Available online at www.sciencedirect.com

ScienceDirect

Journal of Economic Theory 186 (2020) 104979

JOURNAL OF Economic Theory

www.elsevier.com/locate/jet

Costly information acquisition ☆

Christopher P. Chambers [a], Ce Liu [b], John Rehbeck [c,*]

[a] *Department of Economics, Georgetown University, United States of America*
[b] *Department of Economics, Michigan State University, United States of America*
[c] *Department of Economics, Ohio State University, United States of America*

Received 3 May 2018; final version received 12 December 2019; accepted 13 December 2019
Available online 19 December 2019

Abstract

We provide revealed preference characterizations for choices made under various forms of costly information acquisition. We examine nonseparable, multiplicative, and constrained costly information acquisition. In particular, this allows the possibility of unknown time delay for acquiring information. The techniques we use parallel the duality properties in the standard consumer problem.

JEL classification: D01; D80

Keywords: Revealed preference; Information acquisition

1. Introduction

Acquiring information is an integral part of decisions under uncertainty. Most existing research on costly information acquisition studies costs that are additively separable from the expected payoff. This assumes the cost incurred from acquiring information is independent of expected payoff. One can interpret these preferences as an individual having a fixed production

☆ We thank Yaron Azrieli, Mark Dean, and Pietro Ortoleva for useful comments and suggestions. We also thank an anonymous referee for correcting an error in an earlier draft of the paper.
* Corresponding author.
E-mail addresses: cc1950@georgetown.edu (C.P. Chambers), cel013@ucsd.edu (C. Liu), rehbeck.7@osu.edu (J. Rehbeck).

https://doi.org/10.1016/j.jet.2019.104979

technology to acquire information. However, the cost structures of information acquisition can be more complicated.

For example, there may be significant costs incurred from the time delay waiting for information to arrive. Consider when an oil company is deciding between locations to drill for oil. To acquire information, in addition to the monetary expenses to finance geological surveys, there are also significant costs incurred from delayed realization of profits. Suppose the payoff from drilling at a site for each state of the world is given by a net-present-value from the time drilling *begins*. If the oil sites have a higher net-present-value, then this translates into higher costs incurred through discounting. Importantly, costs from time delay now interact with the expected payoff.

In this paper, we study a general model of costly information acquisition that allows for interactions between the information cost and the expected payoff from the decision problem. Apart from the standard assumptions of expected utility maximization and Bayesian updating, the only additional assumption of the model is that the decision maker prefers higher expected payoffs. As special cases of the model, we characterize a representation with multiplicative costs of information and a representation with a constrained information set.

1.1. Importance of non-additive models

The paper of Caplin et al. (2015) investigates the particular case of the model studied here when costs are additive. Our model should be viewed as a direct generalization of this contribution. This being said, the work is motivated by classical economic environments. There are several standard economic frameworks in which we would not expect costs to be additive.

The multiplicative cost model is a particularly interesting case. There are several standard economic environments in which we would expect costs to arise multiplicatively, rather than additively. The most compelling environment is when acquiring information results in a time delay. In standard exponential discounting models, delay enters payoffs multiplicatively (e.g. for discount factor δ, utility takes the form $\delta^{\text{Delay}} \times$ Expected Payoff). Thus, when information acquisition induces different delays, the behavior cannot be rationalized by a model with an additive cost of information acquisition. For a formal representation, see Example 2.

There are also other examples when the costs are multiplicative. For example, the cost of information acquisition may accrue because of some probability of an absolute breakdown. For example, eliciting the information may involve some type of illegal activity where if the acquirer is caught, then they get nothing. In this case, the probability of not being caught enters multiplicatively into a (risk-neutral) individual's utility.

A more straightforward example involves the individual directly contracting with an outside provider of information, who insists on a profit-sharing agreement with the individual. Here the share of profits asked for can depend on the information sold. In such a setup, the profit sharing obviously enters multiplicatively.

More broadly, the class of nonseparable costly information acquisition models nest behavior generated when there are potentially multiple sources of the cost of information acquisition. For example, the decision maker may incur *both* an additively separable cost to access an information structure, as well as a discounting cost from time delay.[1]

[1] The exact content of this particular example remains unknown.

Caplin et al. (2015) provide a revealed preference test for costly information acquisition when costs are additively separable. When there is only a cost from discounting, we show that behavior is characterized essentially by the Homothetic Axiom of Revealed Preference (See Varian (1983)).

1.2. Methods and related literature

We take a revealed preference approach that builds on the recent contribution of Caplin et al. (2015). In particular, the model considers a decision maker facing actions with state-contingent payoffs. The decision maker chooses an information structure and makes stochastic choices conditioning on the signal received from the information structure. Using state dependent stochastic choices, there is a natural revealed information structure that facilitates the analysis. Our main result characterizes the general model of costly information acquisition with an axiom on expected payoffs that resembles the Generalized Axiom of Revealed Preference.[2] To emphasize the potential interaction between expected utility and cost, we refer to such model as a *nonseparable* costly information acquisition model.

Our results generalize the *No Improving Attention Cycles* condition of Caplin et al. (2015) in the same way that the Generalized Axiom of Revealed Preference generalizes the cyclic monotonicity condition of Rockafellar (1966) or the condition of Koopmans and Beckmann (1957).[3] Importantly, cyclic monotonicity is equivalent to rationalization via a quasi-linear utility function, which imposes cardinal restrictions on consumption data. Thus, Caplin et al. (2015) reflects a type of cardinal model, while the model here is ordinal. Using the intuition from the consumer problem, we show that data consistent with nonseparable costly information preferences can be taken to satisfy quasiconcavity and weak Blackwell monotonicity without loss of generality.

The characterizations here exploit results and intuition from classical consumer theory. An experiment, or signal, is a probability distribution over posteriors (as in Blackwell (1953)). Mathematically, up to a normalization, probability measures and normalized *price* vectors can be viewed as the same object. In the consumer setting, expenditure is computed as the inner-product of price and quantity demanded. Similarly, the ex-ante payoff from the experiment can be computed as the inner-product of the information structure and posterior value function. Thus, the ex-ante payoff can be treated as *wealth*.

The similarity between standard consumer theory and costly information acquisition extends beyond the correspondence of primitives. The nonseparable model of costly information acquisition is defined as a preference that is increasing in ex-ante payoff and depends on the information structure. Similarly, in consumer theory, the indirect utility function is increasing in wealth and depends on prices. For costly information acquisition, the utility of a menu is obtained through maximization with respect to information structures; while in consumer theory, the utility of a consumption bundle can be obtained through minimization of the indirect utility over price vectors. While the optimization principle differs, the same underlying duality holds, which leads to the characterization by a condition resembling the General Axiom of Revealed Preference.

While we have highlighted the similarity to standard consumer theory, there are some technical differences. Most importantly, the information structures and posterior value functions are

[2] See Houthakker (1950); Richter (1966); Varian (1982); Chambers and Echenique (2016).

[3] See also Brown Donald and Calsamiglia (2007) and Chambers et al. (2016) for variants of this condition in an explicit revealed preference framework.

objects in infinite dimensional vector spaces. Thus, our proofs utilize the general results on quasi-concave duality that have been fruitfully studied by Chateauneuf and Faro (2009) and Cerreia-Vioglio et al. (2011a,b). However, once one makes this connection the results follow by leveraging existing revealed preference and duality techniques. As such, the paper also serves as a didactic exercise.

This paper is related to other works on costly information acquisition and boundedly rational behavior. Costly information acquisition has received study from various perspectives in Denti et al. (2016), Ellis (2018), and Matejka and McKay (2014). Costly information acquisition has received study from a revealed preference perspective in Caplin and Martin (2015) and Caplin et al. (2015). Boundedly rational behavior has been studied with revealed preference conditions in Fudenberg et al. (2015), Aguiar (2016), and Allen and Rehbeck (2019).

The paper proceeds as follows. Section 2 presents the notation and some useful facts. Section 3.1 introduces and characterizes the nonseparable costly information acquisition model. Section 3.2 presents a model with a multiplicative cost of information acquisition. Section 3.3 presents a variant of the model whereby choice of information structure is costless, but is constrained to lie in some unknown set. Section 4 compares the nonseparable model to the additive model in Caplin et al. (2015), focusing on both the gap between behavioral axioms and out-of-sample predictions. Section 5 discusses some limitations on when violations of the conditions can be detected. Section 6 discusses methods to numerically deal with unknown utility numbers and prior distributions. Finally, Section 7 contains our concluding remarks. Proofs are relegated to the appendix.

2. Preliminaries

2.1. Notation

We study a decision maker facing actions with state-contingent payoffs.[4] Notation is consistent with Caplin et al. (2015) whenever possible for ease of comparison. We study a variety of models that are increasing in ex-ante payoff and satisfy Bayes' law. A decision maker chooses actions whose outcome depends on a finite number of states of the world. Let Ω denote a finite set of states. Let X denote a set of outcomes. Therefore, the set of all actions (state-contingent outcomes) is X^{Ω}.

The set of all finite decision problems is given by $\mathcal{A} = \{A \subset X^{\Omega} \mid |A| < \infty\}$. As in Caplin et al. (2015) we investigate the situation in which a researcher has a state dependent stochastic choice dataset from decision problems in $\mathcal{A}$. For $A \in \mathcal{A}$, $\Delta(A)$ refers to the set of probability distributions over actions in A.[5]

Definition 1. A *state dependent stochastic choice dataset* is a finite collection of decision problems $\mathcal{D} \subset \mathcal{A}$ and a related set of state dependent stochastic choice functions $\mathcal{P} = \{P_A\}_{A \in \mathcal{D}}$ where $P_A : \Omega \to \Delta(A)$. Denote the probability of choosing an action a conditional on state ω in decision problem A as $P_A(a \mid \omega)$.

We assume that the prior beliefs of the decision maker $\mu \in \Gamma = \Delta(\Omega)$ are known. Moreover, we assume that the utility index $u : X \to \Re$ is a known function.

4 The ideas discussed here are broader if one considers general mappings over posteriors.

5 More generally we use $\Delta(S)$ to refer to Borel probability distributions over the set S.

The following example illustrates the notation and primitives of the model. Throughout the paper, we will build on this example in order to illustrate the different testable implications for various models of costly information acquisition.

Example 1. The set of states is $\Omega = \{\omega_1, \omega_2\}$, and the prior is given by $\mu = \left(\frac{1}{2}, \frac{1}{2}\right)$. There are two menus $A = \{a, b\}$ and $A' = \{a', b'\}$. Let the utilities from actions in menu A and A' take the following values:

$$u(a(\omega)) = \begin{cases} 0 & \text{if } \omega = \omega_1 \\ 2 & \text{if } \omega = \omega_2 \end{cases} \quad u(b(\omega)) = \begin{cases} 2 & \text{if } \omega = \omega_1 \\ 0 & \text{if } \omega = \omega_2 \end{cases}$$
$$u(a'(\omega)) = \begin{cases} 0 & \text{if } \omega = \omega_1 \\ 10 & \text{if } \omega = \omega_2 \end{cases} \quad u(b'(\omega)) = \begin{cases} 10 & \text{if } \omega = \omega_1 \\ 0 & \text{if } \omega = \omega_2 \end{cases}.$$

The state dependent choice probabilities are given by

$$P_A(a \mid \omega_1) = \frac{2}{10} \qquad P_A(a \mid \omega_2) = \frac{8}{10}$$
$$P_{A'}(a' \mid \omega_1) = \frac{3}{10} \qquad P_{A'}(a' \mid \omega_2) = \frac{7}{10}$$

where the choice of b and b' in each state are given by the complementary probabilities.

We take an abstract approach to modeling the choice of an information structure. Each subjective signal is identified with its associated posterior beliefs $\gamma \in \Gamma$. Thus, an information structure is given by a finite support distribution over Γ that satisfies Bayes' law.

Definition 2. The set of *information structures*, Π, comprises all Borel probability distributions over Γ, $\pi \in \Delta(\Gamma)$, that have finite support and satisfy Bayes' law. A distribution over posteriors satisfies Bayes' law if the distribution over posteriors is a mean-preserving spread of the prior μ denoted as

$$E_\pi[\gamma] = \sum_{\gamma \in \text{Supp}(\pi)} \gamma \pi(\gamma) = \mu$$

where $\pi(\gamma) = \Pr(\gamma \mid \pi) = \sum_{\omega \in \Omega} \mu(\omega)\pi(\gamma \mid \omega)$.[6]

We now provide definitions necessary to discuss the ex-ante payoff. The ex-ante payoff is the utility an individual expects to receive for a given information structure. Given a utility index, each decision problem $A \in \mathcal{A}$ induces a posterior value function, $f_A : \Gamma \to \Re$, which maps posterior beliefs γ to the maximal utility possible from A under posterior γ. Formally, for any decision problem A and posterior belief γ

$$f_A(\gamma) = \max_{a \in A} \sum_{\omega \in \Omega} \gamma(\omega) u(a(\omega)).$$

[6] A similar notion of Bayesian plausibility is commonly used in the Bayesian persuasion literature. See, for example, Kamenica and Gentzkow (2011).

Definition 3. We denote the *ex-ante payoff* induced by an information structure $\pi \in \Pi$ as

$$\pi \cdot f_A = \sum_{\gamma \in \text{Supp}(\pi)} \pi(\gamma) f_A(\gamma)$$

where $\pi(\gamma) = \Pr(\gamma \mid \pi) = \sum_{\omega \in \Omega} \mu(\omega)\pi(\gamma \mid \omega)$.

This inner-product representation of ex-ante payoff is intuitive since f_A is a continuous function on Γ and the set of continuous functions on Γ is topologically dual to the set of countably additive Borel measures on Γ (Aliprantis and Border (2006), Theorem 14.15). Recall, π is a finite support Borel probability measure on the set of posteriors, Γ.

We now briefly describe the types of costly information acquisition we examine. The nonseparable information cost takes the form $V(\pi \cdot f_A, \pi)$ where V is strictly increasing in the first component. The multiplicative cost model takes the form $R(\pi)(\pi \cdot f_A)$ where $R : \Pi \to \Re_+$. Finally, in the constrained information acquisition model, the decision maker chooses information structures from a set of feasible and free information structures, $\Pi_c \subseteq \Pi$, and maximizes the expected utility.

2.2. Revealed information structures

While we present several models of costly information acquisition, the analysis relies on the recovery of a *revealed information structure* from the state dependent stochastic choice data. Using the procedure from Caplin et al. (2015), we associate each chosen action to a subjective information state. The revealed information structure may not be identical to the true information structure. However, the revealed information structure is a garbling (as defined in Blackwell (1953)) of the true information structure. The relationship between the true information structures and revealed information structures allows us to order the information structures and deduce conditions on revealed information. Without further delay, we define *revealed posteriors* and *revealed information structures*.

Definition 4. Given $\mu \in \Gamma$, $A \in \mathcal{D}$, $P_A \in \mathcal{P}$, and $a \in \text{Supp}(P_A)$, the *revealed posterior* $\bar{\gamma}_A^a \in \Gamma$ is defined as

$$\begin{aligned} \bar{\gamma}_A^a(\omega) &= \Pr(\omega \mid a \text{ is chosen from } A) \\ &= \frac{\mu(\omega) P_A(a \mid \omega)}{\sum_{\nu \in \Omega} \mu(\nu) P_A(a \mid \nu)}. \end{aligned}$$

Definition 5. Given $\mu \in \Gamma$, $A \in \mathcal{D}$, and $P_A \in \mathcal{P}$, the revealed information structure $\bar{\pi}_A \in \Pi$ is defined by

$$\bar{\pi}_A(\gamma \mid \omega) = \sum_{\{a \in \text{Supp}(P_A) \mid \bar{\gamma}_A^a = \gamma\}} P_A(a \mid \omega)$$

and induces a revealed distribution on posteriors $\bar{\pi}_A$ such that

$$\bar{\pi}_A(\gamma) = \sum_{\omega \in \Omega} \mu(\omega) \bar{\pi}_A(\gamma \mid \omega).$$

The revealed information structure for decision problem A is a finite probability measure over the revealed posteriors.

Example 1 *(continued).* The choices in Example 1 generate the following revealed posteriors

$$\bar{\gamma}_A^a = \left(\frac{2}{10}, \frac{8}{10}\right) \quad ; \quad \bar{\gamma}_A^b = \left(\frac{8}{10}, \frac{2}{10}\right)$$
$$\bar{\gamma}_{A'}^{a'} = \left(\frac{3}{10}, \frac{7}{10}\right) \quad ; \quad \bar{\gamma}_{A'}^{b'} = \left(\frac{7}{10}, \frac{3}{10}\right).$$

Each revealed posterior has the same probability of occurring so that

$$\bar{\pi}_A(\bar{\gamma}_A^a) = \bar{\pi}(\bar{\gamma}_A^b) = \bar{\pi}(\bar{\gamma}_{A'}^{a'}) = \bar{\pi}(\bar{\gamma}_{A'}^{b'}) = \frac{1}{2}.$$

The optimal decision rules for these posteriors give

$$f_A(\bar{\gamma}_A^a) = f_A(\bar{\gamma}_A^b) = 1.6 \quad ; \quad f_A(\bar{\gamma}_{A'}^{a'}) = f_A(\bar{\gamma}_{A'}^{b'}) = 1.4$$
$$f_{A'}(\bar{\gamma}_A^a) = f_{A'}(\bar{\gamma}_A^b) = 8 \quad ; \quad f_{A'}(\bar{\gamma}_{A'}^{a'}) = f_{A'}(\bar{\gamma}_{A'}^{b'}) = 7$$

Earlier, we mentioned that the revealed information structure may differ from the true information structure. To capture this idea, we say an information structure π is consistent with a stochastic choice function P_A, when P_A can be generated by the decision maker using information structure π. This is defined formally below.

Definition 6. For $\pi \in \Pi$ and $P_A \in \mathcal{P}$, we say π is *consistent* with P_A if there exists a choice function $C_A : \text{Supp}(\pi) \to \Delta(A)$ such that for all $\gamma \in \text{Supp}(\pi)$,

$$C_A(a \mid \gamma) > 0 \quad \Rightarrow \quad \sum_{\omega\in\Omega} \gamma(\omega)u(a(\omega)) \geq \sum_{\omega\in\Omega} \gamma(\omega)u(b(\omega)) \quad \text{for all } b \in A$$

and for all $\omega \in \Omega$ and $a \in A$

$$P_A(a \mid \omega) = \sum_{\gamma\in\text{Supp}(\pi)} \pi(\gamma \mid \omega)C_A(a \mid \gamma).$$

As mentioned before, we use the notion of garbling to partially order information structures.

Definition 7. The information structure $\pi \in \Pi$ (with posteriors γ^j) is a garbling of $\rho \in \Pi$ (with posteriors η^i) if there exists a $|\text{Supp}(\rho)| \times |\text{Supp}(\pi)|$ matrix $\mathbf{B}$ with non-negative entries such that for all $i \in \{1, \dots, |\text{Supp}(\rho)|\}$ we have $\sum_{\gamma^j\in\text{Supp}(\pi)} b^{i,j} = 1$ and for all $\gamma^j \in \text{Supp}(\pi)$ and $\omega \in \Omega$ that

$$\pi(\gamma^j \mid \omega) = \sum_{\eta^i\in\text{Supp}(\rho)} b^{i,j} \rho(\eta^i \mid \omega).$$

In other words, π is a garbling of ρ when there is a stochastic matrix $\mathbf{B}$ that can be applied to ρ that yields π. We present some properties about revealed information structures and garblings that are used extensively in the analysis.

Lemma 1. *If π is consistent with P_A, then $\bar{\pi}_A$ is a garbling of π.*

Lemma 1 is proved in Caplin et al. (2015). The lemma says that if an information structure is consistent with the state dependent stochastic choice dataset, then the revealed information structure is a garbling.

Lemma 2. *Given a decision problem $A \in \mathcal{A}$ and $\pi, \rho \in \Pi$ with π a garbling of ρ, then*

$$\rho \cdot f_A \geq \pi \cdot f_A.$$

Lemma 2 follows from Blackwell's theorem (Blackwell, 1953). Blackwell's theorem establishes the notion that some information structures are "more valuable" than others. In particular, if π is a garbling of ρ, then ρ yields weakly higher ex-ante payoff in any decision problem.

Lemma 3. *For all decision problems $A, B \in \mathcal{D}$ if π_A is an information structure consistent with choice data P_A, then $f_B \cdot \pi_A \geq f_B \cdot \bar{\pi}_A$ and $f_A \cdot \pi_A = f_A \cdot \bar{\pi}_A$*

The first inequality of Lemma 3 follows since $\bar{\pi}_A$ is a garbling of π_A. The second equality holds since π_A and $\bar{\pi}_A$ induce the same state dependent choices for menu A, so their ex-ante payoffs are identical.

3. Characterizing costly information models

In this section, we introduce three models of costly information acquisition. The nonseparable information cost model is the most general: it assumes only that the decision maker prefers higher ex-ante expected payoffs from choices and that more informative signals are more costly. Both the multiplicative cost model and the constrained information model are special cases of the nonseparable model.

3.1. Nonseparable information cost

We place minimal restrictions on a decision maker's preferences on information structures. The only condition we impose is that preferences are monotone increasing in ex-ante payoff.

Definition 8. Given $\mu \in \Gamma$ and $u : X \to \Re$, a state dependent stochastic choice dataset $(\mathcal{D}, \mathcal{P})$ has a *nonseparable costly information representation* if there exists a function $V : \Re \times \Pi \to \Re \cup \{-\infty\}$, information structures $\{\pi_A\}_{A\in\mathcal{D}}$, and choice functions $\{C_A\}_{A\in\mathcal{D}}$ such that:

1. Monotonicity: For all $\pi \in \Pi$ and for all $t, s \in \Re$, if $t < s$ and $V(t, \pi) > -\infty$, then $V(t, \pi) < V(s, \pi)$.
2. Non-triviality: For all $t \in \Re$, there exists $\pi_t \in \Pi$ such that $V(t, \pi_t) > -\infty$.
3. Information is optimal: For all $A \in \mathcal{D}$, $\pi_A \in \arg\max_{\pi\in\Pi} V(\pi \cdot f_A, \pi)$.
4. Choices are optimal: For all $A \in \mathcal{D}$, the choice function $C_A : \text{Supp}(\pi_A) \to \Delta(A)$ is such that given $a \in A$ and $\gamma \in \text{Supp}(\pi_A)$ with $C_A(a \mid \gamma) = \Pr_A(a \mid \gamma) > 0$, then
$$\sum_{\omega\in\Omega} \gamma(\omega)u(a(\omega)) \geq \sum_{\omega\in\Omega} \gamma(\omega)u(b(\omega)) \quad \text{for all} \quad b \in A.$$
5. The data are matched: For all $A \in \mathcal{D}$, given $\omega \in \Omega$ and $a \in A$,
$$P_A(a \mid \omega) = \sum_{\gamma\in\text{Supp}(\pi_A)} \pi_A(\gamma \mid \omega)C_A(a \mid \gamma).$$

The above definition is a large class of preferences. However, it allows for the presence of unknown discounting and additively separable information costs. We give some examples of functions nested in this class below.

Example 2. We give a special case of V that allows for both unknown discounting from acquiring information and unknown additively separable costs. Consider when the function V takes the form

$$V(\pi \cdot f_A, \pi) = \delta(\pi)(\pi \cdot f_A) - K(\pi)$$

where $\delta(\pi) \in [0, 1]$ gives the fraction of expected utility lost from discounting and $K(\pi)$ specifies the cost of accessing the information. We note the similarity to the polar form from Gorman (1953) which has been characterized using revealed preference by Cherchye et al. (2016).

Example 3. We consider the special case of a non-separable costly information acquisition given by

$$V(\pi \cdot f_A, \pi) = \Phi(\pi \cdot f_A) - K(\pi)$$

where $\Phi : \mathbb{R} \to \mathbb{R}$ is an increasing transformation of the expected utility and $K(\pi)$ is the cost of accessing information. This example takes the utils from expected utility and transforms them to the same units as the cost function. While this is cosmetically similar to the model by Caplin et al. (2015), the characterization there does not apply.

Example 4. A transformation of utils from expected utility may also be pertinent in the presence of discounting. This is represented as

$$V(\pi \cdot f_A, \pi) = \delta(\pi)\Phi(\pi \cdot f_A) - K(\pi)$$

where $\Phi : \mathbb{R} \to \mathbb{R}$ is an increasing transformation of the expected utility, $\delta(\pi) \in [0, 1]$ gives the fraction of utils lost from acquiring information, and $K(\pi)$ is the cost of accessing information.

We now define the properties that completely characterize the model. The first condition is similar to the generalized axiom or revealed preference.

Condition 1 *(Generalized Axiom of Costly Information (GACI)).* We say the dataset $(\mathcal{D}, \mathcal{P})$ satisfies *GACI* if for all sequences $(\bar{\pi}_{A_1}, f_{A_1}), \ldots, (\bar{\pi}_{A_k}, f_{A_k})$ with $A_i \in \mathcal{D}$ for which $\bar{\pi}_{A_i} \cdot f_{A_i} \leq \bar{\pi}_{A_i} \cdot f_{A_{i+1}}$ for all i (with addition modulo k), then equality holds throughout.

Comparing this condition to GARP, we see that the $\bar{\pi}$ play a role similar to prices and the f terms play a role similar to consumption bundles albeit with the inequality reversed. The GACI condition rules out the possibility of cycles in ex-ante payoff across different decision problems. Using this condition, we invoke a version of Afriat's theorem (see Chambers and Echenique (2016)).

Lemma 4 *(Afriat's Theorem). Let $\mathcal{D}$ be finite. For all $(A, B) \in \mathcal{D}^2$, let $\alpha_{A,B} \in \Re$. If for all $A \in \mathcal{D}$ one has $\alpha_{A,A} = 0$ and for any sequence $A_1, A_2, \ldots, A_k \in \mathcal{D}$ with $\alpha_{A_i,A_{i+1}} \leq 0$ for all i (with addition mod k) it follows that $\alpha_{A_i,A_{i+1}} = 0$ for all i, then there exist numbers U_A and $\lambda_A > 0$ such that for all $(A, B) \in \mathcal{D}^2$, $U_A \leq U_B + \lambda_B \alpha_{B,A}$.*

The other condition that characterizes the nonseparable costly information representation is the no improving action switches (NIAS) condition. This condition was first examined in the study of Bayesian decision makers in Caplin and Martin (2015).

Condition 2 *(No Improving Action Switches (NIAS)).* Given $\mu \in \Gamma$ and $u : X \to \mathbb{R}$, a dataset $(\mathcal{D}, \mathcal{P})$ satisfies NIAS if, for every $A \in \mathcal{D}$, $a \in \text{Supp}(P_A)$, and $b \in A$,

$$\sum_{\omega \in \Omega} \mu(\omega) P_A(a \mid \omega)(u(a(\omega)) - u(b(\omega))) \geq 0$$

As we show in Theorem 1 below, the combination of GACI and NIAS completely characterizes the model of nonseparable costly information acquisition; moreover, one can impose additional properties on the nonseparable costly information representation. These conditions are monotonicity, quasiconcavity, and a normalization property on the function $V(\cdot, \cdot)$.

Condition 3. The function $V : \Re \times \Pi \to \Re \cup \{-\infty\}$ satisfies weak monotonicity in information if for any $t \in \Re$ and $\pi, \rho \in \Pi$ with π a garbling of ρ, then

$$V(t, \rho) \leq V(t, \pi).$$

The monotonicity condition says that if one adds noise to a signal ρ, then the noisier signal is cheaper. This is one definition of monotonicity and it agrees with the notion of informativeness introduced in Blackwell (1953).

Condition 4. The function $V : \Re \times \Pi \to \Re \cup \{-\infty\}$ is quasiconcave if for any $(t_1, \pi_1), (t_2, \pi_2) \in \Re \times \Pi$ and $\lambda \in [0, 1]$,

$$V(\lambda t_1 + (1-\lambda)t_2, \lambda\pi_1 + (1-\lambda)\pi_2) \geq \min\{V(t_1, \pi_1), V(t_2, \pi_2)\}.$$

This condition says if there is a mixture between two ex-ante payoffs and information structures, then the utility of the mixture is weakly higher than the worst case of the two environments. In particular, this implies quasiconcavity in information structures if one sets $t_1 = \pi_1 \cdot f$ and $t_2 = \pi_2 \cdot f$.

Condition 5. Define π_0 as the information structure with $\pi_0(\mu|\omega) = 1$ for all $\omega \in \Omega$. The function $V : \Re \times \Pi \to \Re \cup \{-\infty\}$ satisfies the normalization if $V(0, \pi_0) = 0$.

The normalization condition says that utility is normalized to zero when the ex-ante payoff is zero and an individual does not update their prior.

Theorem 1. *Given $\mu \in \Gamma$ and $u : X \to \Re$, the dataset $(\mathcal{D}, \mathcal{P})$ has a nonseparable costly information representation if and only if it satisfies GACI and NIAS. Moreover, if GACI and NIAS are satisfied, then one can find a V that rationalizes the data with a nonseparable costly information representation and satisfies Conditions 3, 4, and 5.*[7]

While we characterize a general model, we show that it is without loss to assume an individual's payoff is quasiconcave in the information structure for a fixed level of expected utility.

[7] As an obvious consequence of Theorem 1, the model is also empirically equivalent to a model in which there is an endogenous (possibly singleton) set H of hidden actions. In particular, the model $\pi \in \max_{h \in H} V(u \cdot \pi, h, \pi)$ is equivalent to ours since one could choose a set H to be a singleton. Thus, unlike Machina (1984), adding the potential for hidden actions does not change the content of observable behavior, and hence is non-testable. This is also true of the model in Caplin et al. (2015).

Quasiconcavity might be interpreted as an informal statement that more informative structures are more costly to achieve. This is not meant in a Blackwell sense. Rather, given two information structures with known costs, taking a convex combination of them leads to a structure which is less costly than the highest cost of the two. The convex combination of information structures is intuitively less informative. While this property is certainly intuitive, the result is mathematical and owes to the structure of data and the same phenomenon whereby Afriat determined that convexity (as a property of preferences over consumption space) is non-testable.

Example 1 *(continued).* One can verify from the earlier information that NIAS holds. To test whether the stochastic choice pattern can be rationalized by the nonseparable costly information acquisition model, it remains to verify that GACI holds. To this end, observe that

$$\bar{\pi}_A \cdot f_A = 1.6 \quad ; \quad \bar{\pi}_{A'} \cdot f_A = 1.4$$
$$\bar{\pi}_A \cdot f_{A'} = 8 \quad ; \quad \bar{\pi}_{A'} \cdot f_{A'} = 7.$$

Now, since

$$\bar{\pi}_A \cdot f_A < \bar{\pi}_A \cdot f_{A'} \quad \text{and} \quad \bar{\pi}_{A'} \cdot f_A < \bar{\pi}_{A'} \cdot f_{A'},$$

there are no cycles that violate GACI. The stochastic choice pattern can be rationalized by the nonseparable costly information acquisition model.

3.2. Multiplicative information cost

We now study a multiplicative costly information representation. In this representation, the cost is interpreted as losing a fraction of the ex-ante payoff. We interpret this cost as resulting from discounting due to unobserved delay when acquiring information.

Definition 9. Given $\mu \in \Gamma$ and $u : X \to \Re_+$, a state dependent stochastic choice dataset $(\mathcal{D}, \mathcal{P})$ has a *multiplicative costly information representation* if there exists a function $R : \Pi \to \Re_+$, information structures $\{\pi_A\}_{A\in\mathcal{D}}$, and choice functions $\{C_A\}_{A\in\mathcal{D}}$ such that:

1. Non-triviality: There exists $\pi \in \Pi$ such that $R(\pi) > 0$.
2. Information is optimal: For all $A \in \mathcal{D}$,
   $\pi_A \in \arg\max_{\pi\in\Pi} [R(\pi)(\pi \cdot f_A)]$.
3. Choices are optimal: For all $A \in \mathcal{D}$, the choice function $C_A : \text{Supp}(\pi_A) \to \Delta(A)$ is such that given $a \in A$ and $\gamma \in \text{Supp}(\pi_A)$ with $C_A(a \mid \gamma) = \Pr_A(a \mid \gamma) > 0$, then
$$\sum_{\omega\in\Omega} \gamma(\omega)u(a(\omega)) \geq \sum_{\omega\in\Omega} \gamma(\omega)u(b(\omega)) \quad \text{for all} \quad b \in A.$$
4. The data are matched: For all $A \in \mathcal{D}$, given $\omega \in \Omega$ and $a \in A$,
$$P_A(a \mid \omega) = \sum_{\gamma\in\text{Supp}(\pi_A)} \pi_A(\gamma \mid \omega)C_A(a \mid \gamma).$$

We note that one difference in the statement of the multiplicative costly information representation is that the utility index u is required to be non-negative. While this is more restrictive than the other cases, this is a common property of multiplicative representations. For example, Chateauneuf and Faro (2009) make such an assumption. The condition that characterizes the multiplicative costly information representation is a version of the homothetic axiom of revealed preference; see Varian (1983).[8]

Condition 6 *(Homothetic Axiom of Costly Information (HACI)).* Given a datatset $(\mathcal{D}, \mathcal{P})$, define $\mathcal{D}_0 = \{A \in \mathcal{D} \mid \sum_{\omega\in\Omega} \mu(\omega)u(a(\omega)) = 0 \text{ for all } a \in A\}$. We say the dataset $(\mathcal{D}, \mathcal{P})$ satisfies *HACI* if for all sequences $(\bar{\pi}_{A_1}, f_{A_1}), \ldots, (\bar{\pi}_{A_k}, f_{A_k})$ with $A_i \in \mathcal{D}\backslash\mathcal{D}_0$, that $\prod_{i=1}^{k} \frac{\bar{\pi}_{A_i}\cdot f_{A_{i+1}}}{\bar{\pi}_{A_i}\cdot f_{A_i}} \leq 1$ where addition for the index i is modulo k.

HACI is essentially the homothetic axiom of revealed preference restricted to decision problems that give positive ex-ante payoff. The decision problems that give zero ex-ante payoff are removed: since utility is non-negative, these decision problems must consist entirely of actions that give zero payoff in every state. Any choice behavior in these decision problems can be trivially rationalized, and they would create an indeterminate fraction above if not removed from the datatset.

As in the case of the nonseparable costly information representation, we are able to put additional properties on the function R. For instance, we can find an R that respects monotonicity with respect to the Blackwell partial order, is concave, and satisfies a normalization property. We now define these properties and then give a statement of the theorem.

Condition 7. The function $R : \Pi \to \mathbb{R}_+$ satisfies weak monotonicity in information if for any $\rho, \pi \in \Pi$ where π is a garbling of ρ,

$$R(\pi) \geq R(\rho).$$

Condition 8. The function $R : \Pi \to \mathbb{R}_+$ is concave in information structures if for any $\pi_1, \pi_2 \in \Pi$ and $\lambda \in [0, 1]$,

$$R(\lambda\pi_1 + (1-\lambda)\pi_2) \geq \lambda R(\pi_1) + (1-\lambda)R(\pi_2).$$

Condition 9. Define π_0 as the information structure with $\pi_0(\mu|\omega) = 1$ for all $\omega \in \Omega$. The function R satisfies normalization if $R(\pi_0) = 1$ and $R : \Pi \to [0, 1]$.

Theorem 2. *Given $\mu \in \Gamma$ and $u : X \to \Re_+$, the dataset $(\mathcal{D}, \mathcal{P})$ has a multiplicative costly information representation if and only if it satisfies HACI and NIAS. Moreover, if HACI and NIAS are satisfied, then one can find an R that rationalizes the data with a multiplicative costly information representation that satisfies Conditions 7, 8, and 9.*

Example 1 *(continued).* In additional to the general nonseparable cost model, the stochastic choice data in Example 1 can also be rationalized by the multiplicative cost model. Recall the multiplicative cost model is a special case of the nonseparable cost model. In fact, the data satisfies HACI since

$$\left(\frac{\bar{\pi}_A \cdot f_{A'}}{\bar{\pi}_A \cdot f_A}\right)\left(\frac{\bar{\pi}_A \cdot f_{A'}}{\bar{\pi}_A \cdot f_A}\right) = \left(\frac{80}{16}\right)\left(\frac{14}{70}\right) = 1.$$

[8] It can also be derived as a relatively easy corollary from the work of Rochet (1987).

Lastly, we note that one could re-parameterize $R(\pi)$ to be $(1 - K(\pi))$ where $K : \Pi \to [0, 1]$ is interpreted as a fractional cost of information relative to the ex-ante payoff. Alternatively, one could re-parameterize $R(\pi)$ to a function $\delta^{T(\pi)}$ where $T(\pi) \geq 0$ represents time delay.

3.3. Constrained information acquisition

We now consider when an individual is constrained to choose an information structure from a fixed set of information structures. The interpretation is that the decision maker does not have access to the full set of information structures. Moreover, all available information structures are costless.

Definition 10. Given $\mu \in \Gamma$ and $u : X \to \Re$, a state dependent stochastic choice dataset $(\mathcal{D}, \mathcal{P})$ has a *constrained costly information representation* if there exists a set $\Pi_c \subseteq \Pi$ of available information structures, information structures $\{\pi_A\}_{A\in\mathcal{D}}$, and choice functions $\{C_A\}_{A\in\mathcal{D}}$ such that:

1. Non-triviality: The set $\Pi_c \neq \emptyset$.
2. Information is optimal: For all $A \in \mathcal{D}$, $\pi_A \in \arg\max_{\pi\in\Pi_c} \pi \cdot f_A$.
3. Choices are optimal: For all $A \in \mathcal{D}$, the choice function $C_A : \text{Supp}(\pi_A) \to \Delta(A)$ is such that given $a \in A$ and $\gamma \in \text{Supp}(\pi_A)$ with $C_A(a \mid \gamma) = \Pr_A(a \mid \gamma) > 0$, then
$$\sum_{\omega\in\Omega} \gamma(\omega)u(a(\omega)) \geq \sum_{\omega\in\Omega} \gamma(\omega)u(b(\omega)) \quad \text{for all} \quad b \in A.$$
4. The data are matched: For all $A \in \mathcal{D}$, given $\omega \in \Omega$ and $a \in A$,
$$P_A(a \mid \omega) = \sum_{\gamma\in\text{Supp}(\pi_A)} \pi_A(\gamma \mid \omega)C_A(a \mid \gamma).$$

A constrained costly information structure is characterized by a condition similar to the Weak Axiom of Profit Maximization (Varian, 1984). Using this intuition, the revealed information structures are analogous to goods and f_A are analogous to prices of goods. To avoid confusion with the Weak Axiom of Revealed Preference, we call this the Binary Axiom of Costly Information.

Condition 10 *(Binary Axiom of Costly Information (BACI)).* The dataset $(\mathcal{D}, \mathcal{P})$ satisfies *BACI* if for all $A, B \in \mathcal{D}$, it follows that

$$\bar{\pi}_A \cdot f_A \geq \bar{\pi}_B \cdot f_A.$$

We note that BACI is a stronger condition than GACI or HACI. In particular, one must check that the ex-ante expected utility in a menu is greater than the ex-ante utility obtained from any other observed revealed information structure. Similar to the nonseparable case, additional structure can be placed on a constrained costly information representation without restricting observable behavior. Using standard arguments, the constraint set Π_c can be made convex.

Theorem 3. *Given $\mu \in \Gamma$ and $u : X \to \Re$, the dataset $(\mathcal{D}, \mathcal{P})$ has a constrained costly information representation if and only if it satisfies BACI and NIAS. Moreover, if BACI and NIAS are* *satisfied, then one can find a convex set Π_c that rationalizes the data with a constrained costly information representation.*

Example 1 *(continued).* The stochastic choice data in Example 1 *cannot* be rationalized by the constrained costly information model, since the dataset violates BACI: $\bar{\pi}_{A'} \cdot f_{A'} = 7 < 8 = \bar{\pi}_A \cdot f_{A'}$.

The nonseparable information cost model generalizes all three alternative models of costly information acquisition. The constrained model, by contrast, is the most special one, and is a special case of both the additive and multiplicative models: it can be regarded as a multiplicative model with function $R(.)$ equals to 1 on Π_c and 0 everywhere else; alternatively, it can also be regarded as an additive model where the additive cost function $K(.)$ equals to 0 on Π_c and $+\infty$ everywhere else.

4. Relationship between nonseparable and additive model

As a point of reference, we examine in detail how the nonseparable costly information representation relates to the additive costly information representation in Caplin et al. (2015).

We first review the definition of an additive costly information model, and show that the nonseparable model generalizes the additive model. We then show that one particular limitation of the additive model is that it forbids individuals from choosing less information whenever the menu provides a higher return to gathering information, even if menus that generate higher returns might also entail higher costs for information.

Definition 11. Given $\mu \in \Gamma$ and $u : X \to \Re$, a state dependent stochastic choice dataset $(\mathcal{D}, \mathcal{P})$ has an *additive costly information representation* if there exists a function $K : \Pi \to \bar{\mathbb{R}} \cup \{\infty\}$, information structures $\{\pi_A\}_{A \in \mathcal{D}}$, and choice functions $\{C_A\}_{A \in \mathcal{D}}$ such that:

1. Non-triviality: There exists $\pi \in \Pi$ such that $K(\pi) < \infty$.
2. Information is optimal: For all $A \in \mathcal{D}$, $\pi_A \in \arg\max_{\pi \in \Pi} [\pi \cdot f_A - K(\pi)]$.
3. Choices are optimal: For all $A \in \mathcal{D}$, the choice function $C_A : \text{Supp}(\pi_A) \to \Delta(A)$ is such that given $a \in A$ and $\gamma \in \text{Supp}(\pi_A)$ with $C_A(a \mid \gamma) = \Pr_A(a \mid \gamma) > 0$, then
$$\sum_{\omega \in \Omega} \gamma(\omega) u(a(\omega)) \geq \sum_{\omega \in \Omega} \gamma(\omega) u(b(\omega)) \quad \text{for all} \quad b \in A.$$
4. The data are matched: For all $A \in \mathcal{D}$, given $\omega \in \Omega$ and $a \in A$,
$$P_A(a \mid \omega) = \sum_{\gamma \in \text{Supp}(\pi_A)} \pi_A(\gamma \mid \omega) C_A(a \mid \gamma).$$

Caplin et al. (2015) showed that an additive costly information representation is characterized by the NIAS condition and a no improving attention cycles (NIAC) condition. The NIAC condition is defined below.

Condition 11 *(No Improving Attention Cycles (NIAC)).* Given $\mu \in \Gamma$ and $u : X \to \mathbb{R}$, a dataset $(\mathcal{D}, \mathcal{P})$ satisfies NIAC if for all sequences $(\bar{\pi}_{A_1}, f_{A_1}), \ldots, (\bar{\pi}_{A_k}, f_{A_k})$ with $A_i \in \mathcal{D}$, then

$$\sum_{i=1}^{k} \bar{\pi}_{A_i} \cdot f_{A_i} \geq \sum_{i=1}^{k} \bar{\pi}_{A_{i+1}} \cdot f_{A_i}$$

where addition of the indices is modulo k.

The interpretation of NIAC is that one cannot cycle through the information structures and improve the ex-ante payoff. From the definition of NIAC and GACI, it is easy to see that if a dataset satisfies NIAC, then the dataset also satisfies GACI with equality.

Proposition 1. *If the dataset $(\mathcal{D}, \mathcal{P})$ satisfies NIAC, then it also satisfies GACI.*

To see why, we show that a violation of GACI implies a violation of NIAC. First, note a violation of GACI implies the existence of a sequence $\bar{\pi}_{A_i} \cdot f_{A_i} \leq \bar{\pi}_{A_i} \cdot f_{A_{i+1}}$, where, say, $\bar{\pi}_{A_k} \cdot f_{A_k} < \bar{\pi}_{A_k} \cdot f_{A_1}$. Subtracting obtains that for each i, $\bar{\pi}_{A_i} \cdot (f_{A_i} - f_{A_{i+1}}) \leq 0$, with one inequality strict, whereby $\sum_i \bar{\pi}_{A_i} \cdot (f_{A_i} - f_{A_{i+1}}) < 0$. Rearranging terms now obtains a violation of NIAC. We come back to Example 1 again as an illustration.

Example 1 *(continued).* Observe that the stochastic choices described earlier cannot be rationalized by the additive cost model since

$$\bar{\pi}_A \cdot f_A + \bar{\pi}_{A'} \cdot f_{A'} = 8.6 < 9.4 = \bar{\pi}_A \cdot f_{A'} + \bar{\pi}_{A'} \cdot f_A$$

and so NIAC fails.

4.1. Gross return from information

We note that an additively separable model forbids an individual from choosing a less informative information structure when there are "higher gross return from information", but this is allowed under the nonseparable cost model. This flexibility may be relevant if a menu generating higher gross return from information may at the same time entail higher costs to gathering information. For example, this occurs when the cost of information is the discounting incurred from waiting. We formally define "higher returns to information" below.

Definition 12. Menu A provides a *higher gross return from information* than menu B if for any information structure π, and π' a garbling of π with $\pi' \neq \pi$, we have[9]

$$\pi \cdot f_A - \pi' \cdot f_A > \pi \cdot f_B - \pi' \cdot f_B.$$

We establish that an individual with an additive costly information representation can never choose a less informative information structure when faced with a menu that has a higher gross return from information.

Proposition 2. *Suppose $\mathcal{D} = \{A, B\}$ for dataset $(\mathcal{D}, \mathcal{P})$ with menu A providing a higher gross return from information than menu B. If $\bar{\pi}_A$ is a garbling of $\bar{\pi}_B$ and $\bar{\pi}_A \neq \bar{\pi}_B$, then the choice data violates NIAC and thus cannot be generated by an additive costly information representation.*

[9] This definition is non-vacuous. In fact, it can be shown that menu A provides a higher gross return from information than menu B if and only if $f_A = f_B + g$ where g is a strictly convex function.

The next result shows that a nonseparable model, on the contrary, always accommodates this behavior if the menu that provides a higher gross return from information also yields higher utility for any posterior.

Proposition 3. *Suppose $\mathcal{D} = \{A, B\}$ for dataset $(\mathcal{D}, \mathcal{P})$ with menu A providing a higher gross return from information than menu B. If $\bar{\pi}_A$ is a garbling of $\bar{\pi}_B$, NIAS is satisfied, and $f_A > f_B$,*[10] *then this dataset is rationalized by a nonseparable costly information representation.*

4.2. Out of sample prediction

One may wonder what type of data will violate GACI, or in other words, the extent to which the nonseparable model puts meaningful constraints on choice behavior. In this section, we first demonstrate the restrictions on choice probabilities for a specific two state environment, with a uniform prior and menus of two acts. This simple environment allows us to obtain a closed-form expression for the restrictions on choice probabilities. We then provide a numerical example as a further illustration.

Let the states be given by $\Omega = \{\omega_1, \omega_2\}$. Let the menus be denoted $A = \{a, b\}$ and $A' = \{a', b'\}$. Assume without loss that $u(a(\omega_1)) > u(b(\omega_1))$ and $u(b(\omega_2)) > u(a(\omega_2))$. Similarly, assume that $u(a'(\omega_1)) > u(b'(\omega_1))$ and $u(b'(\omega_2)) > u(a'(\omega_2))$. We also assume each state is equally likely so that $\mu = (\frac{1}{2}, \frac{1}{2})$.

As is shown in Caplin et al. (2015), NIAS on menu A in this environment is equivalent to

$$P_A(a \mid \omega_1) \geq \max \begin{cases} \frac{u(b(\omega_2))-u(a(\omega_2))}{u(a(\omega_1))-u(b(\omega_1))} P_A(a \mid \omega_2), \\ \frac{u(b(\omega_2))-u(a(\omega_2))}{u(a(\omega_1))-u(b(\omega_1))} P_A(a \mid \omega_2) + \frac{u(a(\omega_1))+u(a(\omega_2))-u(b(\omega_1))-u(b(\omega_2))}{u(a(\omega_1))-u(b(\omega_1))} \end{cases}.$$

The NIAS condition imposes similar restrictions for the choices from menu A'.

We focus on the case when choices are aligned. We say the choices of menu A are aligned with choices from menu A' when

$$a = \arg\max_{c \in \{a,b\}} \sum_{\omega \in \{\omega_1, \omega_2\}} \bar{\gamma}_{A'}^{a'}(\omega) u(c(\omega)),$$
$$b = \arg\max_{c \in \{a,b\}} \sum_{\omega \in \{\omega_1, \omega_2\}} \bar{\gamma}_{A'}^{b'}(\omega) u(c(\omega)).$$

Essentially, choices are aligned when actions that are "similar" across menus are chosen at the revealed posterior of the similar action. Note that action a and a' both have greater payoffs in state ω_1. Thus, a is aligned with a' when it is chosen at the revealed posterior of a'. This assumption is made to make the algebra tractable. The same assumption is also implicitly assumed in Caplin et al. (2015). We also assume that the choice of menu A' are aligned with A using analogous conditions.

Now, there is a violation of GACI if

$$\bar{\pi}_A \cdot f_A \leq \bar{\pi}_A \cdot f_{A'} \quad \text{and} \quad \bar{\pi}_{A'} \cdot f_{A'} \leq \bar{\pi}_{A'} \cdot f_A$$

with one inequality strict. Under the above assumptions of NIAS and aligned choices, a *violation* of GACI is equivalent to the choice probabilities simultaneously satisfying the following two inequalities with one inequality strict:

[10] We say $f_A > f_B$ if $f_A(\gamma) > f_B(\gamma)$ for all $\gamma \in \Gamma$.

$$\begin{aligned} P_A(a \mid \omega_1)\Delta_1 + P_A(a \mid \omega_2)\Delta_2 \le \beta \\ P_{A'}(a' \mid \omega_1)\Delta_1 + P_{A'}(a' \mid \omega_2)\Delta_2 \ge \beta \end{aligned} \tag{1}$$

where

$$\begin{aligned} \Delta_1 &= u(a(\omega_1)) - u(a'(\omega_1)) + u(b'(\omega_1) - u(b(\omega_1)) \\ \Delta_2 &= u(a(\omega_2)) - u(a'(\omega_2)) + u(b'(\omega_2) - u(b(\omega_2)) \\ \beta &= u(b'(\omega_1)) + u(b'(\omega_2)) - u(b(\omega_1)) - u(b(\omega_2)). \end{aligned}$$

Therefore, any probabilities that satisfy these inequalities with at least one strict inequality violate a nonseparable costly information representation.

In general, suppose one has a menu $M \in \mathcal{A}$ such that $M \notin \mathcal{D}$. If the dataset $\mathcal{D}$ satisfies NIAS and GACI, we can use the information to place bounds on the information structures that are consistent with the model using the restrictions of GACI and NIAS. The full set of restrictions is given by a *supporting set* as defined in Varian (1984).

Denote the set of information structures that support the menu M that are consistent with GACI and NIAS by

$$S_{GACI}(M) = \{\pi_M \in \Pi \mid \{(\bar{\pi}_A, f_A)\}_{A \in \mathcal{D}} \cup (\pi_M, f_M) \text{ satisfies NIAS and GACI}\}.$$

This set places restrictions on π_M that can be translated to restrictions on individual state dependent stochastic choices. It is easy to define supporting sets for multiplicatively separable, additively separable, and constrained costly information representation. While the supporting set is often difficult to compute, it provides the full set of π_M consistent with a given representation.

4.3. Numerical example: GACI vs NIAC

Let the payoffs of the actions in menus A and A' take the following values:

$$u(a(\omega)) = \begin{cases} 5 & \text{if } \omega = \omega_1 \\ 0 & \text{if } \omega = \omega_2 \end{cases} \qquad u(b(\omega)) = \begin{cases} 1 & \text{if } \omega = \omega_1 \\ 4 & \text{if } \omega = \omega_2 \end{cases}$$
$$u(a'(\omega)) = \begin{cases} 4 & \text{if } \omega = \omega_1 \\ 1 & \text{if } \omega = \omega_2 \end{cases} \qquad u(b'(\omega)) = \begin{cases} 2 & \text{if } \omega = \omega_1 \\ 3 & \text{if } \omega = \omega_2 \end{cases} \quad .$$

We assume that choices are aligned to compare the restrictions of GACI from the inequalities in (1) to the restrictions of NIAC from Caplin et al. (2015). Substituting the above utility numbers into inequalities (1), we can see that the choice probabilities from A and A' violate GACI if and only if

$$P_A(a \mid \omega_1) + P_A(b \mid \omega_2) \le 1 \;\text{ and }\; P_{A'}(a' \mid \omega_1) + P_{A'}(b' \mid \omega_2) \ge 1 \tag{2}$$

with one inequality strict.

On the other hand, substituting the above utility numbers into the restrictions of NIAC from Caplin et al. (2015), we see that a violation of NIAC for this decision problem is equivalent to

$$P_A(a \mid \omega_1) + P_A(b \mid \omega_2) - [P_{A'}(a' \mid \omega_1) + P_{A'}(b' \mid \omega_2)] < 0. \tag{3}$$

By comparing (2) and (3) above, it is straightforward to see that a violation of GACI implies a violation of NIAC, but not the other way around. In Appendix B, we provide Monte Carlo simulations in the spirit of Bronars (1987) and Beatty and Crawford (2011) to examine the restrictions of GACI, NIAC, HACI, BACI, and NIAS for the experiments run in Dean et al. (2017) and some other illustrative examples.

5. Limitations

The revealed preference conditions for costly information acquisition often provide interesting bounds and intuition for these models. Moreover, we note that an additive costly information representation has the property of being translation invariant in ex-ante payoff. Similarly, a multiplicative costly information representation has the property of being scale invariant in ex-ante payoff.

One may want to look at choices from menus of this type to violate an additively separable or multiplicatively separable costly information representation, respectively. However, a dataset with menus that are additive utility translations of one another always satisfy NIAC. Similarly, a dataset with menus that are scale shifts of one another always satisfy HACI. Thus, these natural changes to environments do not provide any information on preferences. This shows that there are limits to what can be learned about preferences for these environments.

To show this result formally, we provide two definitions. For a menu $A = \{a_1, \ldots, a_n\} \in \mathcal{A}$ and $c \in \Re$ let $A + c = \{a_1 + c, \ldots, a_n + c\}$ be the menu that adds a constant utility c to each act. That is, $u\big((a_i + c)(\omega)\big) = u(a_i(\omega)) + c$ for $i = 1, \ldots, n$ and all $\omega \in \Omega$. Similarly, let $cA = \{ca_1, \ldots, ca_n\}$ be the menu where the utility of all acts is multiplied by c, so $u\big((ca_i)(\omega)\big) = cu(a_i(\omega))$ for $i = 1, \ldots, n$ and all $\omega \in \Omega$.

Proposition 4. *Let $\mu \in \Gamma$ and $u : X \to \Re$. If the dataset $(\mathcal{D}, \mathcal{P})$ satisfies NIAS and $\mathcal{D} = \{A + c_1, A + c_2, \ldots, A + c_M\}$ where $c_m \in \Re$ for all $m = 1, \ldots, M$, then the dataset is rationalized by the additive costly information representation.*

Proposition 5. *Let $\mu \in \Gamma$ and $u : X \to \Re_+$. Suppose the dataset $(\mathcal{D}, \mathcal{P})$ satisfies NIAS and $\mathcal{D} = \{c_1 A, c_2 A, \ldots, c_M A\}$ where $c_m \in \Re_+$ for all $m = 1, \ldots, M$, then the dataset is rationalized by the multiplicative costly information representation.*

6. Unknown utility, unknown prior

The model we examined requires utility and the prior to be known. In fact, we were able to write the revealed information structures in this "reduced-form" only because the prior probabilities are known. With that being said, even if the utility is unknown, then some implications of the model *may* be derived. As a general rule, if utility is totally unrestricted, then the model has no content. This is a relatively standard observation, and owes to the fact that complete indifference can rationalize everything. On the other hand, in our abstract model, it makes sense to ask that utility lies in some set, $\mathcal{U}$.[11]

A violation of the general costly information representation now occurs when GACI is violated for each $u \in \mathcal{U}$. There are nontrivial examples of such violations. For example, there may be monotonicity restrictions on the prizes X that translate to $\mathcal{U}$. In this case, we note that a non-trivial restriction of all models is that a state-wise dominated action cannot be chosen.

When the prior is not known, things can become more complicated. However, if one restricts the prior to be in a set, $\mathcal{M}$, we can perform the tests using a grid search over priors.[12] The complications arise when one looks for a complete characterization without restrictions on the priors.

[11] The notion of a value function f_A now necessarily also depends on $u \in \mathcal{U}$.

[12] Grid search procedures have been used to examine risk preferences using a revealed preference approach in Polisson et al. (2017).

First, we would require a more general model of signal structures where the "reduced-form" of the revealed information structure is unknown. Second, we can no longer use the knowledge of μ to help understand the ex-ante expected utility or which actions are optimal. An interesting study of a related question is due to Oliveira and Lamba (2018).

7. Conclusion

In this paper, we provide revealed preference characterizations for several models of costly information acquisition. The most general form allows for costs from time delay in addition to an additively separable cost. The characterization of these models follows directly from classical revealed preference theory. We also provide examples showing how the information acquisition differs across models.

Appendix A. Proofs of results

Proof of Theorem 1. ($\Rightarrow$) First, we show that a nonseparable costly information representation satisfies NIAS. Fix $A \in \mathcal{D}$, $\pi_A \in \arg\max_{\pi\in\Pi} V(\pi \cdot f_A, \pi)$, and $C_A : \text{Supp}(\pi_A) \to \Delta(A)$ and $a \in \text{Supp}(P_A)$. By definition of a nonseparable costly information representation, we know that the $V(\pi_A \cdot f_A, \pi_A)$ is monotone in $\pi_A \cdot f_A$ and choices are optimal conditional on posteriors. Thus, if a was chosen when γ was realized, then the expected utility must be weakly higher for these γ. For γ such that $C_A(a \mid \gamma) > 0$,

$$\sum_{\omega\in\Omega} \gamma(\omega)u(a(\omega)) \geq \sum_{\omega\in\Omega} \gamma(\omega)u(b(\omega)) \quad \forall b \in A.$$

The proof now follows from arguments in Caplin et al. (2015) that are reproduced here for completeness. Recall that

$$\gamma(\omega) = \frac{\mu(\omega)\pi_A(\gamma \mid \omega)}{\sum_{\nu\in\Omega} \mu(\nu)\pi_A(\gamma \mid \nu)},$$

which can be substituted on both sides and the denominator cancels so

$$\sum_{\omega\in\Omega} \mu(\omega)\pi_A(\gamma \mid \omega)u(a(\omega)) \geq \sum_{\omega\in\Omega} \mu(\omega)\pi_A(\gamma \mid \omega)u(b(\omega)) \quad \forall b \in A.$$

Therefore,

$$\sum_{\gamma\in\text{Supp}(\pi_A)} C_A(a \mid \gamma) \left[\sum_{\omega\in\Omega} \mu(\omega)\pi_A(\gamma \mid \omega)u(a(\omega))\right]$$
$$\geq \sum_{\gamma\in\text{Supp}(\pi_A)} C_A(a \mid \gamma) \left[\sum_{\omega\in\Omega} \mu(\omega)\pi_A(\gamma \mid \omega)u(b(\omega))\right] \quad \forall b \in A$$

since $C_A(a \mid \gamma)$ are either zero or positive multiples of the earlier introduced inequalities. Next, recall from data matching that $P_A(a \mid \omega) = \sum_{\gamma\in\text{Supp}(\pi_A)} \pi_A(\gamma \mid \omega)C_A(a \mid \gamma)$. Therefore, we see that

$$\sum_{\omega\in\Omega} \mu(\omega)u(a(\omega))P_A(a \mid \omega) = \sum_{\omega\in\Omega} \mu(\omega)u(a(\omega)) \left[\sum_{\gamma\in\text{Supp}(\pi_A)} \pi_A(\gamma \mid \omega)C_A(a \mid \gamma)\right]$$

$$= \sum_{\gamma \in \text{Supp}(\pi_A)} C_A(a \mid \gamma) \left[\sum_{\omega \in \Omega} \mu(\omega) u(a(\omega)) \pi_A(\gamma \mid \omega) \right]$$
$$\geq \sum_{\gamma \in \text{Supp}(\pi_A)} C_A(a \mid \gamma) \left[\sum_{\omega \in \Omega} \mu(\omega) u(b(\omega)) \pi_A(\gamma \mid \omega) \right]$$
$$= \sum_{\omega \in \Omega} \mu(\omega) u(b(\omega)) P_A(a \mid \omega)$$

where the first set of equalities follows from substitutions, the inequality follows from optimality conditional on γ, and the last equality follows from the same substitutions above. Rearranging this inequality shows that NIAS is satisfied.

Next, we show that a nonseparable costly information representation implies GACI. Observe $\arg\max_{\pi \in \Pi} V(\pi \cdot f_A, \pi) = V(\pi_A \cdot f_A, \pi_A)$ by definition. We first establish that $V(\pi_A \cdot f_A, \pi_A) > -\infty$ for all $A \in \mathcal{D}$. To see this, notice that for all $A \in \mathcal{D}$, f_A is a continuous function on the compact set Γ, so f_A achieves a minimum value c_A. By non-triviality, there exists π_A^c such that $V(c_A, \pi_A^c) > -\infty$. Observe $\pi_A^c \cdot f_A \geq c_A$. By monotonicity, $V(\pi_A^c \cdot f_A, \pi_A^c) \geq V(c_A, \pi_A^c) > -\infty$. Since π_A is the optimal choice, we have $V(\pi_A \cdot f_A, \pi_A) \geq V(c_A, \pi_A^c) > -\infty$.

Suppose without loss of generality that $\bar{\pi}_{A_i} \cdot f_{A_i} \leq \bar{\pi}_{A_i} \cdot f_{A_{i+1}}$ for $i = \{1, \ldots, k\}$ (with addition modulo k). It follows that

$$\begin{aligned} V(\pi_{A_i} \cdot f_{A_i}, \pi_{A_i}) &= V(\bar{\pi}_{A_i} \cdot f_{A_i}, \pi_{A_i}) \\ &\leq V(\bar{\pi}_{A_i} \cdot f_{A_{i+1}}, \pi_{A_i}) \\ &\leq V(\pi_{A_i} \cdot f_{A_{i+1}}, \pi_{A_i}) \leq V(\pi_{A_{i+1}} \cdot f_{A_{i+1}}, \pi_{A_{i+1}}) \end{aligned} \tag{4}$$

Since $V(\bar{\pi}_{A_i} \cdot f_{A_i}, \pi_{A_i}) = V(\pi_{A_i} \cdot f_{A_i}, \pi_{A_i}) > -\infty$ for all i, strict monotonicity in the first component of V implies that the inequality in (4) is strict if $\bar{\pi}_{A_i} \cdot f_{A_i} < \bar{\pi}_{A_i} \cdot f_{A_{i+1}}$. Suppose there is a strict inequality in the sequence, then we obtain the contradiction $V(\pi_{A_1} \cdot f_{A_1}, \pi_{A_1}) < V(\pi_{A_1} \cdot f_{A_1}, \pi_{A_1})$. Consequently, we must have $\bar{\pi}_{A_i} \cdot f_{A_i} = \bar{\pi}_{A_i} \cdot f_{A_{i+1}}$ for all i.

($\Leftarrow$) The converse is a direct application of Afriat's Theorem. Let $\alpha_{A,B} = -\bar{\pi}_A \cdot (f_B - f_A)$ for all $(A, B) \in \mathcal{D}^2$. Observe that by GACI, the condition in Afriat's Theorem is satisfied. Conclude there is U_A and $\lambda_A > 0$ such that for all $(A, B) \in \mathcal{D}^2$, $U_A \leq U_B - \lambda_B \bar{\pi}_B \cdot (f_A - f_B)$. Taking negatives and letting $\tilde{U}_A = -U_A$, we have

$$\tilde{U}_B + \lambda_B \bar{\pi}_B (f_A - f_B) \leq \tilde{U}_A.$$

Most of the remaining construction follows Afriat's theorem directly. Let $C(\Gamma)$ be the set of continuous, convex functions on Γ. Define $U : C(\Gamma) \to \Re$ by

$$U(f) = \max_{A \in \mathcal{D}} \tilde{U}_A + \lambda_A \bar{\pi}_A \cdot (f - f_A)$$

Clearly, U is convex, continuous, and monotone increasing[13] (as the maximum of a finite number of continuous affine functionals). For every $A \in \mathcal{D}$, $U(f_A) = \tilde{U}_A$ by construction. Moreover, for every $A \in \mathcal{D}$, if $\bar{\pi}_A \cdot f \geq \bar{\pi}_A \cdot f_A$, then $U(f) \geq U(f_A)$, which is also straightforward by construction.

Define $V : \Re \times \Pi \to \Re \cup \{-\infty\}$ by $V(t, \pi) = \inf_{f \in C(\Gamma)} \{U(f) : \pi \cdot f \geq t\}$.

[13] The functional U is monotone increasing in f if $(f - g)(\gamma) > 0$ for all $\gamma \in \Gamma$, then $U(f) > U(g)$.

To see that the monotonicity condition is satisfied for a fixed π, consider two numbers $t_1 > t_2$, we show that for every fixed π, $V(t_1,\pi) > V(t_2,\pi)$ if $V(t_2,\pi) > -\infty$.

Let $\Delta t \equiv t_1 - t_2$. Note that[14]

$$\begin{aligned} V(t_1,\pi) &= \inf_{f\in C(\Gamma)} \{U(f) : \pi\cdot f \geq t_1\} \\ &= \inf_{f\in C(\Gamma)} \{U(f) : \pi\cdot (f - \Delta t) \geq t_2\}. \end{aligned}$$

The second equality follows because π is a probability measure. In addition, since $\bar{\pi}_A$ is a probability measure for every $A \in \mathcal{D}$, by the definition $U(.)$,

$$\begin{aligned} U(\tilde{f} + \Delta t) &= \max_{A\in\mathcal{D}} \{\tilde{U}_A + \lambda_A \bar{\pi}_A \cdot (\tilde{f} + \Delta t - f_A)\} \\ &= \max_{A\in\mathcal{D}} \{\tilde{U}_A + \lambda_A \bar{\pi}_A \cdot (\tilde{f} - f_A)\} + \Delta t \\ &= U(\tilde{f}) + \Delta t \end{aligned}$$

Now, let $\tilde{f} \equiv f - \Delta t$. Then

$$\begin{aligned} V(t_1,\pi) &= \inf_{\tilde{f}+\Delta t\in C(\Gamma)} \{U(\tilde{f} + \Delta t) : \pi\cdot \tilde{f} \geq t_2\} \\ &= \inf_{\tilde{f}\in C(\Gamma)} \{U(\tilde{f} + \Delta t) : \pi\cdot \tilde{f} \geq t_2\} \\ &= \inf_{\tilde{f}\in C(\Gamma)} \{U(\tilde{f}) + \Delta t : \pi\cdot \tilde{f} \geq t_2\} \\ &= \inf_{\tilde{f}\in C(\Gamma)} \{U(\tilde{f}) : \pi\cdot \tilde{f} \geq t_2\} + \Delta t \\ &= V(t_2,\pi) + \Delta t > V(t_2,\pi). \end{aligned}$$

The assumption that for each $t \in \Re$, there exists a $\pi_t \in \Pi$ such that $V(t,\pi_t) > -\infty$ is also satisfied. In fact, we will show $V(t,\bar{\pi}_A) > -\infty$ for any $t \in \Re$ and $A \in \mathcal{D}$. For any $t \in \Re$, let $G^t_- = \{g \in C(\Gamma) \mid U(g) \leq t\}$. G^t_- is closed and convex by the continuity and convexity of $U(\cdot)$.

The set of continuous functions on γ, of which $C(\Gamma)$ is a subset, is the topological dual to the set of signed Borel measures with bounded variation over Γ (Aliprantis and Border (2006) Theorem 14.15). Let $M(\Gamma)$ be the set of such measures on Γ.

Fix $\hat{A} \in \mathcal{D}$. Note that for any $f \in G^t_-$ that

$$\tilde{U}_{\hat{A}} + \lambda_{\hat{A}} \bar{\pi}_{\hat{A}} \cdot (f - f_{\hat{A}}) \leq U(f) = \max_{A\in\mathcal{D}} \tilde{U}_A + \lambda_A \bar{\pi}_A \cdot (f - f_A) \leq t.$$

Rearranging the equation gives

$$\sup_{f\in G^t_-} \bar{\pi}_{\hat{A}} \cdot f \leq \frac{t - \tilde{U}_{\hat{A}} + \lambda_{\hat{A}} \bar{\pi}_{\hat{A}} f_{\hat{A}}}{\lambda_{\hat{A}}}$$

Let $K(t) = \frac{t - \tilde{U}_{\hat{A}} + \lambda_{\hat{A}} \bar{\pi}_{\hat{A}} f_{\hat{A}}}{\lambda_{\hat{A}}}$. Note that the function K is monotonically increasing with domain and range both spanning the reals. The function K^{-1} is well-defined and monotonic, with $K^{-1}(x) > -\infty$ for all $x \in \Re$.

[14] We abuse notation in a standard way by identifying a constant function with the value it takes: we use $f + \Delta t$ to denote the function where $(f + \Delta t)(\gamma) = f(\gamma) + \Delta t$ for all $\gamma \in \Gamma$.

It follows that $G^t_- \subseteq \{f \mid \bar{\pi}_{\hat{A}} \cdot f \leq K(t)\}$. Note that for all f such that $\bar{\pi}_{\hat{A}} \cdot f \geq K(K^{-1}(t))$ it follows that, $U(f) \geq \tilde{U}_{\hat{A}} + \lambda_{\hat{A}} \bar{\pi}_{\hat{A}} \cdot (f - f_{\hat{A}}) \geq K^{-1}(t)$. It follows by definition that for all $t \in \Re$

$$V(t, \bar{\pi}_{\hat{A}}) = \inf_{f \in C(\Gamma)} \{U(f) : \bar{\pi}_{\hat{A}} \cdot f \geq t\} \geq K^{-1}(t) > -\infty.$$

We now assert that for all $A \in \mathcal{D}$, $\bar{\pi}_A \in \arg\max_{\pi \in \Pi} V(\pi \cdot f_A, \pi)$. First, from the monotonicity property of the U function

$$\begin{aligned} V(\bar{\pi}_A \cdot f_A, \bar{\pi}_A) &= \inf_{f \in C(\Gamma)} \{U(f) : \bar{\pi}_A \cdot f \geq \bar{\pi}_A \cdot f_A\} \\ &= U(f_A) \end{aligned}$$

Second, for any $\pi \in \Pi$, we have $V(\pi \cdot f_A, \pi) = \inf_{f \in C(\Gamma)} \{U(f) : \pi \cdot f \geq \pi \cdot f_A\} \leq U(f_A)$, since $\pi \cdot f_A \geq \pi \cdot f_A$. Therefore $V(\pi \cdot f_A, \pi) \leq V(\bar{\pi}_A \cdot f_A, \bar{\pi}_A)$ for all $\pi \in \Pi$. Therefore, the revealed information structure is optimal for V.[15]

The function

$$\tilde{V}(t, \pi) = V(t, \pi) - V(0, \pi_0)$$

satisfies Condition 3, Condition 4, and Condition 5 while maintaining the other properties above. First, note that $\tilde{V}(0, \pi_0) = V(0, \pi_0) - V(0, \pi_0) = 0$ so the normalization condition is satisfied.

Since the difference of V and $\tilde{V}$ is a constant, we can check quasiconcavity and weak monotonicity of V. Next, we check weak monotonicity. If π is a garbling of ρ, then

$$\begin{aligned} V(t, \rho) &= \inf_{f \in C(\Gamma)} \{U(f) \mid \rho \cdot f \geq t\} \\ &\leq \inf_{f \in C(\Gamma)} \{U(f) \mid \pi \cdot f \geq t\} \\ &= V(t, \pi) \end{aligned}$$

since $\pi \cdot f \geq t$ implies that $\rho \cdot f \geq t$ by Lemma 2 so the infimum is taken over a weakly smaller set of functions. Thus, weak monotonicity in the second argument of V holds.

Lastly, we examine quasiconcavity of V. Let $(t_1, \pi_1), (t_2, \pi_2) \in \Re \times \Pi$, then for $\lambda \in [0, 1]$

$$\begin{aligned} &V(\lambda t_1 + (1-\lambda) t_2, \lambda \pi_1 + (1-\lambda)\pi_2) \\ &\quad = \inf_{f \in C(\Gamma)} \{U(f) \mid \lambda \pi_1 \cdot f + (1-\lambda)\pi_2 \cdot f \geq \lambda t_1 + (1-\lambda) t_2\}. \end{aligned}$$

Note that if $\lambda \pi_1 \cdot f + (1-\lambda)\pi_2 \cdot f \geq \lambda t_1 + (1-\lambda) t_2$, then either $\pi_1 \cdot f \geq t_1$ or $\pi_2 \cdot f \geq t_2$. Therefore, for $f \in C(\Gamma)$ we have

$$\{f \mid \lambda \pi_1 \cdot f + (1-\lambda)\pi_2 \cdot f \geq \lambda t_1 + (1-\lambda) t_2\} \subseteq \{f \mid \pi_1 \cdot f \geq t_1\} \cup \{f \mid \pi_2 \cdot f \geq t_2\}.$$

Therefore, the infimum of U over the first set, $V(\lambda t_1 + (1-\lambda) t_2, \lambda \pi_1 + (1-\lambda)\pi_2)$, is greater than or equal to the infimum of U over the second set, $\min\{V(t_1, \pi_1), V(t_2, \pi_2)\}$. Thus, quasiconcavity holds.

We now show data matching and choices are optimal by following Caplin et al. (2015) and using NIAS. Next we show that there exists stochastic choice functions $\{C_A : \text{Supp}(\bar{\pi}_A) \to \Delta(A)\}_{A \in \mathcal{D}}$ that satisfy optimality and matches data.

15 We note that a version of Roy's identity holds (Roy (1947)). Observe that by definition of V, if $\pi \cdot f_A \geq w$ implies $U(f_A) \geq V(w, \pi)$. We conclude that $\pi \cdot f_A \geq \bar{\pi}_A \cdot f_A$ implies $U(f_A) \geq V(\bar{\pi}_A \cdot f_A, \pi)$. We have already shown that $U(f_A) = V(\bar{\pi}_A \cdot f_A, \bar{\pi}_A)$. Thus, if $\pi \cdot f_A \geq \bar{\pi}_A \cdot f_A$, then $V(\bar{\pi}_A \cdot f_A, \bar{\pi}_A) \geq V(\bar{\pi}_A \cdot f_A, \pi)$.

For each $\gamma \in \text{Supp}(\bar{\pi}_A)$, define:

$$C_A(a \mid \gamma) = \begin{cases} \frac{P_A(a)}{\sum_{\{b \in A : \bar{\gamma}_A^b = \gamma\}} P_A(b)} & \text{if } \bar{\gamma}_A^a = \gamma \\ 0 & \text{otherwise} \end{cases}$$

where $P_A(a) = \sum_{\omega \in \Omega} P_A(a \mid \omega)\mu(\omega)$ is the unconditional probability of choosing action a from decision problem A. Note the $C_A(a \mid \gamma) > 0$ only if $\bar{\gamma}_A^a = \gamma$. The NIAS condition implies that

$$\sum_{\omega \in \Omega} \mu(\omega) P_A(a \mid \omega) u(a(\omega)) \geq \sum_{\omega \in \Omega} \mu(\omega) P_A(b \mid \omega) u(b(\omega))$$
$$\Rightarrow \sum_{\omega \in \Omega} \bar{\gamma}_A^a(\omega) u(a(\omega)) \geq \sum_{\omega \in \Omega} \bar{\gamma}_A^a(\omega) u(b(\omega))$$

The second line follows by dividing both sides by $P_A(a)$. Thus, NIAS ensures that the choices are optimal.

It remains to show that the data are matched. In other words, P_A is generated from the information structure $\bar{\pi}_A$ and choices C_A. First, note that for any $b, b' \in A$ such that $\bar{\gamma}_A^b = \bar{\gamma}_A^{b'}$, implies that for any $\omega \in \Omega$ such that $\bar{\gamma}_A^b(\omega) > 0$, then

$$\frac{P_A(b \mid \omega)}{P_A(b' \mid \omega)} = \frac{P_A(b)}{P_A(b')}.$$

Thus, for every $\omega \in \Omega$ and $a \in A$ such that $P_A(a) > 0$, then

$$\begin{aligned}
\sum_{\gamma \in \text{Supp}(\bar{\pi}_A)} \bar{\pi}_A(\gamma|\omega) C_A(a|\gamma) &= \bar{\pi}_A(\bar{\gamma}_A^a|\omega) C_A(a|\bar{\gamma}_A^a) \\
&= \sum_{\{c \in A : \bar{\gamma}_A^c = \bar{\gamma}_A^a\}} P_A(c|\omega) \frac{P_A(a)}{\sum_{\{b \in A | \bar{\gamma}_A^b = \bar{\gamma}_A^a\}} P_A(b)} \\
&= \sum_{\{c \in A | \bar{\gamma}_A^c = \bar{\gamma}_A^a\}} P_A(c|\omega) \frac{P_A(a|\omega)}{\sum_{\{b \in A | \bar{\gamma}_A^b = \bar{\gamma}_A^a\}} P_A(b|\omega)} \\
&= P_A(a|\omega).
\end{aligned}$$

Therefore, the data are matched. □

Proof of Theorem 2. We note that NIAS is equivalent to optimal choices and matched data. Therefore, we focus on non-triviality and optimal information.

($\Rightarrow$) We show that a multiplicatively costly information representation satisfies HACI. For all $A \in \mathcal{D}$, let $\pi_A \in \arg\max_{\pi \in \Pi}[R(\pi)(\pi \cdot f_A)]$.

First, we show for $A \in \mathcal{D} \setminus \mathcal{D}_0$, that $(\pi \cdot f_A) > 0$ for all information structures and $R(\pi_A) > 0$. Let π_0 denote the non-informative information structure with $\pi_0(\mu) = 1$. By assumption, $\pi_0 \cdot f_A > 0$ for any $A \in \mathcal{D} \backslash \mathcal{D}_0$. Since f_A is convex, $\pi \cdot f_A > 0$ for all information structures. In particular, let $\pi' \in \Pi$ be an information structure such that $R(\pi') > 0$, then $\pi' \cdot f_A > 0$ as well. Note that such a $\pi' \in \Pi$ with $R(\pi') > 0$ exists by nontriviality. For any $A \in \mathcal{D} \setminus \mathcal{D}_0$ and for any $\pi \in \Pi$, we have $R(\pi_A)(\pi_A \cdot f_A) \geq R(\pi')(\pi' \cdot f_A) > 0$ since π_A is the optimal information structure. Therefore, for all $A \in \mathcal{D} \backslash \mathcal{D}_0$ we have $R(\pi_A) > 0$.

Next, for any pair $A_i, A_{i+1} \in \mathcal{D}\backslash\mathcal{D}_0$, we have

$$\begin{aligned} R(\pi_{A_i})(\bar{\pi}_{A_i} \cdot f_{A_i}) &= R(\pi_{A_i})(\pi_{A_i} \cdot f_{A_i}) \\ &\geq R(\pi_{A_{i+1}})(\pi_{A_{i+1}} \cdot f_{A_i}) \\ &\geq R(\pi_{A_{i+1}})(\bar{\pi}_{A_{i+1}} \cdot f_{A_i}) > 0 \end{aligned}$$

where the equality follows from equivalent choices, the first inequality follows from optimality, the second inequality follows from Lemma 2, and the last term is greater than zero by the earlier arguments. Rearranging the end terms of the inequalities,

$$\frac{R(\pi_{A_{i+1}})(\bar{\pi}_{A_{i+1}} \cdot f_{A_i})}{R(\pi_{A_i})(\bar{\pi}_{A_i} \cdot f_{A_i})} \leq 1.$$

We can now take any cycle $A_1, \ldots, A_k \in \mathcal{D}\backslash\mathcal{D}_0$ and take products to see that costs will be removed so

$$\prod_{i=1}^{k} \frac{R(\pi_{A_{i+1}})(\bar{\pi}_{A_{i+1}} \cdot f_{A_i})}{R(\pi_{A_i})(\bar{\pi}_{A_i} \cdot f_{A_i})} = \prod_{i=1}^{k} \frac{(\bar{\pi}_{A_{i+1}} \cdot f_{A_i})}{(\bar{\pi}_{A_i} \cdot f_{A_i})} \leq 1$$

where the indices are calculated with addition modulo k. Let $\sigma : \{1, \ldots, k, \} \to \{1, \ldots, k\}$ where $\sigma(1) = 1, \sigma(2) = k, \sigma(3) = k - 1, \ldots, \sigma(k) = 2$.[16] Therefore,

$$\prod_{i=1}^{k} \frac{\bar{\pi}_{A_{\sigma(i)}} \cdot f_{A_{\sigma(i+1)}}}{\bar{\pi}_{A_\sigma(i)} \cdot f_{A_\sigma(i)}} \leq 1$$

and HACI is satisfied.

($\Leftarrow$) Now we show from HACI that we can generate a non-trivial utility function. Following Varian (1983), for all $A \in \mathcal{D}$ let U_A be the maximum of

$$\prod_{i=1}^{k-1} \frac{\bar{\pi}_{A_i} \cdot f_{A_{i+1}}}{\bar{\pi}_{A_i} \cdot f_{A_i}} \tag{5}$$

where the maximization is taken over all finite sequences $\{A_i\}_{i=1}^{k-1} \subseteq \mathcal{D} \setminus \mathcal{D}_0$ with $A_k = A$. Note that if $A \in \mathcal{D}_0$ then $U_A = 0$. Since the number of menus in $\mathcal{D}\backslash\mathcal{D}_0$ is finite, the number of sequences $\{A_i\}_{i=1}^{k-1}$ not containing cycles is also finite. Moreover by HACI, the presence of any cycles in a sequence $\{A_i\}_{i=1}^{k-1}$ only decreases the value of (5). Therefore the maximum in (5) exists for each A. Note that $U_A > 0$ for all $A \in \mathcal{D} \setminus \mathcal{D}_0$, and for all $A, B \in \mathcal{D}$

$$U_B \geq U_A \frac{\bar{\pi}_A \cdot f_B}{\bar{\pi}_A \cdot f_A} \tag{6}$$

by definition.[17] Define

$$U(f) = \begin{cases} \max_{A \in \mathcal{D}} \left[U_A \frac{\bar{\pi}_A \cdot f}{\bar{\pi}_A \cdot f_A} \right] & \text{if } f \in C_+(\Gamma) \\ +\infty & \text{otherwise} \end{cases}$$

[16] Note addition is still modulo k in the index so $\sigma(k + 1) = 1$.

[17] We define $0 \cdot \infty = 0$ as is standard in convex analysis.

where $C_+(\Gamma)$ are nonnegative convex continuous functions on Γ. From the definition of U, it is obvious that $U(\cdot)$ is homogenous of degree 1 (as the supremum of a finite number of linear functionals), and $U(f) \geq 0$ for all $f \in C(\Gamma)$. In addition, inequality (6) implies that for all $A \in \mathcal{D}$ that $U(f_A) = U_A$. It is also straightforward that U is convex, continuous, and monotone increasing over $C_+(\Gamma)$. Finally, we have

$$U(f) \geq U(f_A) \text{ if } \bar{\pi}_A \cdot f \geq \bar{\pi}_A \cdot f_A \tag{7}$$

which is also straightforward by construction.

Let $M_+(\Gamma)$ be the set of non-negative Borel measures over Γ with bounded variation. Define $V : \Re_+ \times M_+(\Gamma) \to \Re_+$ by $V(t, m) = \inf_{f \in C(\Gamma)}\{U(f) : m \cdot f \geq t\}$. Now, we show that $V(\cdot, \cdot)$ is indeed of the multiplicative form. By the definition of V, for any $t > 0$ we have

$$\begin{aligned}
V(t,m) &= \inf_{f \in C(\Gamma)} \{U(f) : m \cdot \frac{f}{t} \geq 1\} \\
&= \inf_{tf' \in C(\Gamma)} \{U(tf') : m \cdot f' \geq 1\} \\
&= \inf_{f' \in C(\Gamma)} \{U(tf') : m \cdot f' \geq 1\} \\
&= t \inf_{f' \in C(\Gamma)} \{U(f') : m \cdot f' \geq 1\} \\
&= t\bar{R}(m)
\end{aligned}$$

where the first equality comes from rearrangement, the second equality comes from $f' = f/t$, the third equality comes since any tf' can be expressed as a function, the fourth equality holds since U is homogeneous degree 1, and the final equality holds by defining the function $\bar{R} : M_+(\Gamma) \to \Re_+$ as $\bar{R}(m) = \inf_{f \in C(\Gamma)}\{U(f) : m \cdot f \geq 1\}$.

Next, if $t = 0$ then $V(t, m) = 0$ which is consistent with the multiplicative form. To see this, consider the constant function $f_0(\gamma) = 0$ for all $\gamma \in \Gamma$ and see that $V(0, \pi) \leq U(f_0) = 0$. Thus, $V(0, \pi) = 0 \cdot \bar{R}(\pi) = 0$. Let $\tilde{R} : \Pi \to \Re_+$ be the restriction of $\bar{R}$ to Π.

Since $U(f) \geq 0$ for all $f \in C(\Gamma)$, we have $\tilde{R}(\pi) = \inf_{f \in C(\Gamma)}\{U(f) : \pi \cdot f \geq 1\} \geq 0$. Moreover, we show that $\tilde{R}(\pi) < \infty$. Consider the constant function $f_1(\gamma) = 1$ for all $\gamma \in \Gamma$ so that $\pi \cdot f_1 = 1$. Therefore, we deduce

$$\tilde{R}(\pi) \leq \max_{A \in \mathcal{D}} \frac{U_A}{\bar{\pi}_A \cdot f_A} < \infty.$$

We also prove that there are $\pi \in \Pi$ such that $\tilde{R} > 0$. For an arbitrary $A \in \mathcal{D} \backslash \mathcal{D}_0$, we have

$$\begin{aligned}
\tilde{R}(\bar{\pi}_A) &= \frac{V(\bar{\pi}_A \cdot f_A, \bar{\pi}_A)}{\bar{\pi}_A \cdot f_A} \\
&= \frac{\inf_{f \in C(\Gamma)}\{U(f) : \bar{\pi}_A \cdot f \geq \bar{\pi}_A \cdot f_A\}}{\bar{\pi}_A \cdot f_A} \\
&= \frac{U(f_A)}{\bar{\pi}_A \cdot f_A} = \frac{U_A}{\bar{\pi}_A \cdot f_A} > 0
\end{aligned}$$

that follows from $\bar{\pi}_A \cdot f_A > 0$ and the definition of $\tilde{R}$. Therefore, $\tilde{R}(\bar{\pi}_A) > 0$.

We note that if π is a garbling of ρ then $\tilde{R}(\rho) \leq \tilde{R}(\pi)$ since $\tilde{R}(\pi) = \inf_{f \in C(\Gamma)}\{U(f) \mid \pi \cdot f \geq 1\}$ and $\pi \cdot f \geq 1$ implies $\rho \cdot f \geq 1$ so the infimum is over a weakly larger set. Let π_0 as the information structure with $\pi_0(\mu|\omega) = 1$ for all $\omega \in \Omega$. Since Π is the set of information sets consistent with Bayes' Law, π_0 is a garbling of any $\pi \in \Pi$. Thus, for all $\pi \in \Pi$, $\tilde{R}(\pi_0) \geq \tilde{R}(\pi) > 0$. Lastly, rescale the function $\tilde{R}(.)$ with $1/\tilde{R}(\pi_0)$, and define

$$R(\pi) = \frac{\tilde{R}(\pi)}{\tilde{R}(\pi_0)}.$$

We now assert that for all $A \in \mathcal{D}$, $\bar{\pi}_A \in \arg\max_{m \in M_+(\Gamma)} V(\pi \cdot f_A, \pi)$. First, from inequality (7) we have

$$\begin{aligned} V(\bar{\pi}_A \cdot f_A, \bar{\pi}_A) &= \inf_{f \in C(\Gamma)} \{U(f) : \bar{\pi}_A \cdot f \geq \bar{\pi}_A \cdot f_A\} \\ &= U(f_A) \end{aligned}$$

Second, for any $m \in M_+(\Gamma)$, we have $V(m \cdot f_A, m) = \inf_{f \in C(\Gamma)}\{U(f) : m \cdot f \geq m \cdot f_A\} \leq U(f_A)$, since $m \cdot f_A \geq m \cdot f_A$. Therefore $V(m \cdot f_A, m) \leq V(\bar{\pi}_A \cdot f_A, \bar{\pi}_A)$ for all $m \in M_+(\Gamma)$. From this, we have that

$$\begin{aligned} \bar{\pi}_A \in \arg\max_{\pi \in \Pi} \frac{V(\pi \cdot f_A, \pi)}{\tilde{R}(\pi_0)} &= \arg\max_{\pi \in \Pi} \frac{\tilde{R}(\pi)}{\tilde{R}(\pi_0)}(\pi \cdot f_A) \\ &= \arg\max_{\pi \in \Pi} R(\pi)(\pi \cdot f_A) \end{aligned}$$

where $\bar{\pi}_A$ is an optimizer since this holds over all of M_+ and thus holds over Π. Therefore π_A is optimal for the rescaled V and has the multiplicative costly representation.

We note that the R was already shown to satisfy weak monotonicity in information and the normalization property. The $\tilde{R}(m)$ defined in Theorem 2 is homogenous of degree one, increasing in m, and quasiconcave by the same arguments used in Theorem 1. By Theorem 1 in Prada (2011), we have that $\bar{R}$ is concave. Therefore, $\tilde{R}$ restricted to Π is the restriction of $\bar{R}$ to a convex set and is thus concave. Finally, R is concave as it is a positive re-scaling. □

Proof of Theorem 3. We note that NIAS is equivalent to optimal choices and matched data. Therefore, we focus on non-triviality and optimal information.

($\Rightarrow$) Suppose the data is represented by a constrained costly information representation and for all $A \in \mathcal{D}$ that $\pi_A \in \arg\max_{\pi \in \Pi_c} \pi \cdot f_A$. Since the utility depends only on ex-ante payoff, then $\pi_A \cdot f_A = \bar{\pi}_A \cdot f_A \geq \pi_B \cdot f_A \geq \bar{\pi}_B \cdot f_A$. The first equality follows from equivalent choices, the next inequality follows from optimality, while the final inequality follows Lemma 2.

($\Leftarrow$) Suppose BACI holds. Let $\bar{\Pi}_c = \bigcup_{A \in \mathcal{D}}\{\bar{\pi}_A\}$. For $\mathcal{D}$ nonempty, $\Pi_c \neq \emptyset$.[18] Moreover, for any $A, B \in \mathcal{D}$, we have

$$\bar{\pi}_A \cdot f_A \geq \bar{\pi}_B \cdot f_A.$$

In other words, for all $A \in \mathcal{D}$ we have $\bar{\pi}_A \in \arg\max_{\pi \in \bar{\Pi}_c} \pi \cdot f_A$. Therefore nontriviality and optimal information hold.

Let $\text{conv}(\bar{\Pi}_c) = \text{conv}\left(\bigcup_{A \in \mathcal{D}}\{\bar{\pi}_A\}\right)$. Here $\text{conv}(\cdot)$ represents the convex hull of information structures. For all $B \in \mathcal{D}$ let $\lambda_B \in [0, 1]$ such that $\sum_{B \in \mathcal{D}} \lambda_B = 1$. Now for fixed $A \in \mathcal{D}$

$$\sum_{B \in \mathcal{D}} \lambda_B \bar{\pi}_B \cdot f_A \leq \sum_{B \in \mathcal{D}} \lambda_B \bar{\pi}_A \cdot f_A = \bar{\pi}_A \cdot f_A$$

[18] If $\mathcal{D} = \emptyset$, then let $\Pi_c = \Pi$.

where the inequality follows from BACI. The result holds for any fixed A and convex combination so that $\bar{\pi}_A \in \arg\max_{\pi \in \text{conv}(\bar{\Pi}_c)} \pi \cdot f_A$. Thus, the constraint set can be chosen convex without loss of generality. □

Proof of Proposition 2. To satisfy NIAC it is required that

$$\bar{\pi}_A \cdot f_A + \bar{\pi}_B \cdot f_B \geq \bar{\pi}_A \cdot f_B + \bar{\pi}_B \cdot f_A$$

or equivalently,

$$\bar{\pi}_B \cdot f_A - \bar{\pi}_A \cdot f_A \leq \bar{\pi}_B \cdot f_B - \bar{\pi}_A \cdot f_B$$

However, since menu A provides a higher return to information than menu B and $\bar{\pi}_A$ is a garbling of $\bar{\pi}_B$, then

$$\bar{\pi}_B \cdot f_A - \bar{\pi}_A \cdot f_A > \bar{\pi}_B \cdot f_B - \bar{\pi}_A \cdot f_B$$

which violates NIAC. □

Proof of Proposition 3. Note that GACI is violated for a dataset of two menus if and only if

$$\bar{\pi}_A \cdot f_A \leq \bar{\pi}_A \cdot f_B \quad \text{and} \quad \bar{\pi}_B \cdot f_B \leq \bar{\pi}_B \cdot f_A$$

with one inequality strict. Since $f_A > f_B$, it follows that

$$\bar{\pi}_A \cdot f_A > \bar{\pi}_A \cdot f_B$$

and there can be no violation of GACI. Since the dataset was assumed to satisfy NIAS, the data is rationalized by a nonseparable costly information representation. □

Proof of Proposition 4. For any sequence $(\bar{\pi}_{A_1}, f_{A_1}), \ldots, (\bar{\pi}_{A_k}, f_{A_k})$ with $A_i \in \mathcal{D}$. Note that $\bar{\pi}_{A_i} \cdot f_{A_i} = \bar{\pi}_{A_i} \cdot f_A + c_{m_i}$ and $\bar{\pi}_{A_{i+1}} \cdot f_{A_i} = \bar{\pi}_{A_{i+1}} \cdot f_A + c_{m_i}$ for some $m_i \in \{1, \ldots, M\}$. This implies that

$$\sum_{i=1}^{k} \bar{\pi}_{A_i} \cdot f_{A_i} = \sum_{i=1}^{k} \bar{\pi}_{A_i} \cdot f_A + c_{m_i} = \sum_{i=1}^{k} \bar{\pi}_{A_{i+1}} \cdot f_A + c_{m_i} = \sum_{i=1}^{k} \bar{\pi}_{A_{i+1}} \cdot f_{A_i}$$

where addition of the index is modulo k. Therefore, NIAC is satisfied in addition to NIAS and the dataset is rationalized by the additive costly information representation. □

Proof of Proposition 5. For any sequences $(\bar{\pi}_{A_1}, f_{A_1}), \ldots, (\bar{\pi}_{A_k}, f_{A_k})$ with $A_i \in \mathcal{D} \backslash \mathcal{D}_0$. Note that $\bar{\pi}_{A_i} \cdot f_{A_i} = c_{m_i} \bar{\pi}_{A_i} \cdot f_A$ and $\bar{\pi}_{A_i} \cdot f_{A_{i+1}} = c_{m_{i+1}} \bar{\pi}_{A_i} \cdot f_A$ for some $m_i, m_{i+1} \in \{1, \ldots, M\}$. This implies that

$$\prod_{i=1}^{k} \frac{\bar{\pi}_{A_i} \cdot f_{A_{i+1}}}{\bar{\pi}_{A_i} \cdot f_{A_i}} = \prod_{i=1}^{k} \frac{c_{m_i}}{c_{m_{i+1}}} = 1$$

since addition of the index is modulo k and each c_{m_i} term appears in the numerator and denominator. Therefore, HACI is satisfied in addition to NIAS and the dataset is rationalized by the multiplicative costly information representation. □

Table 1
Payoffs for Environments 1-4.

Environment	E1	E2	E3	E4
β_1	40	40	30	30
β_2	55	52	55	52

Appendix B. Bronar's power examples

In this appendix, we detail several additional examples and provide Monte Carlo simulations in the spirit of Bronars (1987) and Beatty and Crawford (2011) to check how likely random choices will satisfy or refute the conditions in the main text. We first detail how we generate the choices and then describe the different decision environments. Lastly, we provide the percentage of datasets that satisfy the conditions for each environment. We consider the environments of Dean et al. (2017) which might be of interest to other researchers and expand the example in Section 4.3.

Consider the menu, A, from the dataset $\mathcal{D} \subset \mathcal{A}$. We follow the below procedure to generate the state dependent stochastic choices $P(a \mid \omega)$.

1. For every state $\omega \in \Omega$ and for every $a \in A$, draw a random variable $Z_{a,\omega}$ independently distributed according to the uniform distribution on the unit interval, $U[0, 1]$.
2. Set the state dependent choice probability $P_A(a \mid \omega) = \frac{Z_{a,\omega}}{\sum_{b \in A} Z_{b,\omega}}$.

This procedure gives some sense of how likely a random dataset passes the various conditions. We note this is just one sampling procedure and that others may produce different results. We encourage interested readers to check the sampling distribution that makes sense for the given application of the tests in the main text. A description of the different environments we examine is below.

Example 5. This is the environment of Experiment 1 in Dean et al. (2017). We consider environments where the prior is given by $\mu = (\frac{1}{2}, \frac{1}{2})$. There are actions a, b, and c whose payoffs are

$$u(a(\omega)) = \begin{cases} 50 & \text{if } \omega = \omega_1 \\ 50 & \text{if } \omega = \omega_2 \end{cases} \qquad u(b(\omega)) = \begin{cases} \beta_1 & \text{if } \omega = \omega_1 \\ \beta_2 & \text{if } \omega = \omega_2 \end{cases}$$

$$u(c(\omega)) = \begin{cases} 100 & \text{if } \omega = \omega_1 \\ 0 & \text{if } \omega = \omega_2 \end{cases}$$

and the dataset consists of the menus $\{a, b\}$ and $\{a, b, c\}$. The values of β_1 and β_2 are given in Table 1. We refer to these as environments one through four (E1-E4).

We denote the corresponding datasets as $\mathcal{D}_{E1}, \mathcal{D}_{E2}, \mathcal{D}_{E3}$, and $\mathcal{D}_{E4}$, respectively.

Example 6. This is the environment of Experiment 2 in Dean et al. (2017). We call this environment five (E5). We consider environments where the prior is given by $\mu = (\frac{1}{2}, \frac{1}{2})$. There are actions a_i, b_i whose payoffs are

$$u(a_i(\omega)) = \begin{cases} \alpha_i & \text{if } \omega = \omega_1 \\ 0 & \text{if } \omega = \omega_2 \end{cases} \qquad u(b_i(\omega)) = \begin{cases} 0 & \text{if } \omega = \omega_1 \\ \alpha_i & \text{if } \omega = \omega_2 \end{cases}$$

Table 2
Payoffs for indexes of Environment 5.

Index (i)	1	2	3	4
α_i	5	40	70	90

Table 3
Payoffs for indexes of Environments 6-9.

Index (j)	1	2	3	4
γ_j	0	2	4	6
ξ_j	0	1	2	3

Table 4
Priors for Environments 10-13.

Environment	E10	E11	E12	E13
Prior (μ)	$(\frac{1}{2}, \frac{1}{2})$	$(\frac{3}{5}, \frac{2}{5})$	$(\frac{3}{4}, \frac{1}{4})$	$(\frac{17}{20}, \frac{3}{20})$

and the dataset consists of the menus $A_i = \{a_i, b_i\}$ for $i \in \{1, 2, 3, 4\}$ where the values for α_i are in Table 2.

We denote the dataset with $\mathcal{D}_{\text{E5}}$.

Example 7. Here we use the conditions of the decision problem from Section 4.3 and several others. We consider environments where the prior is given by $\mu = (\frac{1}{2}, \frac{1}{2})$. We call these environments six through (E6-E9). There are actions a_j, b_j, a'_j, b'_j whose payoffs are

$$u(a_j(\omega)) = \begin{cases} 5 + \gamma_j & \text{if } \omega = \omega_1 \\ 0 + \xi_j & \text{if } \omega = \omega_2 \end{cases} \qquad u(b_j(\omega)) = \begin{cases} 1 + \xi_j & \text{if } \omega = \omega_1 \\ 4 + \gamma_j & \text{if } \omega = \omega_2 \end{cases}$$

$$u(a'_j(\omega)) = \begin{cases} 4 + \gamma_j & \text{if } \omega = \omega_1 \\ 1 + \xi_j & \text{if } \omega = \omega_2 \end{cases} \qquad u(b'_j(\omega)) = \begin{cases} 2 + \xi_j & \text{if } \omega = \omega_1 \\ 3 + \gamma_j & \text{if } \omega = \omega_2 \end{cases},$$

where menus are given by $B_j = \{a_j, b_j\}$ and $B'_j = \{a'_j, b'_j\}$ for $j \in \{1, 2, 3, 4\}$ where the values for ξ_j and γ_j are in Table 3. We note that the role of γ_j is to increase the incentive of choosing the higher utility action in each state.

We consider datasets $\mathcal{D}_{\text{E6}} = \{B_1, B'_1\}$, $\mathcal{D}_{\text{E7}} = \bigcup_{j=1}^{2}\{B_j, B'_j\}$, $\mathcal{D}_{\text{E8}} = \bigcup_{j=1}^{3}\{B_j, B'_j\}$ and $\mathcal{D}_{\text{E9}} = \bigcup_{j=1}^{4}\{B_j, B'_j\}$.

Example 8. This is the environment of Experiment 3 in Dean et al. (2017). We call these environments ten through thirteen (E10-E13). Here we use the payoffs of actions a and b given by

$$u(a(\omega)) = \begin{cases} 10 & \text{if } \omega = \omega_1 \\ 0 & \text{if } \omega = \omega_2 \end{cases} \qquad u(b(\omega)) = \begin{cases} 0 & \text{if } \omega = \omega_1 \\ 10 & \text{if } \omega = \omega_2 \end{cases}$$

and consider changing the prior. The different values of the prior used are in Table 4 above. Here there is only one menu, so that we can only present the results for NIAS violations.

Table 5
Percentage of random draws that pass condition.

Condition	E1	E2	E3	E4	E5	E6	E7	E8	E9
GACI	100	100	100	100	100	77.2	61.4	52.8	45.4
NIAC	58.4	58.4	58.4	58.4	4.2	62.5	12.7	1.5	0
HACI	58.3	58.4	58.4	58.4	100	62.5	10.3	1	0.1
BACI	16.5	35.3	32.2	41.5	0	34.3	0.8	0	0

Table 6
Percentage of random draws that pass condition.

Condition	NIAS
E1	1
E2	0.1
E3	0.1
E4	0
E5	4.6
E6	2.9
E7	0.2
E8	0
E9	0
E10	49.3
E11	29.9
E12	12
E13	5.9

For each of the environments above, we generate 1,000 different datasets according to the sampling procedure described above. When the datasets are related such as E1-E4, E6-E9, and E10-E13 we use the same draw of the Monte Carlo distribution to check the conditions in the main text. This guarantees that the differences are driven from the structure of the datasets and not the sampling scheme. The results of the test are presented in Table 5.

We note that in general GACI is a weak test for the experiments in Dean et al. (2017). In particular, E5 we know *a priori* that GACI and HACI will always be satisfied. The environment considered in the main text shows that it is possible to have simple datasets that can detect violations of GACI. In particular, E6 has similar power to detect violations compared to NIAC and HACI. As the number of menus of a similar flavor as the main example, the power to detect violations increases. However, this power increases rather slowly. In Table 6, we examine the violations of NIAS.

We note that it is much easier in general to detect violations of NIAS. In particular, almost all random datasets generated for E1-E9 are refuted by NIAS. We note that for a single menu with more extreme priors lead to higher power to detect violations of NIAS. Thus, it should be more surprising that one satisfies NIAS relative to the power of the experiments in Dean et al. (2017) and perhaps more generally.

References

Aguiar, Victor, 2016. Random Categorization and Bounded Rationality. Working Paper.
Aliprantis, Charalambos D., Border, Kim, 2006. Infinite Dimensional Analysis: A Hitchhiker's Guide. Springer Science & Business Media.

Allen, Roy, Rehbeck, John, 2019. Revealed stochastic choice with attributes. Working Paper.
Beatty, Timothy K.M., Crawford, Ian A., 2011. How demanding is the revealed preference approach to demand? Am. Econ. Rev. 101 (6), 2782–2795.
Blackwell, David, 1953. Equivalent comparisons of experiments. Ann. Math. Stat. 24 (2), 265–272.
Bronars, Stephen G., 1987. The power of nonparametric tests of preference maximization. Econometrica, 693–698.
Brown, Donald J., Calsamiglia, Caterina, 2007. The nonparametric approach to applied welfare analysis. Econ. Theory 31 (1), 183–188.
Caplin, Andrew, Martin, Daniel, 2015. A testable theory of imperfect perception. Econ. J. 125 (582), 184–202.
Caplin, Andrew, Martin, Daniel, Dean, Mark, 2015. Revealed preference, rational inattention, and costly information acquisition. Am. Econ. Rev. 105 (7), 2183–2203.
Cerreia-Vioglio, Simone, Maccheroni, Fabio, Marinacci, Massimo, Montrucchio, Luigi, 2011a. Complete monotone quasiconcave duality. Math. Oper. Res. 36 (2), 321–339.
Cerreia-Vioglio, Simone, Maccheroni, Fabio, Marinacci, Massimo, Montrucchio, Luigi, 2011b. Uncertainty averse preferences. J. Econ. Theory 146 (4), 1275–1330.
Chambers, Christopher P., Echenique, Federico, 2016. Revealed Preference Theory, vol. 56. Cambridge University Press.
Chambers, Christopher P., Echenique, Federico, Saito, Kota, 2016. Testing theories of financial decision making. Proc. Natl. Acad. Sci. 113 (15), 4003–4008.
Chateauneuf, Alain, Faro, José Heleno, 2009. Ambiguity through confidence functions. J. Math. Econ. 45 (9), 535–558.
Cherchye, Laurens, Crawford, Ian, De Rock, Bram, Vermeulen, Frederic, 2016. Gorman revisited: nonparametric conditions for exact linear aggregation. SERIEs 7 (2), 203–220.
Dean, Mark, Neligh, Nate Leigh, 2017. Experimental Tests of Rational Inattention.
Denti, Tommaso, Mihm, Maximilian, de Oliveira, Henrique, Ozbek, Kemal, 2016. Rationally inattentive preferences and hidden information costs. Theor. Econ..
Ellis, Andrew, 2018. Foundations for optimal inattention. J. Econ. Theory 173, 56–94.
Fudenberg, Drew, Iijima, Ryota, Strzalecki, Tomasz, 2015. Stochastic choice and revealed perturbed utility. Econometrica 83 (6), 2371–2409.
Gorman, William M., 1953. Community preference fields. Econometrica, 63–80.
Houthakker, Hendrik S., 1950. Revealed preference and the utility function. Economica 17 (66), 159–174.
Kamenica, Emir, Gentzkow, Matthew, 2011. Bayesian persuasion. Am. Econ. Rev. 101 (6), 2590–2615.
Koopmans, Tjalling C., Beckmann, Martin, 1957. Assignment problems and the location of economic activities. Econometrica, 53–76.
Machina, Mark J., 1984. Temporal risk and the nature of induced preferences. J. Econ. Theory 33 (2), 199–231.
Matejka, Filip, McKay, Alisdair, 2014. Rational inattention to discrete choices: a new foundation for the multinomial logit model. Am. Econ. Rev. 105 (1), 272–298.
Oliveira, Henrique De, Lamba, Rohit, 2018. Rationalizing dynamic choices. Working paper.
Polisson, Matthew, Quah, John K.-H., Renou, Ludovic, 2017. Revealed preferences over risk and uncertainty. https://ideas.repec.org/p/san/wpecon/1706.html.
Prada, David Juan, 2011. A note on concavity, homogeneity and non-increasing returns to scale. Econ. Bull. 31 (1), 100–105.
Richter, Marcel K., 1966. Revealed preference theory. Econometrica, 635–645.
Rochet, Jean-Charles, 1987. A necessary and sufficient condition for rationalizability in a quasi-linear context. J. Math. Econ. 16 (2), 191–200.
Rockafellar, Ralph, 1966. Characterization of the subdifferentials of convex functions. Pac. J. Math. 17 (3), 497–510.
Roy, René, 1947. La distribution du revenu entre les divers biens. Econometrica, 205–225.
Varian, Hal R., 1982. The nonparametric approach to demand analysis. Econometrica, 945–973.
Varian, Hal R., 1983. Non-parametric tests of consumer behaviour. Rev. Econ. Stud. 50 (1), 99–110.
Varian, Hal R., 1984. The nonparametric approach to production analysis. Econometrica, 579–597.