Revealed Preference, Rational Inattention, and Costly Information Acquisition

By Andrew Caplin and Mark Dean*

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 and identify the extent to which information costs can be recovered from choice data.

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, Looney and Kroft (2009)). Buyers of second-hand cars focus their attention on the leftmost digit of the odometer (Lacetera, Pope and Sydnor (2012)). Purchasers limit their attention to a relatively small number of websites when buying over the internet (Santos, Hortacsu and Wildenbeest (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 the 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*

* Caplin: Center for Experimental Social Science and Department of Economics, New York University, 19 West 4th Street, New York, USA, andrew.caplin@nyu.edu. Dean: Department of Economics, Brown University, 64 Waterman Street, Providence, Rhode Island, USA, mark_dean@brown.edu. 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, Michael Woodford and four anonymous referees 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". Dean acknowledges the financial support of National Science Foundation grant 1156090 and the Economic Theory Center at Princeton University at which he was a visitor while working on this paper. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

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

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 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 (2015), henceforth CM15), 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, Looney and Kroft (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 CM15. 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 III 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 key behavioral properties implied by this cost function, which are significantly more restrictive than NIAS and NIAC alone (see also Caplin and Dean (2013)).

As detailed in section 5, 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 (2015)). Our work also fits into a growing literature aimed at identifying the behavioral implications

[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 (2009), van Nieuwerburgh and Veldkamp (2009) Mackowiak and Wiederholt (2010), Matejka (2010), Martin (2013).

of boundedly rational models in which the information state of the decision maker is unknown (Masatlioglu, Nakajima and Ozbay (2012), Bergemann and Morris (2013*b*), Dillenberger et al. (2014), Manzini and Mariotti (2014)). Analogous revealed preference approaches have recently been applied to various behavioral models of individual and group decision making (Crawford (2010), Cherchye, Rock and Vermeulen (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 reviews related literature.

I. A Costly Information Representation

A. 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 II.F.

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.

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

State dependent stochastic choice data can be readily gathered in the laboratory (see for example Caplin and Dean (2014)). 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.

EXAMPLE

We illustrate the concept of state dependent stochastic choice data with an example which we will use throughout the paper (and which forms the basis of the experimental tests reported in Caplin and Dean (2014)). Risk neutral[6] subjects are faced with a screen on which there are 100 balls, each of which may be either red or blue. 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 the empirical frequency with which the subject chooses each action in each state. These frequencies provide an estimate of P_A, the state dependent stochastic choice data for set A.

B. 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

[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.

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 $K : \Pi \to \bar{\mathbb{R}}$ 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)).[7] 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 $G : \mathcal{F} \times \Pi \to \mathbb{R}$ 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.

[7] See section III.C for further details of this model.

C. 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.[8]

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 $C_A : \Gamma(\pi_A) \to \Delta(A)$ is such that, given $a \in A$ and $\gamma \in \Gamma(\pi_A)$ with $C_A(a|\gamma) \equiv \Pr(a|\gamma) > 0$,*
$$\sum_{\omega\in\Omega} \gamma(\omega) u(a(\omega)) \geq \sum_{\omega\in\Omega} \gamma(\omega) u(b(\omega)) \text{ all } b \in A.$$
3) *The data is matched: given $\omega \in \Omega$ and $a \in A$,*
$$P_A(a|\omega) = \sum_{\gamma\in\Gamma(\pi_A)} \pi_A(\gamma|\omega) C_A(a|\gamma).$$

II. 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.

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

A. 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 action. 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 action is chosen.[9] If they choose 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 $\mu \in \Gamma$, $A \in \mathcal{D}$, $P_A \in \mathcal{P}$, and $a \in Supp(P_A)$, the* ***revealed posterior*** $\bar{\gamma}^a_A \in \Gamma$ *is defined by,*

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

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 $\mu \in \Gamma$, $A \in \mathcal{D}$, and $P_A \in \mathcal{P}$, 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 I.B, their revealed information structure may not be the same as their true information structure if they choose the same action following two different signals.[10] However, it must be

[9] Note that 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 true probability.

[10] 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.

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).

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 $\mathbf{B} \geq 0$ 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 information structure ρ is sufficient for information structure π if π can be obtained by applying a stochastic matrix $\mathbf{B}$ 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 $P_A \in P$,*[11] *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.

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

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

[11] i.e. there exists $C : \Gamma(\pi) \to \Delta(A)$ 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 $a \in A$,

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

The conditional probabilities of receiving each signal are:

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

Suppose that after the receipt of signal α the DM chooses action a for sure and after the receipt of β they choose action b for sure. After the receipt of signal δ they randomize between the two actions:

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

This behavior gives rise to state dependent stochastic choice data:

$$\begin{aligned}P_A(a|\omega_1) &= \pi_A(\alpha|\omega_1) + \frac{1}{2}\pi_A(\delta|\omega_1) = \frac{2}{3};\\ P_A(a|\omega_2) &= \pi_A(\alpha|\omega_2) + \frac{1}{2}\pi_A(\delta|\omega_2) = \frac{1}{3};\end{aligned}$$

The resulting revealed information structure has two revealed posteriors, one associated with the choice of a and the other with the choice of b:

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

The corresponding revealed information structure is then given by

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

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 $\mathbf{B}$ *to to π_A, in order to obtain $\bar{\pi}_A$,*

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

The rows relate to signals α, β and δ respectively, while the columns relate to signals

$\bar{\gamma}^{a}_{A}$ and $\bar{\gamma}^{b}_{A}$ respectively. Thus, for example

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

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

B. 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 two state, two action decision problem of section I.A in which the DM earns 10 for choosing action a in state ω_1 or b in state ω_2 and zero otherwise. 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 $A' = \{a', b'\}$ in which the DM earns 2 for choosing the "correct" action and zero otherwise, so $u(a'(\omega_1)) = u(b'(\omega_2)) = 2$ 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 a costly 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,

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

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 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 $A^J = A^1$,

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

In section II.E we demonstrate the application of the NIAC condition to the simple case of two actions and two states.

C. 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. CM15 show that this condition characterizes Bayesian behavior regardless of the rationality of attentional choice. The strategic analog is derived by Bergemann and Morris (2013*b*) 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 $b \in A$,

$$\sum_{\omega \in \Omega} \mu(\omega) P_A(a|\omega) \left(u(a(\omega)) - u(b(\omega))\right) \geq 0.$$

Section II.E contains a simple application of NIAS to the two state, two action case. CM15 contains many further illustrative examples.

D. 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.*

E. 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. 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 I.A). 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,

$$P_A(a|\omega_1) \geq \max\{\alpha P_A(a|\omega_2), \alpha P_A(a|\omega_2) + \beta\}, \tag{4}$$

where

$$\begin{aligned} \alpha &= \frac{u(b(\omega_2)) - u(a(\omega_2))}{u(a(\omega_1)) - u(b(\omega_1))} \\ \beta &= \frac{u(a(\omega_1)) + u(a(\omega_2)) - u(b(\omega_1)) - u(b(\omega_2))}{(a(\omega_1)) - u(b(\omega_1))} \end{aligned}$$

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 $A^i = \left\{a^i, b^i\right\}$ 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,

(5)

$$\Delta P(a|\omega_1)\left(\Delta\left(u(a(\omega_1)) - u(b(\omega_1))\right)\right) + \Delta P(b|\omega_2)\left(\Delta\left(u(b(\omega_2)) - u(a(\omega_2))\right)\right) \geq 0,$$

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.

F. 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).[12] If we further add

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

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).[13]

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 CM15 for the implications of NIAS alone). If the prior is also unknown, then the conditions are non-linear, but still non-vacuous. CM15 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.

III. 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.

A. 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, Rock and Vermeulen (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 $\pi \in \Pi$,

$$K(\bar{\pi}_A) - K(\pi) \leq G(A, \bar{\pi}_A) - G(A, \pi). \tag{6}$$

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 $A,B \in D$,

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

and the costs of unchosen information structures are high enough to satisfy inequality 6.

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 $A^n = B$,

$$\begin{aligned} K(\bar{\pi}_{A^1}) - K(\bar{\pi}_{A^2}) &\leq G(A^1, \bar{\pi}_{A^1}) - G(A^1, \bar{\pi}_{A^2}); \\ K(\bar{\pi}_{A^2}) - K(\bar{\pi}_{A^3}) &\leq G(A^2, \bar{\pi}_{A^2}) - G(A^2, \bar{\pi}_{A^3}); \\ &\vdots \\ K(\bar{\pi}_{A^{n-1}}) - K(\bar{\pi}_{A^n}) &\leq G(A^{n-1}, \bar{\pi}_{A^{n-1}}) - G(A^{n-1}, \bar{\pi}_{A^n}). \end{aligned}$$

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,

$$K(\bar{\pi}_A) - K(\bar{\pi}_B) \leq \min_{\{A^1 ... A^n \in D | A^1 = A, A^n = B\}} \sum_{i=1}^{n-1} \left[G(A^i, \bar{\pi}_{A^i}) - G(A^i, \bar{\pi}_{A^{i+1}}) \right]. \tag{7}$$

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

$$K(\bar{\pi}_A) - K(\bar{\pi}_B) \geq \max_{\{A^1 ... A^n \in D | A^1 = B, A^n = A\}} \sum_{i=1}^{n-1} \left[G(A^i, \bar{\pi}_{A^{i+1}}) - G(A^i, \bar{\pi}_{A^i}) \right]. \tag{8}$$

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.

B. *Untestable 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 the cost of the information structure $\alpha \circ \pi + (1-\alpha) \circ \eta$ 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 $\left(K, \{\pi_A\}_{A \in D}, \{C_A\}_{A \in D}\right)$ of the form produced in the proof of 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 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, strictly 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.

C. The Shannon Cost Function

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

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

Here $H(\mu) = -\sum_{\omega \in \Omega} \mu(\omega) \ln \mu(\omega)$ is the Shannon entropy function[14] 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.

Caplin and Dean (2013) and Matejka and McKay (2015) characterize the pattern of state dependent stochastic choice associated with the model. An "invariant likelihood ratio" (ILR) condition relates revealed posteriors to the utilities of chosen actions. For any $A \in D$ and $a, b \in Supp(P_A)$:

$$\frac{\bar{\gamma}_A^a(\omega)}{\exp(u(a(\omega))/\lambda)} = \frac{\bar{\gamma}_A^b(\omega)}{\exp(u(b(\omega))/\lambda)}. \tag{9}$$

A further complementary slackness condition identifies the set of actions which are chosen with positive probability. For any $A \in D$ and $a \in Supp(P_A)$, $b \notin Supp(P_A)$:

$$\sum_{\omega \in \Omega} \left[\frac{\bar{\gamma}_A^a(\omega)}{\exp(u(a(\omega))/\lambda)} \right] \exp(u(b(\omega))/\lambda) \leq 1. \tag{10}$$

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

Clearly equations 9 and 10 impose significant behavioral restrictions in addition to NIAS and NIAC, as discussed in Caplin and Dean (2013). The ILR condition pins down the rate at which a DM's choice accuracy responds to incentives for making the correct decision. 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.

IV. Existing Literature

Many approaches have been taken to modelling information acquisition in economic applications, including sequential search (e.g. McCall (1970)), selection 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, Lipman and Rustichini (2001) consider the implications for menu preferences, and Gul and Pesendorfer (2006) for stochastic choice.[15] Another distinction lies in our focus on conditions which are necessary and sufficient for finite data sets, and theirs on a data set rich enough to uniquely identify utilities and prior beliefs as well as costs (though see section II.F).

[15] 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.

Other recent literature has considered the behavioral implications of specific models of information acquisition. Caplin and Dean (2013) and Matejka and McKay (2015) 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, Dean and Martin (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 (2015) introduce the NIAS condition to characterize subjective rationality in a single decision problem. Manzini and Mariotti (2014) 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, Nakajima and Ozbay (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. (2014) 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 (2013*a*) and Bergemann and Morris (2013*b*) 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, Rock and Vermeulen (2011)). See also de Clippel and Rozen (2012) for the explicit application of some of these techniques to finite data.

REFERENCES

Afriat, Sydney N. 1967. "The Construction of Utility Functions from Expenditure Data." *International Economic Review*, 8(1): 67–77.

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

Bergemann, Dirk, and Stephen Morris. 2013*a*. "The Comparison of Information Structures in Games: Bayes Correlated Equilibrium and Individual Sufficiency." Cowles Foundation for Research in Economics, Yale University Cowles Foundation Discussion Papers 1909.

Bergemann, Dirk, and Stephen Morris. 2013*b*. "Robust Predictions in Games With Incomplete Information." *Econometrica*, 81(4): 1251–1308.

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

Caplin, Andrew, and Daniel Martin. 2015. "A Testable Theory of Imperfect Perception." *The Economic Journal*, 125(582): 184–202.

Caplin, Andrew, and Mark Dean. 2011. "Search, choice, and revealed preference." *Theoretical Economics*, 6(1): 19–48.

Caplin, Andrew, and Mark Dean. 2013. "Behavioral Implications of Rational Inattention with Shannon Entropy." National Bureau of Economic Research, Inc NBER Working Papers 19318.

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

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

Cherchye, Laurens, Bram De Rock, and Frederic Vermeulen. 2011. "The Revealed Preference Approach to Collective Consumption Behaviour: Testing and Sharing Rule Recovery." *Review of Economic Studies*, 78(1): 176–198.

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

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

de Clippel, Geoffroy, and Kareen Rozen. 2012. "Bounded Rationality and Limited Datasets: Testable Implications, Identifiability, and Out-of-Sample Prediction."

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

de Oliveira, Henrique, Tommaso Denti, Maximilian Mihm, and M. Kemal Ozbek. 2013. "Rationally Inattentive Preferences." SSRN Working Paper.

Dillenberger, David, Juan Sebastian Lleras, Philipp Sadowski, and Norio Takeoka. 2014. "A theory of subjective learning." *Journal of Economic Theory*, 153(0): 287 – 312.

Ellis, Andrew. 2012. "Foundations for Optimal Attention." Boston University Mimeo.

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

Hayek, Friedrich A. 1945. "The Use of Knowledge in Society." *The American Economic Review*, 35(4): 519–530.

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

Koopmans, Tjalling C., and Martin Beckmann. 1957. “Assignment Problems and the Location of Economic Activities.” *Econometrica*, 25(1): 53–76.

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

Lacetera, Nicola, Devin G. Pope, and Justin R. Sydnor. 2012. “Heuristic Thinking and Limited Attention in the Car Market.” *American Economic Review*, 102(5): 2206–36.

Lu, Jay. 2013. “Random Choice and Private Information.” Princeton University Mimeo.

Mackowiak, Bartosz Adam, and Mirko Wiederholt. 2010. “Business Cycle Dynamics under Rational Inattention.” C.E.P.R. Discussion Papers CEPR Discussion Papers 7691.

Manzini, Paola, and Marco Mariotti. 2007. “Sequentially Rationalizable Choice.” *American Economic Review*, 97(5): 1824–1839.

Manzini, Paola, and Marco Mariotti. 2014. “Stochastic Choice and Consideration Sets.” *Econometrica*, 82(3): 1153–1176.

Martin, Daniel. 2013. “Strategic Pricing with Rational Inattention to Quality.” New York University Mimeo.

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

Matejka, Filip. 2010. “Rationally Inattentive Seller: Sales and Discrete Pricing.” The Center for Economic Research and Graduate Education - Economic Institute, Prague CERGE-EI Working Papers wp408.

Matejka, Filip, and Alisdair McKay. 2015. “Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model.” *American Economic Review*, 105(1): 272–98.

McCall, John J. 1970. “Economics of Information and Job Search.” *The Quarterly Journal of Economics*, 84(1): 113–26.

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

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

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

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

Samuelson, Paul A. 1938. "A Note on the Pure Theory of Consumer's Behaviour." *Economica*, 5(17): 61–71.

Santos, Babur De Los, Ali Hortacsu, and Matthijs R. Wildenbeest. 2012. "Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior." *American Economic Review*, 102(6): 2955–80.

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

Sims, Christopher A. 2006. "Rational Inattention: Beyond the Linear-Quadratic Case." *American Economic Review*, 96(2): 158–163.

Stigler, George J. 1961. "The Economics of Information." *Journal of Political Economy*, 69(3): 213–225.

van Nieuwerburgh, Stijn, and Laura Veldkamp. 2009. "Information Immobility and the Home Bias Puzzle." *Journal of Finance*, 64(3): 1187–1215.

Varian, Hal R. 1984. "The Nonparametric Approach to Production Analysis." *Econometrica*, 52(3): 579–97.

Verrecchia, Robert E. 1982. "Information Acquisition in a Noisy Rational Expectations Economy." *Econometrica*, 50(6): 1415–30.

Woodford, Michael. 2009. "Information Constrained State Dependent Pricing." *Journal of Monetary Economics*, 56(S): S100–S124.

Woodford, Michael. 2012. "Inattentive Valuation and Reference-Dependent Choice." Columbia University Mimeo.