Revealed Preference, Rational Inattention, and Costly Information Acquisition*

Andrew Caplin† and Mark Dean‡

August 2014

Abstract

Apparently mistaken decisions are ubiquitous. To what extent does this reflect irrationality, as opposed to a rational trade off between the costs of information acquisition and the expected benefits of learning? We develop a revealed preference test that characterizes all patterns of choice "mistakes" consistent with a general model of optimal costly information acquisition. We identify the extent to which information costs can be recovered from choice data. We experimentally elicit the state dependent stochastic choice data that our tests require. While consistent with the general model, the data is inconsistent with commonly utilized functional forms.

*We thank Roland Benabou, Dirk Bergemann, Laurens Cherchye, Bram De Rock, Thomas Demuynck, Henrique De Oliveira, Federico Echenique, Andrew Ellis, Paola Manzini, Marco Mariotti, Daniel Martin, Filip Matejka, Alisdair McKay, Stephen Morris, Pietro Ortoleva, Kemal Ozbek, Daphna Shohamy, Laura Veldkamp and Michael Woodford for their constructive contributions. We also thank Samuel Brown, Severine Toussaert and Isabel Trevino for their exceptional research assistance. An early version was circulated under the title "Rational Inattention and State Dependent Stochastic Choice".

†Center for Experimental Social Science and Department of Economics, New York University. Email: andrew.caplin@nyu.edu

‡Department of Economics, Brown University. Email: mark_dean@brown.edu

1 Introduction

Limits on attention impact choice. Shoppers may buy unnecessarily expensive products due to their failure to notice whether or not sales tax is included in stated prices (Chetty *et al.* [2009]). Buyers of second-hand cars focus their attention on the leftmost digit of the odometer (Lacetera *et al.* [2012]). Purchasers limit their attention to a relatively small number of websites when buying over the internet (Santos *et al.* [2012]).

While apparently mistaken decisions are ubiquitous, this does not imply that decision makers are irrational. The standard theory of choice asserts only that individuals act optimally, *given what they know*. At least since the work of Hayek [1945] and Stigler [1961], there has been a focus on optimization of *knowledge itself*, with decision makers trading off the cost of learning against improved decision quality. As the universality of knowledge constraints has been increasingly recognized, so the range of information cost functions used to model them has expanded. Verrecchia [1982] models choice of variance of a normal signal; Sims [2003] an unrestricted choice of information structure with costs based on Shannon entropy; and Reis [2006] the binary choice on whether or not to become fully informed.[1]

An important open question is how to test a model of optimal behavior in the face of costly information. Information costs imply that many patterns of apparently mistaken choices can be rationalized. Are there any patterns of choice error that *cannot* be explained by some underlying information cost function, or is such a general theory vacuous? We answer this question by characterizing *all* patterns of stochastic choice consistent with rational decision making in the face of information costs. Since we make no specific assumptions about the costs and constraints that the decision maker faces when gathering information, our tests encompass all existing models of optimal information acquisition.

Our non-parametric approach is motivated by the unobservability of costs

[1] See for example van Nieuwerburgh and Veldkamp [2009] and Woodford [2012] for other informational cost functions.

of information acquisition and acquired knowledge, just as revealed preference theory was motivated by the unobservability of preferences (Samuelson [1938]). To overcome the resulting observational constraint requires rich choice data. The tests that we develop apply to "state dependent" stochastic choice data, which identifies the probability of choosing each available action in each state of the world. Such data allows us to directly observe choice "mistakes", in which a sub-optimal alternative is chosen given the true state. While only recently introduced into revealed preference analysis (see Caplin and Martin [2014], henceforth CM14), this data set is standard in psychometric research on perceptual errors.[2] It is also common in the econometric analysis of discrete choice. For example, it is in just such data that Chetty *et al.* [2009] find evidence of incomplete state awareness among buyers.

We identify two intuitive conditions that render such data consistent with optimal acquisition of costly information. A "no improving action switches" (NIAS) condition ensures that choices are optimal given what was learned about the state of the world, as in CM14. A "no improving attention cycles" (NIAC) condition ensures that total utility cannot be raised by reassigning information structures across decision problems. Our main result is that these conditions are both necessary and sufficient for any arbitrary finite data set to be consistent with a model of costly information acquisition.

In section 4 we show how observed choice data bounds the relative costs of chosen information structures. We also show that adding the assumptions that more information is more costly, that mixed strategies are feasible, and that inattention is costless put no additional restrictions on the data.[3] In contrast, commonly used parametric cost functions have significant additional implications for behavior. We consider the case of information costs based on the expected reduction in Shannon entropy between prior and posterior (Sims [2003]), which has been heavily used in the applied literature.[4] We outline

[2]In the mid-19th century, Ernst Weber pioneered its use in the experimental assessment of how accurately individuals could differentiate between objectively different stimuli (see Murray [1993]).

[3]This result is in the spirit of Afriat [1967].

[4]e.g. Sims [2006], Woodford [2008], van Nieuwerburgh and Veldkamp [2009] Mackowiak

key behavioral properties implied by this cost function, which are significantly more restrictive than NIAS and NIAC alone (see also Caplin and Dean [2013]).

In section 5 we provided an application of our theory to experimentally generated data. We first specialize the NIAS and NIAC conditions to two act, two state decision problems. We then introduce an experimental design for generating state dependent stochastic choice data and present results from a prototypical experiment in which we vary the rewards for making the correct decision. We show that the aggregate state dependent stochastic choice data satisfies the NIAS and NIAC conditions, while rejecting more restrictive models: Our data is inconsistent both with signal detection theory (SDT - Green and Swets [1966]), which fixes the information structure independent of the decision problem, and with the Shannon model, which predicts a specific elasticity of knowledge to with respect to reward. A generalization of the Shannon model introduced by Caplin and Dean [2013] provides a better match for the observed behavior.

As detailed in section 6, our paper is most closely related to that of de Oliveira *et al.* [2013], which derives similar results in the setting of choice over menus. Other authors have considered the implications of more specific models of costly information acquisition (Caplin and Dean [2011], Ellis [2012], Matejka and McKay [2013]). Our work also fits into a growing literature aimed at identifying the behavioral implications of boundedly rational models in which the information state of the decision maker is unknown (Manzini and Mariotti [2012], Masatlioglu *et al.* [2012], Dillenberger *et al.* [2012], Bergemann and Morris [2013b]). Analogous revealed preference approaches have recently been applied to various behavioral models of individual and group decision making (Crawford [2010], Cherchye *et al.* [2011], de Clippel and Rozen [2012]).

Section 2 introduces the costly information representation. Section 3 provides our key characterization theorem. Section 4 establishes limits on the identifiability of information costs. Section 5 details our experimental results. Section 6 reviews related literature.

and Wiederholt [2010], Matejka [2010], Martin [2013].

2 A Costly Information Representation

2.1 Data

We consider a decision maker (DM) who chooses among actions, the outcomes of which depend on which of a finite number of states of the world $\omega \in \Omega$ eventuates. Each action a is a mapping from Ω to a prize space X. We let $F = X^{\Omega}$ denote the grand set of actions and $\mathcal{F} \equiv \{A \subset F||A| < \infty\}$ the set of decision problems (i.e. available alternatives from which the DM must choose).

The behavior of the DM is observed in a finite set of such decision problems. In each decision problem we observe *state dependent stochastic choice* data, which describes the probability of choosing each available action in each state of the world. Such data is richer than standard stochastic choice data (e.g. Gul and Pesendorfer [2006]), as it conditions choice probabilities on the state.

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

In addition to the pair (D, P), the DM's prior beliefs $\mu \in \Gamma = \Delta(\Omega)$ are treated as known. Note that, in our data set, the empirical frequency of each state is observable. Hence an alternative interpretation of the observability of prior beliefs is that we assume μ to be equal to this empirical frequency, in which case our theory incorporates the hypothesis that the DM's prior matches the objective likelihood of each state.[5]

For simplicity we assume that the expected utility function $u : X \longrightarrow \mathbb{R}$ is known, with $u(a(\omega))$ denoting the utility of action a in state ω. This allows us to focus exclusively on the implications of unobserved information costs. We address the case in which beliefs and preferences are unobservable in section 3.5.

[5] Although in principle our model allows for the DM's prior beliefs to be different from the true probability of each state.

Our data set allows us to observe the pattern of choice "errors" made by a decision maker - i.e. cases when they chose one option when another had a higher payoff given the state. Our goal is to characterize what form such errors must take if they are to be consistent with rationality of attentional choice.

We show in section 5 that state dependent stochastic choice data can be readily gathered in the laboratory. By aggregating across individuals, such data can also be extracted from field settings in which fluctuations in an underlying state (for example prices or tax rates) may or may not be fully understood by the DM.

2.1.1 Example

We illustrate the concept of state dependent stochastic choice data with an example which we will use throughout the paper, culminating in the experimental tests described in section 5. Risk neutral[6] subjects are faced with a screen on which there are 100 balls, each of which may be either red or blue.[7] Ex ante, subjects are informed that there is an equal chance that there will be either 49 or 51 red balls on the screen. They choose between action a, which pays \$10 if there are 49 red balls and \$0 otherwise, and action b, which pays \$10 if there are 51 red balls on the screen and \$0 otherwise. In our framework, this setting can be described as a decision problem with two states $\{\omega_1, \omega_2\}$, prior probabilities $\mu(\omega_1) = \mu(\omega_2) = 0.5$, choice set $A = \{a, b\}$, and utility function $u(a(\omega_1)) = u(b(\omega_2)) = 10$ and $u(b(\omega_1)) = u(a(\omega_2)) = 0$.

Subjects make repeated choices in this environment, with a new state and realized array of balls drawn each time. On each trial, the experimenter observes both the true state, and the choice made by the subject. This reveals $\bar{P}_A(i|\omega_j)$, the empirical frequency of the subject choosing action $i \in \{a, b\}$ in state ω_j, $j = 1, 2$. These frequencies provide an estimate of P_A, the state dependent stochastic choice data for set A. We discuss how this estimation can be carried out in practice in section 5.2.1.

[6] An alternative to the assumption of risk neutrality would be to estimate a subject's utility function for money using choices over objective lotteries and use the estimated utility function instead. See Caplin and Dean [2013] for an example of this approach.

[7] An example of such a screen is included in the experimental instructions in the appendix.

2.2 Model

We model the behavior of a DM who can gather information about the state of the world prior to choosing an action. Importantly, the DM can choose what information to gather conditional on the decision problem they are facing. In the example above, the DM observes the contingent payoffs of the two actions they must choose between before deciding how much effort to exert in estimating the number of red balls on the screen. We assume that there are costs associated with gathering information: in our running example, these costs might represent the cognitive effort of counting red balls, or the opportunity cost of the time spent doing so. The DM must therefore trade off these costs against the benefit of better information, and therefore better subsequent choices. We assume that the DM solves this trade off optimally.

We take an abstract approach to modelling the DM's choice of information. In each decision problem, the DM chooses an *information structure*: a stochastic mapping from objective states of the world to a set of subjective signals. Having selected an information structure, the DM can condition choice of action only on these signals. Since we are characterizing expected utility maximizers, we identify each subjective signal with its associated posterior beliefs $\gamma \in \Gamma$, which is equivalent to the subjective information state of the DM following the receipt of that signal. As in Kamenica and Gentzkow [2011], feasible information structures satisfy Bayes' rule.

Definition 2 *The set of* ***information structures*** Π *comprises all mappings* $\pi : \Omega \to \Delta(\Gamma)$ *that have finite support* $\Gamma(\pi) \subset \Gamma$ *and that satisfy Bayes' law, so that for all* $\omega \in \Omega$ *and* $\gamma \in \Gamma(\pi)$*,*

$$\gamma(\omega) = \Pr(\omega|\gamma) = \frac{\Pr(\omega \cap \gamma)}{\Pr(\gamma)} = \frac{\mu(\omega)\pi(\gamma|\omega)}{\sum_{\upsilon \in \Omega} \mu(\upsilon)\pi(\gamma|\upsilon)},$$

where $\pi(\gamma|\omega)$ *is the probability of signal* γ *given state* ω*.*

We assume that there is a cost associated with the use of each information structure.

Definition 3 *An **information cost function** is a mapping $K : \Pi \rightarrow \bar{\mathbb{R}}$ with $K(\pi) \in \mathbb{R}$ for some $\pi \in \Pi$. We let $\mathcal{K}$ denote the class of such functions.*

We put no restrictions on the cost function, meaning that our model nests all standard models of information acquisition. This includes the rational inattention model in which K is proportional to the Shannon mutual information between prior and posterior information states (e.g. Sims [2003]).[8] We allow costs to be infinite to cover hard constraints on information acquisition - as when a bound is imposed on the mutual information between prior and posteriors (Sims [2003]), or when the DM can choose only certain partitional information structures (Ellis [2012]) or specific types of signal (for example Verrecchia [1982], in which the DM can choose only normal signals).

We define $G : \mathcal{F} \times \Pi \rightarrow \mathbb{R}$ as the gross payoff of using a particular information structure in a particular decision problem. This is calculated assuming that actions are chosen optimally following each signal,

$$G(A, \pi) \equiv \sum_{\gamma \in \Gamma(\pi)} \left[\sum_{\omega \in \Omega} \mu(\omega)\pi(\gamma|\omega) \right] \left[\max_{a \in A} \sum_{\omega \in \Omega} \gamma(\omega) u(a(\omega)) \right].$$

Here the first bracketed term is the probability of each signal, and the second is the maximum achievable expected utility from A given the resulting beliefs.

We model a DM who, for any given decision problem, chooses an information structure to maximize gross payoffs net of information costs. We use $\hat{\Pi}(K, A)$ to refer to the set of optimal information structures in decision problem A given cost function K:

$$\hat{\Pi}(K, A) = \arg \max_{\pi \in \Pi} \{G(A, \pi) - K(\pi)\}.$$

While we focus on the case of a static, once-off choice of information structure, we show in Caplin and Dean [2014] that our results extend directly to the case of sequential choice of information.

[8] In section 4.3 below we illustrate specific behavioral features associated with this cost function.

2.3 Representation

Our aim is to understand the conditions under which state dependent stochastic choice data can be represented as resulting from costly information acquisition. Such a representation consists of three unobserved elements: (i) an **information cost function** which captures the subjective cost of different types of information; (ii) an **attention function** which captures the DM's choice of information structure in each decision problem; (iii) a **choice function** which captures the DM's choice of action following the receipt of each signal.[9]

Because we wish to model the behavior of a DM who behaves rationally given their information costs, both the attention function and the choice function must be optimal in order to form part of a costly information acquisition representation. This means that the choice of information structure in each decision problem must be optimal given information costs, and an action can only be chosen with positive probability after the receipt of a signal if it maximizes expected utility given the resulting beliefs. Furthermore, in order to represent a given data set, the information structure and choice function for each decision problem must give rise to the observed pattern of state dependent stochastic choice.

Definition 4 *Given $\mu \in \Gamma$ and $u : X \longrightarrow \mathbb{R}$, a state dependent stochastic choice data set (D, P) has a* ***costly information representation*** *if there exists information cost function $K \in \mathcal{K}$, attention function $\{\pi_A\}_{A \in D}$ and choice function $\{C_A\}_{A \in D}$ such that, for all $A \in D$:*

1. *Information is optimal: $\pi_A \in \hat{\Pi}(K, A) \equiv \arg\max_{\pi \in \Pi} \{G(A, \pi) - K(\pi)\}$.*

2. *Choices are optimal: the choice function $C_A : \Gamma(\pi_A) \rightarrow \Delta(A)$ is such that, given $a \in A$ and $\gamma \in \Gamma(\pi_A)$ with $C_A(a|\gamma) \equiv \Pr(a|\gamma) > 0$,*

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

[9] We allow the DM to randomize their choices of actions conditional on each signal.

3. The data is matched: given $\omega \in \Omega$ and $a \in A$,

$$P_A(a|\omega) = \sum_{\gamma \in \Gamma(\pi_A)} \pi_A(\gamma|\omega) C_A(a|\gamma).$$

3 Characterization

We establish two conditions as necessary and sufficient for a state dependent stochastic choice data set to have a costly information representation. The first ensures optimality of the information structure with regard to some cost function and applies to the collection of decision problems. The second ensures optimality of final choice given an information structure and applies to each decision problem separately.

3.1 The Revealed Information Structure

The key to our approach is the observation that one can learn much about a DM's attention strategy from state dependent stochastic choice data. For each decision problem, we construct a "revealed information structure", which replaces the actual information structure the DM used in a decision problem with an information structure that can be inferred directly from the data. We do this by imagining that each action is chosen in at most one subjective information state. If this assumption holds then the revealed information structure will be identical to the true information structure used by the DM. If not, then the revealed information structure is still related to the true information structure, as we discuss below.

We begin by identifying the revealed posterior beliefs $\bar{\gamma}_A^a$ associated with each chosen act. This specifies probabilities over states of the world conditional on action a being chosen in data set P_A. If the DM chooses each action in at most one subjective information state then the revealed posteriors are the same as their true posterior belief when each act is chosen.[10] If they choose

[10] Note that here we are here assuming that μ specifies both the DM's beliefs and the true probability of each state. If not, the revealed posterior refers to the DM's subjective belief of the likelihood of each state after the choice of each act, which may be different from the

the same action in more than one subjective state then the revealed posterior is the appropriate weighted average of the corresponding beliefs.

Definition 5 *Given $\mu \in \Gamma$, $A \in D$, $P_A \in P$, and $a \in Supp(P_A)$, the* ***revealed posterior*** $\bar{\gamma}_A^a \in \Gamma$ *is defined by,*

$$\begin{aligned} \bar{\gamma}_A^a(\omega) &\equiv \Pr(\omega | a \textit{ chosen from } A) \\ &= \frac{\mu(\omega) P_A(a|\omega)}{\sum_{\upsilon \in \Omega} \mu(\upsilon) P_A(a|\upsilon)}. \end{aligned}$$

In order to construct the revealed information structure, we use the set of revealed posteriors as the set of signals. The probability of signal γ in state of the world ω is then calculated by adding up the choice probabilities in state ω of all actions that have γ as their revealed posterior.

Definition 6 *Given $\mu \in \Gamma$, $A \in D$, and $P_A \in P$, the* ***revealed information structure*** $\bar{\pi}_A \in \Pi$ *satisfies,*

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

Even if the DM is behaving according to the model described in section 2.2, their revealed information structure may not be the same as their true information structure if they choose the same act following two different signals.[11] However, it must be the case that the revealed information structure is weakly less informative (in the sense of statistical sufficiency) than the true information structure, and in fact any information structure consistent with the data. Intuitively, this means that the revealed information structure can be obtained by "adding noise" to the true information structure. This notion is formalized in the following definition, adapted from Blackwell [1953].

true probability.

[11] An optimal DM would never choose to do this if more informative signals (in the sense described below) are more expensive, but might do so if, for example, they are restricted to using normal signals.

Definition 7 *Information structure $\rho \in \Pi$ is **sufficient** for information structure $\pi \in \Pi$ (equivalently π is a **garbling** of ρ) if there exists a $|\Gamma(\rho)| \times |\Gamma(\pi)|$ matrix $B \geq 0$ with $\sum_{\gamma^j \in \Gamma(\pi)} b^{ij} = 1$ all i and such that, for all $\gamma^j \in \Gamma(\pi)$ and $\omega \in \Omega$,*

$$\pi(\gamma^j|\omega) = \sum_{\eta^i \in \Gamma(\rho)} b^{ij} \rho(\eta^i|\omega).$$

This definition states that an information structure ρ is sufficient for information structure π if π can be obtained by applying a stochastic matrix B to ρ. One way to interpret the concept of garbling is by considering a procedure by which π is constructed by first applying ρ, then adding noise by combining the resulting information states together using the weights b^{ij}. Example 1 below includes an application of the concept of sufficiency.

Lemma 1 establishes that any information structure which is consistent with the state dependent stochastic choice data in a given decision problem must be sufficient for the revealed information structure.

Lemma 1 *If $\pi \in \Pi$ is consistent with $P_A \in P$,*[12] *then it is sufficient for $\bar{\pi}_A$.*
Proof. *All proofs can be found in appendix 1.* ■

Thus, while we cannot guarantee that the revealed information structure is the same as the true information structure, we do know that it must be more informative than the true information structure. The following section makes use of this observation to identify a necessary condition for the costly information representation.

The following example demonstrates the construction of the revealed information structure and its relationship with the true information structure.

[12] i.e. there exists a $C : \Gamma(\pi) \to \Delta(A)$ such that, for each $\gamma \in \Gamma(\pi)$,

$$C(a|\gamma) > 0 \Longrightarrow \sum_{\omega \in \Omega} \gamma(\omega) u(a(\omega)) \geq \sum_{\omega \in \Omega} \gamma(\omega) u(b(\omega)) \text{ all } b \in A$$

and for each $\omega \in \Omega$ and $a \in A$,

$$P_A(a|\omega) = \sum_{\gamma \in \Gamma(\pi)} \pi(\gamma|\omega) C(a|\gamma).$$

Example 1 *Consider a DM who, when faced with the decision problem defined in section 2.1.1 employs an information structure with three signals, $\alpha, \beta, \gamma \in \Gamma$ such that the resulting posterior beliefs are:*

$$\alpha = (\alpha(\omega_1), \alpha(\omega_2)) = \left(\frac{3}{4}, \frac{1}{4}\right); \ \beta = \left(\frac{1}{4}, \frac{3}{4}\right); \ \gamma = \left(\frac{1}{2}, \frac{1}{2}\right).$$

and with conditional probabilities of receiving each signal:

$$\begin{aligned} \pi(\alpha|\omega_1) &= \frac{1}{2}; \ \pi(\beta|\omega_1) = \frac{1}{6}; \ \pi(\gamma|\omega_1) = \frac{1}{3}; \\ \pi(\alpha|\omega_2) &= \frac{1}{6}; \ \pi(\beta|\omega_2) = \frac{1}{2}; \ \pi(\gamma|\omega_2) = \frac{1}{3}. \end{aligned}$$

Suppose that after the receipt of signal α the DM chooses act a for sure and after the receipt of β they choose act b for sure. After the receipt of signal γ they randomize between the two acts, so

$$\begin{aligned} C(a|\alpha) &= C(b|\beta) = 1; \\ C(a|\gamma) &= C(b|\gamma) = \frac{1}{2}. \end{aligned}$$

This behavior would give rise to state dependent stochastic choice data,

$$\begin{aligned} P(a|\omega_1) &= \pi(\alpha|\omega_1) + \frac{1}{2}\pi(\gamma|\omega_1) = \frac{2}{3}; \\ P(a|\omega_2) &= \pi(\alpha|\omega_2) + \frac{1}{2}\pi(\gamma|\omega_2) = \frac{1}{3}; \end{aligned}$$

The resulting revealed information structure would have two revealed posteriors, one associated with the choice of a and the other with the choice of b. Application of definition 5 gives,

$$\bar{\gamma}^a = \left(\frac{2}{3}, \frac{1}{3}\right) \text{ and } \bar{\gamma}^b = \left(\frac{1}{3}, \frac{2}{3}\right).$$

The corresponding revealed information structure is then given by

$$\begin{aligned}\bar{\pi}(\bar{\gamma}^a|\omega_1) &= P(a|\omega_1) = \frac{2}{3};\\ \bar{\pi}(\bar{\gamma}^a|\omega_2) &= P(a|\omega_2) = \frac{1}{3}.\end{aligned}$$

Clearly, the revealed information structure is not the same as the true information structure, but the true information structure is sufficient for the revealed information structure. This can be seen by applying the stochastic matrix B to to π_A, in order to obtain $\bar{\pi}_A$, where

$$B = \left\{ \begin{array}{cc} 1 & 0 \\ 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\}.$$

The rows relate to signals α, β and γ respectively, while the columns relate to signals $\bar{\gamma}^a$ and $\bar{\gamma}^b$ respectively. Thus, for example

$$\bar{\pi}(\bar{\gamma}^a|\omega_1) = b^{1,1}\pi(\alpha|\omega_1) + b^{2,1}\pi(\beta|\omega_1) + b^{3,1}\pi(\gamma|\omega_1) = 1 * \frac{1}{2} + \frac{1}{2} * \frac{1}{3} = \frac{2}{3},$$

as required. Note that the stochastic matrix B is closely related to the choice function C: b^{ij} is the probability of choosing the act associated with the revealed posterior in column j following the receipt of the signal associated with row i.

3.2 No Improving Attention Cycles

Our first condition restricts choice of information structure across decision problems to ensure consistency with a fixed underlying information cost function. Essentially, total gross utility cannot be increased by reassigning information structures across decision problems. To illustrate, consider again the decision problem of section 2.1.1 and suppose, as in the example above, that the observed choice behavior is,

$$P_A(a|\omega_1) = \frac{2}{3} = P_A(b|\omega_2).$$

Now consider a second decision problem $A' = \{a', b'\}$ such that $u(a'(\omega_1)) = u(b'(\omega_2)) = 2$ and $u(b'(\omega_1)) = u(a'(\omega_2)) = 0$, with corresponding data,

$$P_{A'}(a'|\omega_1) = \frac{3}{4} = P_{A'}(b'|\omega_2).$$

Intuitively, these data should not have an optimal information representation. Action set A provides greater reward for discriminating between states, yet the DM is more discerning under action set A'. To crystallize the resulting problem, note that, for behavior to be consistent with costly information acquisition for some cost function K it must be the case that DM's true choice of information structures satisfies,

$$\begin{aligned} G(A, \pi_A) - K(\pi_A) &\geq G(A, \pi_{A'}) - K(\pi_{A'}); \\ G(A', \pi_{A'}) - K(\pi_{A'}) &\geq G(A', \pi_A) - K(\pi_A). \end{aligned}$$

Hence,

$$G(A, \pi_A) - G(A, \pi_{A'}) \geq K(\pi_A) - K(\pi_{A'}) \geq G(A', \pi_A) - G(A', \pi_{A'}),$$

implying finally that,

$$G(A, \pi_A) + G(A', \pi_{A'}) \geq G(A, \pi_{A'}) + G(A', \pi_A). \quad (1)$$

We conclude that, for this data to be rationalizable, gross benefit must be maximized by the assignment of the chosen information structure to the corresponding decision problem. We translate this into a testable condition by noting that the corresponding inequality remains valid when we replace the true with the revealed information structure. To do so, we make use of Blackwell's theorem (Blackwell [1953]), which establishes the equivalence of the statistical notion of sufficiency and the economic notion "more valuable than". If information structure π is sufficient for strategy ρ, then it yields (weakly) higher gross payoffs in any decision problem.

Remark 1 *Given decision problem $A \in \mathcal{F}$ and $\pi, \rho \in \Pi$ with ρ sufficient for*

π,

$$G(A,\rho) \geq G(A,\pi).$$

Combined with observation that the true information structure must be sufficient for the revealed information structure, this implies that $G(i,\pi_j) \geq G(i,\bar{\pi}_j)$ for $i.j \in \{A,A'\}$. Furthermore, it is clear that $G(i,\pi_i) = G(i,\bar{\pi}_i)$ for $i \in \{A,A'\}$ since the resulting state dependent choices are identical. As a result, we can replace equation 1 with a testable condition,

$$G(A,\bar{\pi}_A) + G(A',\bar{\pi}_{A'}) \geq G(A,\bar{\pi}_{A'}) + G(A',\bar{\pi}_A). \tag{2}$$

In the above example $G(A,\bar{\pi}_A)+G(A,\bar{\pi}_{A'}) = 8\frac{1}{6}$, while $G(A,\bar{\pi}_{A'})+G(A',\bar{\pi}_A) = 8\frac{5}{6}$. Thus, there is no cost function that can be used to rationalize this data.

Our explanation so far has considered only bilateral reassignments of information structures. The NIAC condition ensures that gross utility cannot be increased by reassigning information structures along any *cycle* of decision problems. It is analogous to the cyclical monotonicity condition discussed in Rockafellar [1970], and has been used in other recent work examining the revealed preference implications of behavioral models (see for example Crawford [2010]).

Condition D1 (No Improving Attention Cycles) Given $\mu \in \Gamma$ and $u : X \to \mathbb{R}$, (D,P) satisfies NIAC if, for any set of decision problems $A^1, A^2, ..., A^J \in D$ with $A^J = A^1$,

$$\sum_{j=1}^{J-1} G(A^j,\bar{\pi}_j) \geq \sum_{j=1}^{J-1} G(A^j,\bar{\pi}_{j+1}),$$

where $\bar{\pi}_j = \bar{\pi}_{A^j}$.

In section 5 we demonstrate the application of the NIAC condition to the simple case of two acts and two states.

3.3 No Improving Actions Switches

Our second condition is based on the fact that a DM's choices must be optimal given posterior beliefs. Thus when one identifies in the data the revealed posterior associated with any chosen action, that action must be optimal given those beliefs. This implies that, for any $A \in D$, $a \in Supp(P_A)$, and $b \in A$

$$\sum_{\omega \in \Omega} \bar{\gamma}_A^a(\omega)\, u(a(\omega)) \geq \sum_{\omega \in \Omega} \bar{\gamma}_A^a(\omega)\, u(b(\omega)). \tag{3}$$

This statement follows directly from the optimality of choice if each action is chosen in at most one state, and so the revealed posterior is equal to the true posterior. It also holds true if the same action is chosen in many given information states: the action must be optimal at each information state at which it is chosen, and so also must be optimal at any convex combination of those beliefs, including the revealed posterior. CM14 show that this condition characterizes Bayesian behavior regardless of the rationality of attentional choice. The strategic analog is derived by Bergemann and Morris [2013b] in characterizing Bayesian correlated equilibria.

Equation 3 can be rewritten directly in terms of state dependent stochastic choice data:

Condition D2 (No Improving Action Switches) Given $\mu \in \Gamma$ and $u : X \to \mathbb{R}$, data set (D, P) satisfies NIAS if, for every $A \in D$, $a \in Supp(P_A)$, and $b \in A$,

$$\sum \mu(\omega) P_A(a|\omega)\left(u(a(\omega)) - u(b(\omega))\right) \geq 0.$$

Section 5 contains a simple application of NIAS to the two state, two action case. CM14 contains many further illustrative examples.

3.4 Characterization

The above analysis shows that both NIAC and NIAS are necessary for the existence of a costly information representation. Our central result is that they are also sufficient. We establish this by following the approach that Koopmans and Beckmann [1957] developed to solve the problem of locating indivisible factories across sites so as to maximize total profits. They show that the solution to this allocation problem can be found by solving a linear program in which one imagines the factories to be divisible. Their key observation is that there is an extreme point solution, which corresponds to placing each factory in one and only one location. Associated with the solution to the linear programming problem are shadow prices (either rents on locations or prices of factories) that decentralize the allocation. By direct analogy, the NIAC conditions states that the DM has allocated revealed information structures to decision problems in such a manner as to maximize total gross expected utility. The cost function K that we introduce is based directly on the shadow prices that decentralize this optimal allocation (see also Rochet [1987]).

Theorem 1 *Given $\mu \in \Gamma$ and $u : X \to \mathbb{R}$, data set (D, P) has a costly information acquisition representation if and only if it satisfies NIAS and NIAC.*

3.5 Unobservable Utility and Prior Beliefs

So far we have assumed that the DM's expected utility function and prior beliefs over states of the world are both known to the researcher - only information structures, choice functions and costs are not directly observable. We now outline two ways to adapt our approach to allow for an unknown utility function and/or prior.

One approach is to enrich the data set to allow for the recovery of beliefs and preferences from choices that are unaffected by information costs. These beliefs and preferences could then be used as a starting point for our representation. In order to recover utility, we could replace the "Savage style" actions we use in this paper (which map deterministically from states of the world to prizes) with "Anscombe-Aumann" acts that map states of the world to probability distributions over the prize space. Assuming the DM does maximize expected utility, u could then be recovered by observing choices over degenerate acts (i.e. acts whose payoffs are state independent).[13] If we further add to our data set the choices of the DM over acts *before* the state of the world is determined (or at least in a situation in which they cannot exert any effort to determine that state) then we can also recover the DM's prior over objective states (again assuming expected utility maximization). This method is pursued in de Oliveira *et al.* [2013].[14]

A second approach is to directly identify testable implications when utility, prior beliefs, and information costs are all unobserved. In Caplin and Dean [2014] we show that, in such cases, the model is consistent with the data if and only if there *exists* a utility function and set of prior beliefs such that NIAC and NIAS hold. This gives rise to a set of inequality constraints to which a solution must exist if the data is to be rationalizable with a costly information representation (such a result is similar in spirit to Crawford [2010]). In the case in which the prior is known but the utility function is not, these constraints are linear and easy to check (see CM14 for the implications of NIAS alone). If the prior is also unknown, then the conditions are non-linear, but still non-vacuous. CM14 provide an example of data that is incompatible with NIAS for any utility function and prior. Caplin and Dean [2014] provide an example of behavior that is commensurate with NIAS but is not commensurate with NIAC for any non-degenerate utility function and prior.

4 The Information Cost Function

In this section we discuss what can be learned about information costs from state dependent stochastic choice data, as well as the behavioral implications of placing further restrictions on the information cost function.

[13] An applied variant of this approach involves seperately identifying a subject's utility function for money using, for example, a multiple price list approach - see for example Caplin and Dean [2013].

[14] Ellis [2012] also uses this method to identify the DM's utility function, but takes a different approach to identifying prior beliefs.

4.1 Recoverability and Uniqueness

Theorem 1 tells us the conditions under which there exists an information cost function that will rationalize the data. We now identify all such cost functions, in the spirit of Varian [1984] and Cherchye *et al.* [2011]. We restrict ourselves to cost functions in which more information is at least weakly more costly, so that we can treat revealed information structures as optimal. The key observation is that the choice of $\bar{\pi}_A$ in decision problem A puts an upper bound on its cost relative to that of any other strategy $\pi \in \Pi$,

$$K(\bar{\pi}_A) - K(\pi) \leq G(A, \bar{\pi}_A) - G(A, \pi). \tag{4}$$

This directly implies an upper and lower bound on the relative costs of any two revealed information structures $\bar{\pi}_A$, $\bar{\pi}_B$ for $A, B \in D$,

$$G(B, \bar{\pi}_A) - G(B, \bar{\pi}_B) \leq K(\bar{\pi}_A) - K(\bar{\pi}_B) \leq G(A, \bar{\pi}_A) - G(A, \bar{\pi}_B).$$

An obvious corollary of theorem 1 is that a weakly monotonic information cost function can rationalize a data set if and only it satisfies this inequality for every A,$B \in D$, and the costs of unchosen information structures are high enough to satisfy inequality 4.

This condition implies potentially tighter bounds on the relative cost of any two revealed information structures. Consider the corresponding inequalities in the sequence $A^1 ... A^n \in D$ with $A^1 = A$ and $A^n = B$,

$$\begin{aligned}
K(\bar{\pi}_1) - K(\bar{\pi}_2) &\leq G(A^1, \bar{\pi}_1) - G(A^1, \bar{\pi}_2); \\
K(\bar{\pi}_2) - K(\bar{\pi}_3) &\leq G(A^2, \bar{\pi}_2) - G(A^2, \bar{\pi}_3); \\
&\vdots \\
K(\bar{\pi}_{n-1}) - K(\bar{\pi}_n) &\leq G(A^{n-1}, \bar{\pi}_{n-1}) - G(A^{n-1}, \bar{\pi}_n).
\end{aligned}$$

Summing these inequalities yields a bound on $K(\bar{\pi}_A) - K(\bar{\pi}_B)$. This relative cost must obey such bounds for all cycles,

$$K(\bar{\pi}_A) - K(\bar{\pi}_B) \leq \min_{\{A^1 \ldots A^n \in D | A^1 = A, A^n = B\}} \sum_{i=1}^{n-1} \left[G(A^i, \bar{\pi}_i) - G(A^i, \bar{\pi}_{i+1}) \right]. \quad (5)$$

Considering the reverse sequence $A^1, ..., A^n \in D$ with $A^1 = B$ and $A^n = A$,

$$K(\bar{\pi}_A) - K(\bar{\pi}_B) \geq \max_{\{A^1 \ldots A^n \in D | A^1 = B, A^n = A\}} \sum_{i=1}^{n-1} \left[G(A^i, \bar{\pi}_{i+1}) - G(A^i, \bar{\pi}_i) \right]. \quad (6)$$

Note also that if one considers cost functions for which inattention is free (as discussed below), the above inequalities can be used to place absolute bounds on the level of costs. Moreover if one ever sees a switch in information structure for decision problems that are "close together", in that available vectors of state dependent payoffs always fall within $\epsilon > 0$, then one can bound cost differences to within ϵ. Hence, with a rich enough data set, arbitrarily tight bounds can be placed on costs in models in which the data is generated by a finite set of revealed information structures.

4.2 Unobservable Restrictions on Information Costs

We now introduce three natural restrictions on K: weak monotonicity with respect to sufficiency, feasibility of mixed strategies, and costless inattention. In principle these restrictions might tighten requirements for rationalizability of stochastic choice data, since they constrain the costs of unchosen strategies. Theorem 2 establishes that this is not the case: if state dependent stochastic choice is rationalizable, then it is rationalizable by a cost function that satisfies these three conditions.

A partial ranking of the informativeness of information structures is provided by the notion of statistical sufficiency (see definition 7). Our first, apparently natural condition for an information cost function is that more information is (weakly) more costly. This is implied, for example, by free disposal of information.

Condition K1 $K \in \mathcal{K}$ satisfies **weak monotonicity in information** if, for any $\pi, \rho \in \Pi$ with ρ sufficient for π,

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

A second natural condition is that DMs can choose to mix information structures and pay the corresponding expected costs. For example, they could flip a coin and choose strategy π if the coin comes down heads and strategy η if it comes down tails. In expectation the cost of this strategy would be half that of π and half that of η. Note that the resulting mixing is not of the posteriors themselves, but of the odds of the given posteriors. To illustrate, consider again a case with two equiprobable states. Let information structure π be equally likely to produce posteriors $(.3, .7)$ and $(.7, .3)$, with η equally likely to produce posteriors $(.1, .9)$ and $(.9, .1)$. Then the mixture strategy $0.5\circ \pi + 0.5 \circ \eta$ is equally likely to produce all four posteriors.

Definition 8 *Given information structures $\pi, \eta \in \Pi$, and $\alpha \in [0, 1]$, the **mixture strategy** $\alpha \circ \pi + (1 - \alpha) \circ \eta \equiv \psi \in \Pi$ is defined by*

$$\psi(\gamma|\omega) = \alpha\pi(\gamma|\omega) + (1 - \alpha)\eta(\gamma|\omega),$$

all $\omega \in \Omega$ and $\gamma \in \Gamma(\pi) \cup \Gamma(\eta)$.

Allowing mixtures between strategies $\pi, \eta \in \Pi$ puts an upper bound on $\alpha \circ \pi + (1 - \alpha) \circ \eta$ in terms of $K(\pi)$ and $K(\eta)$. However, it does not pin down the cost precisely, since there may be a more efficient way of constructing the mixed information structure.

Condition K2 Mixture Feasibility: for any two strategies $\pi, \eta \in \Pi$ and $\alpha \in (0, 1)$, the cost of the mixture strategy $\psi = \alpha\circ \pi + (1 - \alpha) \circ \eta \in \Pi$ satisfies,

$$K(\psi) \leq \alpha K(\pi) + (1 - \alpha)K(\eta).$$

It is typical in the applied literature to allow inattention at no cost, and otherwise to have costs be non-negative. This is our third condition.

Condition K3 Define $I \in \Pi$ as the strategy in which $\pi(\mu|\omega) = 1$ all $\omega \in \Omega$. Information cost function $K \in \mathcal{K}$ satisfies **normalization** if it is non-negative where real-valued, with $K(I) = 0$.

Theorem 2 states that, whenever a costly information representation exists, one also exists in which the cost function satisfies conditions K1 through K3. Even if any of the above conditions is false, any data set that can be rationalized can equally be rationalized by a cost function that satisfies them all.

Theorem 2 *Given $\mu \in \Gamma$ and $u : X \to \mathbb{R}$, data set (D, P) satisfies NIAS and NIAC if and only if it has a costly information representation with conditions K1 to K3 satisfied.*

Necessity is immediate from theorem 1. As detailed in the appendix, the proof of sufficiency proceeds in three steps, starting with a costly information acquisition representation (K, π, C) of the form produced in theorem 1, which assigns infinite information costs to all non-used information structures. The first step is to expand the domain on which K is real-valued to be closed under mixtures and garbling. The second step is to define a candidate function $\mathring{K}$ on this larger domain that satisfies conditions K1 through K3. The final step is to confirm that this function provides a costly information representation.

Theorem 2 has the flavor of the Afriat characterization of rationality of choice from budget sets (Afriat [1967]), which states that choices can be rationalized by a non-satiated utility function if and only if they can be rationalized by a non-satiated, continuous, monotone, and concave utility function. Not all restrictions on the form of the cost function can be so readily absorbed. For example, we cannot strengthen condition K1 to cover the case of strict monotonicity with respect to sufficiency. We show in appendix 2 that there are data sets satisfying NIAS and NIAC for which there exists no cost function that produces a costly information acquisition representation with a cost function that is strictly monotonic with the informativeness of the information structure.

Theorem 2 has interesting implications for identification. For example, individuals may in reality be curious to learn in various contexts, meaning that the "utility cost" of becoming better informed is negative. Theorem 2 implies that our data set is insufficiently rich to identify such curiosity should it exist: such a person would be indistinguishable from one whose costs were weakly monotonic with respect to informativeness.

4.3 The Shannon Cost Function

The Shannon mutual information cost function for an information structure is defined as,

$$K^S(\pi) = \lambda \left[H(\mu) - \sum_{\gamma \in \Gamma(\pi)} \left(\sum_{\omega \in \Omega} \pi(\gamma|\omega) \right) H(\gamma) \right].$$

Here $H(\mu) = -\sum_{\omega \in \Omega} \mu_\omega \ln \mu_\omega$ is the Shannon entropy function[15] and $\lambda > 0$ scales the cost of information. This highly parameterized special case of the costly information model was introduced into the economics literature by Sims [2003], and has been justified on both information theoretic and axiomatic grounds. As noted in the introduction, this model has been widely used in applied work.

Matejka and McKay [2013] and Caplin and Dean [2013] characterize the pattern of state dependent stochastic choice associated with the model as satisfying

$$P(a|\omega) = \frac{P(a)z(a(\omega))}{\sum_{b \in A} P(b)z(b(\omega))}, \tag{7}$$

where $z(a(\omega)) \equiv \exp(u(a(\omega))/\lambda)$ and $P(a) = \sum_{\omega \in \Omega} \mu(\omega) P(a|\omega)$ is the unconditional probability of choosing action a. The unconditional probabilities themselves are pinned down by the complementary slackness conditions:

$$\sum_{\omega \in \Omega} \mu(\omega) \left(\frac{P(a)z(a(\omega))}{\sum_{b \in A} P(b)z(b(\omega))} \right) \leq 1 \text{ all } a; \tag{8}$$

[15] Extended to boundary points using the limit condition $\lim_{\gamma \searrow 0} \gamma \ln \gamma = 0$.

with equality if $P(a) > 0$.

Clearly equations 7 and 8 impose significant behavioral restrictions in addition to NIAS and NIAC. One necessary condition pins down the rate at which a DM's choice accuracy responds to incentives for making the correct decision. We test this condition using our experimental data in the next section. The Shannon model also imposes cross-prior restrictions: once one identifies the revealed posteriors for one prior, these same posteriors will be used for any prior in their convex hull. Restrictions of this form are entirely absent in the general case, which requires only validity of NIAS and NIAC for each prior.

5 Applying NIAC and NIAS

5.1 NIAC and NIAS: the 2×2 Case

We now provide a concrete application of the NIAC and NIAS conditions to the simple case of two actions and two equally likely states which we will subsequently apply to experimental data. Consider first a single decision problem $A = \{a, b\}$ with two equally likely states of the world, and with action a better than action b in state ω_1, and vice versa in state 2: $u(a(\omega_1)) > u(b(\omega_1))$ and $u(b(\omega_2)) > u(a(\omega_2))$ (this represents a generalization of the example from section 2.1.1). To apply NIAS, note that posterior beliefs can be summarized by $\gamma(\omega_1)$, the probability of state 1. The value of choosing action a is increasing and the value of choosing action b is decreasing in this probability. There is therefore a threshold on $\gamma(\omega_1)$ such that above the threshold it is optimal to choose a and below it is optimal to choose action b. NIAS translates this observation into the following condition on the state dependent stochastic choice data,

$$P(a|\omega_1) \geq \max\left\{\alpha P(a|\omega_2), \alpha P(a|\omega_2) + \beta\right\}, \tag{9}$$

where $\alpha = \frac{b_2 - a_2}{a_1 - b_1}$ and $\beta = \frac{a_1 + a_2 - b_1 - b_2}{a_1 - b_1}$. Thus the relative cost of mistakes in the two states puts a bound on the relative likelihood of choosing action a in the two states

Given two analogous decision problems $A^i = \{a^i, b^i\}$ for $i = 1, 2$ with

$u(a^i(\omega_1)) > u(b^i(\omega_1))$ and $u(b^i(\omega_2)) > u(a^i(\omega_2))$, the NIAC condition reduces to,

$$\Delta P(a|\omega_1)\left(\Delta\left(u(a(\omega_1)) - u(b(\omega_1))\right)\right) + \Delta P(b|\omega_2)\left(\Delta\left(u(b(\omega_2)) - u(a(\omega_2))\right)\right) \geq 0 \tag{10}$$

where Δ indicates the change in the relevant variable between the two decision problems. The first term is equal to the product of change in the probability of making the correct choice in state ω_1, with the change in the value of making the correct choice in that state. The second term is the same product for state ω_2.

5.2 An Experimental Implementation

We now present an experiment which we use to illustrate the implementation of our tests.[16] As in the running example, subjects are presented with a screen shot with 100 balls that may either red or blue. Prior to seeing the screen, subjects are informed that there are two equally likely states of the world: ω_1 in which there are 49 red balls on the screen, and ω_2 in which there are 51. Subjects choose among actions whose payoffs are state dependent. There is neither an external limit (such as a time constraint) nor an extrinsic cost associated with understanding the state of the world. Information constraints derive from agents' unwillingness to trade time and cognitive effort for monetary reward.

Table 1: Decision Problems

	Payoffs (US$)			
DP	$u(a(\omega_1))$	$u(a(\omega_2))$	$u(b(\omega_1))$	$u(b(\omega_1))$
1	2	0	0	2
2	10	0	0	10
3	20	0	0	20
4	30	0	0	30

[16] See Caplin and Dean [2014] for related experiments. Given the psychological precedents, it is perhaps surprising that there has been so little experimental work on state dependent stochastic choice data within economics. One related paper is Cheremukhin *et al.* [2011], which estimates a rationally inattentive model of lottery choice.

Subjects face four decision problems described in Table 1. Each problem consists of two symmetric actions: a, which pays off in state ω_1 and b which pays off in state ω_2. Decision problems vary only in the payoff for choosing the correct action. A subject faces each decision problem 50 times.[17] All occurrences of the same problem are grouped, but the order of the problems is block-randomized. 46 Subjects were recruited from the New York University student population. Each subject answered 200 questions as well as 1 practice question. At the end of each session, one question was selected at random for payment, the result of which was added to the show up fee of \$10. Subjects took on average approximately 45 minutes to complete a session. Instructions are included in the appendix 4.

5.2.1 Matching Data to Theory

The NIAC and NIAS conditions are stated with regard to the state dependent stochastic choice function P. An experimentally generated data set will instead report the frequency with which each action is chosen in each state of the world in each decision problem. Under the assumption that each repetition represents an independent draw from the relevant distribution in P, we can use this data to provide an estimate of P: for example, the point estimate of $P_{A^1}(a|\omega_1)$ is the proportion of times that action a was chosen in state ω_1 in decision problem 1. We also know the distribution of this estimate given the true choice probabilities and number of observations. This allows us to make statistical statements about the likelihood that the axioms are satisfied by the underlying P, given the observed data. This is the approach we take below, using a linear probability model to estimate the relevant choice probabilities.

In order to test the NIAS and NIAC conditions it is generally necessary to observe the utility function of the subject. However, this is not the case given the simple structure of our experiments: both conditions can be tested using only the assumption of strict monotonicity - i.e. more money is preferred to

[17] To prevent subjects from learning to recognize patterns, we randomize the position of the balls. The implicit assumption is that the perceptual cost of determining the state is the same for all realized configurations of balls.

less.

We assume that subjects' prior beliefs are equal to the true probabilities of each state, as this information was presented to them before each repetition of the decision problem.

5.2.2 Data Overview

Our experimental design produces choice data that is both stochastic and state dependent. Subjects gather some information prior to choice, but this information is incomplete. In the aggregate, subjects made "mistakes", choosing the inferior action on 32% of all trials. However, choice behavior is also significantly different across states (Fisher's exact test, $p < 0.0001$): averaging across all 4 decision problems, a was chosen 75% of the time in state 1 and 38% of the time in state 2. These patterns hold true at the individual level: of the 46 subjects, 15% made mistakes in less than 10% of questions, while 76% had choice behavior that was significantly different between the two states at the 10% level.[18] These results imply that our subjects are absorbing some information about the state of the world, but are not fully informed when they make their choices.

5.2.3 Testing NIAS and NIAC

Applying equation 9 to the symmetric environment of our experiment shows that NIAS requires only that $P_j(a|\omega_1) \geq P_j(a|\omega_2)$ for each decision problem j. Table 2 reports the point estimates for these values for each decision problem

[18] Statistical test carried out using OLS regression of the choice of act on a dummy associated with the state.

from the aggregate data

Table 2: NIAS Test

DP	$P_j(a\|\omega_1)$	$P_j(a\|\omega_2)$	Prob
1	0.71	0.48	0.000
2	0.75	0.39	0.000
3	0.76	0.37	0.000
4	0.80	0.29	0.000

For all decision problems, the point estimates satisfy the relevant condition. We can also test whether these estimated values allow us to reject the hypothesis that NIAS is violated. To do so, we test the hypothesis that $P_j(a|\omega_1) = P_j(a|\omega_2)$ in each decision problem. If this is rejected, then we can also reject any null hypothesis of the form $P_j(a|\omega_1) < P_j(a|\omega_2)$ - i.e. any. that implies violation of NIAS. Such a test can be performed for each decision problem by regressing the choice of action in each trial on a dummy taking the value 1 if the true state is ω_1 and 0 otherwise. We do so using OLS regressions, with standard errors clustered at the individual level. The final column reports the result of the test of the restriction that the coefficient on the dummy is equal to zero. In each case we are indeed able to reject this hypothesis at the 0.1% level. Note that this also indicates that our aggregate data has power to differentiate between subject behavior and a random choice benchmark (in the style of Beatty and Crawford [2011]).

In order to understand the implications of NIAC for our data, we can apply equation 10 noting in addition that the cost of making the incorrect decision is the same in the two states. This implies that the NIAC condition is satisfied for our data if and only if the total probability of making the correct decision is increasing with the reward of doing so. Figure 1 shows the aggregate probability of making the correct choice in each decision problem, along with the associated standard errors.[19]

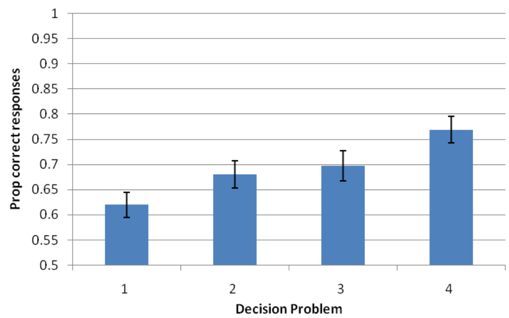


Figure 1

The point estimates are in line with the NIAC condition. Again, we can test whether we can reject the hypothesis that the NIAC conditions are violated by testing for equality in the probability of making the correct choice between any pair of decision problems. We do so using the data from all decision problems by regressing a dummy equal to one if the correct choice is made on a set of dummies indicating the decision problem. We then perform tests of equality of the resulting coefficients for all pairs of dummies. All such hypotheses can be rejected at the 1% level, apart from the comparison of decision problem 2 and 3 for which the coefficient is not significant ($P = 0.44$).[20] The aggregate data from our experiment is therefore in line with the model of costly information acquisition.[21]

[19] Clustered at the individual level.

[20] Again using OLS regression with standard errors clustered at the individual level.

[21] Individual level data is significantly more noisy than the aggregate data. The point estimates of many (24%) subjects violate NIAS, and most (82%) show some violation of NIAC. However, this is largely due to the relatively small number of observations at the individual level, and the subsequent lack of precision in the point estimates. For only 2% of subjects is there a decision problem in which the point estimates of their stochastic choice function indicate a violation of NIAS such that we can reject the hypothesis that the condition holds at equality at the 10% level. Similarly, for only 18% can we find a "significant" violation of NIAS. Moreover, as we show in appendix 2, even taking the point estimates as equal to the true choice probabilities, individual level violations of NIAS and NIAC are small in that they result in small financial losses.

We can use table 2 to identify the point estimates of the revealed posteriors and the revealed information structure in each decision problem. For example, in decision problem 1, $\bar{\gamma}_1^a = (0.60, 0.40)$, $\bar{\gamma}_1^b = (0.36, 0.64)$, while $\bar{\pi}\left(\bar{\gamma}_1^a|\omega_1\right) = 0.71$ and $\bar{\pi}\left(\bar{\gamma}_1^a|\omega_1\right) = 0.48$. Using the techniques described in section 4.1 we can also put bounds on the relative costs of the information structures revealed in the different decision problems.[22] Because the probability of being paid for any given round is small, so is the expected utility of any information structure used. This means that the estimated cost differences are also small: for example, our data implies that,

$$\$0.003 \leq K\left(\bar{\pi}_2\right) - K\left(\bar{\pi}_1\right) \leq \$0.013.$$

5.2.4 Discussion

Given that our aggregate data is consistent with a general model of costly information acquisition, we can ask whether it is also consistent with specific models within that class. We discuss two examples.[23]

First we consider a refinement in which the information structure used by the DM is invariant with the decision problem. While more typical in psychology (for example signal detection theory (SDT) - Green and Swets [1966]), fixed information models have attracted recent attention in the economics literature (e.g. Lu [2013]). Such models are clearly a special case of the costly information acquisition model, and so imply both NIAC and NIAS. However the same signal structure must rationalize behavior in all decision problems. If information does not change as a function of incentives, neither should choice behavior vary across the decision problems in this experiment, as the optimal action in each posterior state is independent of the precise value of $u(a(\omega_1))$ and $u(b(\omega_2))$, as long as they are equal.[24] This is contradicted by the fact that

[22] Treating these point estimates as true underlying information structures.

[23] Caplin and Dean [2014] discuss further experiments which also rule out the possibility that a standard random utility model could explain our data.

[24] Assuming that the tie-breaking rule for the case of $\gamma_1 = 0.5$ also does not change as a function of x.

the proportion of correct choices significantly increases with reward.[25]

Consider now the Shannon model described in section 4.3. While this model does allow for an increase in accuracy with rewards, the rate at which this occurs is tightly constrained by equations 7 and 8. Intuitively, a single decision problem is enough to pin down the parameter λ, which then determines behavior in all subsequent decision problems. Caplin and Dean [2013] show that the data from this experiment is not commensurate with this constraint: choice accuracy responds too slowly with respect to reward. This holds true even controlling for the risk aversion of subjects.

In response to the rejection of the Shannon model, Caplin and Dean [2013] introduce a class of alternative cost functions that significantly improve our ability to match the data. This class generalizes the Shannon cost function to the wider class,

$$K^G(\pi) = \lambda \left[G(\mu) - \sum_{\gamma \in \Gamma(\pi)} \left(\sum_{\omega \in \Omega} \pi(\gamma|\omega) \right) G(\gamma) \right],$$

with G itself belonging to a parametric class of strictly convex functions of which the Shannon model is a special case. This class of cost function matches the experimental data significantly better than does the Shannon model. The reason for this is that it allows for different elasticities of attentional effort with respect to incentives. Costs are also naturally formulated in terms of the posterior beliefs associated with an information structure. As for production functions and utility functions, one can anticipate future development of more elaborate functional forms as the behavioral evidence demands.

6 Existing Literature

Many approaches have been taken to modelling information acquisition in economic applications, including sequential search (e.g. McCall [1970]), selection

[25] The fact that choice accuracy does respond to incentives also rules out models in which there is a hard constraint to the amount of information processing instead of a marginal cost (e.g. Sims [2003]).

of the variance of a normal signal (Verrecchia [1982]), the binary choice to either be fully informed or not (Reis [2006]), and rational inattention with information costs based on Shannon mutual information (Sims [2003]). Our approach allows for all of the above costs functions. The costs of feasible attention strategies can be captured by K, while the cost of inadmissible strategies can be set to infinity. The NIAS and NIAC conditions therefore provide a test of the entire class of costly information acquisition models currently in use.

The paper closest in spirit to ours is de Oliveira *et al.* [2013] (henceforth DDMO), which also identifies the behavioral implications of costly information acquisition without making strong assumptions about the form of information costs. Rather than state dependent stochastic choice, DDMO use preference over menus as their evidentiary base. They show in this setting that a model of optimal costly information acquisition is characterized by a preference for flexibility and for early resolution of uncertainty. DDMO show also that a result similar to Theorem 2 holds in this setting, essentially for the same reasons. For example, in both cases, if information structure ρ is more informative than π, the latter will (weakly) never be chosen at any cost $K(\pi) \geq K(\rho)$ and so it is without loss of generality to assume for all such cases that the relationship holds at equality.

We see the two approaches as complementary. The main difference between our work and DDMO involves the underlying data set. We consider only data on patterns of final choice from available actions, without considering how the set of available actions themselves may have been chosen at an earlier stage. In contrast, they consider only preference over menus without considering the resulting patterns of final choice. Our approach therefore focuses directly on patterns of observed mistakes rather than how anticipation of such mistakes impacts choice of menu. An analogy can be drawn with the random utility literature, in which Kreps [1979] and Dekel *et al.* [2001] consider the implications for menu preferences, and Gul and Pesendorfer [2006] for stochastic choice.[26]

[26] The task analogous to that of Ahn and Sarver [2013] - i.e. to understand when both stochastic choice and menu preference can be modelled as coming from the same underlying optimization problem, is an interesting avenue for future work.

Another distinction lies in our focus on conditions which are necessary and sufficient for finite data sets, and theirs on a data set large enough to uniquely identify utilities and prior beliefs as well as costs (though see section 3.5).

Other recent literature has considered the behavioral implications of specific models of information acquisition. Matejka and McKay [2013] and Caplin and Dean [2013] analyze the ramifications of rational inattention with Shannon mutual information costs for state dependent stochastic choice data. Ellis [2012] works with state dependent deterministic choice data to characterize choice among available information partitions. Caplin and Dean [2011] and Caplin *et al.* [2011] consider the case of optimal sequential information search, using an extended data set to derive behavioral restrictions. Again our work nests all these models as special cases.

Our work forms part of a broader effort to characterize choice behavior when the internal information state of the agent is not directly observable. Caplin and Martin [2014] introduce the NIAS condition to characterize subjective rationality in a single decision problem. Manzini and Mariotti [2012] consider a model in which the decision maker has a stochastic consideration set, and makes choices to optimize preferences given what they have paid attention to. Masatlioglu *et al.* [2012] characterize "revealed attention", using the identifying assumption that removing an unattended item from the choice set does not affect attention. Lu [2013] models the stochastic choice of a DM who has some unobserved (but fixed) information structure. Dillenberger *et al.* [2012] consider a dynamic problem in which the DM receives information in each period which is externally unobservable, characterizing the resulting preference over menus. In a strategic setting, Bergemann and Morris [2013b] and Bergemann and Morris [2013a] consider the related problem of identifying all patterns of play that are consistent with some underlying information structure for all players.

In approach, our work is related to the recent resurgence in use of revealed preference methods to understand the observable implications of models of behavior - examples include sequential application of criteria (Manzini and Mariotti [2007]), habit formation (Crawford [2010]), and collective consumption behavior (Cherchye *et al.* [2011]). See also de Clippel and Rozen [2012] for the explicit application of some of these techniques to finite data.

References

S. N. Afriat. The Construction of Utility Functions from Expenditure Data. *International Economic Review*, 8(1), 1967.

David S. Ahn and Todd Sarver. Preference for Flexibility and Random Choice. *Econometrica*, 81(1):341–361, 01 2013.

Timothy K. M. Beatty and Ian A. Crawford. How demanding is the revealed preference approach to demand? *American Economic Review*, 101(6):2782–95, October 2011.

Dirk Bergemann and Stephen Morris. The comparison of information structures in games: Bayes correlated equilibrium and individual sufficiency. Cowles Foundation Discussion Papers 1909, Cowles Foundation for Research in Economics, Yale University, September 2013.

Dirk Bergemann and Stephen Morris. Robust predictions in games with incomplete information. *Econometrica*, 81(4):1251–1308, 07 2013.

David Blackwell. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, 24(2):265–272, 06 1953.

Andrew Caplin and Mark Dean. Search, choice, and revealed preference. *Theoretical Economics*, 6(1), January 2011.

Andrew Caplin and Mark Dean. Behavioral implications of rational inattention with shannon entropy. NBER Working Papers 19318, National Bureau of Economic Research, Inc, August 2013.

Andrew Caplin and Mark Dean. Revealed Preference, Rational Inattention, and Costly Information Acquisition. NBER Working Papers 19876, National Bureau of Economic Research, Inc, January 2014.

Andrew Caplin and Daniel Martin. A testable theory of imperfect perception. *Economic Journal*, forthcoming, 2014.

Andrew Caplin, Mark Dean, and Daniel Martin. Search and satisficing. *American Economic Review*, 101(7):2899–2922, December 2011.

Laurens Cherchye, Bram De Rock, and Frederic Vermeulen. The revealed preference approach to collective consumption behaviour: Testing and sharing rule recovery. *Review of Economic Studies*, 78(1):176–198, 2011.

Anton Cheremukhin, Anna Popova, and Antonella Tutino. Experimental evidence on rational inattention. Technical report, 2011.

Raj Chetty, Adam Looney, and Kory Kroft. Salience and taxation: Theory and evidence. *American Economic Review*, 99(4):1145–77, September 2009.

Ian Crawford. Habits revealed. *Review of Economic Studies*, 77(4):1382–1402, 2010.

Geoffroy de Clippel and Kareen Rozen. Bounded rationality and limited datasets: Testable implications, identifiability, and out-of-sample prediction. Technical report, 2012.

Henrique de Oliveira, Tommaso Denti, Maximilian Mihm, and M. Kemal Ozbek. Rationally inattentive preferences. Ssrn working paper, 2013.

Eddie Dekel, Barton L Lipman, and Aldo Rustichini. Representing Preferences with a Unique Subjective State Space. *Econometrica*, 69(4):891–934, 2001.

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

Andrew Ellis. Foundations for optimal attention. Mimeo, Boston University, 2012.

D. M. Green and J. A. Swets. *Signal detection theory and psychophysics*. Wiley, New York, 1966.

Faruk Gul and Wolfgang Pesendorfer. Random Expected Utility. *Econometrica*, 74(1):121–146, 2006.

F. A. Hayek. The use of knowledge in society. *The American Economic Review*, 35(4):pp. 519–530, 1945.

Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101(6):2590–2615, September 2011.

Tjalling C. Koopmans and Martin Beckmann. Assignment problems and the location of economic activities. *Econometrica*, 25(1):pp. 53–76, 1957.

David Kreps. A Representation Theorem for "Preference for Flexibility". *Econometrica*, 47(3):565–578, 1979.

Nicola Lacetera, Devin G. Pope, and Justin R. Sydnor. Heuristic thinking and limited attention in the car market. *American Economic Review*, 102(5):2206–36, September 2012.

Jay Lu. Random choice and private information. Mimeo, Princeton University, 2013.

Bartosz Adam Mackowiak and Mirko Wiederholt. Business cycle dynamics under rational inattention. CEPR Discussion Papers 7691, C.E.P.R. Discussion Papers, February 2010.

Paola Manzini and Marco Mariotti. Sequentially Rationalizable Choice. *American Economic Review*, pages 1824–1839, 2007.

Paola Manzini and Marco Mariotti. Stochastic choice and consideration sets. IZA Discussion Papers 6905, Institute for the Study of Labor (IZA), October 2012.

Daniel Martin. Strategic pricing with rational inattention to quality. Mimeo, New York University, 2013.

Yusufcan Masatlioglu, Daisuke Nakajima, and Erkut Y. Ozbay. Revealed attention. *American Economic Review*, 102(5):2183–2205, August 2012.

Filip Matejka and Alisdair McKay. Rational inattention to discrete choices: A new foundation for the multinomial logit model. Mimeo, CERGE, 2013.

Filip Matejka. Rationally inattentive seller: Sales and discrete pricing. CERGE-EI Working Papers wp408, The Center for Economic Research and Graduate Education - Economic Institute, Prague, March 2010.

John J McCall. Economics of information and job search. *The Quarterly Journal of Economics*, 84(1):113–26, February 1970.

David J. Murray. A perspective for viewing the history of psychophysics. *Behavioral and Brain Sciences*, 16:115–137, 2 1993.

Ricardo Reis. Inattentive Consumers. *Journal of Monetary Economics*, 53(8):1761–1800, 2006.

Jean-Charles Rochet. A necessary and sufficient condition for rationalizability in a quasi-linear context. *Journal of Mathematical Economics*, 16(2):191–200, April 1987.

R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970.

P. A. Samuelson. A note on the pure theory of consumer's behaviour. *Economica*, 5(17):pp. 61–71, 1938.

Babur De Los Santos, Ali Hortacsu, and Matthijs R. Wildenbeest. Testing models of consumer search using data on web browsing and purchasing behavior. *American Economic Review*, 102(6):2955–80, October 2012.

Christopher Sims. Implications of Rational Inattention. *Journal of Monetary Economics*, 50(3):665–690, 2003.

Christopher A. Sims. Rational inattention: Beyond the linear-quadratic case. *American Economic Review*, 96(2):158–163, May 2006.

George J. Stigler. The economics of information. *Journal of Political Economy*, 69:213, 1961.

Stijn van Nieuwerburgh and Laura Veldkamp. Information immobility and the home bias puzzle. *Journal of Finance*, 64(3):1187–1215, 06 2009.

Hal R Varian. The nonparametric approach to production analysis. *Econometrica*, 52(3):579–97, May 1984.

Robert E Verrecchia. Information acquisition in a noisy rational expectations economy. *Econometrica*, 50(6):1415–30, November 1982.

Michael Woodford. Information-constrained state-dependent pricing. NBER Working Papers 14620, National Bureau of Economic Research, Inc, December 2008.

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

Revealed Preference, Rational Inattention, and Costly Information Acquisition

Andrew Caplin and Mark Dean

Online Appendix: Not for Publication

1 Appendix 1: Proofs

1.1 Lemma 1

Lemma 1 *If $\pi \in \Pi$ is consistent with $P_A \in P$,*[1] *then it is sufficient for $\bar{\pi}_A$.*

Proof. Let $\pi \in \Pi$ be an information structure that is consistent with $P_A \in P$. First, we list in order all distinct posteriors $\eta^i \in \Gamma(\pi)$ for $1 \leq i \leq I = |\Gamma(\pi)|$. Given that π is consistent with P_A, there exists a corresponding optimal choice strategy $C_A : \{1, ..., I\} \to \Delta(A)$, with $C_A(a|i)$ denoting the probability of choosing action $a \in A$ following posterior η^i, such that the information structure and choice functions match the data,

$$P_A(a|\omega) = \sum_{i=1}^{I} \pi(\eta^i|\omega) C_A(a|i).$$

We also list in order all possible posteriors associated with the corresponding revealed information structure, $\gamma^j \in \bar{\Gamma} \equiv \Gamma(\bar{\pi}_A)$, $1 \leq j \leq |\bar{\Gamma}|$, and identify all chosen actions that are associated with posterior γ^j as $\bar{F}^j$,

$$\bar{F}^j \equiv \{a \in A | \bar{\gamma}^a_A = \gamma^j\}.$$

The garbling matrix b^{ij} sets the probability of $\gamma^j \in \bar{\Gamma}$ given $\eta^i \in \Gamma(\pi)$ as the probability of all choices associated with actions $a \in \bar{F}^j$.

$$b^{ij} = \sum_{a \in \bar{F}^j} C_A(a|i).$$

Note that this is indeed a $|\Gamma(\pi)| \times |\bar{\Gamma}|$ stochastic matrix $B \geq 0$ with $\sum_{j=1}^{J} b^{ij} = 1$ all i. Given

[1] i.e. there exists a $C : \Gamma(\pi) \to \Delta(A)$ such that, for each $\gamma \in \Gamma(\pi)$,

$$C(a|\gamma) > 0 \Longrightarrow \sum_{\omega \in \Omega} \gamma_\omega u(a(\omega)) \geq \sum_{\omega \in \Omega} \gamma_\omega u(b(\omega)) \text{ all } b \in A,$$

and for each $\omega \in \Omega$ and $a \in A$,

$$P_A(a|\omega) = \sum_{\gamma \in \Gamma(\pi)} \pi(\gamma|\omega) C(a|\gamma).$$

$\gamma^j \in \Gamma(\pi)$ and $\omega \in \Omega$, note that,

$$\sum_{i=1}^{I} b^{ij}\pi(\eta^i|\omega) = \sum_{i=1}^{I} \pi(\eta^i|\omega) \sum_{a\in\bar{F}^j} C_A(a|i) = \sum_{a\in\bar{F}^j} P_A(a|\omega),$$

by the data matching property. It is definitional that $\bar{\pi}_A(\gamma^j|\omega)$ is precisely equal to this, as the observed probability of all actions associated with posterior $\gamma^j \in \bar{\Gamma}$. Hence,

$$\bar{\pi}_A(\gamma^j|\omega) = \sum_{i=1}^{I} b^{ij}\pi(\eta^i|\omega),$$

as required for sufficiency. ■

1.2 Theorem 1 and 2

Theorem 1 *Given $\mu \in \Gamma$ and $u : X \to \mathbb{R}$, data set (D, P) has a costly information representation if and only if it satisfies NIAS and NIAC.*

Proof of Necessity. Necessity of NIAS follows much as in CM14. Fix $A \in D$, $\pi_A \in \Pi$ and $C_A : \Gamma(\pi_A) \to \Delta(A)$ in a costly information representation, and $a \in Supp(P)$. By definition of a costly information representation,

$$\sum_{\gamma\in\Gamma(\pi_A)} C_A(a|\gamma) \left[\sum_{\omega\in\Omega} \gamma(\omega)\, u(a(\omega))\right] \geq \sum_{\gamma\in\Gamma(\pi_A)} C_A(a|\gamma) \left[\sum_{\omega} \gamma(\omega) u(b(\omega))\right] \text{ all } b \in A.$$

Substituting,

$$\gamma(\omega) = \frac{\mu(\omega)\pi_A(\gamma|\omega)}{\sum_{\upsilon} \mu(\nu)\pi_A(\gamma|\nu)},$$

cancelling the common denominator $\sum_{\upsilon} \mu(\nu)\pi_A(\gamma|\nu)$ in the inequality and substituting $P_A(a|\omega) = \sum_{\gamma\in\Gamma(\pi_A)} \pi_A(\gamma|\omega)C_A(a|\gamma)$, we derive,

$$\begin{aligned} \sum_{\omega} \mu(\omega)P_A(a|\omega)u(a(\omega)) &= \sum_{\gamma\in\Gamma(\pi_A)} C_A(a|\gamma) \left[\sum_{\omega\in\Omega} \gamma(\omega)\, u(a(\omega))\right] \geq \\ \sum_{\gamma\in\Gamma(\pi_A)} C_A(a|\gamma) \left[\sum_{\omega\in\Omega} \gamma(\omega)\, u(b(\omega))\right] &= \sum_{\omega} \mu(\omega)P_A(a|\omega)u(b(\omega)) \end{aligned}$$

establishing necessity of NIAS.

To confirm necessity of NIAC consider any sequence $A_1, A_2, ...A_J \in D$ with $A_J = A_1$ and corresponding information structure π_j for $1 \leq j \leq J$. By optimality,

$$G(A_j, \pi_j) - K(\pi_j) \geq G(A_j, \pi_{j+1}) - K(\pi_{j+1}).\forall\ j \in \{1, .., J\},$$

so that,

$$\sum_{j=1}^{J-1} G(A_j,\pi_j) - K(\pi_j) \geq \sum_{j=1}^{J-1} G(A_j,\pi_{j+1}) - K(\pi_{j+1}).$$

Given that $K(\pi^1) = K(\pi^J)$, note that,

$$\sum_{j=1}^{J-1} G(A_j,\pi_j) - G(A_j,\pi_{j+1}) \geq \sum_{j=1}^{J-1} K(\pi_j) - K(\pi_{j+1}) = 0,$$

so that,

$$\sum_{j=1}^{J-1} G(A_j,\pi_j) \geq \sum_{j=1}^{J-1} G(A_j,\pi_{j+1}).$$

To establish that this is inherited by the revealed information structures $\bar{\pi}_j$ for $1 \leq j \leq J$, note from lemma 1 that with π_j sufficient for $\bar{\pi}_j$, $G(B,\pi_j) \geq G(B,\bar{\pi}_j)$ for all $B \in \mathcal{F}$. For $B = A_j$ this is an equality since both information structures give rise to the same state dependent stochastic demand,

$$G(A_j,\pi_j) = G(A_j,\bar{\pi}_j) = \sum_{a\in A_j}\sum_{\omega} \mu(\omega)P_{A_j}(a|\omega)u(a(\omega)).$$

Hence,

$$\sum_{j=1}^{J-1} G(A_j,\bar{\pi}_j) = \sum_{j=1}^{J-1} G(A_j,\pi_j) \geq \sum_{j=1}^{J-1} G(A_j,\pi_{j+1}) \geq \sum_{j=1}^{J-1} G(A_j,\bar{\pi}_{j+1}),$$

establishing NIAC. ■

Proof of Sufficiency. There are three steps in the proof that the NIAS and NIAC conditions are sufficient for (D,P) to have a costly information representation. The first step is to establish that the NIAC conditions ensures that there is no global reassignment of the revealed information structures observed in the data to decision problems $A \in D$ that raises total gross surplus. The second step is use this observation to define a candidate cost function on information structures, $\bar{K} : \Pi \to \mathbb{R} \cup \infty$. The key is to note that, as the solution to the classical allocation problem of Koopmans and Beckmann [1957], this assignment is supported by "prices" set in expected utility units. It is these prices that define the proposed cost function. The final step is to apply the NIAS conditions to show that $(\bar{K},\bar{\pi})$ represents a costly information representation of (D,P), where $\bar{\pi}$ comprises revealed information structures.

Enumerate decision problems in D as A_j for $1 \leq j \leq J$. Define the corresponding revealed information structures $\bar{\pi}_j$ for $1 \leq j \leq J$ as revealed in the corresponding data and let $\bar{\Pi} \equiv \cup_{j=1}^{J}\bar{\pi}_j$ be the set of all such structures across decision problems, with a slight enrichment to ensure that there are precisely as many structures as there are decision problems. If all revealed information structures are different, the set as just defined will have cardinality J. If there is repetition, then retain the decision problem index with which identical revealed information structures are associated so as to make them distinct. This ensures that the resulting set Π has precisely J elements. Index elements $\bar{\pi}_j \in \bar{\Pi}$ in order of the decision problem A_j with which they are associated.

We now allow for arbitrary matchings of information structures to decision problems. First, let b_{jl} denote the gross utility of decision problem j combined with revealed information structure l,

$$b_{jl} = G(A_j,\bar{\pi}_l),$$

with B the corresponding matrix. Define $\mathcal{M}$ to be the set of all matching functions $\theta : \{1, .., J\} \rightarrow \{1, .., J\}$ that are 1-1 and onto and identify the corresponding sum of gross payoffs,

$$S(\theta) = \sum_{j=1}^{J} b_{j\theta(j)}.$$

It is simple to see that the NIAC condition implies that the identify map $\theta^I(j) = j$ maximizes the sum over all matching functions $\theta \in \mathcal{M}$. Suppose to the contrary that there exists some alternative matching function that achieves a strictly higher sum, and denote this match $\theta^* \in \mathcal{M}$. In this case construct a first sub-cycle as follows: start with the lowest index j_1 such that $\theta^*(j_1) \neq j_1$. Define $\theta^*(j_1) = j_2$ and now find $\theta(j_2)$, noting by construction that $\theta(j_2) \neq j_2$. Given that the domain is finite, this process will terminate after some $K \leq J$ steps with $\theta^*(j_K) = j_1$. If it is the case that $\theta^*(j) = j$ outside of the set $\cup_{k=1}^{K} j_k$, then we know the increase in the value of the sum is associated only with this cycle, hence,

$$\sum_{k=1}^{K-1} b_{j_k j_k} < \sum_{j=1}^{K-1} b_{j_k j_{k+1}},$$

directly in contradiction to NIAC. If this inequality does not hold strictly, then we know that there exists some j' outside of the set $\cup_{k=1}^{K} j_k$ such that $\theta^*(j') \neq j'$. We can therefore iterate the process, knowing that the above strict inequality must be true for at least one such cycle to explain the strict increase in overall gross utility. Hence the identity map $\theta^I(j) = j$ indeed maximizes $S(\theta)$ amongst all matching functions $\theta \in \mathcal{M}$.

With this, we have established that the identity map solves an allocation problem of precisely the form analyzed by Koopmans and Beckmann [1957]. They characterize those matching functions $\theta : \{1, ..J\} \rightarrow \{1, ..J\}$ that maximize the sum of payoffs defined by a square payoff matrix such as B that identifies the reward to matching objects of one set (decision problems in our case) to a corresponding number of objects in a second set (revealed information structures in our case). They show that the solution is the same as that of the linear program obtained by ignoring integer constraints,

$$\max_{x_{jl} \geq 0} \sum_{j,l} b_{jl} x_{jl} \text{ s.t. } \sum_{j=1}^{J} x_{jl} = \sum_{l=1}^{J} x_{jl} = 1.$$

Standard duality theory implies that the optimal assignment $\theta^I(j) = j$ is associated with a system of prices on revealed information structures such that the increment in net payoff from any move of any decision problem is not more than the increment in the cost of the information structure.

Defining these prices as $\bar{K}_j$, their result implies that,

$$b_{jl} - b_{jj} = G(A_j, \bar{\pi}_l) - G(A_j, \bar{\pi}_j) \leq \bar{K}_l - \bar{K}_j;$$

or,

$$G(A_j, \bar{\pi}_j) - \bar{K}_j \geq G(A_j, \bar{\pi}_l) - \bar{K}_l.$$

The result of Koopmans and Beckmann therefore implies existence of a function $\bar{K} : \Pi \longrightarrow \mathbb{R}$ that decentralizes the problem from the viewpoint of the owner of the decision problems, seeking to identify surplus maximizing information structures to match to their particular problems. Note that if there are two distinct decision problems with the same revealed posterior, the result directly implies that they must have the same cost, so that one can in fact ignore the reference to the decision problem and retain only the posterior in the domain. Set $K(\pi) = \infty$ if $\pi \neq \bar{\pi}^A$. We have now completed construction of a qualifying cost function $\bar{K} : \Pi \to \mathbb{R} \cup \infty$ that satisfies $\bar{K}(\pi) \in \mathbb{R}$ for some $\pi \in \Pi$. The entire construction was aimed at ensuring that the observed information structure choices were always maximal, $\bar{\pi}^A \in \hat{\Pi}(K, A)$ for all $A \in D$. It remains to prove that $\bar{\pi}^A$ is consistent with P_A for all $A \in D$. This requires us to show that, for all $A \in D$, the choice rule that associates with each $\gamma \in \Gamma(\bar{\pi}^A)$ the certainty of choosing the associated action $a \in F(A)$ as observed in the data is both optimal and matches the data. That it is optimal is the precise content of the NIAS constraint,

$$\sum_{\omega} \gamma_A^a(\omega)\, u(a(\omega)) \geq \sum_{\omega} \gamma_A^a(\omega)\, u(b(\omega)),$$

for all $b \in A$. That this choice rule and the corresponding revealed information structure match the data holds by construction. ■

1.3 Theorem 2

Theorem 2 *Given $\mu \in \Gamma$ and $u : X \to \mathbb{R}$, data set (D, P) satisfies NIAS and NIAC if and only if it has a costly information representation with conditions K1 to K3 satisfied.*

Proof. The proof of necessity is immediate from theorem 1. The proof of sufficiency proceeds in four steps, starting with a costly information representation $\left(\bar{K}, \bar{\pi}\right)$ of (D, P) of the form produced in theorem 1 based on satisfaction of the NIAS and NIAC conditions. A key feature of this function is that the function $\bar{K}$ is real-valued only on the revealed information structures $\bar{\Pi} \equiv \{\bar{\pi}_A | A \in D\}$ associated with all corresponding decision problems, otherwise being infinite. The first step is the proof is to construct a larger domain $\mathring{\Pi} \supset \bar{\Pi}$ to satisfy three additional properties: to include the inattentive strategy, $I \in \mathring{\Pi}$; to be closed under mixtures so that $\pi, \eta \in \mathring{\Pi}$ and $\alpha \in (0,1)$ implies $\alpha \circ \pi + (1 - \alpha) \circ \eta \in \mathring{\Pi}$; and to be "closed under garbling," so that if $\pi \in \mathring{\Pi}$ is sufficient for information structure $\rho \in \Pi$, then $\rho \in \mathring{\Pi}$. The second step is to define a new function $\mathring{K}$ that preserves the essential elements of $\bar{K}$ while being real-valued on the larger domain $\mathring{\Pi} \supset \bar{\Pi}$, and thereby to construct the full candidate cost function $\mathring{K} : \Pi \to \mathbb{R} \cup \infty$. The third step is to confirm that $\mathring{K} \in \mathcal{K}$ and that $\mathring{K}$ satisfies the required conditions K1 through K3. The final step is to confirm that $\left(\mathring{K}, \bar{\pi}\right)$ forms a costly information representation of (D, P).

We construct the domain $\mathring{\Pi}$ in two stages. First, we define all information structures for which some revealed information structure $\bar{\pi} \in \Pi$ is sufficient;

$$\bar{\Pi}_S = \{\rho \in \Pi | \exists \pi \in \bar{\Pi} \text{ sufficient for } \rho\}.$$

Note that this is a superset of $\bar{\Pi}$ and that it contains I. The second step is to identify $\mathring{\Pi}$ as the smallest mixture set containing $\bar{\Pi}_S$: this is itself a mixture set since the arbitrary intersection of mixture sets is itself a mixture set.

By construction, $\mathring{\Pi}$ has three of the four desired properties: it is closed under mixing; it contains $\bar{\Pi}$, and it contains the inattentive strategy. The only condition that needs to be checked is that it retains the property of being closed under garbling:

$$\pi \in \mathring{\Pi} \text{ sufficient for } \rho \in \Pi \Longrightarrow \rho \in \mathring{\Pi}\ .$$

To establish this, it is useful first to establish certain properties of $\bar{\Pi}_S$ and of $\mathring{\Pi}$. The first is that $\bar{\Pi}_S$ is closed under garbling:

$$\pi \in \bar{\Pi}_S \text{ sufficient for } \rho \in \Pi \Longrightarrow \rho \in \bar{\Pi}_S.$$

Intuitively, this is because the garbling of a garbling is a garbling. In technical terms, the product of the corresponding garbling matrices is itself a garbling matrix. The second is that one can explicitly express $\mathring{\Pi}$ as the set of all finite mixtures of elements of $\bar{\Pi}_S$,

$$\mathring{\Pi} = \left\{ \pi = \sum_{j=1}^{J} \lambda_j \circ \pi_j | J \in \mathbb{N}, (\lambda_1, ..\lambda_J) \in S^{J-1}, \pi_j \in \bar{\Pi}_S \right\},$$

where S^{J-1} is the unit simplex in $\mathbb{R}^{J.}$ To make this identification, note that the set as defined on the RHS certainly contains $\bar{\Pi}_S$ and is a mixture set, hence is a superset of $\mathring{\Pi}$. Note moreover that all elements in the RHS set are necessarily contained in any mixture set containing $\bar{\Pi}_S$ by a process of iteration, making is also a subset of $\mathring{\Pi}$, hence finally one and the same set.

We now establish that if $\rho \in \Pi$ is a garbling of some $\pi \in \mathring{\Pi}$, then indeed $\rho \in \mathring{\Pi}$. The first step is to express $\pi \in \mathring{\Pi}$ as an appropriate convex combination of elements of $\bar{\Pi}_S$ as we now know we can,

$$\pi = \sum_{j=1}^{J} \lambda_j \circ \pi_j.$$

with all weights strictly positive, $\lambda_j > 0$ all j. Lemma 2 below establishes that in this case there exist garblings ρ_j of $\pi_j \in \bar{\Pi}_S$ such that,

$$\rho = \sum_{j=1}^{J} \lambda_j \circ \rho_j,$$

establishing that indeed $\rho \in \mathring{\Pi}$ since, with $\bar{\Pi}_S$ closed under garbling, $\pi_j \in \bar{\Pi}_S$ and ρ_j a garbling of π_j implies $\rho_j \in \bar{\Pi}_S$.

We define the function $\mathring{K}$ on $\mathring{\Pi}$ in three stages. First we define the function $\bar{K}_S$ on the domain $\bar{\Pi}_S$ by identifying for any $\rho \in \bar{\Pi}_S$ the corresponding set of revealed information structures $\bar{\pi} \in \bar{\Pi}$ of which ρ is a garbling, and assigning to it the lowest such cost. Formally, given $\rho \in \bar{\Pi}_S$,

$$\bar{K}_S(\rho) \equiv \min_{\{\pi \in \bar{\Pi} | \pi \text{ sufficient for } \rho\}} \bar{K}(\pi).$$

Note that $\bar{K}_S(\pi) = \bar{K}(\pi)$ all $\pi \in \bar{\Pi}$. To see this, consider $A, A' \in D$ with $\bar{\pi}_{A'}$ sufficient for $\bar{\pi}_A$. By the Blackwell property, expected utility is at least as high using $\bar{\pi}_{A'}$ as using $\bar{\pi}_A$ for which it is sufficient,

$$G(A, \bar{\pi}_{A'}) \geq G(A, \bar{\pi}_A).$$

At the same time, since $(\bar{K}, \bar{\pi})$ is a costly information representation of (D, P), we know that $\bar{\pi}_A \in \hat{\Pi}(K, A)$, so that,

$$G(A, \bar{\pi}_A) - K(\bar{\pi}_A) \geq G(A, \bar{\pi}_{A'}) - K(\bar{\pi}_{A'}).$$

Together these imply that $K(\bar{\pi}_A) \leq K(\bar{\pi}_{A'})$, which in turn implies that $\bar{K}_S(\pi) = \bar{K}(\pi)$ all $\pi \in \bar{\Pi}$.

Note that $\bar{K}_S(\pi)$ also satisfies weak monotonicity on this domain, since if we are given $\rho, \eta \in \bar{\Pi}_S$ with ρ sufficient for η, then we know that any information structure $\pi \in \bar{\Pi}$ that is sufficient for ρ is also sufficient for η, so that the minimum defining $\bar{K}_S(\rho)$ can be no lower than that defining $\bar{K}_S(\eta)$.

The second stage in the construction is extend the domain of the cost function from $\bar{\Pi}_S$ to $\mathring{\Pi}$.

As noted above, this set comprises all finite mixtures of elements of $\bar{\Pi}_S$,

$$\mathring{\Pi} = \left\{ \pi = \sum_{j=1}^{J} \lambda_j * \pi_j | J \in \mathbb{N}, (\lambda_1, ..\lambda_J) \in S^{J-1}, \text{ and } \pi_j \in \bar{\Pi}_S \right\}.$$

Given $\pi \in \mathring{\Pi}$, we take the set of all such mixtures that generate it and define $\mathring{K}(\pi)$ to be the corresponding infimum,

$$\mathring{K}(\pi) = \inf_{\left\{ J \in \mathbb{N}, \lambda \in S^{J-1}, \{\pi^j\}_{j=1}^{J} \in \bar{\Pi}_S | \pi = \sum_{j=1}^{J} \lambda_j * \pi_j \right\}} \sum_{j=1}^{J} \lambda_j \bar{K}_S(\pi_j).$$

Note that this function is well defined since $\bar{K}_S$ is bounded below by the cost of inattentive structures and the feasible set is non-empty by definition of $\mathring{\Pi}$. We establish in Lemma 3 that the infimum is achieved. Hence, given $\pi \in \mathring{\Pi}$, there exists $J \in \mathbb{N}, \lambda \in S^{J-1}$, and elements $\pi_j \in \bar{\Pi}_S$ with $\pi = \sum_{j=1}^{J} \lambda_j \circ \pi_j$ such that,

$$\mathring{K}(\pi) = \sum_{j=1}^{J} \lambda_j \bar{K}_S(\pi_j).$$

We show now that $\mathring{K}$ satisfies K2, mixture feasibility. Consider distinct structures $\pi \neq \eta \in \mathring{\Pi}$. We know by Lemma 3 that we can find $J^{\pi,\eta} \in \mathbb{N}$, corresponding probability weights $\lambda^{\pi,\eta} \in S^{\pi,\eta}$ and elements $\eta_j, \pi_j \in \bar{\Pi}_S$ with $\eta = \sum_{j=1}^{J^\eta} \lambda_j \circ \eta_j$, $\pi = \sum_{j=1}^{J^\pi} \lambda_j * \pi_j$, and such that,

$$\begin{aligned} \mathring{K}(\eta) &= \sum_{j=1}^{J^\eta} \lambda_j^\eta \bar{K}_S(\eta_j); \\ \mathring{K}(\pi) &= \sum_{j=1}^{J^\pi} \lambda_j^\pi \bar{K}_S(\pi_j). \end{aligned}$$

Given $\alpha \in (0, 1)$, consider now the mixture strategy defined by taking each strategy π_j with probability $\alpha\lambda_j^\pi$ and each strategy η_j with probability $(1-\alpha)\lambda_j^\eta$. By construction, this mixture strategy generates $\psi = [\alpha * \pi + (1-\alpha) * \eta] \in \Pi$ and hence we know by the infimum feature of $\mathring{K}(\psi)$ that,

$$\mathring{K}(\psi) \leq \sum_{j=1}^{J^\pi} \alpha\lambda_j^\pi \bar{K}_S(\pi_j) + \sum_{j=1}^{J^\eta} (1-\alpha)\lambda_j^\eta \bar{K}_S(\eta_j) = \alpha\mathring{K}(\pi) + (1-\alpha)\mathring{K}(\eta),$$

confirming mixture feasibility.

We show also that $\mathring{K}$ satisfies K1, weak monotonicity in information. Consider $\pi, \eta \in \mathring{\Pi}$ with π sufficient for η. We know by Lemma 3 that we can find $J \in \mathbb{N}$, $\lambda \in S^{J-1}$, and corresponding elements $\{\pi_j\}_{j=1}^{J} \in \bar{\Pi}_S$ such that $\pi = \sum_{j=1}^{J} \lambda_j * \pi_j$ and such that,

$$\mathring{K}(\pi) = \sum_{j=1}^{J} \lambda_j \bar{K}_S(\pi_j).$$

We know also from Lemma 2 that we can construct $\left\{\eta_j\right\}_{j=1}^{J} \in \bar{\Pi}_S$ such that $\eta = \sum_{j=1}^{J} \lambda_j \circ \eta_j$ and such that each η_j is a garbling of the corresponding π_j. Given that $\bar{K}_S$ satisfies weak monotonicity on its domain $\bar{\Pi}_S$, we conclude that,

$$\bar{K}_S(\pi_j) \geq \bar{K}_S(\eta_j).$$

By the infimum feature of $\mathring{K}(\eta)$ we therefore know that,

$$\mathring{K}(\eta) \leq \sum_{j=1}^{J} \lambda_j \bar{K}_S(\eta_j) \leq \sum_{j=1}^{J} \lambda_j \bar{K}_S(\pi_j) = \mathring{K}(\pi),$$

confirming weak monotonicity.

We show now that we have retained the properties that made $\left(\bar{K}, \bar{\pi}\right)$ a costly information representation of (D, P). Given $A \in D$, it is immediate that $\bar{\pi}$ and the choice function that involves picking action $a \in Supp(P_A)$ for sure in revealed posterior $\bar{\gamma}_A^a$ is consistent with the data, since this was part of the initial definition. What needs to be confirmed is only that the revealed revealed information structures are optimal. Suppose to the contrary that there exists $A \in D$ such that,

$$G(A, \pi) - \mathring{K}(\pi) > G(A, \bar{\pi}_A) - \mathring{K}(\bar{\pi}_A),$$

for some $\pi \in \mathring{\Pi}$. By Lemma 3 we can find $J \in \mathbb{N}$, a strictly positive vector $\lambda \in S^{J-1}$, and corresponding elements $\{\pi_j\}_{j=1}^{J} \in \bar{\Pi}_S$, such that $\pi = \sum_{j=1}^{J} \lambda_j * \pi_j$ and such that,

$$\mathring{K}(\pi) = \sum_{j=1}^{J} \lambda_j \bar{K}_S(\pi_j).$$

By the fact that $\pi = \sum_{j=1}^{J} \lambda_j * \pi_j$ and by construction of the mixture strategy,

$$G(A, \pi) = \sum_{j=1}^{J} \lambda_j G(A, \pi_j),$$

so that,

$$\sum_{j=1}^{J} \lambda_j \left[G(A, \pi_j) - \bar{K}_S(\pi_j)\right] > G(A, \bar{\pi}_A) - \mathring{K}(\bar{\pi}_A).$$

We conclude that there exists j such that,

$$G(A, \pi_j) - \bar{K}_S(\pi_j) > G(A, \bar{\pi}_A) - \mathring{K}(\bar{\pi}_A).$$

Note that each $\pi_j \in \bar{\Pi}_S$ inherits its cost $\bar{K}_S(\pi_j)$ from an element $\bar{\pi}_j \in \bar{\Pi}$ that is the lowest cost revealed information structure according to $\bar{K}$ on set $\bar{\Pi}$ that is sufficient for π_j,

$$\bar{K}_S(\pi_j) = \bar{K}(\bar{\pi}_j),$$

where the last equality stems from the fact (established above) that $\bar{K}_S(\pi) = \bar{K}(\pi)$ on $\bar{\pi} \in \bar{\Pi}$. Note by the Blackwell property that each strategy $\bar{\pi}_j \in \bar{\Pi}$ offers at least as high gross value as the strategy $\pi_j \in \bar{\Pi}_S$ for which it is sufficient, so that overall,

$$G(A, \bar{\pi}_j) - \bar{K}(\bar{\pi}_j) \geq G(A, \pi_j) - \bar{K}_S(\pi_j) > G(A, \bar{\pi}_A) - \mathring{K}(\bar{\pi}_A).$$

To complete the proof it is sufficient to show that,

$$\mathring{K}(\pi) = \bar{K}(\pi),$$

on $\pi \in \bar{\Pi}$. With this we derive the contradiction that,

$$G(A, \bar{\pi}_j) - \bar{K}(\bar{\pi}_j) > G(A, \bar{\pi}_A) - \bar{K}(\bar{\pi}_A),$$

in contradiction to the assumption that $\left(\bar{K}, \bar{\pi}\right)$ formed a costly information representation of (D, P).

To establish that $\mathring{K}(\pi) = \bar{K}(\pi)$ on $\pi \in \bar{\Pi}$, note that we know already that $\bar{K}_S(\pi) = \bar{K}(\pi)$ on $\bar{\pi} \in \bar{\Pi}$. If this did not extend to $\mathring{K}(\pi)$, then we would be able to identify a mixture strategy $\psi \in \bar{\Pi}$ sufficient for $\bar{\pi}_A$ with strictly lower expected costs, $\mathring{K}(\psi) < \mathring{K}(\pi)$. To see that this is not possible, note first from Lemma 1 that all structures that are consistent with A and P_A are sufficient for $\bar{\pi}_A$. Weak monotonicity of $\mathring{K}$ on $\mathring{\Pi}$ then implies that the cost $\mathring{K}(\psi)$ of any mixture strategy sufficient for $\bar{\pi}_A$ satisfies $\mathring{K}(\psi) \geq \mathring{K}(\pi)$, as required.

The final and most trivial stage of the proof is to ensure that normalization (K3) holds. Note that $I \in \bar{\Pi}_S$, so that $\mathring{K}_S(I) \in \mathbb{R}$ according to the rule immediately above. If we renormalize this function by subtracting $\mathring{K}(I)$ from the cost function for all information structures then we impact on no margin of choice and do not interfere with mixture feasibility, weak monotonicity, or whether or not we have a costly information representation. Hence we can avoid pointless complication by assuming that $\mathring{K}(I) = 0$ from the outset so that this normalization is vacuous. In full, we define the candidate cost function $\mathring{K} : \mathring{\Pi} \to \mathbb{R} \cup \infty$ by,

$$\mathring{K}(\pi) = \begin{cases} \mathring{K}(\pi) \text{ if } \pi \in \mathring{\Pi} \\ \infty \text{ if } \pi \notin \mathring{\Pi}. \end{cases}$$

Note that weak monotonicity implies that the function is non-negative on its entire domain.

It is immediate that $\mathring{K} \in \mathcal{K}$, since $\mathring{K}(\pi) = \infty$ for $\pi \notin \mathring{\Pi}$ and the domain contains the corresponding inattentive strategy I on which $\mathring{K}(\pi)$ is real-valued. It is also immediate that $\mathring{K}$ satisfies K3, since $\mathring{K}(I) = 0$ by construction. It also satisfies K1 and K2, and represents a costly information representation, completing the proof. ■

Lemma 2 If $\pi = \sum_1^J \lambda_j \circ \pi_j$ with $J \in \mathbb{N}$, $\lambda \in S^{J-1}$ with $\lambda_j > 0$ all j, and $\{\pi_j\}_{j=1}^J \in \Pi$, then for any garbling ρ of π, there exist garblings ρ_j of $\pi_j \in \Pi$ such that,

$$\rho = \sum_{j=1}^{J} \lambda_j * \rho_j,$$

Proof. By assumption, there exists a $|\Gamma(\pi)| \times |\Gamma(\rho)|$ matrix B with $\sum_k b^{ik} = 1$ all i and such that, for all $\gamma^k \in \Gamma(\rho)$,

$$\rho(\gamma^k|\omega) = \sum_{\eta^i \in \Gamma(\pi)} b^{ik} \pi(\eta^i|\omega).$$

Since $\pi = \sum_{1}^{J} \lambda_j \circ \pi_j$, we know that $\Gamma(\pi^j) \subset \Gamma(\pi)$. Now define compressed matrix B^j as the unique submatrix of B obtained by first deleting all rows corresponding to posteriors $\eta^i \in \Gamma(\pi) \backslash \Gamma(\pi_j)$, and then deleting all columns corresponding to posteriors γ^k such that $b^{ik} = 0$ all $\eta^i \in \Gamma(\pi) \backslash \Gamma(\pi_j)$. Define $\rho_j \in \Pi$ to be the strategy that has as its support the set of all posteriors that are possible given the garbling ρ_j of π_j,

$$\Gamma(\rho_j) = \{\gamma^k \in \Gamma(\rho) | b^{ik} > 0 \text{ some } \eta^i \in \Gamma(\pi_j)\},$$

and in which state dependent probabilities of all posteriors are generated by the compressed matrix B^j,

$$\rho^j_\omega(\gamma^k) = \sum_{\eta^i \in \Gamma(\pi^j)} b^{ik} \pi_i(\eta^i|\omega),$$

for all $\gamma^k \in \Gamma(\rho_j)$.

Note by construction that each information structure ρ_j is a garbling of the corresponding $\pi_j \in \Pi$, since each B^j is itself a garbing matrix for which $\sum_k b^{ik} = 1$ for all $\eta^i \in \Gamma(\pi_j)$. It remains only to verify that $\rho = \sum_{j=1}^{J} \lambda_j * \rho_j$. This follows since,

$$\rho(\gamma^k|\omega) = \sum_{\eta^i \in \Gamma(\pi)} b^{ik} \pi(\eta^i|\omega) = \sum_{\eta^i \in \Gamma(\pi)} b^{ik} \sum_{j=1}^{J} \lambda_j \pi_j(\eta^i|\omega) = \sum_{j=1}^{J} \lambda_j \sum_{\eta^i \in \Gamma(\pi^j)} b^{ik} \pi_j(\eta^i|\omega) = \sum_{j=1}^{J} \lambda_j \rho_j(\gamma^k|\omega).$$

■

Lemma 3 Given $\pi \in \mathring{\Pi}$, there exists $J \in \mathbb{N}$, $\lambda \in S^{J-1}$, and elements $\pi_j \in \bar{\Pi}_S$ with $\pi = \sum_{j=1}^{J} \lambda_j \circ \pi_j$ such that,

$$\mathring{K}(\pi) = \sum_{j=1}^{J} \lambda_j \bar{K}_S(\pi_j).$$

Proof. By definition $\mathring{K}(\pi)$ is the infimum of $\sum_{j=1}^{J} \lambda_j \bar{K}_S(\pi_j)$ over all lists $\{\pi_j\}_{j=1}^{J} \in \bar{\Pi}_S$ such that $\pi = \sum_{j=1}^{J} \lambda_j * \pi_j$. We now consider a sequence of such lists, indicating the order in this sequence in parentheses, $\{\pi_j(n)\}_{j=1}^{J(n)}$, such that in all cases there are corresponding weights $\lambda(n) \in S^{J(n)-1}$ with $\pi = \sum_{j=1}^{J(n)} \lambda_j(n) * \pi_j(n)$ and that achieve a value that is heading in the limit to the infimum,

$$\lim_{n \longrightarrow \infty} \sum_{j=1}^{J(n)} \lambda_j(n) \bar{K}_S(\pi_j(n)) = \mathring{K}(\pi).$$

A first issue that we wish to avoid is limitless growth in the cardinality $J(n)$. The first key observation is that, by Charateodory's theorem, we can reduce the number of strictly positive weights in a convex combination $\pi = \sum_{j=1}^{J^*(n)} \lambda_j^*(n) * \pi_j(n)$ to have cardinality $J^*(n) \leq M+1$. We confirm now that we can do this without raising the corresponding costs, $\sum_{j=1}^{J^*(n)} \lambda_j^*(n) \bar{K}_S(\pi_j(n))$. Suppose that there is some integer n such that the original set of information structures has strictly higher cardinality $J(n) > M+1$. Suppose further that the first selection of $J^1(n) \leq M+1$ such posteriors for which there exists a strictly positive probability weights $\delta_j^1(n)$ such that $\pi = \sum_{j=1}^{J^1(n)} \delta_j^1(n) * \pi_j(n)$ has higher such costs (note WLOG that we are treating these as the first $J^1(n)$ information structures in the original list). It is convenient to define $\delta_j^1(n) = 0$ for $J^1(n) + 1 \leq j \leq J(n)$ so that we can express this inequality in the simplest terms,

$$\sum_{j=1}^{J(n)} \delta_j^1(n) \bar{K}_S(\pi_j(n)) > \sum_{j=1}^{J(n)} \lambda_j(n) \bar{K}_S(\pi_j(n)).$$

This inequality sets up an iteration. We first take the smallest scalar $\alpha^1 \in (0,1)$ such that,

$$\alpha^1 \delta_j^1(n) = \lambda_j(n).$$

That such a scalar exists follows from the fact that $\sum_{j=1}^{J^1(n)} \delta_j^1(n) = \sum_{j=1}^{J(n)} \lambda_j(n) = 1$, with all components in both sums strictly positive and with $J(n) > J^1(n)$. We now define a second set of probability weights $\lambda_j^2(n)$,

$$\lambda_j^2(n) = \frac{\lambda_j(n) - \alpha^1 \delta_j^1(n)}{1 - \alpha^1},$$

for $1 \leq j \leq J(n)$. Note that these weights have the property that $\pi = \sum_{j=1}^{J(n)} \lambda_j^2(n) * \pi_j(n)$ and that,

$$\sum_{j=1}^{J(n)} \lambda_j^2(n) \bar{K}_S(\pi_j(n)) = \sum_{j=1}^{J(n)} \left[\frac{\lambda_j(n) - \alpha^1 \delta_j^1(n)}{1 - \alpha^1} \right] \bar{K}_S(\pi_j(n)) < \sum_{j=1}^{J(n)} \lambda_j(n) \bar{K}_S(\pi_k(n)).$$

By construction, note that we have reduced the number of strictly positive weights $\lambda_j^2(n)$ by at least one to $J(n)-1$ or less. Iterating the process establishes that indeed there exists a set of no more than $M+1$ posteriors such that a mixture produces that first strategy π and in which this mixture has no higher weighted average costs than the original strategy. Given this, there is no loss of generality in assuming that $J(n) \leq M+1$ in our original sequence.

With this bound on cardinality, we know that we can find a subsequence of information structures $\pi_j(n)$ which all have precisely the same cardinality $J(n) = J \leq M+1$ all n. Going further, we can impose properties on all of the J corresponding sequences $\{\pi_j(n)\}_{n=1}^{\infty}$. First, we can select subsequences in which the ranges of all corresponding information structures have the same cardinality independent of n,

$$|\Gamma(\pi_j(n))| = K^j,$$

for $1 \leq j \leq J$. Note we can do this because, for all j and n, the number of posteriors in the information structure $\pi_j(n)$ is bounded above by the number of posteriors in the strategy π, which is finite.

With this, we can index the possible posteriors $\gamma^{jk}(n) \in \Gamma(\pi_j(n))$ in order, $1 \leq k \leq K^j$ and then select further subsequences in which these posteriors themselves converge to limit posteriors,

$$\gamma^{jk}(L) = \lim_{n\to\infty} \gamma^{jk}(n) \in \Gamma.$$

which is possible posteriors lie in a compact set, and so have a convergent subsequence.

We ensure also that both the associated state dependent probabilities themselves and the weights $\lambda_j(n)$ in the expression $\pi = \sum_{j=1}^{J(n)} \lambda_j(n) * \pi_j(n)$ converge,

$$\begin{aligned} \lim_{n\to\infty} \pi\left(\gamma^{jk}(n)|\omega\right) &= \pi_{jk}(L|\omega); \\ \lim_{n\to\infty} \lambda_j(n) &= \lambda_j(L). \end{aligned}$$

Again, this is possible because the state dependent probabilities and weights lie in compact sets.

The final selection of a subsequence ensures that, given $1 \leq j \leq J$, each $\pi_j(n) \in \bar{\Pi}_S$ has its value defined by precisely the same revealed information structure $\bar{\pi}_j \in \bar{\Pi}$ as the least expensive among those that were sufficient for it and hence whose cost it was assigned in function $\bar{K}_S$. Technically, for each $1 \leq j \leq J$,

$$\bar{K}_S(\pi_j(n)) = \bar{K}(\bar{\pi}_j),$$

for $1 \leq n \leq \infty$: this is possible because the data set and hence the number of revealed information structures is finite.

We first use these limit properties to construct a list of limit information structures $\pi_j(L) \in \bar{\Pi}_S$ with $\pi = \sum_{j=1}^{J} \lambda_j \circ \pi_j$ for $1 \leq j \leq J$. Strategy $\pi_j(L)$ has range,

$$\Gamma(\pi_j(L)) = \cup_{k=1}^{K^j} \gamma^{jk}(L),$$

with state dependent probabilities,

$$[\pi_j(L)]_\omega \left(\gamma^{jk}(L)\right) = \pi_{ik}(L|\omega).$$

Note that the construction ensures that $\pi = \sum_{j=1}^{J} \lambda_j(L) \circ \pi_j(L)$. To complete the proof we must establish only that,

$$\mathring{K}(\pi) = \sum_{j=1}^{J} \lambda_j(L) \bar{K}_S(\pi_j(L)).$$

We know from the construction that, for each n,

$$\sum_{j=1}^{J} \lambda_j(n) \bar{K}_S(\pi_j(n)) = \sum_{j=1}^{J} \lambda_j(n) \bar{K}(\bar{\pi}_j).$$

Hence the result is established provided only,

$$\bar{K}_S(\pi_j(L)) \leq \bar{K}(\bar{\pi}_j),$$

which is true provided $\bar{\pi}_j$ being sufficient for all $\pi_j(n)$ implies that $\bar{\pi}_j$ is sufficient for the corresponding limit vector $\pi^j(L)$. That this is so follows by defining $B^j(L) = [b^{ik}(L)]^j$ to be the limit of any subsequence of the $|\Gamma(\bar{\pi}_j)| \times K^j$ stochastic matrices $B^j(n) = [b^{ik}(n)]^j$ which have the defining property of sufficiency,

$$[\pi_j(n)]_\omega (\gamma^{jk}(n)) = \sum_{\bar{\gamma}^i \in \Gamma(\bar{\pi}_j)} [b^{ik}(n)]^j * \bar{\pi}(\bar{\gamma}^i|\omega),$$

for all $\gamma^{jk}(n) \in \Gamma(\pi_j(n))$ and $\omega \in \Omega$. It is immediate that the equality holds up in the limit, establishing that indeed $\bar{\pi}_j$ is sufficient for each corresponding limit vector $\pi_j(L)$, confirming finally that $\bar{K}_S(\pi_j(L)) \leq \bar{K}(\bar{\pi}_j)$ and with it establishing the Lemma. ■

2 Appendix 2: No Strong Blackwell

A simple example with data on one decision problem with two equally likely states illustrates that one cannot further strengthen the result in this dimension. Suppose that there are three available actions $A = \{a, b, c\}$ with corresponding utilities,

$$(u(a(\omega_1)), u(a(\omega_2))) = (10, 0)\,; (u(b(\omega_1)), u(b(\omega_2))) = (0, 10)\,; (u(c(\omega_1)), u(c(\omega_2))) = (7.5, 7.5)\,.$$

Consider the following state dependent stochastic choice data in which the only two chosen actions are a and b,

$$P(a|\omega_1) = P(b|\omega_2) = \frac{3}{4} = 1 - P(b|\omega_1) = 1 - P(a|\omega_1).$$

Note that this data satisfies NIAS; given posterior beliefs when a is chosen, a is superior to b and indifferent to c, and when b is chosen it is superior to a and indifferent to c. It trivially satisfies NIAC since there is only one decision problem observed. We know from theorem 2 that is has a costly information representation with the cost of the revealed information structure $K(\bar{\pi}) \geq 0$ and that of the inattentive strategy being zero, $K(I) = 0$. Note that $\bar{\pi}$ is sufficient for I but not vice versa, hence any strictly monotone cost function would have to satisfy $K(\bar{\pi}) > 0$. In fact it is not possible to find a representation with this property. To see this, note that both structures have the same gross utility,

$$G(A, \pi) = \frac{1}{2} * \frac{3}{4} * 10 + \frac{1}{2} * \frac{3}{4} * 10 = 1 * 7.5 = G(A, I),$$

where we use the fact that the inattentive strategy involves picking action c for sure. In order to rationalize selection of the inattentive strategy, it must therefore be that $\bar{\pi}$ is no more expensive than I, contradicting strict monotonicity.

3 Appendix 3: Further Details of NIAS and NIAC Tests

Figures s1 and s2 show the subject level distribution of losses due to NIAC and NIAS violations in dollar terms, compared to a benchmark simulated distribution of random choice. Losses due to NIAC are calculated assuming that the point estimate of posterior beliefs upon the the choice of each act are the subject's true posterior beliefs, and then comparing the expected value of the chosen act to that of the optimal act at each posterior. Losses below are summed across all chosen acts in all decision problems. NIAS losses are calculated by treating each subject's estimated choice probabilites in each decision problem as their true choice probabilities, and calculating the maximal surplus that could be obtained by correctly assigning information structures to decision problems. This is compared to the surplus obtained from the subject's actual assignment, assuming NIAS to be satisfied.

In both cases, the distributions of actual and simulated subjects are significantly different at the 0.001% level (Kolmogorov-Smirnov test).

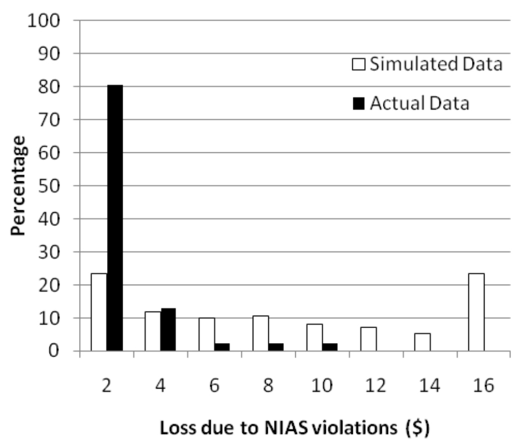


Figure S1: NIAS Losses

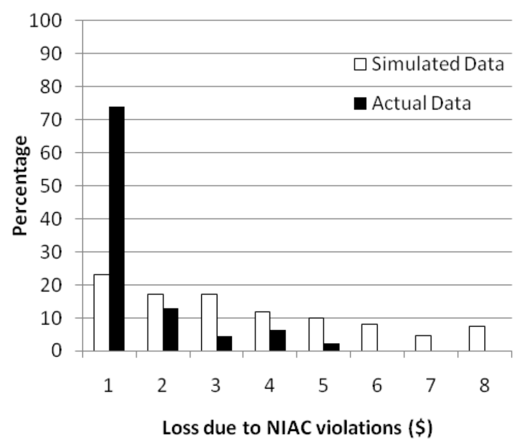


Figure S2: NIAC losses

References

Tjalling C. Koopmans and Martin Beckmann. Assignment problems and the location of economic activities. *Econometrica*, 25(1):pp. 53–76, 1957.

Appendix 4: Example Instructions

Individual Decision-Making Experiment

Instructions

This experiment is designed to study decision making, and consists of 4 sections. Each section will consist of 50 questions. At the end of the experiment, one question will be selected at random from those you answered. The amount of money that you get at the end of the experiment will depend on your answer to this question. Anything you earn from this part of this experiment will be added to your show-up fee of $10.

Please turn off cellular phones now.

The entire session will take place through your computer terminal. Please do not talk or in any way communicate with other participants during the session.

Please **do NOT use the forward and back buttons in your browser** to navigate. Only use the links at the bottom of each page to move forward or back.

We will start with a brief instruction period. During this instruction period, you will be given a description of the main features of the session and will be shown how to use the program. If you have any questions during this period, please raise your hand.

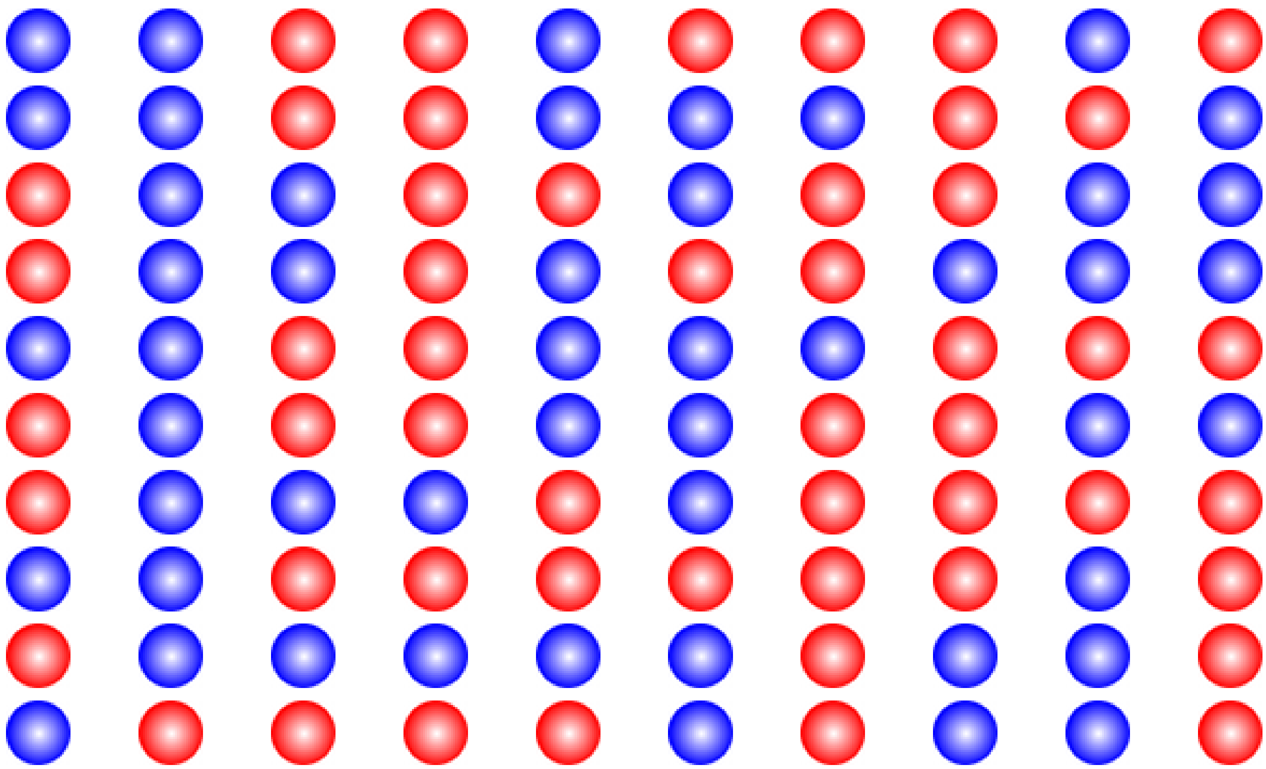
After you have completed the experiment, **please remain quietly seated until everyone has completed the experiment**.

← Home Next →

Individual Decision-Making Experiment

Instructions

For each question you will be shown 100 dots on a screen. Some of these dots will be red, while some will be blue. Here is an example of such a screen:



The number of red dots will be determined at random. You will be told how likely each number of red dots is. So, for example you might be told that there is a 75% chance of there being 49 red dots and a 25% chance of there being 51 red dots. In this case there is a 3/4 chance that there will be 49 red dots on the screen, and a 1/4 chance that there will be 51 red dots. There will never be any other number of red dots on the screen. The number of red dots in each question is determined independently of the number of red dots that have appeared in previous questions.

← Previous Next →

Individual Decision-Making Experiment

Instructions

You will be asked to make a choice between two or more options. Each of these options will pay out different amounts of money, depending on how many red dots are on the screen.

	Option	Pay if there are 49 red dots	Pay if there are 51 red dots
○	A	10	0
○	B	0	10
○	C	5	5

In this case, if you chose option A (and this question was the one selected for payment) then you would get $10 if there were 49 red dots on the screen and $0 if there were 51 red dots. If you chose option B you would get $10 if there were 51 red dots on the screen and $0 if there were 49 red dots. If you chose option C you would receive $5 regardless of the number of red dots on the screen.

You will now have the chance to try an example question. You will not be paid depending on your answer to this question - it is just for practice.

← Previous Next →

Individual Decision-Making Experiment

Instructions

Example Question

You are about to see a screen with 100 dots on it. These dots will be either red or blue. The likelihood of the number of red dots is as follows:

- With 50% probability there will be 49 red dots
- With 50% probability there will be 51 red dots

You will then be asked to choose between a number of alternatives. These alternatives will pay money depending on the number of dots on the screen.

← Previous Next →

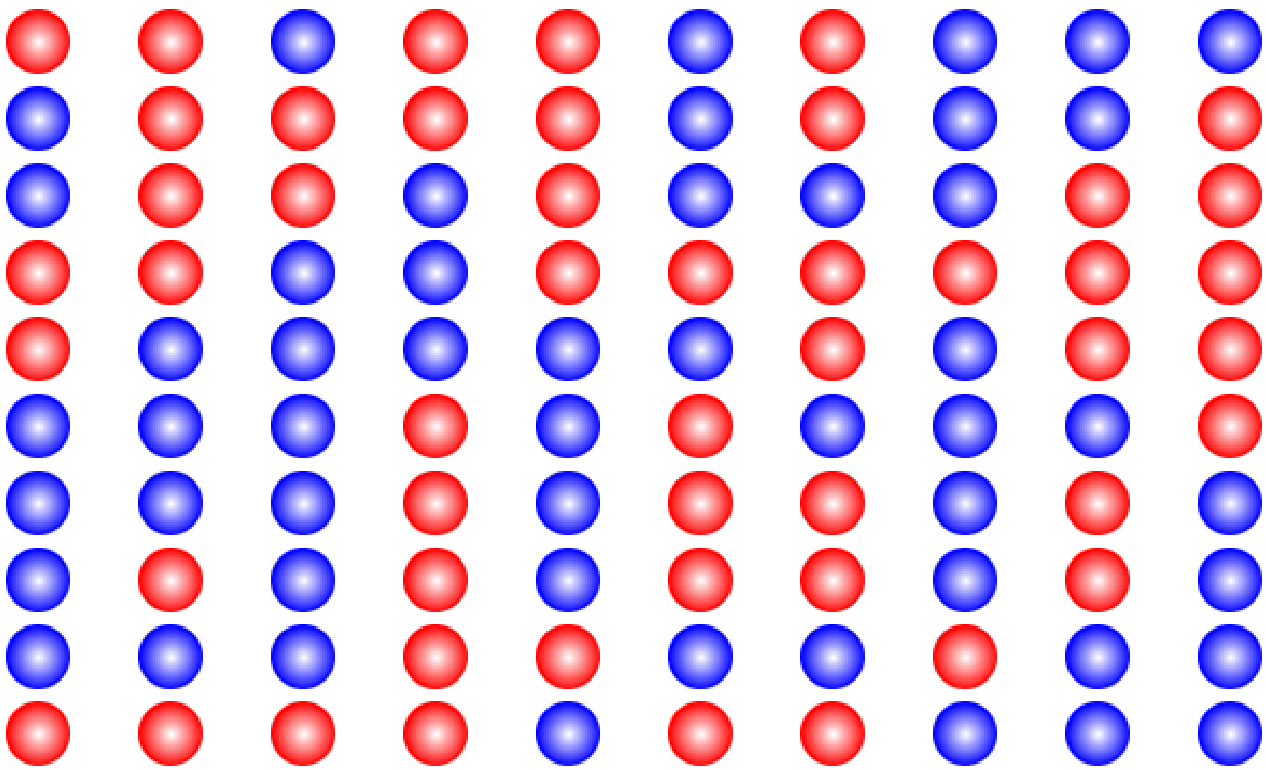
Individual Decision-Making Experiment

Instructions

Example Question

Remember:

- With 50% probability there will be 49 red dots
- With 50% probability there will be 51 red dots



Please select from the following options:

	Option	Pay if there are 49 red dots	Pay if there are 51 red dots
	A	10	0
	B	0	10
	C	5	5

← Previous Next →

Individual Decision-Making Experiment

Instructions

Payment

For this question, you chose the following option:

	Option	Pay if there are 49 red dots	Pay if there are 51 red dots
◉	A	10	0

There were 49 red dots on the screen.

If this were the question that had been selected for payment, you would have received $10 in addition to your show up fee.

← Previous Next →

Individual Decision-Making Experiment

Instructions

Here is a description of the questions that you will face in each of the 4 sections of the experiment.

Block 1

- With 50% probability there will be 49 red dots
- With 50% probability there will be 51 red dots

You will be asked to choose between the following options:

	Option	Pay if there are 49 red dots	Pay if there are 51 red dots
○	A	23	23
○	B	21	25
○	C	40	0

Block 2

- With 50% probability there will be 49 red dots
- With 50% probability there will be 51 red dots

You will be asked to choose between the following options:

	Option	Pay if there are 49 red dots	Pay if there are 51 red dots
○	A	23	23
○	B	21	25

Block 3

- With 50% probability there will be 49 red dots
- With 50% probability there will be 51 red dots

You will be asked to choose between the following options:

	Option	Pay if there are 49 red dots	Pay if there are 51 red dots
○	A	23	23
○	B	21	25
○	C	35	5

Block 4

- With 50% probability there will be 49 red dots
- With 50% probability there will be 51 red dots

You will be asked to choose between the following options:

	Option	Pay if there are 49 red dots	Pay if there are 51 red dots
○	A	23	23
○	B	21	25
○	C	30	10

REMEMBER: Each section consists of 50 questions, each with the same probabilities and available options. You will be reminded in each question what the probabilities and available options are for that question.

Again, please **do NOT use the forward and back buttons in your browser** to navigate. Only use the links at the bottom of each page to move forward or back.

If you have any questions, please raise your hand now, otherwise click to the lower right to return to the Home Page and begin the experiment.

← Previous   Home →