrained by the insurance company, only has time to pay attention to results of a single test. Merck's drug is always moderately effective, and prescribing either of the other two leads to either a good outcome or a very bad one, depending on whether or not it is the most effective. If it is more likely that Pfizer's drug is best, then the doctor should attend to the test for state ϕ when treating the good-insurance patient and prescribe Pfizer's drug in that state. She would like to follow the same strategy for the bad-insurance patient, but she cannot prescribe it. Instead, she tests for the effectiveness of Glaxo's drug.

However, not all choices are compatible with optimal inattention. Consider a second doctor who chooses according to $c'(\cdot)$, where $c'(\cdot)$ is the same as $c(\cdot)$ except $c'(\{g, m, f\}|\phi) = \{m\}$. The second doctor cannot have an optimal inattention representation, as $c'(\cdot)$ violates INRA – Pfizer's drug is never chosen for the patient with good insurance, yet removing it affects her choice. To see this more directly, observe that the second doctor chooses Merck's drug when the patient has good insurance, regardless of the state of the world. As above, the doctor's choices reveal that she 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

[5] While she may have paid attention to finer information, it resulted in the same choices. See Section 5.1.

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.[6]

INRA is perhaps best viewed as a *normative* rather than a descriptive axiom.[7] That is, given that the doctor finds attention costly, what properties should her choices satisfy if they are to be consistent but take into account costs? WARP fails the latter desideratum, as obeying WARP requires that she always pay attention to the same information. Her choices, while coherent, would not take into account costs. The doctor can safely adopt INRA and make coherent choices. For the above choices, consistency requires that she makes the same choices from $\{m, f\}$ as from $\{g, m, f\}$. Otherwise, two distinct sets of choices are each revealed better than the other, as is the case for the choices described by c' above. However, consistency does not require that she choose m in state γ from $\{g, m\}$ because the benefits of the subjective information at $\{g, m, f\}$ have changed. This consistency is exactly the content of INRA.

This paper contributes to a developing literature that studies choice when the information processed is unobserved. In particular, related papers by Caplin and Dean (2015) and de Olivera et al. (2016) study related models of inattention in a stochastic conditional choice framework and a preference over menus framework, respectively. The three papers are complementary: Caplin and Dean (2015) focus on deriving implications that are easily tested in a lab rather than identification and interpretation of the parameters, while de Olivera et al. (2016) focus on ex-ante preference over menus rather than ex-post choice of acts. The relationship between the three papers is further developed in Section 7, as well as that with the literature studying decision making with a fixed but unobserved information structure, e.g. Dillenberger et al. (2014) or Lu (2016).

The remainder of the paper proceeds as follows. Section 2 presents the model in detail. In Section 3, I formally introduce the six axioms that characterize the model. Section 4 contains the main results. Theorems 1 and 2 show that these axioms characterize an optimally inattentive DM's choices. It also contains characterizations of the fully attentive model, the fixed attention model, and the constrained attention model. In Section 5, I study the uniqueness properties of the model and interpret changes in parameters in terms of changes in the DM's behavior. Theorem 3 provides a uniqueness result, and Theorems 4 and 5 provide behavioral meaning to changes in the attention cost function. Section 6 shows how to infer the conditional choice correspondence from ex ante choice of menu. Section 7 concludes by discussing the relationship with the literature. Proofs are collected in the appendices.

---

[6]Technically, indifference between the subjective information partitions must be ruled ruled out by other choices. One set of choices that rules out indifference is $c'(\{m, f\}|\gamma) = \{f\}$, $c'(\{m, f\}|\phi) = c'(\{m, f\}|\mu) = \{m\}$, $c'(\{g, f\}|\gamma) = \{g\}$, and $c'(\{g, f\}|\phi) = c'(\{g, f\}|\mu) = \{f\}$.

[7]I thank an anonymous referee for discussion of this interpretation.

2. Setup and Model

2.1. **Setup.** I adopt the following version of the classic Anscombe and Aumann (1963) setting. Uncertainty is captured by a finite set of states 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 functions $f : \Omega \to X$. Let $\mathcal{F}$ be the set of all acts, endowed with the topology of uniform convergence.

The DM chooses from a compact set of acts, 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.[8] Endow $K(\mathcal{F})$ with the topology generated by the Hausdorff metric, $d(\cdot)$, generated by a compatible metric on $\mathcal{F}$.

Prior to making a choice, the DM observes (but may not pay attention to) the realization of information that perfectly reveals the state of the world.[9] As choice may depend on the realization of Ω, I follow the literature in taking the DM's behavior conditional on each state as a primitive.[10] Motivated by the noted violations of WARP, I take a *conditional choice correspondence* $c(\cdot)$ as my primitive. The DM is willing to choose any of the acts in $c(B|\omega)$ from the problem B when the state is ω. Formally, $c(\cdot)$ is a 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 $\omega \in \Omega$.[11]

I adopt the following notation throughout. Let $\mathbb{P}$ be the set of partitions of Ω, denoting $Q(\omega)$ for the cell of $Q \in \mathbb{P}$ containing ω. For any $Q, Q' \in \mathbb{P}$, write $Q \gg Q'$ if Q is finer than Q'. 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 $x(\omega) = x \ \forall \omega \in \Omega$. 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, i.e. the act taking the value $\alpha f(\omega) + (1-\alpha)g(\omega)$ in state ω, with the usual mixture operation on lotteries. For any $A, B \in K(\mathcal{F})$ and $\alpha \in [0, 1]$, let $\alpha A + (1-\alpha)B$ equal $\{\alpha a + (1-\alpha)b : a \in A, b \in B\}$. For an event E and two acts f, g, identify fEg with the act so that $fEg(\omega) = f(\omega)$ if $\omega \in E$ and $fEg(\omega) = g(\omega)$ if $\omega \notin E$.

2.2. **Model.** Definition 1 formalizes what it means for a conditional choice correspondence to have an optimal inattention representation. First, the DM chooses her subjective information optimally, i.e. it gives at least as high expected utility net of attention cost as any other partition (Equation (1)). Second, her choice maximizes expected utility conditional on the realized cell of her subjective information (Equation (2)).

[8]The results remain true as stated if one studies choice from finite rather than compact subsets.
[9]Alternatively, one could explicitly model information P as a partition of Ω, with minor changes in results.
[10]Section 2.3 provides citations and remarks on observability.
[11]This conditional choice correspondence can be represented by a family of conditional preference relations if and only if it satisfies WARP in every state (see Section 4.1 for formal definition).

Definition 1. The conditional choice correspondence $c(\cdot)$ has *an optimal inattention representation* if there exists $(u, \pi, \gamma, \hat{P})$ so that

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

and

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

for every choice problem B and state ω, where

- $u : X \to \mathbb{R}$ is unbounded, continuous, and affine,
- $\pi : 2^\Omega \to [0, 1]$ is a full-support probability measure,
- $\gamma : \mathbb{P} \to \mathbb{R}_+ \bigcup \{\infty\}$ satisfies $\gamma(\{\Omega\}) = 0$ and $Q \gg R \implies \gamma(Q) \geq \gamma(R)$, and
- $\hat{P} : K(\mathcal{F}) \to \mathbb{P}$.

The collection $(u, \pi, \gamma, \hat{P})$ is said *to represent* $c(\cdot)$.

The *utility index* u and *prior* π have familiar interpretations. Neither varies with the problem, so an optimally inattentive DM has stable tastes and beliefs. The *attention cost function* γ maps each partition to an extended real number representing the cost that the agent incurs if she pays attention to that partition. Many examples exist in the literature; for instance, Simon (1971) suggests cost proportional to processing time and Sims (2003) argues for cost proportional to Shannon entropy. If Q contains strictly more information than R, i.e. $Q \gg R$, then Q costs at least as much as R, i.e. $\gamma(Q) \geq \gamma(R)$. 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.

Another special case is a DM who always pay attention to the same 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 so that

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

for every B and ω. This model corresponds to the special case where $\gamma(Q) = 0$, any R strictly finer than Q has $\gamma(R) = \infty$, and $\hat{P}(B) = Q$ for every problem B.

A final special case is a DM with constrained attention. Such a DM has an *attention constraint* rather than a cost function, i.e. she can costlessly pay attention to one of several partitions but no others. Such a DM decides *to which* information she pays attention but not *how much* attention she should pay. Formally, $c(\cdot)$ has a *constrained attention representation* if $\gamma(Q) = 0$ or $\gamma(Q) = \infty$ for every $Q \in \mathbb{P}$.

In addition to the above, one can specialize γ to admit other instances in the literature. For instance, $\gamma(Q)$ may be proportional to to the number of elements in Q (as in Rubinstein (1993)) or to the mutual information between Q and P (similar to Sims (2003)).[12]

2.3. **Remarks.** This subsection discusses issues regarding the model's interpretation.

2.3.1. *Consideration and costly information acquisition.* An optimally inattentive DM considers all available acts. In contrast, Masatlioglu et al. (2012) (or in a stochastic setting, Manzini and Mariotti (2014)) study 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. While DMs conforming to either model may violate WARP, the reason for such violations is different.[13]

Interpreting the representation as inattention requires that perfect information is freely available to the DM. If this assumption is dropped, the model can be reinterpreted as costly, optimal and unobserved information acquisition. In an information acquisition problem, the cost of information is external, e.g. technological, rather than internal, e.g. psychological. Van Zandt (1996) considers this model in a related setting, assuming that choice observed in only a single state, and shows the model has no testable implications. It is also similar to costly contemplation, e.g. Ergin and Sarver (2010). In such a model, the states are subjective and correspond to the DM's taste for the various consequences; the information available plays no role.[14]

2.3.2. *Information.* In the model, all information is partitional, including the DM's subjective information. Partitional information is flexible, analytically convenient and conforms with the traditional approach in information economics.[15] However, partitional information does require certain modeling tradeoffs, which are explored in this subsection.

In a choice theoretic context, the state space consists of only payoff relevant states by construction. With partitional information, each signal must correspond to at least one state, so expanding the set of signals necessarily expands the set of states. Hence, enlarging

[12]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.

[13]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.

[14]While not explicitly modeled, the interpretation suggests fixed "true" tastes and thus deterministic choice.

[15]Indeed, the traditional interpretation of the state space is that each $\omega \in \Omega$ contains a complete description of the world. With such a state space, the restriction to an information partition is without loss; for instance, Aumann (1976) writes, "Included within the full description of a state ω is the manner in which information is imparted." Expanding the set of states in this way may require an undesirable expansion of the set of acts available to the DM. Theorem 2, which shows the necessity of my axioms, applies even if not all acts are available.

the state space also entails enlarging the space of acts and enriching the choice data on which the model is based.[16]

Since signals must be perfectly correlated with the state, stochastic conditional choices are ruled out. As a consequence, fixing a state space necessarily fixes the possible information structures. Thus, modeling information as a partition is especially appropriate when the payoff relevant states are a close approximation of the information available to the DM. This is likely to be satisfied in an environment wherein the information to which the DM can pay attention is sparse, such as in an experiment where the information available to subjects is carefully controlled. An experimenter can, at least in principle, offer subjects acts that vary in an arbitrary fashion with the realized signal, and the subject cannot, much as she would like to, acquire information not provided to her. For instance, in an experiment with exactly two states, the subject may prefer to pay attention to a less costly information structure with two signals, the first of which occurs with probability $p \in (0, 1)$ in state 1 and probability 0 in state 2, rather than a partitional one that distinguishes the states perfectly, but if no such information is provided by the experimenter, then she is forced to choose between the two possible partitional structures.

2.3.3. *Timing.* While time does not play a formal role in the model, the behavior is interpreted as resulting from a dynamic process. To understand dynamic behavior, I follow the literature by taking choices in distinct states of the world as observable. For instance, Epstein and Schneider (2003), Maccheroni et al. (2006b), Klibanoff et al. (2009), Ghirardato (2002), and Ortoleva (2012) all take a family of complete and transitive preference relations indexed by either every state-time pair or every event. This data corresponds to that typically considered in applications, as an agent chooses (either deterministically or stochastically) from a single feasible set conditional on each state in each of the papers cited in Footnote 2. For reasons familiar from static decision theory, choice in an isolated situation reveals little about underlying behavior without additional assumptions, so the primitive expands this data in the natural way. Section 6 shows the data can be obtained from ex ante preference over menus under a consistency assumption.

2.3.4. *Infinite regress.* The optimal inattention model may be subject to the infinite regress critique, e.g. Lipman (1991). Namely, "to what does the agent pay attention when deciding to what she pays attention when deciding...?" By taking her conditional choice correspondence, rather than the model, as primitive, my approach bypasses this critique. According to Theorems 1 and 2, if $c(\cdot)$ has certain properties, then the DM acts "as if" she has an optimal inattention representation. If an infinite regress occurs and the agent fails to carry out the maximization above, then she violates my axioms.

[16]As standard, one need not have all acts available to falsify any of the axioms. Of course, choices from any (finite) subset of acts are needed to establish sufficiency.

3. Foundations

I impose six axioms, Independence of Never Relevant Acts (INRA), Attention Constrained Independence (ACI), Monotonicity, Subjective Consequentialism (SC), Continuity, and Unboundedness. Each relaxes or is equivalent to one of the axioms that characterize the standard subjective expected utility model with an unbounded utility index.[17] The quantifier "for all $f, g, h \in \mathcal{F}$, $A, B \in K(\mathcal{F})$, $\omega \in \Omega$ and $\alpha \in (0, 1]$" is suppressed throughout.

A DM satisfies *WARP*, sometimes referred to as Independence of Irrelevant Acts, if $A \subset B$ and $c(B|\omega) \bigcap A \neq \emptyset$ imply that $c(A|\omega) = c(B|\omega) \bigcap A$. 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 $c(B|\omega') \bigcap A \neq \emptyset$ for every $\omega' \in \Omega$, then*

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

In the example, 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. $\{f\} \neq c(B|\omega')$ for all ω'), and let $A = B \backslash \{f\}$.[18] Suppose that her choices from B are conditioned on the subjective information Q. Because she never chooses only 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 $c(B|\omega') \bigcap A \neq \emptyset$ 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 from A and B should not violate WARP.[19]

[17]Specifically, INRA relaxes weak order, ACI relaxes Independence, SC relaxes Consequentialism, Continuity relaxes its counterpart, and both Monotonicity and Unboundedness hold in both models.

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

[19]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. $c(B|\phi) \bigcap A = \emptyset$. 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 $c(A|\gamma) \neq c(B|\gamma) \bigcap A$.

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

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

That is, if the DM chooses g over each h in A and f over each h' in B, then she chooses $\alpha g + (1-\alpha) f$ over each $\alpha h + (1-\alpha) h'$ in $\alpha A + (1-\alpha) B$.[20] 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)**.** *If* $\alpha g + (1-\alpha) f \in c(\alpha\{g\} + (1-\alpha) B|\omega)$,
then $\alpha h + (1-\alpha) f \in c(\alpha\{h\} + (1-\alpha) B|\omega)$.

In the example, this says that if there is a state-independent chance α that the patient will take some drug h regardless of what the doctor actually prescribes, then her choice of prescription is unaffected by the identity of h. However, it leaves open the possibility that the magnitude of α affects what the doctor prescribes. For instance, if α is close to 1, then the probability that the patient follows the doctor's advice is very small, so the doctor may pay attention to less costly information than when the patient follows her advice for sure, i.e. where $\alpha = 0$.

To interpret the axiom, fix problems B, $\{g\}$, and $\{h\}$. Because $\{g\}$ and $\{h\}$ are singletons, the DM makes the same choice from either no matter what her subjective information is. Therefore, the difference between the benefits of any two subjective information partitions is the same for the problem $\alpha\{g\} + (1-\alpha) B$ as it is for the problem $\alpha\{h\} + (1-\alpha) B$. Intuitively, one can think of $\alpha\{g\} + (1-\alpha) B$ ($\alpha\{h\} + (1-\alpha) B$) as choosing from B then flipping a coin and getting one's choice if the coin comes up heads and otherwise receiving g (h), where the DM must choose her subjective information before observing the outcome of the coin-flip. Information only has value if the coin comes up heads, and conditional on heads, a given partition has the same value in either problem. Since the probability of heads is the same in either case, if paying attention to Q is *optimal* when facing $\alpha\{g\} + (1-\alpha) B$, then paying attention to Q is *also optimal* when facing $\alpha\{h\} + (1-\alpha) B$. Consequently, an optimally inattentive DM conditions her choices on the same subjective information when facing $\alpha\{g\} + (1-\alpha) B$ as she does when facing $\alpha\{h\} + (1-\alpha) B$. Because her conditional preferences satisfy Independence, she chooses the mixture of f with g from $\alpha\{g\} + (1-\alpha) B$

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

if and only if she chooses the mixture of f with h from $\alpha\{h\} + (1-\alpha)B$. Mathematically, it ensures consistency between the DM's choices in problems B and B', when the utility of each act in B' results from applying the same translation to each act in B.

The next axiom adapts the standard Monotonicity axiom to the present setting. To state the axiom concisely, for any $x, y \in X$, say that x is *revealed preferred* to y, denoted $x \succeq^R y$, if there exists a state ω so that $x \in c(\{x,y\}|\omega)$, and x is *revealed strictly preferred* to y, denoted $x \succ^R y$, if $x \succeq^R y$ and not $y \succeq^R x$.[21] If tastes are state-independent, then x is revealed preferred to y only if the DM regards x as a better consequence than y.

Axiom 3 (Monotonicity)**.** *If $f, g \in B$ and $f(\omega') \succeq^R g(\omega')$ for every $\omega' \in \Omega$, then $g \in c(B|\omega) \implies f \in c(B|\omega)$; if in addition $f(\omega) \succ^R g(\omega)$, then $g \notin c(B|\omega)$.*

In the example, this says that if one drug gives a better consequence in every state than another, then the DM never prescribes the inferior drug. To interpret the axiom, 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 state ω, then the DM does not choose g in that state. Thus, an inattentive DM will never pick a dominated act. Monotonicity also implies state-independent tastes: if she chooses x over y in state ω, then she also chooses x over y in state ω'.

The next axiom, *Subjective Consequentialism* or *SC*, requires that choice between any two acts is unaffected by their outcomes in states that the DM knows did not occur. It weakens the standard property known as Consequentialism; see Section 4.1.

Axiom 4 (SC)**.** *If $f(\omega) = g(\omega)$ and for all $\omega' \neq \omega$ either $f(\omega') = g(\omega')$ or $c(B|\omega') \neq c(B|\omega)$, then $f \in c(B|\omega) \iff g \in c(B|\omega)$ whenever $f, g \in B$.*

To interpret SC, 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 $c(B|\omega) = c(B|\omega')$. Therefore if $c(B|\omega') \neq c(B|\omega)$, then ω and ω' must be in different cells of her subjective information. By hypothesis, either f and g give the same consequence in state ω' or ω' is in a different cell of the DM's subjective information than ω. In state ω, the DM knows that she receives the same consequence regardless of whether she chooses f or g, so she chooses f if and only if she chooses g.

The final two axioms are technical conditions. The first ensures the continuity of the underlying preference relation, and the second ensures unboundedness of the utility index. Complicating the statement of the first is that the DM's choices from different problems may

[21]Note that $\succeq^R$ is a binary relation over X rather than $\mathcal{F}$.

be conditioned on different information. Consequently, her choices may appear discontinuous to the modeler.[22] 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 one preliminary definition.

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, $c(B_{i+1}|\omega) \bigcap B_i \neq \emptyset$.

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.

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 can be extended to indirect comparisons as well using $A\, IS\, B$. Because this indirect ranking compares more sets of choices, indirect selections are important for characterizing optimal inattention. Continuity ensures a minimal consistency between the limits of indirect comparisons and the direct comparisons.

Axiom 5 (Continuity)**.**

(1) *For any* $\{B_n\}_{n=1}^{\infty} \subseteq K(\mathcal{F})$ *and* $\{f_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$*, if* $B_n \to B$*,* $f_n \to f$*,* $f_n \in c(B_n|\omega)$*, and for* $c(B|\omega') \neq c(B|\omega) \iff c(B_n|\omega') \neq c(B_n|\omega)$ *for every* n *and* ω'*, then* $f \in c(B|\omega)$*.*

(2) *For any* $x, y \in X$ *and* $f, g_1, ..., g_n \in \mathcal{F}$ *such that* $y \succ^R x$*,* $x \in c(\{f, x\}|\omega')$ *for all* ω'*, and* $f \notin c(\{f, x\}|\omega'')$ *for some* ω''*: if* $\{g_i\}\ IS\ \{x\}$ *for* $i = 1, ..., n$*, then for any* $\alpha_1, ..., \alpha_n \in [0, 1]$ *with* $\sum_{i=1}^{n} \alpha_i = 1$*, there exists* $\epsilon > 0$ *and* ω^* *such that* $\sum \alpha_i g_i(\omega^*) \succ^R \epsilon y + (1-\epsilon) f(\omega^*)$*.*

Both parts of the require continuity only in places where attention is fixed. In the first part, the information that DM has in state ω remains fixed. In the second part, when choosing from $\{f, g\}$, the DM chooses f in each state of the world, so paying attention to information has no value when facing either $\{f, g\}$, $\{f\}$, or $\{g\}$. Observe that the axiom is implied by combining WARP, Independence, and upper-hemicontinuity.

[22]The key issue is that the DM may be indifferent between paying attention to two partitions, and the modeler only observes her choices conditional on one of them. Consequently, the choices from the limit of a sequence of problems may be conditioned on different information than those along the sequences. Consider a sequence of choice problems B_n with limit B. If the DM pays attention to Q when facing every B_n, then paying attention to Q is optimal when facing B. However, another partition, say R, may also be optimal when facing B. If the DM pays attention to R when facing B, and $Q(\omega) \neq R(\omega)$, then her choice in state ω typically fails upper-hemicontinuity.

The first condition is simply a restriction of upper-hemicontinuity. It requires that if attention converges for a sequence of choice problems, then the DM's sequence of choices from those problems also converges.

The second part of the axiom ensures continuity of the DM's revealed ex ante preference. To interpret it, fix f, g_i, x and y as in the statement.[23] The DM reveals that y is strictly better than x, which is in turn strictly better than f. Continuity of her ex ante ranking implies she should also strictly prefer x to $\epsilon y + (1-\epsilon)f$ when ϵ is small enough. Now, $\{g_i\}$ IS $\{x\}$ reveals that g_i has at least as much value as x, and thus also at least as much as $\epsilon y + (1-\epsilon)f$. Therefore, g_i should yield at least as good an outcome as $\epsilon y + (1-\epsilon)f$ in some state. As the DM has an expected utility ex ante preference, this must also hold for any mixture between $g_1, ..., g_n$.

The final axiom guarantees that the utility index is unbounded.

Axiom 6 (Unboundedness)**.** *There exist $x, y \in X$ with $x \succ^R y$ such that for any $\beta \in (0,1)$ there exist $z^*, z_* \in X$ with $\beta z^* + (1-\beta)y \succ^R x$ and $y \succ^R \beta z_* + (1-\beta)x$.*

This condition is equivalent to the range of $u(\cdot)$ equals $\mathbb{R}$. Versions appear in other work, e.g. Kopylov (2001), Maccheroni et al. (2006a) and de Olivera et al. (2016). Unlike in Maccheroni et al. (2006a) but as in Kopylov (2001) or de Olivera et al. (2016), Unboundedness plays a role in the sufficiency part of my proof. To understand the axiom, assume an affine utility representation over X, $u(\cdot)$.[24] The lottery $\beta z^* + (1-\beta)y \succ^R x$ if and only if $\beta u(z^*) + (1-\beta)u(y) > u(x)$, and $y \succ^R \beta z_* + (1-\beta)x$ if and only if $u(y) > \beta u(z_*) + (1-\beta)u(x)$. Since $u(x) > u(y)$, as β goes to zero both $u(z^*)$ and $-u(z_*)$ must approach infinity.

4. Characterization

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

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

Theorem 1 shows that the above axioms are sufficient for the DM to have an optimal inattention representation. The key idea of the proof is to map the agent's choices onto a different domain where they behave better. In particular, I consider the space of "plans." A

[23]There is a natural trade-off between the type of independence and continuity; strengthening one allows weakening the other. By assuming Strong ACI (below) or requiring that for any x, there exists $\epsilon > 0$ so that $c(B|\cdot)$ is constant when $d(B, \{x\}) < \epsilon$, one only needs the conclusion of Continuity (2) to hold when $n = 1$. That is, instead assume "For any $x, y \in X$ and $f, g \in \mathcal{F}$ where $y \succ^R x$, $x \in c(\{f, x\}|\omega')$ for all ω', and $f \notin c(\{f, x\}|\omega'')$ for some ω'': if $\{g\}$ IS $\{x\}$, then there exists $\epsilon > 0$ and ω^* such that $g(\omega^*) \succ^R \epsilon y + (1-\epsilon)f(\omega^*)$." However, both restrict the cost function; see Corollaries 2 and 4.

[24]INRA, ACI, Monotonicity, and Continuity guarantee the existence of such a representation; see Lemma 1.

plan F maps each state to an act, with the interpretation that if the DM chooses F, then she plans to choose the act $F(\omega)$ in state ω. Each set of conditional choices from a given problem defines a plan. In the introductory example, the doctor chooses the plan "pick f in state ϕ, otherwise pick m" from $\{g, m, f\}$ and chooses the plan "pick g in state γ, otherwise pick m" from $\{g, m\}$. Although choice in a given state may violate WARP, INRA guarantees that her choice of plan maximizes a preference relation. For instance, the choices above reveal that the doctor prefers the former plan over the latter one. Given the other axioms, this relation can be taken to be well behaved enough to have a representation similar to Equation (1). To complete the proof, I show that choosing a plan corresponds to choosing her subjective information.[25]

Necessity is more complicated because of tie-breaking. That is, what happens when two or more partitions are optimal for the same choice problem? INRA and ACI impose some restrictions on how these ties are broken, and if the DM breaks these ties non-systematically, then she may violate the two axioms.[26] The axioms are necessary if conditions on tie-breaking are imposed when defining the model, such as requiring that ties are broken according to a linear order. Nevertheless, 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.

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

(i) $c(\cdot)$ and $c'(\cdot)$ are represented by $(u, \pi, \gamma, \hat{P})$ and $(u, \pi, \gamma, \hat{Q})$, 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 cost, there are always attention rules that satisfy INRA and ACI. The characterization is tight; for each of the axioms, I provide counter-examples satisfying all the others in the Supplemental Appendix.

[25]Optimal inattention admits foundations if preference over plans is a primitive; see the Supplementary Appendix. Observing the DM's 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., 2017)). However, this has some significant drawbacks in terms of observability, interpretation, and connection to the objects of economic interest. Moreover, if the DM follows through with her choice of plan, then her final conditional choices of acts satisfy my axioms.

[26]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.

4.1. **Special cases.** The axioms allow a clearer picture of how the model relates to others in terms of behavior. To better understand these connections, I characterize the three special cases of optimal inattention introduced in Section 2.2.

For a given choice problem, all conditional choices by any optimally inattentive DM can be replicated by a DM that conforms to either of the first two, the fixed attention model and the constrained attention model.[27] However, varying the problem allows the three models to be distinguished. Corollary 1 shows a DM with fixed attention must satisfy independence, and Corollary 2 shows a DM with constrained attention may violate independence but satisfies a stronger version of ACI. Taken together, the results show that the patterns of violations of independence distinguish the three models.

Corollary 1. *If $c(\cdot)$ satisfies the axioms of Theorem 1, then $c(\cdot)$ satisfies Independence if and only if $c(\cdot)$ satisfies WARP if and only if $c(\cdot)$ has a fixed attention representation.*

It immediately follows that WARP and Independence are equivalent for an optimally inattentive DM. The intuition behind Corollary 1 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.

The key property that differentiates constrained attention from costly attention is *Strong ACI*, which $c(\cdot)$ satisfies if for any $\omega \in \Omega$, $\alpha \in (0,1]$, $B \in K(\mathcal{F})$, and $f, g \in \mathcal{F}$,

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

ACI requires that when a problem is mixed with a singleton, the DM's choice is independent of the identity of the singleton. Strong ACI requires that in addition to being independent of the identity of the singleton, it is also independent of the *magnitude* of the weight given to that singleton. In particular, the DM's choice from B is the same whether she faces B for sure or she faces it with very small probability and gets g otherwise. Strengthening ACI to Strong ACI characterizes a DM with constrained attention.

Corollary 2. *If $c(\cdot)$ satisfies INRA, Strong ACI, Monotonicity, SC, Continuity, and Unboundedness, then $c(\cdot)$ has a constrained attention representation. Moreover, the axioms are generically necessary, in the same sense as Theorem 2.*

The final special case I turn to is the full attention model, the standard model of dynamic choice. In these models, an information partition (or a filtration) is taken as given. For the

[27]Fix an optimally inattentive DM with cost γ and a problem B. Let $Q = \hat{P}(B)$. Consider a fixed attention DM who always pays attention to Q and a constrained attention DM with constraint given by $\{Q' : \gamma(Q') \leq \gamma(Q)\}$. All three DMs make the same choices in every state when facing B.

remainder of this section only, I assume that there is an information partition P and that for each B, $c(B|\cdot)$ is P-measurable, i.e. she never distinguishes between two states that P does not. A DM has a full attention representation if she has a fixed attention representation with attention fixed at Q.

An optimally attentive DM has a full attention representation if and only if she satisfies the standard *Consequentialism* axiom. This property, adapted to my context, states that if $f(\omega) = g(\omega)$ and $f(\omega') = g(\omega')$ for all $\omega' \in P(\omega)$, then $f \in c(B|\omega) \iff g \in c(B|\omega)$ whenever $f, g \in B$. It has a standard interpretation: if f and g are identical in every objectively possible state, then she chooses one if and only if she chooses the other. Note Consequentialism implies SC.

Corollary 3. *$c(\cdot)$ satisfies Consequentialism in addition to INRA, ACI, Monotonicity, Continuity and Unboundedness if and only if $c(\cdot)$ has a full attention representation.*

Corollary 3 provides a novel characterization of the standard subjective expected utility model. If the DM satisfies Consequentialism and the optimal inattention axioms, then the DM must also satisfy WARP, Independence and the other expected utility axioms. Moreover, violations of Consequentialism behaviorally distinguish an optimally inattentive DM from a fully attentive DM, as the former need only satisfy Subjective Consequentialism. 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 full attention representation.

5. Identification and Comparative Behavior

5.1. **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. Even though the modeler does not directly observe ex-ante preference, subjective information, or attention cost, Theorem 3 shows the utility index, attention cost, subjective information and beliefs are suitably unique, when every non-trivial partition has a positive cost, i.e. $\gamma(P) > 0$ for all $P \neq \{\Omega\}$. This assumption rules out particular cases where the likelihood of certain events does not affect the DM's choices.[28] In the Supplementary Appendix, I give a full characterization of all representations consistent with choices even when this condition is violated.

First, I provide a behavioral characterization of positive cost of information.

[28]For one case, consider a fully attentive DM. Because the modeler observes only ex-post choice, *any prior* represents the agent's choices: her choice in state ω is the act with the best outcome in state ω, and the prior plays no role in her decision.

Corollary 4. *If c has an optimal inattention representation $(u, \pi, \gamma, \hat{P})$, then $\gamma(P) > 0$ for all $P \neq \{\Omega\}$ if and only if for any $x \in X$, there exists $\epsilon > 0$ so that $c(B|\cdot)$ is constant for any $B \in K(\mathcal{F})$ with $d(B, \{x\}) < \epsilon$.*

Before stating Theorem 3, I introduce the following normalization: an attention rule is *canonical* if no strictly coarser attention rule also represents choice. For instance, if B contains only constant acts, then $\hat{P}(B)$ could in principal be any partition with zero cost, but a canonical rule requires that $\hat{P}(B) = \{\Omega\}$. Such a normalization is unnecessary if each strictly finer partition has a strictly higher cost.

Theorem 3. *If $(u, \pi, \gamma, \hat{P})$ and $(u', \pi', \gamma', \hat{Q})$ represent $c(\cdot)$ and $\gamma(P) > 0$ for all $P \neq \{\Omega\}$:*
1. *$\pi = \pi'$,*
2. *$\hat{Q}(B) = \hat{P}(B)$ for all $B \in K(\mathcal{F})$ whenever $\hat{P}$ and $\hat{Q}$ are canonical, and*
3. *there exist $\alpha > 0$ and $\beta \in \mathbb{R}$ so that $u'(x) = \alpha u(x) + \beta$ and $\gamma'(Q) = \alpha\gamma(Q)$ for all $x \in X$ and $Q \in \mathbb{P}$.*

Theorem 3 establishes that an optimally inattentive DM's utility index, attention rule, attention cost and prior are unique, as long as all information is costly. To see how choices identify tastes, note that information is irrelevant for choosing constant acts. Consequently, the DM acts as an expected utility maximizer when facing a problem containing only lotteries, and identification of the utility index is a standard exercise.

The canonical attention rule is the partition

$$\hat{P}(B) = \{\{\omega' : c(B|\omega') = c(B|\omega)\} : \omega \in \Omega\}. \tag{3}$$

The DM's subjective information when facing B must be at least as fine as the above. If not, then the information to which she pays attention does not distinguish two states on which she makes different conditional choices. Although she may pay attention to a partition strictly finer than it, she makes the same conditional choices on at least two cells if she does, so this partition also represents her choices.

To see how choices identify the cost of attention, consider first an optimally inattentive DM who can only pay attention to either Q or $\{\Omega\}$. Define a choice problem B that consists of bets on each element of Q with stakes x and y as well as one sure option z, so that $u(x) > u(z) > u(y)$. If y is a sufficiently undesirable consequence, then the DM strictly prefers to pay attention to Q rather than $\{\Omega\}$ whenever $u(x) - \gamma(Q) > u(z) - \gamma(\{\Omega\})$, and she does not choose z only if she pays attention to Q. One can then identify $\gamma(Q) - \gamma(\{\Omega\})$ by finding the smallest x for which the DM does not choose z; this identifies $\gamma(Q)$ uniquely because $\gamma(\{\Omega\}) = 0$. When the DM may pay attention to more partitions, the above identification procedure generalizes, but it must take into account that the DM may opt to pay attention to a third partition. There exist partitions $Q_1, Q_2, ..., Q_n$ with $Q_1 = \{\Omega\}$ and

$Q_n = Q$ for which similar choice problems allow determination of $\gamma(Q_i) - \gamma(Q_{i-1})$. Since $\gamma(Q_1) = 0$, the sum of these cost differences equals $\gamma(Q)$.

5.2. **Comparative Attention.** In this section, I develop two notions of what it means for one DM to be more attentive than another, and interpret these behaviors in terms of the model's parameters. First, DM1 pays more attention than DM2 if for every choice problem B, DM2 distinguishes ω from ω' when facing B only if DM1 also distinguishes ω from ω' when facing B. Second, DM1 has a higher capacity for attention than DM2 if for any choice problem B, there is a second choice problem B' so that DM2 distinguishes ω and ω' when facing B if and only if DM1 also distinguishes ω and ω' when facing B'. The first condition implies that DM1's marginal cost of attention is lower than DM2's, and the second condition implies that any partition that is feasible for DM2 is also feasible for DM1.

Definition 3. $c_1(\cdot)$ *pays more attention than* $c_2(\cdot)$ if for every $B \in K(\mathcal{F})$ that has a unique optimal partition for c_1 and all $\omega, \omega' \in \Omega$, $c_2(B|\omega) \neq c_2(B|\omega') \implies c_1(B|\omega) \neq c_1(B|\omega')$.[29]

To interpret the definition, consider DM1 and DM2 with conditional choice correspondences given by $c_1(\cdot)$ and $c_2(\cdot)$, and suppose that $c_1(\cdot)$ pays more attention $c_2(\cdot)$. Because $c(B|\omega') \neq c(B|\omega)$ only if the DM's subjective information distinguishes ω from ω', whenever DM2's subjective information distinguishes ω and ω', so does DM1's. Consequently, DM1 pays attention to finer information than DM2 does, no matter which choice problem they face.

Assuming DM1 and DM2 conform to the optimal inattention model with the same ex-ante preference, if c_1 pays more attention than c_2, then DM1 receives a higher expected utility than DM2 when facing every choice problem. In fact, it is necessary and sufficient for the above whenever any two partitions in the support of either γ_1 or γ_2 can be compared by the "finer than" relation. Such DMs chooses how precision their subjective information should be. Theorem 4 characterizes this behavioral property in terms of the representation for the case where the two DMs have the same ex-ante preference and any two partitions in the support of either γ_1 or γ_2 can be compared by the "finer than" relation.

Theorem 4. *If $(u, \pi, \gamma_1, \hat{P}_1)$ represents $c_1(\cdot)$, $(u, \pi, \gamma_2, \hat{P}_2)$ represents $c_2(\cdot)$, and either $Q \gg R$ or $R \gg Q$ for any $Q, R \in supp(\gamma_1) \bigcup supp(\gamma_2)$, then $c_1(\cdot)$ pays more attention than $c_2(\cdot)$ if and only if $Q \gg R \implies \gamma_1(Q) - \gamma_1(R) \leq \gamma_2(Q) - \gamma_2(R)$.*

Theorem 4 says that DM1 pays more attention than DM2 if and only if her marginal cost of attention is lower. Intuitively, each DM refines her information until the marginal

[29]From behavior, B has a unique optimal partition i there exists an $\epsilon > 0$ so that

$$c(B|\omega) \neq c(B|\omega') \iff c(B'|\omega) \neq c(B'|\omega')$$

for all B' with $d(B, B') < \epsilon$.

cost exceeds the marginal benefit. Whenever DM2 evaluates the marginal benefit of more information as lower than the marginal cost, DM1 does so as well. Because the two DMs have the same ex-ante preference, they also have the same marginal benefit and thus DM1 must have a lower marginal cost of paying attention to finer information.

It should be noted that under the assumptions of Theorem 4, if c_1 pays more attention than c_2, then $\gamma_1(Q) \leq \gamma_2(Q)$ for all Q. However, the converse is false. While $\gamma_1(Q) \leq \gamma_2(Q)$ for all Q implies that DM1's net ex-ante utility is larger than DM2's net ex-ante utility for every problem, it does not admit a natural comparison in terms of the two DMs' conditional choices. In fact, a DM with a lower cost of attention sometimes selects coarser subjective information than one with a higher cost of attention.[30]

The second comparison considers two DM's capacities for attention.

Definition 4. $c_1(\cdot)$ *has a higher capacity for attention than* $c_2(\cdot)$ if for any B, there exists a B' so that for all $\omega, \omega' \in \Omega$, $c_2(B|\omega) \neq c_2(B|\omega') \iff c_1(B'|\omega) \neq c_1(B'|\omega')$.

To interpret the definition, consider DM1 and DM2 with conditional choice correspondences given by $c_1(\cdot)$ and $c_2(\cdot)$, and suppose that $c_1(\cdot)$ has a higher capacity for attention than $c_2(\cdot)$. For any problem B, there exists a B' so that DM2 distinguishes ω and ω' when facing B if and only if DM1 distinguishes ω from ω' when facing B'. Therefore, DM1 must have the ability to pay attention to the information to which DM2 paid attention. Theorem 5 shows that for any two optimally inattentive DMs, this comparison is equivalent to the support of DM2's cost function being contained in the support of DM1's cost function.

Theorem 5. *If* $(u_1, \pi_1, \gamma_1, \hat{P}_1)$ *represents* $c_1(\cdot)$ *and* $(u_2, \pi_2, \gamma_2, \hat{P}_2)$ *represents* $c_2(\cdot)$*, then* $c_1(\cdot)$ *has a higher capacity for attention than* $c_2(\cdot)$ *if and only if* $supp(\gamma_1) \subset supp(\gamma_2)$.

Note that $supp(\gamma_i)$ is the set of partitions to which the DM may choose to pay attention. Theorem 5 thus formalizes that $supp(\gamma_i)$ reflects the DM's capacity for attention. Unlike Theorem 4, the hypothesis of the result does not require that u_1 or π_1 are related to u_2 or π_2 in any way. The behavioral comparison does not require identical ex-ante preference to be meaningful.

6. Observing Conditional Choice

Our goal in the section is to establish that ex ante data is often sufficient to elicit ex post conditional choices. While conditional choices are often used in applications, they are notoriously difficult to observe. To do so often requires seeing choice from the same menu in the same setting multiple times or by multiple identical individuals, along with auxiliary

[30]Consider three equally likley states $1,2,3$, and for $Q = \{\{1\},\{2,3\}\}$, $R = \{\{1\},\{2\},\{3\}\}$, $\gamma_1(\{\Omega\}) = \gamma_2(\{\Omega\}) = 0$, $\gamma_1(Q) = 4$, $\gamma_1(R) = 8$, $\gamma_2(Q) = 7$ and $\gamma_2(R) = 9$ (both have infinite cost for all other partitions). One can easily construct an example where DM 1 pays less attention than DM 2.

assumptions regarding independence. Indeed, when a decision is not repeated, how a given individual would choose in unrealized states cannot be observed. Ex ante choices, however, are often easy to observe.

When conditional choices are unobservable, the best one can do is assume consistency between ex ante and ex post choice and elicit the former from ex ante behavior. Roughly, consistency requires that her ex ante choice of menu reflects what she actually chooses ex post. While such an assumption is not necessarily a desirable property of a model of bounded rationality (Spiegler, 2010), it overcomes limitations in observability. This section formalizes the needed assumption and shows how to infer conditional choices from the DM's ex ante choice of a menu of acts.

One well-known parallel in the literature is dynamic choice under expected utility (Savage, 1954; Ghirardato, 2002). It considers a DM who chooses an act after learning that the realized state lies in the event $E \subset \Omega$; formally, her ex ante choice maximizes $\succeq$, while her ex post choice maximizes $\succeq_E$. The two are related via *Dynamic Consistency (DC)*: $f \succeq_E g$ if and only if $fEg \succeq g$. Observe that DC can only be tested with both ex ante and ex post choice data. If one is nevertheless willing to assume DC, then the DM's behavior ex post behavior can be predicted even if one only observes $\succeq$. Specifically, one elicits that $f \succeq_E g$ whenever the DM expresses that $fEg \succeq g$.[31] The resulting family of preference relations, jointly with the observed ex ante relation $\succeq$, satisfies DC.

I proceed along similar lines to elicit an optimally inattentive DM's conditional choices. The DM ex ante maximize a preference $\succsim$ over menus and ex post chooses according to a conditional choice correspondence c. I propose a consistency condition that relates $\succsim$ to c. As is the case for DC, one cannot directly test the required consistency condition from ex ante data alone. If one is nevertheless willing to assume consistency, Theorem 6 shows that the choices of an optimally inattentive DM can be elicited from $\succsim$.[32]

Formally, the DM has a complete and transitive binary relation $\succsim$ on $K(\mathcal{F})$, with $\succ$ denoting strict preference and $\sim$ denoting indifference. Her ex ante choice of menu maximizes this preference relation. Our main goal in this section is to elicit the DM's conditional choice correspondence from $\succsim$, not provide a representation theorem; for an elegant representation theorem in this framework, see de Olivera et al. (2016).

Conditional choices are elicited as follows. For any menu B and state ω, let

$$B_\omega = \Big\{ x_B\{\omega\}g : g \in B \Big\}$$

for any $x_B \in X$ such that $\{f(\omega')\} \succ \{x_B\}$ for all $f \in B$ and $\omega' \in \Omega$.

[31]This is well-defined provided that $\succeq$ satisfies Completeness, Transitivity and Savage's P2.

[32]The Supplemental Appendix gives assumptions on $\succsim$ alone that allow elicitation of a conditional choice correspondence that is consistent with $\succsim$.

Definition 5. The DM's *anticipated choice from the menu B in the state ω* is the set

$$c^A(B|\omega) = \left\{ f \in B : \{f\} \bigcup B_\omega \sim B \right\}.$$

For an intuition, observe that the menu B_ω contains only acts that give a lower payoff in state ω than anything in B. Thus, she is strictly worse off when facing B_ω than when facing B. If she anticipates choosing f in state ω, then she can nevertheless get the same outcome in every state when facing $B_\omega \bigcup\{f\}$ as when facing B, so she should express $B_\omega \bigcup\{f\} \sim B$. When she does not anticipate choosing f, she gets a worse outcome in state ω from $B_\omega \bigcup\{f\}$ than from B, so she expresses $B \succ B_\omega \bigcup\{f\}$. Therefore, she expresses $B_\omega \bigcup\{f\} \sim B$ only if she anticipates choosing f from B in state ω.

For this to be meaningful, the ex ante preference must reflect her ex post choices. I propose the following relationship between the preference relation $\succsim$ and the conditional choice correspondence $c(\cdot)$.

Definition 6. The pair $(\succsim, c)$ is *consistent* when for any $A, B \in K(\mathcal{F})$, if $c(B|\omega') \bigcap A \neq \emptyset$ for each $\omega' \in \Omega$, then $A \succsim B$; if, in addition, there exists ω^* such that $c(A|\omega^*) \bigcap B = \emptyset$ and

$$c(A|\omega) \neq c(A|\omega') \iff c(B|\omega) \neq c(B|\omega')$$

for every $\omega, \omega' \in \Omega$, then $A \succ B$.

For an intuition, suppose that the DM always chooses an act in A when facing B. When facing A, she can replicate her choice from B, state-by-state, or perhaps make better choices. If she correctly anticipates this, then she should weakly prefer the menu A to the menu B. Moreover, if some of the acts she chooses from A are not available in B and her choices from B require the same information to make as those from A, then she is strictly better off with A than with B.[33]

Theorem 6. *If $(\succsim, c)$ is consistent and $c(\cdot)$ has an Optimal Inattention Representation, then there exists an open, dense $K \subset K(\mathcal{F})$ such that for each $B \in K$ and all ω,*

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

Theorem 6 provides conditions under which anticipated choice agrees with actual choice from almost every menu. Moreover, the definition of $c^A(\cdot)$ depends only on ex ante preference. Thus, an optimally inattentive DM's conditional choices can be inferred from her ex ante preference over menus via $c^A(\cdot)$, as long as one is willing to take consistency as given.

The set K contains the menus with unique optimal information partitions to which DM anticipates paying attenion. In the remaining menus, one can only identify a set of potential choices, all of which are equally good. This identification relies on similar ideas, but must

[33]This second condition accounts for "ties" in the optimal information. If A has a unique optimal partition, then the second clause can be replaced with $c(A|\omega^*) \bigcap B = \emptyset$ for some $\omega^* \in \Omega$ implies $A \succ B$.

be formalized differently. See the Supplementary Appendix for details, including how to use anticipated choices to construct a representation of $\succsim$.

7. Discussion

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), Gul et al. (2017), and Saint-Paul (2011) all consider consumers who conform exactly to the optimal inattention model.

Van Zandt (1996) studies choice behavior under hidden information acquisition, which is readily reinterpreted as inattention. He provides a negative result with a sparser set of primitives, namely that the model has no testable implications. Specifically, he takes as given any choice correspondence on a finite collection of alternatives. He shows that one can construct a state space, utility function, and an information acquisition problem so that in a fixed state, the resulting best alternatives equal the choice correspondence. At the cost of taking a richer set of primitives, I not only derive testable implications but also achieve substantial uniqueness.

de Olivera et al. (2016) study rational inattention as revealed by a DM's ex-ante preference over menus of acts.[34] The 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. These approaches are complementary and highlight different aspects of the problem. To illustrate the distinction, consider a consumption-savings problem where the agent faces two choices: first, an ex-ante choice of how much to save and second, an ex-post choice of what to consume, and in between she receives information about her consumption options. The approach of de Olivera et al. (2016) imposes properties over the savings decision and infers, via the representation, what she will consume in each state. In contrast, the present paper takes as primitive her state-by-sate consumption and infers what she would choose to save. Some axioms are similar; for instance, their Weak Singleton Independence axiom is very similar to ACI. However, some of the key axioms have no analogs. For instance, their "Aversion to Randomization" axiom plays a central role in their characterization but has no analog in mine, just as INRA plays a central role in my characterization but has no analog in theirs.

Caplin and Dean (2015) develop a related framework for testing rational inattention. Their work is also complementary to the present paper, in that it is better designed for testing in the laboratory but does not achieve as precise identification. As in this paper, they

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

study conditional choice, but choice is stochastic and the DM's prior and utility index are also primitives. The conditions that characterize the model are very different, namely two inequalities that relate the agent's choices, utility, prior, and inferred attention.

Another recent literature studies choice with fixed but unobserved information. Dillenberger et al. (2014) study the ex-ante preference over menus of acts who has or anticipates receiving information before making her choice from the menu. Lu (2016) studies the unconditional distribution of choices from a menu of acts when the DM acquires fixed but unobserved information. As in this paper, the information must be inferred from choice data, but the information is fixed. Consequently, the models are most closely related to the fixed attention special case, but neither focuses on interpreting this behavior as inattention.

By way of conclusion, I compare the optimal inattention model with some non-axiomatic 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.[35] The interpretation that fits best with the framework of the present paper is that the agent has access to arbitrarily precise, and arbitrarily imprecise, signals about the state of the world on which she conditions her action, and but the modeler does not observe the realization of this information. If the information in the economy were explicitly modeled and the possible signals incorporated within the state space, then conditional choices are deterministic and the interpretation fits within my model. Another interpretation is that the agent has perfect information but her perception of this information is stochastic and by exerting effort can decreases the randomness of her perception. This second interpretation cannot be accommodated within my framework.

Mankiw and Reis (2002) introduce the sticky information model.[36] It postulates that agents update their information infrequently, and when they update, they obtain perfect information. If, as in Reis (2006), the agent chooses whether to update at a cost, this model is the special case of optimal inattention in a static setting. Specifically, the cost of the information the agent had in the previous period is zero and the costs of all other partitions are equal but positive. If, as in Mankiw and Reis (2002), updating is exogenous, then the model is a special case of fixed attention in a static setting, given a realization of the update process. However, the key interest of this line of research is in a dynamic, not static, setting.

Gabaix (2014) introduces the sparse max model and shows it has interesting implications for consumer theory. The realized true parameter in sparse max plays the same role as the

[35]Recently, Matejka and McKay (2015) 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.

[36]This formalization is based on Gabaix and Laibson (2001) which can be interpreted similarly, though they do not focus on interpreting their stickiness as information.

true state in optimal inattention. Rather than choosing a coarser information partition, the agent chooses a sparse parametrization of the model. Unlike optimal inattention, the choice of the sparse parametrization is only approximately optimal in the sparse max, making the model very tractable. Both agents then make the optimal choice given the realization of the subjective information or sparse parameter.

Appendix A. Proofs

A.1. **Proofs from Section 4.**

A.1.1. *Sketch for Proof of Theorem 1:* Before turning to the details, I provide a detailed sketch of the arguments that will be used to establish sufficiency of the axioms for the representation. The full proof can be found in the next subsection. I focus, for this subsection only, on the case where $\Omega = \{\omega_1, \omega_2\}$.

The goal will be to map conditional choices into a preference relation that has the noted representation. The preference is constructed over "plans", or mappings from states to acts representing what the DM will do in each state. The key to doing so is INRA, which ensures that we can define such a preference in a way that makes sense.

As a preliminary step, there exists an expected utility representation for choice over lotteries, with an unbounded utility index u.

Given two states, one can write the plan to choose $f_i \in \mathcal{F}$ in state ω_i as (f_1, f_2) or as F with $F(\omega_i) = f_i$. The elicitation of preference over plans is based on the idea that if $f \in c(B|\omega_1)$ and $g \in c(B|\omega_2)$, then the plan (f, g) is preferred to any plan (h_1, h_2) with $h_1, h_2 \in B$. INRA implies that this preference is well-defined and actually captures the choices of the DM: given the choices above, if $f, g, h_1, h_2 \in B'$, then (h_1, h_2) is not chosen from B'. INRA alone is enough to define such a preference, and the remainder of the proof shows the other axioms imply that it has enough regularity properties to represent as optimal inattention.

Because I do not directly observe preference over plans, some plans are never chosen from any menu. For instance, if $u(f_1(\omega)) \geq u(f_2(\omega))$ for all ω, then (f_1, f_2) is never chosen from any problem. For any plan F, I construct a stand-in plan $\bar{F}$ that is chosen from $\{\bar{F}(1), \bar{F}(2)\}$ and will be chosen from $\{\bar{F}(1), \bar{F}(2)\} \bigcup B$ whenever F is chosen from B (Lemma 4). Roughly, $\bar{F}(\omega)$ differs from $F(\omega)$ by decreasing its payoff in the states where it would not be not chosen. INRA and Monotonicity ensure that adding these modified acts to the menu do not change the DM's choices, and SC ensures that $\bar{F}(\omega)$ is chosen whenever $F(\omega)$ is. As a consequence, only the realized outcomes from a plan and the information used to implement it matter for preference, i.e. the vector $(u(f_1(1)), u(f_2(2)))$ and the algebra $\sigma((f_1, f_2))$ suffice for ranking (f_1, f_2).

I further extend the preference via sequences similar to those used in the strong axiom of revealed preference (the relation IS from the text) and also using plans chosen from "nearby" menus (the closure of IS, denoted $\bar{IS}$). I denote the resulting relation $\succeq$.

While $\succeq$ is typically incomplete, the axioms imply it has useful regularity properties. INRA, Monotonicity, and ACI imply that $\succeq$ is transitive (Lemma 6). ACI implies that $\succeq$ satisfies "translation invariance": for any acts f, g and plans F, G, if $\alpha F + (1-\alpha)(f,f) \succeq \alpha G + (1-\alpha)(f,f)$, then $\alpha F + (1-\alpha)(g,g) \succeq \alpha G + (1-\alpha)(g,g)$ (Lemma 7). Intuitively, this means that if the utility of two plans is shifted by the same amount in each state, then the DM does not reverse her preference. Finally, Continuity, along with the above two properties, allows me to show that, restricted to constant plans, there is a probability measure that represents choice in the following sense: if $(f,f) \succeq (g,g)$, then $\int u \circ f d\pi \geq \int u \circ g d\pi$ (Lemma 9).

Since $\succeq$ may be incomplete, I directly use the above properties to construct the cost function (Lemma 10). Incompleteness rules out taking the cost function equal to the concave conjugate. To find the cost of the partition Q, I compute the expected utility, according to u and π, of the outcomes given by the worst plan requiring Q to implement that is preferred to the lottery that gives zero utility for sure. If no such plan exists, then the cost of paying attention to Q is infinite. To conclude, I use Continuity and INRA to establish that the representation of plans maps back into choices.

A.1.2. ***Proof of Theorem 1:*** First, I introduce some notation. Let $K(X)$ be the set of compact, non-empty subsets of X, noting that $K(X) \subset K(\mathcal{F})$. Let Σ be the set of all subsets of Ω. For any $f, g \in \mathcal{F}$ and any event $E \subset \Omega$, let $fEg \in \mathcal{F}$ be such that $fEg(\omega) = f(\omega)$ if $\omega \in E$ and $fEg(\omega) = g(\omega)$ if $\omega \notin E$. Let $\mathcal{F}^\Omega$ be the set of functions from Ω to $\mathcal{F}$; recall $\sigma(\{E_i\}_{i\in I})$ is the smallest $\sigma-$algebra containing $\{E_i\}_{i\in J}$ and $\sigma(f)$ is the smallest $\sigma-$algebra by which f is measurable. I refer to elements of $\mathcal{F}^\Omega$ as "plans" with the interpretation that the DM plans to choose $F(\omega)$ in state ω. I denote elements of X by x, y, z , elements of $\mathcal{F}$ by f, g, h and elements of $\mathcal{F}^\Omega$ by F, G, H and elements of $K(\mathcal{F})$ by A, B, C. With usual abuse of notation, identify $\mathcal{F}$ with the subset of $\mathcal{F}^\Omega$ that does not vary with the state, so $X \subset \mathcal{F} \subset \mathcal{F}^\Omega$. Define mixtures between elements of $\mathcal{F}^\Omega$ state-by-state, so $[\alpha F + (1-\alpha)G](\omega) = \alpha F(\omega) + (1-\alpha)G(\omega)$.

The proof proceeds as a sequence of Lemmas.

Lemma 1. *There exists an affine, continuous, unbounded $u : X \to \mathbb{R}$ 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. Observe that by definition, $x \succeq^R y$ if and only if $x(\omega') \succeq^R y(\omega')$ for all ω'. First, for any $B \in K(X)$, $c(B|\omega') = \max_{\succeq^R} B$ for all ω': by Monotonicity if $y \in c(B|\omega)$ and $x \succeq^R y$, then $x \in c(B|\omega)$, and if $x \succ^R y$, then $y \notin c(B|\omega)$. Second, $c(\cdot)$ satisfies WARP when restricted to X. Now, fix any $A \subset B \in K(X)$. By the claim, $c(B|\omega) \bigcap A \neq \emptyset \iff c(B|\omega') \bigcap A \neq \emptyset$. So if $c(B|\omega) \bigcap A \neq \emptyset$, $c(B|\omega) \bigcap A = c(A|\omega)$ by INRA. Conclude $\succeq^R$ is complete and transitive.

The weak order $\succeq^R$ satisfies the remaining axioms of Herstein and Milnor (1953); continuity follows from Continuity. Denoting the symmetric part of $\succeq^R$ by $\sim^R$, it remains to show that $a \sim^R a' \iff \frac{1}{2}a + \frac{1}{2}b \sim^R \frac{1}{2}a' + \frac{1}{2}b$. If not, then there exist a, a', b with $a \sim^R a'$ and $\frac{1}{2}a + \frac{1}{2}b \succ^R \frac{1}{2}a' + \frac{1}{2}b$. By ACI, $a = \frac{1}{2}a + \frac{1}{2}a \succ^R \frac{1}{2}a' + \frac{1}{2}a$ and $\frac{1}{2}a + \frac{1}{2}a' \succ^R \frac{1}{2}a' + \frac{1}{2}a' = a'$. Transitivity gives a contradiction, and similar arguments yield the converse. By Unboundedness, $\succeq^R$ is non-degenerate. Therefore, $\succeq^R$ also satisfies the assumptions of (Grandmont, 1972, Thm 2). Therefore an affine, continuous $u : X \to \mathbb{R}$ exists. Because of the argument that follows the statement of Unboundedness in the main text of the paper, u must be unbounded. □

For any $B \in K(\mathcal{F})$, let $\hat{P}(B)$ be the partition defined by Equation (3).

Lemma 2. *For any $B \in K(\mathcal{F})$, all $f, h \in B$ and act g such that $u \circ f \geq u \circ g$, $h \in c(B|\omega) \iff h \in C(B \bigcup \{g\}|\omega)$; moreover, if $f \in c(B|\omega)$ and $u(h(\omega')) = u(f(\omega'))$ for all $\omega' \in \hat{P}(\omega)$, then $h \in c(B|\omega)$.*

Proof. Let $f \in B$ and suppose $u \circ f \geq u \circ g$. Then $f(\omega)$ is revealed preferred to $g(\omega)$ for all ω. By monotonicity, $g \in c(B \bigcup \{g\}|\omega')$ implies $f \in c(B \bigcup \{g\}|\omega')$. Therefore, $c(B \bigcup \{g\}|\omega') \bigcap B \neq \emptyset$ for all ω. Hence $c(B|\omega') = c(B \bigcup \{g\}|\omega') \bigcap B$.

For the second part, fix any $f \in c(B|\omega)$ and $h \in B$ with $u(h(\omega')) = u(f(\omega'))$ for all $\omega' \in \hat{P}(B)(\omega)$. Consider $g \in \mathcal{F}$ such that $g(\omega') = h(\omega')$ for all ω' for all $\omega' \in \hat{P}(B)(\omega)$ and $g(\omega') = f(\omega')$ otherwise. Then, $u \circ f \geq u \circ g$, and applying the first part gives that $f \in c(B \bigcup \{g\}|\omega)$. By Monotonicity, $g \in c(B \bigcup \{g\}|\omega)$, and by SC, $h \in c(B \bigcup \{g\}|\omega)$. By INRA, $h \in c(B|\omega)$. □

Define $\hat{c} : K(\mathcal{F}) \twoheadrightarrow \mathcal{F}^\Omega$ by $F \in \hat{c}(B)$ if and only if both $F(\omega) \in c(B|\omega)$ for every ω and $\sigma(F) \subset \sigma(\hat{P}(B))$. Since $\sigma(\hat{P}(B)) \subset \sigma(P)$, any $\sigma(\hat{P}(B))$-measurable selection from $c(B|\cdot)$ is in $\mathcal{F}^\Omega$. Measurability ensures that every plan in $\hat{c}(B)$ requires no more information to follow than $\hat{P}(B)$ and maintains indifference between $f, g \in c(B|\omega)$. For any $F \in \mathcal{F}^\Omega$, let $Im(F)$ be the image F, i.e. $Im(F) = \{F(\omega) : \omega \in \Omega\}$. By INRA, if $F \in \hat{c}(B)$ then $F \in \hat{c}(Im(F))$; I sometimes write ImF instead of $Im(F)$.

Define $\mathbb{P}^* = \{\hat{P}(B) : B \in K(\mathcal{F})\}$, the set of partitions to which the DM sometimes attends. For the act g, we construct a plan g_Q that is chosen from $Im(g_Q)$, the DM pays attention to Q when facing $Im(g_Q)$, and she gets an outcome yielding a utility of $\frac{1}{2}u(g(\omega)) + \frac{1}{2}u(0)$ in the state 0. One can adjust g to mimic the outcome, state-by-state, of choosing any plan that requires paying attention to Q.

Lemma 3. *For each $Q \in \mathbb{P}^*$, there exists $x \in X$, $B_Q = \{f_\omega : \omega \in \Omega\} \subset \mathcal{F}$ with $f_\omega = f_{\omega'}$ whenever $\omega' \in Q(\omega)$ such that:*

(i) $u\left(\frac{1}{2}f_\omega + \frac{1}{2}x\right)(\omega') = 0$ for all $\omega' \in Q(\omega)$, and

(ii) $\{\frac{1}{2}f_\omega + \frac{1}{2}x\} = c(\frac{1}{2}B_Q + \frac{1}{2}x|\omega)$ for all ω.

Then, for any $g \in \mathcal{F}$, $\{\frac{1}{2}f_\omega + \frac{1}{2}g\} = c(\frac{1}{2}B_Q + \frac{1}{2}g|\omega)$ for all ω. Call the resulting plan g_Q.

Proof. Pick any $Q \in \mathbb{P}^*$ and B such that $\hat{P}(B) = Q$. By Lemma 2, INRA and SC, it is without loss to assume that $c(B|\omega)$ is a singleton, so for each ω, there exists g_ω so that $\{g_\omega\} = c(B|\omega)$. Let $u(x) = 0$ and define an act f'_ω so that $u(f'_\omega(\omega')) = 4u(g_\omega(\omega'))$, and observe that $u \circ [\frac{1}{2}\left[\frac{1}{2}f'_\omega + \frac{1}{2}x\right] + \frac{1}{2}x] = u \circ g_\omega$. Set $B'_Q = \{\frac{1}{2}f'_\omega + \frac{1}{2}x : \omega \in \Omega\}$ and then by Lemma 2 and INRA,

$$g_{\omega'} \in c(B|\omega) \iff \frac{1}{2}\left[\frac{1}{2}f'_\omega + \frac{1}{2}x\right] + \frac{1}{2}x \in c\left(\frac{1}{2}B'_Q + \frac{1}{2}x|\omega\right).$$

Let h be an act such that $u(h(\omega')) = -4u(g_{\omega'}(\omega'))$. For $f_\omega = \frac{1}{2}f'_\omega + \frac{1}{2}h$, $B_Q = \{f_\omega : \omega \in \Omega\}$ has the desired properties, since ACI gives that

$$\begin{aligned}\frac{1}{2}f_\omega + \frac{1}{2}x = \frac{1}{2}\left[\frac{1}{2}f'_\omega + \frac{1}{2}h\right] + \frac{1}{2}x &\in c\left(\frac{1}{2}B_Q + \frac{1}{2}x|\omega\right)\\ \iff \frac{1}{2}\left[\frac{1}{2}f'_\omega + \frac{1}{2}x\right] + \frac{1}{2}x &\in c\left(\frac{1}{2}B'_Q + \frac{1}{2}x|\omega\right)\end{aligned}$$

and $u(f_\omega(\omega)) = 0$ for all ω. The final statement follows by using ACI to replace x with g. ☐

Now, let

$$\mathcal{H} = \bigcup_{Q \in \mathbb{P}^*} \{F \in \mathcal{F}^\Omega : F \text{ } is \text{ } \sigma(Q) - measurable\}.$$

For any $F \in \mathcal{H}$, let $\bar{F} = g_Q$ for the $Q \in \mathbb{P}^*$ such that $\sigma(Q) = \sigma(F)$ and an act g such that $u(g(\omega)) = 2u(F(\omega)(\omega))$ for all ω.[37] Thus, $\bar{F} \in \hat{c}(Im(\bar{F}))$, $u(\bar{F}(\omega)(\omega)) = u(F(\omega)(\omega))$ for all ω, and $\sigma(\bar{F}) = \sigma(Q)$. The next lemma shows that when F is chosen from $Im(F)$, then the plan $\bar{F}$ is indifferent to the plan F.

Lemma 4. *If $F \in \hat{c}(Im(F))$, then $Im(\bar{F})$ IS $Im(F)$ and $Im(F)$ IS $Im(\bar{F})$.*

Proof. Suppose $F \in \hat{c}(Im(F))$ and that $\hat{P}(Im(F)) = Q$. Let x be the element of $\{f(\omega) : \omega \in \Omega, f \in Im(F) \bigcup Im(\bar{F})\}$ that minimizes $u(\cdot)$. Let $F^*_x \in \mathcal{H}$ be such that

$$F^*_x(\omega)(\omega') = \begin{cases} F(\omega)(\omega') & if\ \omega' \in Q(\omega) \\ x & if\ \omega' \notin Q(\omega) \end{cases}.$$

Note that $\bar{F} \in \hat{c}(Im(\bar{F}))$, so the first part of Lemma 2 implies that $\bar{F} \in \hat{c}(Im(\bar{F}) \bigcup Im(F^*_x))$. Moreover, $\hat{P}(Im(\bar{F}) \bigcup Im(F^*_x)) \gg \hat{P}(Im(\bar{F}))$. By the second part of Lemma 2, $F^*_x \in$

[37]By Lemma 2, whenever $u \circ g_1 = u \circ g_2$, $\frac{1}{2}B_Q + \frac{1}{2}g_i$ *IS* $\frac{1}{2}B_Q + \frac{1}{2}g_j$ for any $i, j \in \{1, 2\}$.

$\hat{c}(Im(\bar{F})\bigcup Im(F_x^*))$ and $F_x^*, F \in \hat{c}(Im(F)\bigcup Im(F_x^*))$. Hence

$$Im(\bar{F})\ IS\ Im(F)\ IS\ Im(\bar{F}),$$

completing the proof. □

Let $\bar{IS}$ be the sequential closure of IS, that is $A\,\bar{IS}\,B$ if and only if there exists $A_n \to A$, $B_n \to B$, and $A_n\,IS\,B_n$ for all n. Let $\succeq$ be defined by $F \succeq G$ if and only if either $Im(\bar{F})\ \bar{IS}\ Im(\bar{G})$ or $G \notin \mathcal{H}$. $\succeq$ relates to choice as follows.

Lemma 5. *Suppose $F \in \hat{c}(B)$. If $Im(G) \subset B$, then $F \succeq G$.*

Proof. If $G \notin \mathcal{H}$, then $F \succeq G$. Otherwise, by Lemma 2, $F \in \hat{c}(B\bigcup\{\bar{G}\})$ so $Im(F)\,IS\,\{\bar{G}\}$. Since $\bar{F}\,IS\,F$, $\bar{F}\,IS\,\bar{G}$, so $F \succeq G$. □

The remainder of the proof collects properties of $\succeq$ and then proves a representation theorem. The following Corollary of Lemma 4 shows that knowing the utility of the outcome given by a plan and the information required to follow it suffices for preference.

Corollary 5. *For $F, G \in \mathcal{H}$: if $u\left(F(\omega)(\omega)\right) = u\left(G(\omega)(\omega)\right)$ for all $\omega \in \Omega$ and $\sigma(F) = \sigma(G)$, then $F \sim G$.*

Proof. Define F_x^* as in Lemma 4 for $u(x) < \min\{u(f(\omega)) : f \in Im(\bar{F})\bigcup Im(\bar{G}),\ \omega \in \Omega\}$. Then, the arguments in the Lemma show $Im(H)\ IS\ Im(F_x^*)\ IS\ Im(H')$ for $H, H' \in \{\bar{F}, \bar{G}\}$, since $H'' \in \hat{c}(Im(H''))$ for each $H'' \in \{\bar{F}, \bar{G}\}$. The conclusion follows immediately. □

Lemma 6. *$\succeq$ is a preorder.*

Proof. Reflexivity is trivial. To see $\succeq$ is transitive, fix any finite $A, B, C \in K(\mathcal{F})$. Suppose $A\,\bar{IS}\,B$ and $B\,\bar{IS}\,C$, so there are sequences A_n, B_n, B'_n, C_n converging to A, B, B, C respectively with $A_n\,IS\,B_n$ and $B'_n\,IS\,C_n$ for all n. Pick $0, (-2), 2 \in X$ such that $u(0) = 0$, $u(-2) = -2$ and $u(2) = 2$. Pick $\bar{x}, \underline{x} \in X$ so that $u(\bar{x}) = \max\{2u(f(\omega)) : \omega \in \Omega, f \in A\bigcup C\} + 1$ and $u(\underline{x}) = \min\{2u(f(\omega)) : \omega \in \Omega, f \in A\bigcup C\} - 1$. For any $f \in \mathcal{F}$ with $\frac{1}{2}u(\underline{x}) \leq u \circ f \leq \frac{1}{2}u(\bar{x})$, let $f^* \in \mathcal{F}$ be so that $f^*(\omega) \in co\{\bar{x}, \underline{x}\}$ and $\frac{1}{2}f^*(\omega) + \frac{1}{2}0 \sim f(\omega)$.

Because $u(\cdot)$ is continuous, A, B, C are finite, and each act in $A\bigcup B\bigcup C$ is simple, for every $\epsilon > 0$ there exists N_ϵ such that

$$\max_{f\in D}\min_{f_n\in D_n}\max_{\omega\in\Omega} u(f_n(\omega)) - u(f(\omega)) < \epsilon$$

for all $n > N_\epsilon$ and for $D = A, B, C$. WLOG (by taking a subsequence), $N_{(\frac{1}{n})} = n$. Now, consider the sequence $D_1^n, ..., D_m^n$ that reveals $A_n\,IS\,B_n$. By Monotonicity and Lemma 2, instead consider $\frac{1}{2}\hat{D}_1^n + \frac{1}{2}0, ..., \frac{1}{2}\hat{D}_m^n + \frac{1}{2}0$, where each $\hat{D}_i^n$ contains acts with exactly twice the utility of the acts in D_i^n. Similarly, consider $\frac{1}{2}\hat{E}_1^n + \frac{1}{2}0, ..., \frac{1}{2}\hat{E}_k^n + \frac{1}{2}0$ instead of $E_1^n, ..., E_k^n$ that reveals $B'_n\,IS\,C_n$. Note that in either sequence, one can replace 0 with another $x \in X$ and shift the utility up or down by $\frac{1}{2}u(x)$ via ACI.

Now, define

$$A'_n(\alpha) = \{(1-\alpha)f + \frac{\alpha}{2}f^* + \frac{\alpha}{2}(2) : f \in A_n\}$$

and

$$C'_n(\alpha) = \{(1-\alpha)f + \frac{\alpha}{2}f^* + \frac{\alpha}{2}(-2) : f \in A_n\}.$$

Note $u([(1-\alpha)f + \frac{\alpha}{2}f^* + \frac{\alpha}{2}(2)](\omega)) = u(f(\omega)) + \alpha$. By ACI Lemma 2, $f \in c(A_n|\omega)$ if and only if $(1-\alpha)f + \frac{\alpha}{2}f^* + \frac{\alpha}{2}(2) \in c(A'_n(\alpha)|\omega)$. Similarly for C'_n.

For $\epsilon \in [-2, 2]$, let ϵ correspond to the mixture between -2 and 2 that yields $u(\epsilon) = \epsilon$. Applying ACI successively to each $\frac{1}{2}\hat{D}^n_i + \frac{1}{2}(0)$ and $\frac{1}{2}\hat{D}^n_{i+1} + \frac{1}{2}(0)$ shows that $\frac{1}{2}\hat{D}^n_1 + \frac{1}{2}(\frac{2}{n+1})\ IS\ \frac{1}{2}\hat{D}^n_m + \frac{1}{2}(\frac{2}{n+1})$ and by Monotonicity and INRA, $A'_n(\frac{1}{n})\ IS\ \frac{1}{2}\hat{D}^n_1 + \frac{1}{2}(\frac{2}{n+1})$. Similarly, $\frac{1}{2}E^n_1 + \frac{1}{2}(\frac{-2}{n+1})\ IS\ \frac{1}{2}E^n_k + \frac{1}{2}(\frac{-2}{n+1})$ and $\frac{1}{2}E^n_k + \frac{1}{2}(\frac{-2}{n+1})\ IS\ C'_n(\frac{1}{n})$. Moreover, by Lemma 2, $\frac{1}{2}\hat{D}^n_m + \frac{1}{2}(\frac{2}{2n})\ IS\ \frac{1}{2}E^n_1 + \frac{1}{2}(\frac{-2}{n})$. Since IS is transitive, $A'_n(\frac{1}{n})\ IS\ C'_n(\frac{1}{n})$ and thus since $A'_n(\frac{1}{n}) \to A$ and $C'_n(\frac{1}{n}) \to C$, conclude that $A\ \bar{IS}\ C$.

To see that $\succeq$ is a preorder, pick any $F, G, H \in \mathcal{H}$ so that $F \succeq G \succeq H$. Note $Im(\bar{F}), Im(\bar{G}), Im(\bar{H})$ are finite and apply the above. Conclude $F \succeq H$. The result is trivial if one of F, G, H does not belong $\mathcal{H}$. □

Recall that if $f \in \mathcal{F}$ and $F \in \mathcal{F}^\Omega$, then $\alpha F + (1-\alpha)f$ is the plan that gives $\alpha F(\omega) + (1-\alpha)f$ in state ω.

Lemma 7. *For any $F, G \in \mathcal{F}^\Omega$ and any $f, g \in \mathcal{F}$:*
if $\alpha F + (1-\alpha)f \succeq \alpha G + (1-\alpha)f$, then $\alpha F + (1-\alpha)g \succeq \alpha G + (1-\alpha)g$.

Proof. I show first that for any finite $A, B \in K(\mathcal{F})$,

$$\alpha A + (1-\alpha)\{g\}\, IS\, \alpha B + (1-\alpha)\{g\} \implies \alpha A + (1-\alpha)\{h\}\, IS\, \alpha B + (1-\alpha)\{h\}.$$

If $\alpha A + (1-\alpha)\{g\}\, IS\, \alpha B + (1-\alpha)\{g\}$, then there exists a sequence $C_1, ..., C_m$ so that $C_1 = \alpha A + (1-\alpha)\{g\}$ and $C_m = \alpha B + (1-\alpha)\{g\}$ satisfying $c(C_i|\omega) \subseteq C_{i-1}$ for all ω and $i > 1$. By Lemma 1, for every h there exists $\hat{h}$ so that $u(\hat{h}(\omega)) = \frac{1}{\alpha}u(f(\omega)) - \frac{1-\alpha}{\alpha}u(g(\omega))$. Therefore, $u \circ [\alpha\bar{h} + (1-\alpha)g] = u \circ h]$.

Consider C_i and $\hat{C}_i = \{\hat{h} : h \in C_i\}$ for $i > 1$, with $\hat{C}_1 = \{\hat{h} : h \in \bigcup_\omega c(C_2|\omega)$. Since $\hat{c}(C_2) \in C_1$, $f \in c(C_1|\omega)$ implies $f \in c(C_1 \bigcup \hat{C}_1|\omega)$ by Lemma 2. For $i > 1$, Lemma 2 gives $c(C_i|\omega) = c(C_i \bigcup \hat{C}_i|\omega) \bigcap C_i$ and $c(\hat{C}_i|\omega) = c(\hat{C}_i \bigcup C_i|\omega) \bigcap \hat{C}_i$; moreover, Monotonicity and Lemma 1 give that $h \in c(C_i \bigcup \bar{C}_i|\omega) \iff \hat{h} \in c(C_i \bigcup \hat{C}_i|\omega)$. Hence $h \in c(C_i|\omega) \iff \hat{h} \in c(\hat{C}_i|\omega)$.

Now, $D_1 = C_1, D_2 = \hat{C}_1, ..., D_{n+1} = \hat{C}_n, D_{n+2} = C_n$ satisfy $c(D_i|\omega) \subset D_{i-1}$ for all ω and $i > 1$. Moreover, each D_i can be written as $\alpha D'_i + (1-\alpha)\{g\}$. ACI implies that $\alpha f + (1-\alpha)g \in c(\alpha D'_i + (1-\alpha)\{g\}|\omega)$ if and only if $\alpha f + (1-\alpha)h \in c(\alpha D'_i + (1-\alpha)\{h\}|\omega)$. Therefore, $\alpha A + (1-\alpha)\{h\}\, IS\, \alpha B + (1-\alpha)\{h\}$.

Now, if $\alpha A + (1-\alpha)\{g\} \, \overline{IS} \, \alpha B + (1-\alpha)\{g\}$, there is a sequence A_n and B_n of choice problems so that $A_n \to \alpha A + (1-\alpha)\{g\}$, $B_n \to \alpha B + (1-\alpha)\{g\}$, and $A_n \; IS \; B_n$. By Lemmas 1 and 2, it is without loss of generality to take $A_n = \alpha A'_n + (1-\alpha)\{g\}$ and $B_n = \alpha B'_n + (1-\alpha)\{g\}$ where $A'_n \to A$ and $B'_n \to B$. By the above, $\alpha A'_n + (1-\alpha)\{g\} \; IS \; \alpha B'_n + (1-\alpha)\{g\}$ implies $\alpha A'_n + (1-\alpha)\{h\} \; IS \; \alpha B'_n + (1-\alpha)\{h\}$, and since $A'_n \to A$ and $B'_n \to B$, $\alpha A + (1-\alpha)\{h\} \; \overline{IS} \; \alpha B + (1-\alpha)\{h\}$.

Pick any plans F, G and acts f, g with $\alpha F + (1-\alpha) f \succeq \alpha G + (1-\alpha) f$ with $\sigma(F) = \sigma(Q)$ and $\sigma(G) = \sigma(Q)$. Recall $Im(\overline{\alpha F + (1-\alpha) f}) = \frac{1}{2} B_Q + \frac{1}{2} f^*$ and $Im(\overline{\alpha G + (1-\alpha) f}) = \frac{1}{2} B_R + \frac{1}{2} g^*$ for some $f^*, g^* \in \mathcal{F}$, and that both are finite. By unbounded, there exist $f', g', \hat{f} \in \mathcal{F}$ so that $u \circ f^* = u \circ \left[\alpha f' + (1-\alpha)\hat{f}\right]$, $u \circ g^* = u \circ \left[\alpha g' + (1-\alpha)\hat{f}\right]$ and $u \circ \hat{f} = 2u \circ f$. By Corollary 5,

$$\frac{1}{2} B_Q + \frac{1}{2}\left[\alpha f' + (1-\alpha)\hat{f}\right] \; \bar{IS} \; \frac{1}{2} B_Q + \frac{1}{2} f^*$$

and

$$\frac{1}{2} B_R + \frac{1}{2} g^* \; \bar{IS} \; \frac{1}{2} B_R + \frac{1}{2}\left[\alpha g' + (1-\alpha)\hat{f}\right].$$

Now, $\alpha F + (1-\alpha) f \succeq \alpha G + (1-\alpha) f$ and Lemma 6 imply that

$$\frac{1}{2} B_Q + \frac{1}{2}\left[\alpha f' + (1-\alpha)\hat{f}\right] \; \bar{IS} \; \frac{1}{2} B_R + \frac{1}{2}\left[\alpha g' + (1-\alpha)\hat{f}\right].$$

Let $\hat{g} \in \mathcal{F}$ be so that $u \circ \hat{g} = 2u \circ g$. By above,

$$\frac{1}{2} B_Q + \frac{1}{2}\left[\alpha f' + (1-\alpha)\hat{g}\right] \; \bar{IS} \; \frac{1}{2} B_R + \frac{1}{2}\left[\alpha g' + (1-\alpha)\hat{g}\right],$$

and by Corollary 5,

$$Im(\overline{\alpha F + (1-\alpha) g}) \; \bar{IS} \; \frac{1}{2} B_Q + \frac{1}{2}[\alpha f' + (1-\alpha)\hat{g}]$$

and

$$\frac{1}{2} B_R + \frac{1}{2}[\alpha g' + (1-\alpha)\hat{g}] \; \bar{IS} \; Im(\overline{\alpha G + (1-\alpha) g}).$$

Applying Lemma 6 gives implies that

$$Im(\overline{\alpha F + (1-\alpha) g}) \; \bar{IS} \; Im(\overline{\alpha G + (1-\alpha) g}),$$

so $\alpha F + (1-\alpha) g \succeq \alpha G + (1-\alpha) g$. □

Corollary 5 and Lemma 7 imply the following useful observation.

Corollary 6. *For any $\phi \in \mathbb{R}^\Omega$ and any $F, G \in \mathcal{F}^\Omega$: if $F \succeq G$, then $F + \phi \succeq G + \phi$ for any $F + \phi, G + \phi \in \mathcal{F}^\Omega$ such that $u((F+\phi)(\omega)(\omega)) = u(F(\omega)(\omega)) + \phi(\omega)$ and $u((G+\phi)(\omega)(\omega)) = u(G(\omega)(\omega)) + \phi(\omega)$, $\sigma(F) = \sigma(F+\phi)$ and $\sigma(G) = \sigma(G+\phi)$.*

The next two lemmas establish existence of the underlying probability measure π.

Lemma 8. *If $\{x\} \, \bar{IS} \, \{f\}$, then there exist $f_n \to f$ such that $\{x\} \, IS \, \{f_n\}$.*

Proof. Le $y \in X$ be such that $u(y) = u(x) - 1$ and $g \in \mathcal{F}$ such that $u(g_n(\omega)) = 2u(f(\omega)) - u(y) + 1$ for all ω. Define $x_n = \frac{n-1}{n}x + \frac{1}{n}y$, $f'_n = \frac{n-2}{n}f(\omega) + \frac{1}{n}(g_n(\omega)) + \frac{1}{n}y$, noting that $x_n \to x$, $f'_n \to f$, and $u(x_n) = u(x) - \frac{1}{n}$ and $u(f'_n) = u(f(\omega)) + \frac{1}{n}$. Since u is continuous and f is simple, for every n there exists a $\delta_n > 0$ such that $d(z', z) < \delta_n$ implies $|u(z') - u(z)| < \frac{1}{n}$ for $z \in \{x\} \bigcup \{f(\omega) : \omega \in \Omega\}$. Since $\{f\}\, \bar{IS}\, \{x\}$, there are $A_m \to \{x\}$ and $B_m \to \{f\}$ such that $B_m\, IS\, A_m$. Then for every n there exists an M_n such that $m > M_n$ implies $d(b, x) < \delta_n$ and $d(a, f) < \delta_n$. In particular, $m > M_n$ implies $u(x_n) < u(a(\omega))$ and $u(b(\omega)) < u(f'_n(\omega))$ for all $b \in B_m$, $a \in A_m$ and $\omega \in \Omega$. By Lemma 2, $\{f'_n\}\, IS\, \{x_n\}$ for all n. But then defining $f''_n = \frac{n-2}{n-1}f + \frac{1}{n-1}(g_n)$ and $f_n = \frac{n-1}{n}f''_n + \frac{1}{n}x$, $\{\frac{n-1}{n}f''_n + \frac{1}{n}y\}\, IS\, \{\frac{n-1}{n}x + \frac{1}{n}y\}$ by Lemma 2. But then using ACI, $\{\frac{n-1}{n}f''_n + \frac{1}{n}x\}\, IS\, \{\frac{n-1}{n}x + \frac{1}{n}x\} = \{x\}$. Defining $f_n = \frac{n-1}{n}f''_n + \frac{1}{n}x$, $\{f_n\}\, IS\, \{x\}$ for all n and $f_n \to f$. □

Lemma 9. *There is a full-support probability measure π so that for any $f, g \in \mathcal{F}$, if $f \succeq g$, then $\int u \circ f d\pi \geq \int u \circ g d\pi$.*

Proof. Let 0 be the act with $u(0)(\omega) = 0$ for all ω. Define

$$K = \{u \circ f : f \succeq 0, f \in \mathcal{F}\}.$$

Let $K^* = \bar{co}(K)$, the convex closure of K, and define $f\, R\, g$ on $\mathcal{F}$ by $f\, R\, g \iff u \circ f - u \circ g \in K^*$, with R^I the symmetric and R^S the assymetric parts. It is easy to verify that R is reflexive, transitive, monotonic, and suitably continuous. By e.g. Gilboa et al. (2010), there exists a weak* closed and convex set Π of probability measures on Ω so that fRg if and only if $\int u \circ f d\hat{\pi} \geq \int u \circ g d\hat{\pi}$ for all $\hat{\pi} \in \Pi$. By Lemma 7, $f \succeq g$ implies fRg.

Now, adapt the argument of Dubra et al. (2004) to show the existence of a π such that $fR(R^S)g \implies \int u \circ f d\pi \geq (>) \int u \circ g d\pi$. The probability measures on Ω are separable and metrizable by Theorem 15.12 of Aliprantis and Border (2006) (henceforth, AB). Let $\{\pi_i : i \in \mathbb{N}\}$ be a countable dense subset of Π, and define $\pi = \sum \frac{1}{2^i}\pi_i$. Then if fRg, then $\int u \circ f d\pi_i \geq \int u \circ g d\pi_i$ for all i, so $\int u \circ f d\pi \geq \int u \circ g d\pi$. If $fR^S g$, then $\int u \circ f d\hat{\pi} \geq \int u \circ g d\hat{\pi}$ all $\hat{\pi} \in \Pi$ without equality for at least one $\hat{\pi}$. By continuity of the integral and that $\{\pi_i : i \in \mathbb{N}\}$ is dense, there exists $j \in \mathbb{N}$ such that $\int u \circ f d\pi_j > \int u \circ g d\pi_j$ and thus $\int u \circ f d\pi > \int u \circ g d\pi$.

The measure π has full support if and only if $xE0R^S0$ for any arbitrary $x > 0$ and $E \in P$. Suppose not. By Monotonicity, $xE0R0$, so when $\theta = -(xE0)$, it must holds that $\theta R0$. Let f and y be any acts such that $u \circ f = \theta$ and $u(y) = 1$. Note $f \notin c(\{0, f\}|\omega)$ when $\omega \in E$ by Monotonicity, and that $0 \in c(\{0, f\}|\omega)$ and y revealed strictly preferred to $f(\omega)$ for all ω. By Continuity (ii), there exists $\epsilon > 0$ such that for all $f_1, ..., f_m$ and $\alpha_1, ..., \alpha_m$ with $f_i\, IS\, 0$, $\sum \alpha_i u \circ f_i(\omega') > [\epsilon y + (1 - \epsilon)\theta](\omega')$ for some ω'. Now, if $\theta R0$, then $\theta \in K^*$ and for every n, there is a $\theta_n \in co(K)$ such that $\max |\theta_n(\omega) - \theta| < \frac{1}{n}$. Picking n such that $\frac{1}{n} < \min_\omega[(\epsilon y + (1 - \epsilon)\theta)(\omega) - \theta(\omega)]$, there are $f_1^n, ..., f_{m_n}^n \in K$ (and thus slightly abusing notation, also in $\mathcal{F}$) and $\alpha_1^n, ..., \alpha_{m_n}^n \in [0, 1]$ with $\{f_i^n\}\, IS\, \{0\}$ and $\sum \alpha_i^n u \circ f_i^n = \theta_n$. But by the above, there is ω' such that $\theta_n(\omega') = \sum \alpha_i u \circ f_i(\omega') > [\epsilon y + (1-\epsilon)\theta](\omega') > \theta_n(\omega')$, a contradiction. Thus $xE0R^S 0$ and $\pi(E) > 0$. □

The next lemma establishes existence of the cost function. For any $F \in \mathcal{F}$, let $F^* \in \mathcal{F}$ be such that $F^*(\omega) = F(\omega)(\omega)$ for all $\omega \in \Omega$.

Lemma 10. *There is a cost function $\gamma : \mathbb{P} \to \bar{\mathbb{R}}_+$ so that $F \succeq G$ implies that $V(F) = \int u \circ F^* d\pi - \gamma(\hat{P}(Im\bar{F})) \geq \int u \circ G^* d\pi - \gamma(\hat{P}(Im\bar{G})) = V(G)$.*

Proof. Label $0 \in X$ a lottery with $u(0) = 0$. For $Q, R \in \mathbb{P}$, let $M_{Q,R} = \{f \in \mathcal{F} : f_Q \succeq 0_R\}$; recall f_Q is defined before Lemma 4. Define

$$\gamma(Q) = \inf_{\phi \in M_{Q,\{\Omega\}}} \int \phi d\pi$$

(as is standard, the infimum of the empty set is taken to be ∞) and let

$$V(F) = \int u \circ F^* d\pi - \gamma(\hat{P}(Im\bar{F})).$$

The Lemma is true if $F \succeq G$ implies $V(F) \geq V(G)$.

I claim $\inf_{g \in M_{Q,R}} \int u \circ g d\pi \geq \gamma(Q) - \gamma(R)$ when $\gamma(Q) < \infty$. If $\gamma(R) = \infty$, this is trivial, so assume $M_{R,\{\Omega\}} \neq \emptyset$. Let $\hat{g}^n = \frac{1}{2}g^n + \frac{1}{2}0$ be a sequence in $M_{R,\{\Omega\}}$ so that $\int u \circ \hat{g}^n d\pi$ approaches $\gamma(R)$, and let $\hat{h}^n = \frac{1}{2}h^n + \frac{1}{2}0$ be a sequence in $M_{Q,R}$ so that $\int u \circ \hat{h}^n d\pi$ approaches $\inf_{f \in M_{Q,R}} \int u \circ d\pi$. By definition, Lemma 7 and Monotonicity,

$$[\hat{h}^n]_Q \succeq 0_R \text{ and } [\hat{g}^n]_R \succeq 0_{\{\Omega\}}.$$

Let g^n_* be an act so that $u \circ g^n_* = -u \circ g^n$. By Lemmas 2 and 7

$$\begin{aligned} & [\hat{g}^n]_R \succeq 0_{\{\Omega\}} \\ \iff & [\frac{1}{2}g^n + \frac{1}{2}0]_R \sim \frac{1}{2}[g^n]_R + \frac{1}{2}0 \succeq \frac{1}{2}0_{\{\Omega\}} + \frac{1}{2}0 \\ \iff & 0_R \sim [\frac{1}{2}g^n + \frac{1}{2}g^n_*]_R \sim \frac{1}{2}[g^n]_R + \frac{1}{2}g^n_* \succeq \frac{1}{2}0_{\{\Omega\}} + \frac{1}{2}g^n_* \sim \frac{1}{2}[g^n_*]_{\{\Omega\}} + \frac{1}{2}0 \end{aligned}$$

and using Transitivity in addition to the two lemmas,

$$\begin{aligned} & \frac{1}{2}[h^n]_Q + \frac{1}{2}0 \sim [\frac{1}{2}h^n + \frac{1}{2}0]_Q \succeq \frac{1}{2}[g^n_*]_{\{\Omega\}} + \frac{1}{2}0 \\ \iff & \frac{1}{2}[h^n]_Q + \frac{1}{2}g^n \succeq \frac{1}{2}[g^n_*]_{\{\Omega\}} + \frac{1}{2}g^n \\ \iff & [\frac{1}{2}h^n + \frac{1}{2}g^n]_Q \succeq 0_{\{\Omega\}}. \end{aligned}$$

Since $u \circ [\frac{1}{2}h^n + \frac{1}{2}g^n] = [u \circ \hat{g}^n + u \circ \hat{h}^n]$, we have

$$\gamma(Q) = \inf_{f \in M_{Q,\{\Omega\}}} \int u \circ f d\pi \leq \int [u \circ \hat{g}^n + u \circ \hat{h}^n] d\pi.$$

As $n \to \infty$, the right hand side goes to $\inf_{f \in M_{Q,R}} \int u \circ f d\pi + \gamma(R)$, proving the claim.

To complete the proof, fix any arbitrary $F, G \in \mathcal{F}^\Omega$ so that $F \succeq G$, and let $Q = \hat{P}(Im\bar{F})$ and $R = \hat{P}(Im\bar{G})$. If $G \notin \mathcal{H}$, then the conclusion follows trivially. Otherwise, by Lemma 4, there are acts f, g so that $u \circ f = 2u \circ F^*$ and $u \circ g = 2u \circ G^*$ and

$$[\frac{1}{2}f + \frac{1}{2}0]_Q \sim F \succeq G \sim [\frac{1}{2}g + \frac{1}{2}0]_R.$$

As above,

$$[\frac{1}{2}f + \frac{1}{2}0]_Q \sim \frac{1}{2}[f]_Q + \frac{1}{2}0 \succeq \frac{1}{2}[g]_R + \frac{1}{2}0 \sim [\frac{1}{2}g + \frac{1}{2}0]_R.$$

Let g_* be any act so that $u \circ g_* = -u \circ g$. By Lemmas 2 and 7,

$$\frac{1}{2}[f]_Q + \frac{1}{2}g_* \succeq \frac{1}{2}[g]_R + \frac{1}{2}g_* \sim 0_R$$

Hence, $\frac{1}{2}f + \frac{1}{2}g_* \in M_{Q,R}$; observe $u \circ [\frac{1}{2}f + \frac{1}{2}g_*] = u \circ F^* - u \circ G^*$ and so

$$\int u \circ F^* d\pi - \int u \circ G^* d\pi \geq \inf_{f \in M_{Q,R}} \int u \circ f d\pi \geq \gamma(Q) - \gamma(R)$$

implying

$$\int u \circ F^* d\pi - \gamma(\hat{P}(Im\bar{F})) \geq \int u \circ G^* d\pi - \gamma(\hat{P}(Im\bar{G})),$$

or equivalently $V(F) \geq V(G)$. ☐

Lemma 11. *For any $Q, R \in \mathbb{P}$: if $R \gg Q$, then $\gamma(R) \geq \gamma(Q)$, and if $Q \gg P$ and $Q \neq P$, then $\gamma(Q) = \infty$.*

Proof. The first implication follows from Lemma 2. In particular, for $u(x)$ sufficiently low, $0Q(\omega)x \in c(B|\omega)$ for all ω when $B = \{0Q(\omega)x, 0R(\omega)x : \omega \in \Omega\}$, so $V(0_Q) \geq V(0_R)$, equivalently

$$0 - \gamma(Q) \geq 0 - \gamma(R),$$

completing the proof. The second implication follows from noting that if $Q \gg P$ and $Q \neq P$, then $M_{Q,\{\Omega\}} = \emptyset$. Since the infimum of the empty set is ∞, it follows that $\gamma(Q) = \infty$. ☐

Lemmas 5 and 10 give that

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

It remains to show that choices agree with the preference over plans.

Lemma 12. $c(B|\omega) = \arg\max_{f \in B} \int u \circ f d\pi(\cdot|\hat{P}(B)(\omega))$.

Proof. Fix $\omega \in \Omega$, $B \in K(\mathcal{F})$, and $E = \hat{P}(B)(\omega)$.

I show first that $c(B|\omega) \subset \arg\max_{f \in B} \int u \circ f d\pi(\cdot|E)$. Suppose $f \in c(B|\omega)$ and set $F \in \hat{c}(B)$ so that $F(\omega) = f$ for all $\omega \in E$. By Lemma 5, $F \succeq G$ for all G so that $Im(G) \subset B$. If $f \notin \arg\max_{g \in B} \int u \circ g d\pi(\cdot|E) \ni f'$, then the plan G defined so that $G(\omega) = f'$ for all $\omega \in E$

and $G(\omega) = F(\omega)$ for all $\omega \notin E$ has $V(G) > V(F)$, contradicting Lemma 10. Therefore, $f \in \arg\max_{g\in B} \int u \circ g d\pi(\cdot|E)$, and consequently $c(B|\omega) \subset \arg\max_{f\in B} \int u \circ f d\pi(\cdot|E)$.

I now show that $\arg\max_{f\in B} \int u \circ f d\pi(\cdot|E) \subset c(B|\omega)$. Pick any $F \in \hat{c}(B)$ and $g \in \arg\max_{h\in B} \int u \circ h d\pi(\cdot|E)$. Set $x \in X$ so that

$$u(x) < -\frac{u(x^*) - V(F)}{\min_{\omega'} \pi(\omega')}$$

where $x^* \in \arg\max_{y\in\{f(\omega):f\in B,\omega\in\Omega\}} u(y)$, and define $\hat{F}, \hat{G} \in \mathcal{F}^\Omega$ by $\hat{F}(\omega') = F(\omega')Ex$ for all ω' and $\hat{G}(\omega') = F(\omega')Ex$ for all $\omega' \notin E$ and $\hat{G}(\omega') = gEx$ for all $\omega' \in E$. Consider $B' = B \bigcup Im(\hat{F}) \bigcup Im(\hat{G})$ and $B'' = Im(\hat{F}) \bigcup Im(\hat{G})$. Subjective Consequentialism and Monotonicity imply $\hat{F} \in \hat{c}(B')$. By INRA, $\hat{F} \in \hat{c}(B'')$.

Let $f' = F(\omega)Ex$. Define $\hat{g} \in \mathcal{F}$ such that $u \circ \hat{g} = u \circ g + 1$, and

$$B_n = (B''\backslash\{f', gEx\}) \bigcup \{\frac{1}{n}\hat{g} + \frac{n-1}{n} gEx\}$$

for every $n \in \mathbb{N}$. Let $G_n(\omega') = \hat{F}(\omega')E(\omega')x$ for all $\omega \notin E$ and $G_n(\omega') = \frac{1}{n}\hat{g} + \frac{n-1}{n} gEx$ for all $\omega' \in E$. By Lemmas 5 and 10, $V(G) \geq V(F')$ for all F' so that $\{F'\} \subset B$. Hence $V(G_n) > V(F')$ for all $F' \neq G_n$ so that $\{F'\} \subset B_n$. Note $V(\hat{F}) = V(\hat{G}) = V(F)$, and if $H(\omega') \in B''$ for all ω' and $\hat{P}(Im\bar{H}) \not\gg \hat{P}(B)$, then $V(H) < V(F)$; to see why, note $H(\omega') = x$ for all $\omega' \in E' \in P$ and $u(H(\omega'')) < u(x^*)$, so $V(H) < u(x^*) + \pi(E')u(x) < u(x^*) - u(x^*) + V(F)$.

By Lemmas 5 and 10, $G_n \in \hat{c}(B_n)$ and $\frac{1}{n}\hat{g} + \frac{n-1}{n} gEx \in c(B_n|\omega)$ for every n. By construction, $\hat{P}(B_n) = \hat{P}(B'')$ for all n. Since $\frac{1}{n}\hat{g} + \frac{n-1}{n} gEx \in c(B_n|\omega)$ and $\frac{1}{n}\hat{g} + \frac{n-1}{n} gEx \to gEx$, it follows from Continuity that $gEx \in c(B''|\omega)$. By INRA, $c(B''|\omega) = c(B'|\omega) \bigcap B''$. Since $u \circ g \geq u \circ gEx$, $gEx \in c(B'|\omega) \implies g \in c(B'|\omega)$ by Monotonicity. By INRA, $c(B'|\omega) = c(B'|\omega) \bigcap B$, so $g \in c(B|\omega)$. Therefore, $\arg\max_{f\in B} \int u \circ f d\pi(\cdot|E) \subset c(B|\omega)$, completing the proof. □

Lemma 12 completes the proof.

A.1.3. ***Proof of Theorem 2:*** Suppose $c(\cdot)$ is represented by $(u(\cdot), \pi(\cdot), \gamma(\cdot), \hat{P}(\cdot))$; WLOG, take $\hat{P}(\cdot)$ to be canonical. First, I show monotonicity. Observe $x \succeq^R y$ if and only if $u(x) \geq u(y)$. Now, suppose $f(\omega') \succeq^R g(\omega')$ for every ω'. Then $u \circ f \geq u \circ g$. Consequently, for any $E \subseteq \Omega$, $\int u \circ f d\pi(\cdot|E) \geq \int u \circ g d\pi(\cdot|E)$. Therefore, if $g \in c(B|\omega)$ and $f \in B$, then $f \in c(B|\omega)$. If in addition $f(\omega) \succ^R g(\omega)$, then since $\pi(\omega) > 0$, $\int u \circ f d\pi(\cdot|\hat{P}(B)(\omega)) > \int u \circ g d\pi(\cdot|\hat{P}(B)(\omega))$ so $g \notin c(B|\omega)$. Unboundedness follows from the argument following the statement of the axiom.

Turn to SC. Fix B and consider $\hat{P}(B)$. Note that $c(B|\omega') \neq c(B|\omega)$ implies $\omega' \notin \hat{P}(B)(\omega)$. If $f(\omega) = g(\omega)$ and $\omega' \neq \omega$ implies either $f(\omega') = g(\omega')$ or $c(B|\omega') \neq c(B|\omega)$, then $f(\omega') = g(\omega')$ for all $\omega' \in \hat{P}(B)(\omega)$. Therefore, $\int u \circ f d\pi(\cdot|\hat{P}(B)(\omega)) = \int u \circ g d\pi(\cdot|\hat{P}(B)(\omega))$, so $f \in c(B|\omega) \iff g \in c(B|\omega)$.

Now, to see continuity (i), let $B_n \to B$ and $f_n \to f$ be as above. Note that $E = \hat{P}(B_n)(\omega) = \hat{P}(B)(\omega)$ for all n. By the Theorem of the Maximum, $\arg\max_{f \in B'} \int u \circ f d\pi(\cdot|E)$ is upper hemi-continuous as a function of B'. Since f_n is in each argmax and $f_n \to f$, $f \in c(B|\omega)$.

To see continuity (ii), define $V(\cdot)$ and $\hat{c}(\cdot)$ as in the proof of Theorem 1 and fix any x, y, f as in the hypothesis, noting this requires $u(y), u(x) > V(f)$. Pick any $\epsilon \in (0,1)$ such that

$$u(x) > \epsilon u(y) + (1-\epsilon)V(f) = V(\epsilon y + (1-\epsilon)f).$$

If $\{f_i\}\, IS\, \{x\}$, then $V(f_i) \geq V(x)$. If this holds for all i, then $V(\sum \alpha_i f_i) > V(\epsilon y + (1-\epsilon)f)$, an impossibility if $u(\epsilon y + (1-\epsilon)f(\omega)) \geq u(\sum \alpha_i f_i(\omega))$ for all ω.

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

$$W(B|Q) = \sum_{E \in Q} \pi(E) \max_{f \in B} \int u \circ f d\pi(\cdot|E) - \gamma(Q). \tag{4}$$

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

Define K by

$$K = \{B \in K(\mathcal{F}) : \# \arg\max_{Q \in \mathbb{P}} W(B|\cdot) = 1\}.$$

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

Lemma 13. $cl(K) = K(\mathcal{F})$

Proof. Pick any $B \in K(\mathcal{F})$ and any $\epsilon > 0$. Take $Q \in \arg\max_{Q \in \mathbb{P}} W(B, \cdot)$ so that no $Q' \in \arg\max_{Q \in \mathbb{P}} W(B|\cdot)$ is strictly finer than 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 $u(f^{**}(\omega)) = u(f^*(\omega)) + k$ for some $k > 0$.

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

$$B' = B \bigcup \{f^i_{\alpha^*}\}_{i=1}^n.$$

By construction, $W(B'|Q) > W(B'|Q')$ for all $Q' \in \mathbb{P}$ so that $Q' \neq Q$. Therefore, $B' \in K$.

Since B and ϵ are arbitrary, there is such a $B' \in K$ arbitrarily close to any $B \in K(\mathcal{F})$. In particular, for every n, one can find $B_n \in K$ so that $d(B_n, B) < \frac{1}{n}$. The sequence $B_n \to B$. Therefore, $cl(K) = K(\mathcal{F})$. □

Lemma 14. *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 $B_n \to B$. Let (Q_n^1, Q_n^2) be two distinct elements in $\arg\max_{R\in\mathbb{P}^*} W(B_n|R)$. There exists a subsequence n_k where $(Q_{n_k}^1, Q_{n_k}^2) = (Q^1, Q^2)$ for all k because $\mathbb{P}^*$ is finite. By the Theorem of the Maximum, $Q^1, Q^2 \in \arg\max_{Q\in\mathbb{P}^*} V(Q, B)$. Conclude that $B \in K^c$, so K^c is closed and K is open. □

Let $>$ be an arbitrary linear order on $\mathbb{P}$ and set

$$\hat{Q}(B) = \max_{>} \arg\max_{Q\in\mathbb{P}} W(B|Q)$$

for every $B \in K(\mathcal{F})$. 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 15. *$c'(\cdot)$ satisfies ACI.*

Proof. Fix any $B \in K(\mathcal{F})$, $f, g, h \in \mathcal{F}$, and $\omega \in \Omega$. Suppose $\alpha f + (1-\alpha)g \in c(\alpha B + (1-\alpha)\{g\}|\omega)$ and let $B' = \alpha B + (1-\alpha)\{g\}$. Note that $W(\alpha B + (1-\alpha)\{g\}, Q)$ equals

$$\begin{aligned}
&= \sum_{E\in Q} \pi(E) \max_{f'\in B'} \int u \circ f' d\pi(\cdot|E) - \gamma(Q) \\
&= \sum_{E\in Q} \pi(E) [\max_{f\in B} \int [\alpha u \circ f + (1-\alpha) u \circ g] d\pi(\cdot|E)] - \gamma(Q) \\
&= \alpha \sum_{E\in Q} \pi(E) \max_{f\in B} \int u \circ f d\pi(\cdot|E) + (1-\alpha) \int u \circ g d\pi - \gamma(Q).
\end{aligned}$$

Similarly, $W(\alpha B + (1-\alpha)\{h\}, Q)$ equals

$$\alpha \sum_{E\in Q} \pi(E) \max_{f\in B} \int u \circ f d\pi(\cdot|E) + (1-\alpha) \int u \circ h d\pi - \gamma(Q).$$

Therefore,

$$\hat{Q}(\alpha B + (1-\alpha)\{g\}) = \hat{Q}(\alpha B + (1-\alpha)\{h\}),$$

so clearly ACI holds and the proof is complete. □

Lemma 16. *$c'(\cdot)$ satisfies INRA.*

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

$$\arg\max_{Q\in\mathbb{P}} W(A, Q) \subset \arg\max_{Q\in\mathbb{P}} W(B|Q)$$

since

$$W(A, \hat{Q}(B)) = W(B|\hat{Q}(B)) \geq W(B|Q') \geq W(A, Q')$$

for all $Q' \in \mathbb{P}$. By construction of $c'(\cdot)$, $\hat{Q}(B) > Q'$ for all $Q' \in \arg\max_{Q \in \mathbb{P}} W(B|Q)$. Therefore, $\hat{Q}(B) > Q'$ for all $Q' \in \arg\max_{Q \in \mathbb{P}} W(A|Q)$. Conclude that $\hat{Q}(A) = \hat{Q}(B)$, implying $c'(B|\omega) \bigcap A = c'(A|\omega)$. □

These lemmas complete the proof.

A.1.4. ***Proof of Corollary 1:*** I show that Independence implies fixed attention and that WARP implies independence. That fixed attention implies WARP and Indepedendence is trivial.

Suppose Independence, which implies ACI. Then $c(\cdot)$ has an costly attention representation $(u, \pi, \gamma, \hat{P})$, where $\hat{P}(B)$ is canonical. Let Q be coarsest common refinement of $\{\hat{P}(B)\}_{B \in K(\mathcal{F})}$. Pick any B and any ω. There is a finite collection $\{B_1, ..., B_n\} \subset K(\mathcal{F})$ so that $[\bigcap_{i=1}^{n} \hat{P}(B_i)(\omega)] = Q(\omega)$ and $c(B_i|\omega) \neq c(B_j|\omega)$ for all $i \neq j$. Set $B^* = \prod_{i=1}^{n} \frac{1}{n} B_i$ and note that $\hat{P}(B^*)(\omega) = Q(\omega)$ by Independence.

It follows that

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

Independence implies that $c(\frac{1}{2}B^* + \frac{1}{2}B|\omega) = \frac{1}{2}c(B^*|\omega) + \frac{1}{2}c(B|\omega)$, and since Q is the coarsest common refinement of $\{\hat{P}(B)\}_{B \in K(\mathcal{F})}$ and $\hat{P}(B^*)(\omega) = Q(\omega)$, $Q(\omega) = \hat{P}(\frac{1}{2}B^* + \frac{1}{2}B)(\omega)$. Therefore, $c(\frac{1}{2}B^* + \frac{1}{2}B|\omega)$ equals

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

which requires that $c(B|\omega) = \arg\max_{f \in B} \int u \circ f d\pi(\cdot|Q(\omega))$. Since B and ω were chosen arbitrarily, this holds for all B and all ω.

Suppose WARP, which implies INRA. Theorem 1 implies that $c(\cdot)$ has a representation $(u, \pi, \gamma, \hat{P})$. Take arbitrary $\alpha \in (0, 1)$, B, A, ω, $f \in c(B|\omega)$, $g \in c(A|\omega)$, and $\alpha f' + (1-\alpha)g' \in c(\alpha B + (1-\alpha)A|\omega)$. Given WARP, ACI, Monotonicity and Continuity, there is an subjective expected utility preference relation $\succeq_\omega$ so that $\succeq_\omega$ rationalizes $c(\cdot|\omega)$ (using the Mixture Space Theorem). Then $f \succeq_\omega h$ for all $h \in B$ and $g \succeq_\omega h$ for all $h \in A$. Therefore $\alpha f + (1-\alpha)g \succeq_\omega h$ for all $h \in \alpha B + (1-\alpha)A$, so $\alpha f + (1-\alpha)g \in c(\alpha B + (1-\alpha)A|\omega)$. Moreover, $f' \succeq_\omega h$ for all $h \in B$; if not, there exists $h' \in B$ so that $h' \succ f'$, which would imply $\alpha h' + (1-\alpha)g' \succ_\omega \alpha f' + (1-\alpha)g'$, contradicting that $\succeq_\omega$ rationalizes $c(\cdot|\omega)$. Similarly, $g' \succeq_\omega h$ for all $h \in A$. Therefore $f' \in c(B|\omega)$ and $g' \in c(A|\omega)$. Conclude independence holds, completing the proof.

A.1.5. ***Proof of Corollary 2:*** Strong ACI implies ACI, so by Theorem 1, $c(\cdot)$ has a costly attention representation $(u, \pi, \gamma, \hat{P})$. Choose any $Q \in$ *support*(γ). For some $x, y \in X$ where $u(x) > u(y)$, the problem $B = \{xEy : E \in Q\}$ is such that $\hat{P}(B) = Q$ (otherwise, $Q \notin$ *support*(γ)). By strong ACI, $\hat{P}(B) = \hat{P}(\alpha B + (1-\alpha)\{x\})$ for all $\alpha \in (0, 1]$. For W as in Eq. (4),

$$W(\alpha B + (1-\alpha)\{x\}|\{\Omega\}) \to_{\alpha \to 0} u(x)$$

and that

$$W(\alpha B + (1-\alpha)\{x\}|Q) \to_{\alpha \to 0} u(x) - \gamma(Q).$$

Because $W(\alpha B + (1-\alpha)\{x\}|Q) \geq W(\alpha B + (1-\alpha)\{x\}|\{\Omega\})$ for all $\alpha \in (0, 1]$, $\gamma(Q) = 0$. Since Q is arbitrary, $\gamma(Q) = 0$ for all $Q \in$ *support*(γ).

Necessity follows from Theorem 2 and trivially adapting Lemma 15, completing the proof.

A.1.6. ***Proof of Corollary 3:*** Pick any B, any ω, and any $g \in c(B|\omega)$. Because $\gamma(P) = 0$ and $\gamma(Q) < \infty \implies P \gg Q$, $g \in \arg\max_{f \in B} \int u \circ f d\pi(\cdot|P(\omega))$. Define $B' = B \bigcup \{gP(\omega)x\}$, where $x \in X$ satisfies $u(x) < u(g(\omega))$ for all ω. By Lemma 2, $g \in c(B'|\omega)$. By Consequentialism, $gP(\omega)x \in c(B'|\omega)$, since $c(B'|\omega) = \arg\max_{f \in B'} \int u \circ f d\pi(\cdot|\hat{P}(B')(\omega))$, $\hat{P}(B')(\omega) = P(\omega)$. By INRA,

$$\begin{aligned} c(B|\omega) &= c(B'|\omega) \bigcap B = c(B'|\omega) \backslash \{gP(\omega)x\} \\ &= \arg\max_{f \in B} \int u \circ f d\pi(\cdot|P(\omega)). \end{aligned}$$

Hence $\hat{P}(B)(\omega) = P(\omega)$. Since B and ω are arbitrary, conclude $\hat{P}(B)(\omega) = P(\omega)$ for every B and every ω, completing the proof.

A.2. **Proofs from Section 5.1.**

A.2.1. ***Proof of Corollary 4***. For necessity, fix x and pick ϵ such that $|u(y) - u(x)| < \min_{P \neq \{\Omega\}} \gamma(P)$. Observe that if $d(B, \{x\}) < \epsilon$, then the maximum utility expected from any partition is less than $\min_{P \neq \{\Omega\}} \gamma(P) + u(x)$, and so the DM finds it optimal to pay attention to $\{\Omega\}$.

Conversely, suppose not: $\gamma(Q) = 0$ and $Q \neq \{\Omega\}$. Pick $x, y \in X$ with $u(y) > u(x)$. Pick $E \in Q$ and let $B_n = \{(\frac{1}{n}y + \frac{n-1}{n}x)Ex, xE(\frac{1}{n}y + \frac{n-1}{n}x)\}$. Note $B_n \to \{x\}$, that Q is optimal for each, that any other optimal partition is finer than $\{E, E^c\}$, and that thus $c(B_n|\omega) \neq c(B_n|\omega')$ for each n when $\omega \in E$ and $\omega' \in E^c$, a contradiction, completing the proof.

A.2.2. ***Proof of Theorem 3:*** Suppose $(u_i, \pi_i, \gamma_i, \hat{P}_i)$ represent c for $i = 1, 2$. It is standard to show $u_2(x) = \alpha u_1(x) + \beta$. The proof that $\hat{P}_1 = \hat{P}_2$ is in the text.

I begin by showing that $\pi_1 = \pi_2 \equiv \pi$. Since $\gamma(P) > 0$ for all P, fix x and ϵ so that $d(x, y) < \epsilon$ implies $|u(y) - u(x)| < \frac{1}{3} \min \gamma(P)$. Consider $y, z \in X$ with $d(y, x) + d(z, x) < \epsilon$

and $u_1(y) > u_1(z)$. For any $E \subseteq \Omega$, observe $\hat{P}(\{yEz, \alpha y + (1-\alpha)z\}) = \{\Omega\}$ for all $\alpha \in [0,1]$. A simple continuity argument shows that for some α, $\{yEz, \alpha y + (1-\alpha)z\} = c(\{yEz, \alpha y + (1-\alpha)z\}|\omega)$ for all ω. Thus $\alpha u_j(y) + (1-\alpha)u_j(z) = \pi_j(E)u_j(y) + (1-\pi_j(E))u_j(z)$ for $j = 1,2$ so $\pi_1(E) = \pi_2(E)$. Since E is arbitrary, $\pi_1 = \pi_2$.

Now, I show that $\gamma_1 = \alpha\gamma_2$. Let

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

for $j = 1,2$, noting $V_2(B,Q') = \alpha V_1(B,Q') + \beta$. Pick any Q with $\gamma_1(Q) < \infty$. Define $y(x)$ for all x such that $u_1(y(x)) = -\frac{u_1(x) + \max_{R \in \mathbb{P}^*} \gamma_1(R) + 1}{\min \pi(\omega)}$. Identify $x \in \mathbb{R}$ with consequence that gives $u_1(x) = x$. Let $x_0 = 0$, $Q_0 = \{\Omega\}$, and

$$B_x = \{0\} \bigcup \{xEy(x) : E \in Q\}.$$

By construction, $xEy(x)$ is chosen from B_x if and only if $\hat{P}(B_x)(\omega) \subseteq Q(\omega)$. Hence, if $R \not\gg Q$, then for some $\rho \in [0,1)$, $V_j(B_x, R) = \rho u_j(x) < u_j(x) = V_j(B_x, Q)$. Therefore, there exists an $\bar{x}$ so that $\hat{P}(B_x) = Q$ for all $x > \bar{x}$. For $i \geq 1$ and as long as $Q_i \neq Q$, define x_i and Q_i recursively by $x_i = \sup_x\{\hat{P}(B_x) = Q_{i-1}\}$ for all ω, Q_i an arbitrary limit point of $\hat{P}(B^i_{x_i + \frac{1}{n}})$, and

$$E_i = \{\omega : Q_i(\omega) \subseteq Q(\omega)\}. \tag{5}$$

Let I be the number of iterations to get to Q, i.e. $Q_I = Q$.

For $j = 1,2$ and $R \in \mathbb{P}$, $V_j(B_x, Q_i) = \pi(E_i)u_j(x)$. This is continuous in x. Since $\hat{P}(B_x)$ is optimal and V_j continuous in x, $V_j(B_{x_i}, Q_i) - V_j(B_{x_i}, Q_{i+1}) = \gamma_j(Q_i) - \gamma_j(Q_{i+1})$. Therefore,

$$\begin{aligned}\sum_{i=0}^{I-1}[\gamma_2(Q_i) - \gamma_2(Q_{i+1})] &= \sum_{i=0}^{I-1}[V_2(B_{x_i}, Q_i) - V_2(B_{x_i}, Q_{i+1})] \\ &= \sum_{i=0}^{I-1} \alpha[V_1(B_{x_i}, Q_i) - V_1(B_{x_i}, Q_{i+1})] \\ &= \sum_{i=0}^{I-1} \alpha[\gamma_1(Q_i) - \gamma_1(Q_{i+1})].\end{aligned}$$

Since $\sum_{i=0}^{I-1} \gamma_j(Q_i) - \gamma_j(Q_{i+1}) = \gamma_j(Q_0) - \gamma_j(Q_I) = -\gamma_j(Q)$, one has $\gamma_1(Q) = \alpha\gamma_2(Q)$. Since Q was arbitrary, $\gamma_1 = \alpha\gamma_2$, completing the proof.

A.2.3. ***Proof of Theorem 4:*** Note that $c_1(\cdot)$ *pays more attention than* $c_2(\cdot)$ iff $\tilde{P}_{c_1}(B) \gg \tilde{P}_{c_2}(B)$ for every B so that the optimal partition for c_1 is unique. Suppose $Q \gg R \implies \gamma_1(Q) - \gamma_1(R) \leq \gamma_2(Q) - \gamma_2(R)$ for all Q, R. If c_1 does not pay more attention than c_2, then there is B with a unique c_1-optimal partition so that $\tilde{P}_{c_1}(B) \not\gg \tilde{P}_{c_2}(B)$; since both are comparable by $\gg$, it must be that $\tilde{P}_{c_2}(B) \gg \tilde{P}_{c_1}(B)$. Letting $V(B,Q')$ be as in the proof of Theorem 3, $V(B, \tilde{P}_{c_1}(B)) \leq V(B, \tilde{P}_{c_2}(B))$. Since $\tilde{P}_{c_1}(B)$ is the unique c_1-optimal partition, it must be that

$$\begin{aligned} V(B,\tilde{P}_{c_1}(B)) - \gamma_1(\tilde{P}_{c_1}(B)) &> V(B,\tilde{P}_{c_2}(B)) - \gamma_1(\tilde{P}_{c_2}(B)) \\ \iff V(B,\tilde{P}_{c_1}(B)) - V(B,\tilde{P}_{c_2}(B)) &> \gamma_1(\tilde{P}_{c_1}(B)) - \gamma_1(\tilde{P}_{c_2}(B)) \\ &\geq \gamma_2(\tilde{P}_{c_1}(B)) - \gamma_2(\tilde{P}_{c_2}(B)). \end{aligned}$$

In particular, it must be the case that paying attention to $\tilde{P}_{c_1}(B)$ is strictly better for DM2 when she faces B, a contradiction.

Suppose c_1 pays more attention than c_2. Label $supp(\gamma_1) = \{R_0, ..., R_n\}$ so that $R_i \gg R_{i-1}$ for all i and for convenience label X so that $u(x) = x$ for all $x \in X$. Pick $i \geq 1$. For any $x \in \mathbb{R}_+$, define

$$B^i_x = \{xEz(x)|E \in R_i \backslash R_{i-1}\} \bigcup \{0Ez(x)|E \in R_{i-1}\},$$

where $u(x) = x$ and $z(x) = \frac{-\gamma(R_{i-1})}{\min \pi(\omega)}$, and $E_i = \bigcup\{E \in R_i \backslash R_{i-1}\}$. When facing B^i_x, it is either optimal pay attention R_i or to R_{i-1} for either c_1 or c_2. In fact, c_j pays attention to R_i only if $\frac{\gamma_j(R_i)-\gamma_j(R_{i-1})}{\pi(E_i)} \geq x$ and to R_{i-1} only if $\frac{\gamma_j(R_i)-\gamma_j(R_{i-1})}{\pi(E_i)} \leq x$. Picking x^* so that

$$x^* = \frac{\gamma_1(R_i) - \gamma(R_{i-1})}{\pi(E_i)},$$

DM1 is exactly indifferent between paying attention to R_i and R_{i-1}. In particular, taking $x_n = x^* + \frac{1}{n}$ for every $n \in \mathbb{N}$,

$$\tilde{P}_{c_1}(B^i_{x_n}) = R_i$$

for all n, and this optimum is unique for all n. Moreover $\tilde{P}_{c_2}(B^i_{x_n}) \in \{R_i, R_{i-1}\}$ also. As n goes to zero, conclude that

$$\gamma_2(R_i) - \gamma_2(R_{i-1}) \geq \gamma_1(R_i) - \gamma_1(R_{i-1}).$$

Since i is arbitrary, this holds for all i. To compare R_i and R_j with j less than i, note that

$$\gamma_k(R_i) - \gamma_k(R_j) = \sum_{j'=j+1}^{i} \gamma_k(R_{j'}) - \gamma_k(R_{j'-1}).$$

Since each term in the sum is smaller when $k = 1$ than when $k = 2$, Therefore, whenever $Q \gg R$, $\gamma_1(Q) - \gamma_1(R) \leq \gamma_2(Q) - \gamma_2(R)$, completing the proof.

A.2.4. ***Proof of Theorem 5:*** Note that $c_1(\cdot)$ has a higher capacity for attention than $c_2(\cdot)$ iff for any B, there exists a B' so that $\tilde{P}_{c_1}(B') = \tilde{P}_{c_2}(B)$. Suppose c_1 has a higher capacity for attention than c_2. If $supp(\gamma_2) \nsubseteq supp(\gamma_1)$, take $Q \in supp(\gamma_2) \backslash supp(\gamma_1)$. Since $Q \notin supp(\gamma_1)$, there is no $R \in supp(\gamma_1)$ so that $R \gg Q$. Therefore, $\tilde{P}_{c_1}(B) \neq Q$ for every B. However, by taking $x, y \in X$ so that $u_2(y) < -\frac{\gamma(Q)}{\min \pi(\omega)}$ and $u_2(x) = 0$, $\tilde{P}_{c_2}(\{xEy : E \in Q\}) = Q$. To see this, note that paying attention to Q yields $-\gamma(Q)$. Paying attention to any partition coarser than Q yields strictly less than $-\gamma(Q)$ because she must choose y with probability no smaller than $\min \pi(\omega)$. Paying attention to any partition finer than Q results in the same choices as paying attention to Q. Moreover, if she pays attention Q, then she makes a different choice on each cell of Q.

Now, suppose $supp(\gamma_2) \subset supp(\gamma_1)$. As above, for any Q in the support γ_i one can construct a set of bets so that c_i pays attention to Q when she faces it. Therefore, for every $Q \in supp(\gamma_i)$, there is a problem so that $\tilde{P}_{c_i}(B) = Q$. Therefore, if $\tilde{P}_{c_2}(B) = Q$, then there is a B' so that $\tilde{P}_{c_2}(B) = Q$, completing the proof.

A.2.5. ***Proof of Theorem 6***. The set K consists of the choice problems for which the optimal information partition is unique. From Theorem 2, this set is open and dense. Pick any $B \in K$ and let $Q = \arg\max W(B|P)$ where W is defined in Eq. (4). Fix ω and suppose first that $f \in c(B|\omega)$. Then $Q = \arg\max W(B_\omega \bigcup\{f\}|P)$, and similarly for $\arg\max W(B_\omega \bigcup B|P)$. Moreover, it is easy to see $B \sim B \bigcup B_\omega \sim B_\omega \bigcup\{f\}$. By transitivity, $B \sim B_\omega \bigcup\{f\}$, so $f \in c^A(B|\omega)$.

Now, suppose that $f \notin c(B|\omega)$. Pick $g \in c(B|\omega)$ and define g' so that $g'(\omega')$ equals $g(\omega')$ for $\omega' \neq \omega$ and equals $(1-\epsilon)g(\omega)+\epsilon x_B$ for $\epsilon > 0$ small enough that $Q = \arg\max W((B_\omega \bigcup\{f, g'\}|P)$ and $g' \in c(B_\omega \bigcup\{f, g'\}|\omega)$.[38] Observe that the DM finds Q the unique optimal partition when facing B, $B \bigcup B_\omega$, $B_\omega \bigcup\{f, g, g'\}$ and $B_\omega \bigcup\{f, g'\}$. By Consistency, $B \sim B \bigcup B_\omega$ and $B \bigcup B_\omega \sim B_\omega \bigcup\{f, g\}$. Since $g(\omega) \succeq^R g'(\omega)$ for all ω, adding it does not change her choices. Thus $B_\omega \bigcup\{f, g\} \sim B_\omega \bigcup\{f, g, g'\}$. Now, since ϵ is small enough, she pays attention to Q when facing $B_\omega \bigcup\{f, g'\}$, and thus, $B_\omega \bigcup\{f, g, g'\} \succ B_\omega \bigcup\{f, g'\}$. By Consistency, $B_\omega \bigcup\{g', f\} \succsim B_\omega \bigcup\{f\}$. By transitivity, $B \succ B_\omega \bigcup\{f\}$ and thus $f \notin c^A(B|\omega)$, completing the proof.

References

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

Anscombe, F. J. and R. J. Aumann (1963), "A definition of subjective probability." *The Annals of Mathematical Statistics*, 34, 199–205.

Aumann, Robert J. (1976), "Agreeing to disagree." *The Annals of Statistics*, 4, 1236–1239.

Caplin, Andrew and Mark Dean (2015), "Revealed preference, rational inattention, and costly information acquisition." *American Economic Review*, 105, 2183–2203.

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

[38] Specifically, when

$$\epsilon\,[u(g(\omega)) - u(x_B)] < \min\left\{W(B|Q) - \max_{P\neq Q} W(B|P), \frac{\min_{h\in B_\omega \bigcup\{f\}} \int [u\circ g - u\circ h]\, d\pi(\cdot|Q(\omega))}{\pi(\omega|Q(\omega))}\right\}.$$

Coibion, Olivier and Yuriy Gorodnichenko (2015), "Information rigidity and the expectations formation process: A simple framework and new facts." *American Economic Review*, 105, 2644–78.

de Olivera, Henrique, Tommasso Denti, Maximilian Mihm, and M. Kemal Ozbek (2016), "Rationally inattentive preferences and hidden information costs." *Theoretical Econimcs*, Forthcoming.

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

Dillenberger, David, Juan Sebastian Lleras, Philipp Sadowski, and Norio Takeoka (2014), "A theory of subjective learning." *Journal of Economic Theory*, 153, 287–312.

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

Dubra, Juan, Fabio Maccheroni, and Efe A. Ok (2004), "Expected utility theory without the completeness axiom." *Journal of Economic Theory*, 115, 118–133.

Epstein, Larry G. and Martin Schneider (2003), "Recursive multiple priors." *Journal of Economic Theory*, 113.

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

Gabaix, Xavier (2014), "A sparsity-based model of bounded rationality, applied to basic consumer and equilibrium theory." *Quarterly Journal of Economics*, 130, 1369–1420.

Gabaix, Xavier and David Laibson (2001), "The 6d bias and the equity premium puzzle." *NBER Macroeconomics Annual*, 16, 257–312.

Ghirardato, Paolo (2002), "Revisiting savage in a conditional world." *Economic Theory*, 20, 83–92.

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

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

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

Gul, Faruk, Wolfgang Pesendorfer, and Tomasz Strzalecki (2017), "Behavioral competitive equilibrium and extreme prices." *American Economic Review*, 107, 109–137.

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

Herstein, I. N. and John Milnor (1953), "An axiomatic approach to measurable utility." *Econometrica*, 21, pp. 291–297.

Klibanoff, Peter, Massimo Marinacci, and Sujoy Mukerji (2009), "Recursive smooth ambiguity preferences." *Journal of Economic Theory*, 144, 930–976.

Kopylov, Igor (2001), "Procedural rationality in the multiple priors model." *mimeo*.

Lipman, Barton L. (1991), "How to decide how to decide how to...: Modeling limited rationality." *Econometrica*, 59, pp. 1105–1125.

Lu, Jay (2016), "Random choice and private information." *Econometrica*, 84.

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

Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini (2006a), "Ambiguity aversion, robustness, and the variational representation of preferences." *Econometrica*, 74, 1447–1498.

Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini (2006b), "Dynamic variational preferences." *Journal of Economic Theory*, 128, 4–44.

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

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

Manzini, Paola and Marco Mariotti (2014), "Stochastic choice and consideration sets." *Econometrica*, 82, 1153–1176.

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

Matejka, Filip and Alisdair McKay (2015), "Rational inattention to discrete choices: A new foundation for the multinomial logit model." *American Economic Review*, 105.

Meyer, Margaret A. (1991), "Learning from coarse information: Biased contests and career profiles." *Review of Economic Studies*, 58, 15–41.

Ortoleva, Pietro (2012), "Modeling the change of paradigm: Non-bayesian reactions to unexpected news." *American Economic Review*, 102, 2410–2436.

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

Reis, Ricardo (2006), "Inattentive producers." *The Review of Economic Studies*, 73, 793–821.

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

Saint-Paul, Gilles (2011), "A "quantized" approach to rational inattention." *mimeo*.

Savage, Leonard J. (1954), *The Foundations of Statistics*, 2nd edition. Dover, Toronto, Ontario, Candada.

Simon, Herbert (1971), "Designing organizations for an information-rich world." In *Computers, communications, and the public interest* (Martin Greenberger, ed.), Johns Hopkins University.

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

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

Spiegler, Ran (2010), "Comments on "behavioral" decision theory: Discussion of surveys in the econometric society 2010 world congress." *mimeo*.

Van Nieuweburgh, Stijn and Laura Veldkamp (2010), "Information acquisition and under-diversification." *Review of Economic Studies*, 77, 779–805.

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

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