Revealed Preference, Rational Inattention, and Costly Information Acquisition*

Andrew Caplin†and Mark Dean‡

January 2014

Abstract

We develop a revealed preference test for optimal acquisition of costly information. The test encompasses models of rational inattention, sequential signal processing, and search. We provide limits on the extent to which attention costs can be recovered from choice data. We experimentally elicit state dependent stochastic choice data of the form the tests require. In simple cases, tests confirm that subjects adjust their attention in response to incentives as the theory dictates.

1 Introduction

Modelling behavior when information is costly to acquire has been central to economic analysis since the seminal work of Stigler [1961]. As the importance

*We thank Roland Benabou, Dirk Bergemann, Laurens Cherchye, Bram De Rock, Thomas Demuynck, Federico Echenique, Andrew Ellis, Paola Manzini, Marco Mariotti, Daniel Martin, Filip Matejka, Alisdair McKay, Stephen Morris, Pietro Ortoleva, 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

of information constraints have been increasingly recognized,[1] so an ever wider array of information gathering technologies have been modelled. For example, McCall [1970] considered the case of sequential search; Verrecchia [1982] the choice of variance of a normal signal; and Sims [2003] an unrestricted choice of information structure with costs based on Shannon entropy. Many other alternatives have been implemented in the literature.[2]

The information costs faced by a decision maker are rarely known to an outside observer. This makes it of interest to test a general model of optimal information acquisition that makes minimal assumptions about these costs. In this paper we introduce precise behavioral tests covering all standard theories of costly information acquisition, including rational inattention theory as well as static and sequential signal acquisition theories.[3] We establish limits on what choice data can reveal about the costs of information. In an experimental implementation, we confirm that attentional adjustments are well-modeled as rationally responsive to costs.

The purpose of our tests is to introduce non-parametric methods into the theory of information acquisition. Just as unobservability of preferences motivated the revealed preference approach to utility (Samuelson [1938], Afriat [1967]), so the unobservability of information acquisition costs motivates our approach. In the revealed preference spirit, our tests can be readily applied in practice, and fully characterize optimal behavior for an arbitrary finite data set.

Enriched choice data plays a central role in our tests. We utilize "state dependent" stochastic choice data, which describes the decision maker's prob-

[1] For example 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]), while purchasers limit their attention to a relatively small number of websites when buying over the internet (Santos *et al.* [2012]).

[2] See, for example Reis [2006], van Nieuwerburgh and Veldkamp [2009] and Woodford [2012].

[3] That our conditions characterize so many distinct microeconomic models is striking. It echoes the finding of Manzini and Mariotti [2007] that identical conditions ("weak WARP") capture the behavioral content of many apparently distinct procedural models of boundedly rational behavior.

ability of choosing each available action in each state of the world.[4] While only recently introduced into revealed preference analysis (see Caplin and Martin [2014], henceforth CM14), it is standard in the econometric analysis of discrete choice. For example, Chetty *et al.* [2009] study how choice distributions are impacted by an observable state: the inclusion or exclusion of sales taxes in stated prices. They find evidence of incomplete state awareness among buyers.[5]

Our key theoretical insight is that the decision maker's attention strategy is largely revealed by their state dependent stochastic choice data. Using this observation, we describe two conditions that render such data consistent with optimal acquisition of costly information. A "no improving action switch" (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 attention strategies across decision problems. Our main result is that these conditions, clearly necessary for rationality, are also sufficient. Any data set in which they are satisfied can be rationalized with a standard model of costly information acquisition.

We provide limits on the identification of attention costs from choice data, providing bounds on the relative costs of all chosen attention strategies. We show also that the assumptions that more informative strategies (in the sense of Blackwell) are more costly, that mixed strategies are feasible, and that inattention is costless put no additional restrictions on the data.[6]

We implement our tests by experimentally eliciting state dependent stochastic choice data. We test for three important forms of optimal information gathering behavior: responsiveness on the extensive margin; responsiveness on the intensive margin; and the presence of spillover effects. Our experiments

[4] The key role of data enrichment has arisen previously in our use of "choice process" data to test theories of sequential search (Caplin *et al.* [2011]).

[5] The data set has a long history in psychometric research. It is essential to the formulation of the Weber-Fechner laws on limits to perceptual discrimination (see Murray [1993]).

[6] This result is in the spirit of Afriat [1967], and pinpoints limits on the identifiability of cost functions in behavioral data.

demonstrate all three forms of behavior: when incentives are increased, more is learned; when the relative importance of differentiating between different states changes, subjects focus their attention appropriately; and the introduction of an action that increases the returns to attention has spillover effects on the choice probabilities of previously available alternatives. Our data indicates that subjects actively modify their attention in response to incentives in line with the optimizing model. Alternative theories in which learning is unresponsive to attentional incentives are clearly rejected.

Our paper is related to a recent literature analyzing specific models of costly information acquisition. Matejka and McKay [2011] study the implications of rational inattention with Shannon entropy costs for state dependent stochastic choice data, while Ellis [2012] uses deterministic state dependent choice to study a model in which a decision maker has a fixed set of information partitions to choose from. de Oliveira *et al.* [2013] consider a more general model of attention using choice over menus as their data. Unlike these papers, our revealed preference approach provides necessary and sufficient conditions in any arbitrary finite data set. Such 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]). Our work also fits into a growing literature aimed at identifying the behavioral implications of 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]).

Section 2 introduces the basic model of costly information acquisition. Section 3 provides our characterization. Section 4 establishes limits on identification of attention costs. Section 5 provides model extensions. Section 6 details our experimental design, with results in section 7. Section 8 relates our work to the broader literature. Section 9 concludes.

2 A Model of Costly Information Acquisition

2.1 Actions, Prizes, States, and Beliefs

A decision maker (DM) is initially unaware of the consequences attached to available options. With attention, this uncertainty can be reduced. To formalize, we consider choice between actions the payoff of which depend on the realization of a state ω from a set Ω of cardinality M. Prior beliefs are captured by $\mu \in \Gamma = \Delta(\Omega)$. An action a is a mapping from Ω to a prize space X. We use F to denote the grand set of actions.

Our goal is to identify conditions under which choice data can be rationalized as resulting from maximization of net utility by a Bayesian expected utility maximizer with costs of acquiring information. For the next two sections we focus squarely on costs of information acquisition. Hence we treat as known to the outside observer or econometrician both prior beliefs $\mu \in \Gamma$ and the expected utility function $U : X \longrightarrow \mathbb{R}$, with U_ω^a denoting the expected utility of action a in state ω. In section 5.2 we allow for unknown utility and for an unknown prior.

2.2 State Dependent Stochastic Choice Data

Let $D \subset \mathcal{F} \equiv \{A \in 2^F/\varnothing | A \text{ is finite}\}$ be a finite set of decision problems (defined by the set of available alternatives) with generic element $A \in D$. The idealized data set that we use to test the model of costly information acquisition is state dependent stochastic choice data. This describes for each decision problem the likelihood of choosing each available action in each state of the world. We define $\mathcal{Q}$ to be the set of mappings from Ω to probability distributions over F with finite support. Given $q \in \mathcal{Q}$, we let q_ω^a denote the probability of the DM choosing action a in state ω and denote as $F(q) \subset F$ the set of actions chosen with non-zero probability in some state of the world under state dependent stochastic choice function q. For $A \in \mathcal{F}$, we define $\mathcal{Q}^A$ as all data sets with $F(q) \subset A$.

Definition 1 *A state dependent stochastic choice* ***data set*** (D, Q) *comprises* *a finite set of decision problems* $D \subset \mathcal{F}$ *and a function* $Q : D \to \mathcal{Q}$*, with* $Q(A) \in \mathcal{Q}^A$*.*

2.3 Attention Strategies and Attention Costs

The DM chooses an *attention strategy* for each given decision problem that defines the effort that they put into learning the state of the world. Initially we assume that information processing is static (the extension to sequential choice of information is considered in section 5.1). We define an attention strategy as a stochastic mapping from states of the world to subjective 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. Having selected an attention strategy, the DM can condition choice of action only on these signals. Feasible attention strategies satisfy Bayes' rule.

Definition 2 *Given prior* $\mu \in \Gamma$*, feasible* ***attention strategies*** $\Pi(\mu)$ *comprise 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* $\gamma \in \Gamma(\pi)$*,*

$$\gamma_\omega = \frac{\mu_\omega \pi_\omega(\gamma)}{\sum_{\upsilon \in \Omega} \mu_\upsilon \pi_\upsilon(\gamma)},$$

where $\pi_\omega(\gamma) \equiv \pi(\omega)(\{\gamma\})$ *is the probability of posterior beliefs* γ *given state* ω*. Let* $\tilde{\Pi} \equiv \cup_\mu \Pi(\mu)$ *denote all attention strategies.*

An attention cost function maps attention strategies to the corresponding level of disutility.

Definition 3 *Given prior* $\mu \in \Gamma$*, an* ***attention cost function*** *is a mapping* $K : \Pi(\mu) \to \bar{\mathbb{R}}$ *with* $K(\pi) \in \mathbb{R}$ *for some* $\pi \in \Pi(\mu)$*. We let* $\mathcal{K}$ *denote the class of such functions.*

Note that we put no restrictions on the cost function K, meaning that our model nests all standard models of static information acquisition, including a 'rational inattention' model in which K would equal the Shannon mutual information between prior and posterior information states (e.g. Sims [2003]). We allow costs to be infinite to nest constraints on information acquisition - as when a hard limit 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 normal signals (Verrecchia [1982]). To avoid triviality we assume that finite-cost feasible attention strategies exists.

In sections 2 through 4 we impose no cross-prior restrictions on behavior. Until that point it simplifies notation to specify arbitrary $\mu \in \Gamma$, to limit the state space to satisfy $\mu_\omega > 0$, and to let Π identify feasible attention strategies given this prior.

2.4 Costly Information Representations

We model a DM who chooses an attention strategy to maximize gross payoffs net of information costs. The gross payoff associated with attention strategy $\pi \in \Pi$ in decision problem $A \in \mathcal{F}$ is calculated assuming that actions are chosen optimally in each posterior state. Let $G : \mathcal{F} \times \Pi \to \mathbb{R}$ denote the gross payoff of using a particular attention strategy in a particular decision problem:

$$G(A, \pi) = \sum_{\gamma \in \Gamma(\pi)} \left[\sum_{\omega} \mu_\omega \pi_\omega(\gamma) \right] g(\gamma, A);$$

where $g(\gamma, A) = \max_{a \in A} \sum_\omega \gamma_\omega U^a_\omega$. We make the standard assumption that attention costs are additively separable from the prize-based utility derived from the actions taken. We let $\hat{\Pi} : \mathcal{K} \times \mathcal{F} \to \Pi$ map cost functions and decision problems into rationally inattentive strategies. These are the strategies (if any) that maximize gross payoff net of attention costs,

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

The choice of the DM conditional on the signal received is captured by the function $C : \Gamma(\pi) \to \Delta(A)$, with $C^a(\gamma)$ the probability of action $a \in A$ given

$\gamma \in \Gamma(\pi)$. An attention strategy is consistent with observed state dependent stochastic choice data in some decision problem if optimal choice contingent on the signal received could produce such a pattern of data.

Definition 4 *For decision problem A, attention strategy $\pi \in \Pi$ is **consistent with** $q \in \mathcal{Q}^A$ if there exists $C : \Gamma(\pi) \to \Delta(A)$ such that:*

1. *Final choices are optimal:*

$$C^a(\gamma) > 0 \Longrightarrow \sum_{\omega} \gamma_\omega U_\omega^a \geq \sum_{\omega} \gamma_\omega U_\omega^b \quad \text{all } b \in A.$$

2. *The attention and choice functions match the data:*

$$q_\omega^a = \sum_{\gamma \in \Gamma(\pi)} \pi_\omega(\gamma) C^a(\gamma).$$

A data set admits a costly information representation if there exists an information cost function such that behavior in each decision problem is consistent with an optimal information strategy given those costs.

Definition 5 *Data set (D, Q) has a **costly information representation** $(\tilde{K}, \tilde{\pi})$ if there exists $\tilde{K} \in \mathcal{K}$ and $\tilde{\pi} : D \to \Pi$ such that, for all $A \in D$, $\tilde{\pi}(A) \equiv \tilde{\pi}^A$ is consistent with $Q(A)$ and satisfies $\tilde{\pi}^A \in \hat{\Pi}(\tilde{K}, A)$.*

3 Characterization

We establish two conditions as necessary and sufficient for (D, Q) to have a costly information acquisition representation. The first ensures optimality of final choice given an attention strategy and applies to each decision problem separately. The second ensures optimality of the attention strategy and applies to the collection of decision problems.

3.1 Minimal Attention Strategies

The key to our approach is the observation that, if a DM is behaving optimally, then one can learn much about their attention strategy from state dependent stochastic choice data. In particular, one can identify the average posterior beliefs that a DM must have had when choosing each act.

Definition 6 *Given $q \in \mathcal{Q}$ and $a \in F(q)$ define the* ***revealed posterior*** $r^a(q) \in \Gamma$,

$$r^a_\omega(q) = \frac{\mu_\omega q^a_\omega}{\sum_\upsilon \mu_\upsilon q^a_\upsilon},$$

all $\omega \in \Omega$.

The revealed posterior $r^a_\omega(q)$ is the probability of state of the world ω conditional on action a being chosen given state dependent stochastic choice data q. If the DM chooses each action in at most one subjective information state then the revealed posteriors are the same as their true posteriors when a was chosen. If they choose the same action in more than one subjective state then the revealed posterior is the corresponding weighted average.

We can use the revealed posteriors to construct a "revealed" attention strategy for each decision problem. We do so by assuming that any action is chosen in at most one subjective state. Under this assumption we can identify the resulting attention strategy directly from the data. The probability of posterior γ in state of the world ω is calculated by adding up probabilities of choosing all actions that have that revealed posterior.

Definition 7 *Given $q \in \mathcal{Q}$, $\omega \in \Omega$, and $\gamma = r^a(q)$ for some $a \in F(q)$, define the* ***minimal attention strategy*** $\bar{\pi}^q \in \Pi$ *to satisfy,*

$$\bar{\pi}^q_\omega(\gamma) = \sum_{\{a \in F(q) | r^a(q) = \gamma\}} q^a_\omega.$$

While the minimal attention strategy may not be the same as the DM's true attention strategy, a key observation is that it must be weakly less informative (in the sense of statistical sufficiency) than any attention strategy consistent with the data. Intuitively, this means that the minimal attention strategy can be obtained by "adding noise" to the true attention strategy.

Definition 8 *Attention strategy $\rho \in \Pi$ is **sufficient** for attention strategy $\pi \in \Pi$ (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 $\omega \in \Omega$,*

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

Lemma 1 establishes that any consistent attention strategy must be sufficient for the minimal attention strategy.

Lemma 1 *If $\pi \in \Pi$ is consistent with $q \in \mathcal{Q}$, then it is sufficient for $\bar{\pi}^q$.*
Proof. *All proofs can be found in online appendix 1.* ■

Blackwell's theorem establishes the equivalence of the statistical notion of sufficiency and the economic notion "more valuable than". If attention strategy π is sufficient for strategy ρ, then it yields (weakly) higher gross payoffs in any decision problem. This result plays a significant role in our characterization.

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

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

3.2 No Improving Actions Switches

Our first condition ensures that the DM's choices are optimal given posterior beliefs. It specifies that, when one identifies in the data the revealed posterior associated with any chosen action, this action must be optimal at that 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.

Condition D1 (No Improving Action Switches) Data set (D, Q) satisfies NIAS if, for every $A \in D$ and $a \in F(Q(A))$,

$$\sum_{\omega} r_{\omega}^{a}\left(Q(A)\right) U_{\omega}^{a} \geq \sum_{\omega} r_{\omega}^{a}\left(Q(A)\right) U_{\omega}^{b},$$

all $b \in A$.

3.3 No Improving Attention Cycles

Our second condition restricts choice of attention strategy across decision problems. Essentially, it cannot be the case that total gross utility can be increased by reassigning attention strategies across decision problems. To illustrate, consider a decision problem with two equiprobable states and two available actions, $A = \{a, b\}$, and with the state dependent payoffs,

$$(U_1^a, U_2^a) = (10, 0); (U_1^b, U_2^b) = (0, 20).$$

Suppose now that the observed choice behavior is,

$$Q_1^a(A) = 1 - Q_2^a(A) = \frac{2}{3} = 1 - Q_1^b(A) = Q_2^b(A).$$

Now consider a second decision problem differing only in that the action set is $B = \{a, c\}$, with $(U_1^c, U_2^c) = (0, 10)$, with corresponding data,

$$Q_1^a(B) = 1 - Q_2^a(B) = \frac{3}{4} = 1 - Q_1^b(B) = Q_2^b(B).$$

The specified data looks problematic with respect to optimal information acquisition. Action set A provides greater reward for discriminating between states, yet the DM is more discerning under action set B. 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,

$$\begin{aligned} G(A, \pi^A) - K(\pi^A) &\geq G(A, \pi^B) - K(\pi^B); \\ G(B, \pi^B) - K(\pi^B) &\geq G(B, \pi^A) - K(\pi^A). \end{aligned}$$

While we do not observe attention strategies directly, it is immediate that $G(i, \pi^i) = G(i, \bar{\pi}^i)$ for $i \in \{A, B\}$. Furthermore, as π^i is sufficient for $\bar{\pi}^i$, Blackwell's theorem tells us that $G(i, \pi^j) \geq G(i, \bar{\pi}^j)$ for $i.j \in \{A, B\}$ (see Remark 1). Thus we can insert the minimal attention strategies in the calculation of gross benefits to conclude,

$$G(A, \bar{\pi}^A) - G(A, \bar{\pi}^B) \geq K(\bar{\pi}^A) - K(\bar{\pi}^B) \geq G(B, \bar{\pi}^A) - G(B, \bar{\pi}^B).$$

For there to exist a cost function that can rationalize this data, the left hand side of this inequality must be no lower than the right hand side. Using this condition we conclude that, for this data to be rationalizable, gross benefit must be maximized by the assignment of minimal attention strategies to decision problems observed in the data,

$$G(A, \bar{\pi}^A) + G(B, \bar{\pi}^B) \geq G(A, \bar{\pi}^B) + G(B, \bar{\pi}^A). \tag{1}$$

In the above example $G(A, \bar{\pi}^A) + G(B, \bar{\pi}^B) = 17\frac{1}{2}$, while $G(A, \bar{\pi}^B) + G(B, \bar{\pi}^A) = 17\frac{11}{12}$. Thus, there is no cost function that can be used to rationalize this data. The NIAC condition ensures precisely that no such cycles of attention strategy raise gross utility.

Condition D2 (No Improving Attention Cycles) Data set (D, Q) 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 $\bar{\pi}^j = \bar{\pi}^{Q(A_j)}$.

The NIAC condition 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]).

3.4 Characterization

Our first main result is that, while clearly necessary conditions, NIAC and NIAS together are also sufficient for (D, Q) to have a costly information acquisition representation. We establish this by applying the results of Koopmans and Beckmann [1957] concerning the linear allocation problem. The cost function that we introduce is based on the shadow prices that decentralize the optimal allocation in their model (see also Rochet [1987]).

Theorem 1 *Data set (D, Q) has a costly information acquisition representation if and only if it satisfies NIAS and NIAC.*

4 Identification

In this section we establish limits on identification of the cost function. We open by considering three natural restrictions on attention cost functions: 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 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. Following this we provide limits on the recoverability of the cost function by characterizing all weakly monotonic cost functions that can rationalize a given data set.

4.1 Weak Monotonicity

A partial ranking of the informativeness of attention strategies is provided by the notion of statistical sufficiency (see definition 8). A natural condition for an attention cost function is that more information is (weakly) more costly. Free disposal of information would imply this property, as would a ranking based on Shannon mutual information.[7]

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

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

4.2 Mixture Feasibility

In addition to using pure attention strategies, it may be feasible for the DM to mix these strategies using some randomizing device.

Definition 9 *Given attention strategies $\pi, \eta \in \Pi$, and $\alpha \in [0,1]$, the* ***mixture strategy*** *$\alpha \circ \pi + (1-\alpha) \circ \eta \equiv \psi \in \Pi$ is defined by,*

$$\psi_\omega(\gamma) = \alpha \pi_\omega(\gamma) + (1-\alpha)\eta_\omega(\gamma),$$

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

The definition implies that the 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 attention strategy 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.

A natural assumption is that DMs can choose to mix attention strategies and pay the corresponding expected costs. They could flip a coin and choose strategy π if the coin comes down heads and strategy η if it comes down tails.

[7] While in many ways intuitively attractive, this assumption may not be universally valid. In a world with discrete signals it may be very costly or even impossible to generate continous changes in information. Moreover the DM may be restricted to some fixed set of signals in which case less informative structures are essentially disallowed. It may not be possible to automatically and freely dispose of information once learned.

In expectation the cost of this strategy would be half that of π and half that of η. Allowing such mixtures puts an upper bound on the cost of the strategy $0.5\circ\, \pi + 0.5 \circ \eta$. However, it does not pin down the cost precisely, since there may be a more efficient way of constructing the mixed attention strategy.

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,

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

4.3 Normalization

It is typical in the applied literature to allow inattention at no cost, and otherwise to have costs be non-negative. Given weak monotonicity, non-negativity of the entire function follows immediately if one ensures that inattention is costless.

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

4.4 Theorem 2

Theorem 2 states that, whenever a costly information acquisition representation exists, one also exists in which the cost function satisfies conditions K1 through K3. Whatever one thinks of the above assumptions on intuitive grounds, even if any one or all of them are in fact false, any data set that can be rationalized can equally be rationalized by a function that satisfies them all.

Theorem 2 *Data set (D, Q) satisfies NIAS and NIAC if and only if it has a costly information acquisition representation with conditions K1 to K3 satisfied.*

This result 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 online appendix 2 that there are data sets satisfying NIAS and NIAC yet 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.

4.5 Recoverability

Theorem 1 tells us the conditions under which there exists an attentional cost function that will rationalize the data. We now provide conditions that identify the set of all such cost functions, in the spirit of Varian [1984] and Cherchye *et al.* [2011]. We restrict ourselves to cost functions that satisfy weak monotonicity, so that we can treat minimal attention strategies 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 $\pi \in \Pi$,

$$K(\bar{\pi}^A) - K(\pi) \leq G(A, \bar{\pi}^A) - G(A, \pi). \tag{2}$$

This directly implies an upper and lower bound on the relative costs of any two revealed attention strategies $\bar{\pi}^A$, $\bar{\pi}^B$ 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 attentional 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 attention strategies are high enough to satisfy inequality 2.

This condition implies potentially tighter bounds on the relative cost of any two revealed attention strategies. Consider the corresponding inequalities in string $A_1...A_n \in D$ 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, implying

$$K(\bar{\pi}^A) - K(\bar{\pi}^B) \leq \min_{\{A_1...A_n \in D | A_1 = A, A_n = B\}} \sum_{i=1}^{n-1} \left[G(A_i, \bar{\pi}^{A_i}) - G(A_i, \bar{\pi}^{A_{i+1}}) \right]; \quad (3)$$

Considering the reverse string $A_1...A_n \in D$ with $A_1 = B$ and $A_n = A$ yields

$$K(\bar{\pi}^A) - K(\bar{\pi}^B) \geq \max_{\{A_1...A_n \in D | A_1 = B, A_n = A\}} \sum_{i=1}^{n-1} \left[G(A_i, \bar{\pi}^{A_{i+1}}) - G(A_i, \bar{\pi}^{A_i}) \right]. \quad (4)$$

Note also that if one considers cost functions for which inattention is free, the above inequalities can be used to place absolute bounds on the level of costs. Moreover if one ever sees a switch in attention strategy 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 possible attention strategies.

5 Extensions

Sequential sampling has been the central focus in models of information acquisition since the work of Wald [1947]. In this section we extend our results to a model of sequential attention, assuming that only final choices are seen: evolution of learning before choice is not directly observable. We also extend our results to allow for unobservability of the utility function and of the prior. In both cases, NIAS and NIAC are unchanged in essentials.

5.1 Sequentially Rational Inattention

We fix a time interval within which a decision is to be made and divide it into $T \geq 1$ sub-periods. In each such sub-period the DM must choose whether or not to collect more information conditional on what has already been learned. In the former case they must decide what additional information to collect. In the latter, information gathering finishes and they must choose one of the available actions. The sequence of attention and action choices is made to maximize the net undiscounted value of final prize utility less sequential attention costs. Neither attentional inputs nor decision time are observed.

The DM at the start of period 0 is endowed with prior beliefs $\mu \in \Gamma$. We use the time indexed sets G_t and S_t to define the DM's deterministic stopping rule by identifying the states in which they respectively continue to search and stop searching. If the DM stops searching immediately then $S_0 = \{\mu\}$ and $G_0 \subset 2^\Gamma$ is empty. If not we define $G_0 = \{\mu\}$ and $S_0 \subset 2^\Gamma$ as empty. In this case a first attention strategy $\pi_1 \in \Pi(\mu)$ is selected at the start of period 1, with posterior $\gamma_1 \in \Gamma(\pi_1)$ realized instantly. The stopping rule in period 1 is defined by $G_1 \subset \Gamma(\pi_1)$ which contains all the posteriors at which further information is gathered, and $S_1 = \Gamma(\pi_1)/G_1$ the corresponding stopping set.

The process iterates from this point forward. We allow for history dependence by defining the continuation set G_t for $t \geq 1$ on sequences of posteriors $\gamma^t = (\mu, \gamma_1, .., \gamma_t) \in \Gamma^{t+1}$. The first time that G_t is empty identifies the maximal stopping time, $\tau = t \leq T$. In all earlier periods, the DM picks a period $t+1$ attention strategy,

$$\pi_{t+1} : G_t \to \tilde{\Pi} \text{ with } \pi_{t+1}(\gamma^t) \in \Pi(\gamma_t).$$

The ensuing continuation set and attention strategy are correspondingly defined:

$$G_{t+1} \subset \left\{ \gamma^{t+1} \in \Gamma^{t+2} | \gamma^t \in G_t, \gamma_{t+1} \in \Gamma(\pi_t(\gamma_t)) \right\},$$

with all other sequences of posteriors in $S_{t+1} \subset \Gamma^{t+2}$. The above fully specifies a sequential attention strategy. Given prior μ, we let $\Sigma(\mu)$ be the set of such strategies, with generic element $\sigma \in \Sigma(\mu)$.

Note that each strategy $\sigma \in \Sigma(\mu)$ induces a probability distribution over sequences of posteriors. Define $\gamma^0 = \gamma_0 = \mu$, and let $\rho^\sigma(\gamma^t)$ be the probability of $\gamma^t \in \Gamma^{t+1}$ given strategy σ, with $\rho^\sigma_\omega(\gamma^t)$ being the corresponding state dependent probability,

$$\rho^\sigma(\gamma^t) = \sum_\omega \mu_\omega \rho^\sigma_\omega(\gamma^t).$$

We turn to the evaluation of strategies $\sigma \in \Sigma(\mu)$. We assume that choices are made optimally given posteriors. Thus, when faced with decision problem A, for each $\gamma^t \in S_t$ for $0 \leq t \leq T$, the decision maker will receive utility $g(\gamma_t, A)$, maximal expected utility at the posterior relevant for action choice. To count against reward utility are the attentional costs which we assume to be independent of preferences over prize lotteries and additively separable across periods.

Definition 10 *Given $\mu \in \Gamma$, an admissible attention cost function $E \in \mathcal{E}$ specifies for each $t \geq 0$ cost $E(\gamma^t, \pi) \in \bar{\mathbb{R}}$ on $\gamma^t = (\mu, \gamma_1, .., \gamma_t) \in \Gamma^{t+1} \in G_t$, with $E(\gamma^t, \pi) = \infty$ for $\pi \notin \Pi(\gamma_t)$.*

The above covers all standard sequential models with additive attention costs. In fact one can enrich the domain of the period attention cost functions to include all past attention levels as well as the past posteriors without changing in any way the ensuing analysis.[8] We define a strategy σ as sequentially rational for decision problem A if it solves the following problem:

$$\sigma \in \arg\max_{\sigma \in \Sigma(\mu)} \sum_{t=0}^{T} \left[\sum_{\gamma^t \in S_t} \rho^\sigma(\gamma^t) g(\gamma_t, A) - \sum_{\gamma^t \in G_t} \rho^\sigma(\gamma^t) E(\gamma^t, \pi^\sigma_{t+1}(\gamma^t)) \right].$$

[8]While substantively enriching the model by allowing for tiredness resulting from past effort etc., including these effects greatly complicates notation.

Where such optima exist, $\hat{\Sigma} : \mathcal{E} \times \mathcal{F} \to \Sigma$ identifies all sequentially rationally inattentive strategies.

Our goal is to identify all data sets (D, Q) that can be rationalized by some fixed $E \in \mathcal{E}$ as consistent with sequentially optimal behavior in the face of costly attention. Given $A \in \mathcal{F}$, the definition of consistency of $\sigma \in \Sigma(\mu)$ with $q \in \mathcal{Q}$ is essentially unchanged: it requires existence of a choice function $C : \Gamma \to \Delta(A)$ such that final choices are optimal for all $1 \leq t \leq \tau$ and $\gamma^t \in S_t$, and that the attention and choice functions match the data,

$$q_\omega^a = \sum_{t=0}^{\tau} \sum_{\gamma^t \in S_t} \rho_\omega^\sigma(\gamma^t) C^a(\gamma_t).$$

Definition 11 *Data set (D, Q) has a* ***sequential costly information (SCI) representation*** *$(\tilde{E}, \tilde{\sigma})$ if there exists $\tilde{E} \in \mathcal{E}$ and $\tilde{\sigma} : D \to \Sigma$ such that, for all $A \in D$, $\tilde{\sigma}(A)$ is consistent with $Q(A)$ and satisfies $\tilde{\sigma}(A) \in \hat{\Sigma}(\tilde{E}, A)$.*

The key result is that NIAS and NIAC remain necessary (as well as sufficient) for such a representation despite the richer class of learning behaviors covered. We establish this in the appendix as a corollary to theorem 1. Intuitively, time consistency reduces the dynamic problem of sequential choice to a static problem of choice of strategy.

Corollary 1 *Data set (D, Q) has an SCI representation if and only if it satisfies NIAS and NIAC.*

5.2 Unobservability of Prior and Utility Function

Returning to the case of static information acquisition, we now consider the observable implications of optimal behavior when preferences and prior beliefs are unknown. As in CM14, theorem 1 extends directly to cases in which the utility function $U : X \to \mathbb{R}$ and the prior are both unknown, provided the prior assigns strictly positive probability to all states, $\mu \in \Gamma^I \equiv \{\mu \in \Gamma | \mu_\omega > 0$ all $\omega\}$. To establish this we correspondingly amend key definitions. We treat

(μ, U) as unknowns and define:

$$G^{(\mu,U)}(A,\pi) \equiv \sum_{\gamma\in\Gamma(\pi)} \left[\sum_{\omega} \mu_\omega \pi_\omega(\gamma)\right] g^U(\gamma, A);$$
$$\hat{\Pi}^{(\mu,U)}(K,A) \equiv \arg\sup_{\pi\in\Pi(\mu)} \left\{G^{(\mu,U)}(A,\pi) - K(\pi)\right\};$$

with $g^U(\gamma, A) = \max_{a\in A} \sum_\omega \gamma_\omega U^a_\omega$. We amend the definition of a costly information representation to avoid the conditions being satisfied by a constant utility function.

Definition 12 *Data set (D, Q) has a* ***costly information representation with unknown prior and utility function*** *if there exists $\tilde{U} : X \to \mathbb{R}$, $\tilde{\mu} \in \Gamma^I$, $\tilde{K} \in \mathcal{K}$ and $\tilde{\pi} : D \to \Pi$ such that, for all $A \in D$, $\tilde{\pi}(A) \equiv \tilde{\pi}^A$ is consistent with $Q(A)$, $\tilde{\pi}^A \in \hat{\Pi}^{(\tilde{\mu},\tilde{U})}(\tilde{K}, A)$, and such that $\exists A \in D$, $a \in A$, $\gamma \in \Gamma$ with $C^a(\gamma) > 0$, and $b \in A$ such that,*

$$\sum_\omega \gamma_\omega \tilde{U}^a_\omega > \sum_\omega \gamma_\omega \tilde{U}^b_\omega.$$

The conditions for existence of such a representation are the precise analog of the NIAS and NIAC conditions, with the additional requirement of some strict inequality in the value of acts.

Condition D3 (NIAS*) Data set (D, Q) satisfies NIAS* with respect to $\mu \in \Gamma^I$ and $U : X \to \mathbb{R}$ if, for every $A \in D$ and $a \in A$,

$$\sum_\omega r^a_\omega\left(\mu, Q(A)\right) U^a_\omega \geq \sum_\omega r^a_\omega\left(\mu, Q(A)\right) U^b_\omega,$$

all $b \in A$, where,

$$r^a_\omega\left(\mu, q\right) = \frac{\mu_\omega q^a_\omega}{\sum_\upsilon \mu_\upsilon q^a_\upsilon}.$$

and there exists $A \in D$, $a \in F(Q(A))$, and $b \in A$ such that,

$$\sum_{\omega} r_{\omega}^{a}\left(\mu, Q(A)\right) U_{\omega}^{a} > \sum_{\omega} r_{\omega}^{a}\left(\mu, Q(A)\right) U_{\omega}^{b}.$$

Condition D4 (NIAC*) Data set (D, Q) satisfies NIAC* with respect to $\mu \in \Gamma^I$ and $U : X \to \mathbb{R}$ if, for any set of decision problems $A_1, A_2, ..., A_J \in D$ with $A_J = A_1$,

$$\sum_{j=1}^{J-1} G^{(\mu,U)}(A_j, \bar{\pi}^j) \geq \sum_{j=1}^{J-1} G^{(\mu,U)}(A_j, \bar{\pi}^{j+1}),$$

where $\bar{\pi}^j = \bar{\pi}^{Q(A_j)}$.

The characterization is precisely as expected, as follows from a careful reading of the original proof: hence this is treated as a corollary.

Corollary 2 *Data set (D, Q) has a costly information representation with unknown prior and utility function if and only if there exists $\tilde{U} : X \to \mathbb{R}$ and $\tilde{\mu} \in \Gamma^I$ with respect to which (D, Q) satisfies both NIAS* and NIAC*.*

NIAS* and NIAC* therefore identify inequality constraints to which a solution must exist if the data is to be rationalizable with a costly information representation. 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*). 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. The following example demonstrates behavior that is commensurate with NIAS* but is not commensurate with NIAC* for any non-degenerate utility function and prior.

Example 1 *Let $X = \{x, y\}$, $\Omega = \{1, 2\}$ $A = \{a, b, c\}$ $B = \{a, b\}$ with actions* *defined as follows*

	State	
Action	*1*	*2*
a	x	y
b	y	x
c	x	x

We show now that data set $Q_1^a(A) = Q_2^b(A) = 1$, $Q_1^b(A) = Q_1^c(A) = Q_2^a(A) = Q_2^c(A) = 0$, $Q_1^a(B) = Q_2^a(B) = 1$, *and* $Q_1^b(B) = Q_2^b(B) = 0$ *does not permit a costly information representation with unknown prior and utility function.*

Intuitively, the DM in the 3 action case (choice set A*) perfectly identifies the state and chooses the action that gives prize* x *rather than* y*. Since indifference is not allowed, NIAS* requires* $U(x) > U(y)$*. In the two action case (choice set* B*) the DM always chooses action* a*, whether it yields* x *or* y*. The problem with this is that the availability in set* A *of an action that yields the better prize without attention implies that the cost of being perfectly informed must be zero, making it impossible to understand why the fully informative attention strategy was not chosen when facing choice set* B*. Technically, NIAC* fails since, with* $\mu_2 > 0$,

$$\begin{aligned} & G^{(\mu,U)}(A, \bar{\pi}^A) + G^{(\mu,U)}(B, \bar{\pi}^B) - G^{(\mu,U)}(A, \bar{\pi}^B) - G^{(\mu,U)}(B, \bar{\pi}^A) \\ = \quad & U(x) + [\mu_1 U(x) + \mu_2 U(y)] - U(x) - U(x) = \mu_2 \left(U(y) - U(x)\right) < 0. \end{aligned}$$

6 Experimental Design

6.1 Design

We introduce an experimental design that generates state dependent stochastic choice data with which we test our conditions. Subjects are shown a screen with 100 balls that are either red or blue. The state of the world is the number of red balls on the screen. Prior to seeing the screen, subjects are informed of the probability distribution over states. They 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 cognitive effort for monetary reward.

A decision problem is defined by the set of available actions. A subject faces each decision problem 50 times.[9] We estimate state dependent stochastic choice functions at the individual and aggregate level using the observed frequency of choosing each action in each state. In any given experiment, the subject faces four different problems. All occurrences of the same problem are grouped, but the order of the problems is block-randomized. In estimating the state dependent stochastic choice function we treat the 50 times that a subject faces the same decision making environment as 50 independent repetitions of the same event.

The aim of our experiments is to generate environments in which subjects may actively alter their attention in response to incentives. We focus on three such cases: changes to the overall reward for attention which should lead to behavioral changes on the extensive margin; the introduction of new actions to the choice set, which should generate spillover effects; and changes in the states between which it important to differentiate, which should lead to behavioral changes on the intensive margin. These experiments provide a testing ground for the NIAS and NIAC conditions, as well as allowing us to rule out models in which attention is fixed.

Subjects were recruited from the New York University student population.[10] 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 online appendix 4.

[9] 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 each possible configuration of balls.

[10] 46 subjects took part in experiment 1, 45 in experiment 2, and 24 in experiment 3. Each subject took part in only one experiment.

6.2 Experiment 1: The Extensive Margin

Experiment 1 tests whether subjects increase overall attentional effort as incentives increase. It comprises four decision problems with two equally likely states: in state 1 there are 49 red balls and in state 2 there are 51. In each decision problem there are two actions available $\{a^i, b^i\}$ with i indexing the decision problem. In each case, a is superior in state 1 while b is superior in state 2. Across decision problems the reward for making the correct choice in each state varies.[11] Table 1 describes the available actions in the four decision problems in this experiment (payoffs are in US$).

Table 1: Experiment 1

DP	Payoffs U_1^a	U_2^a	U_1^b	U_2^b
1	2	0	0	2
2	10	0	0	10
3	20	0	0	20
4	30	0	0	30

Table 2: Experiment 2

DP	Payoffs U_1^a	U_2^a	U_1^b	U_2^b	U_1^c	U_2^c
5	23	23	21	25	n/a	n/a
6	23	23	21	25	30	10
7	23	23	21	25	35	5
9	23	23	21	25	40	0

Experiment 1 allows us to differentiate between models in which attention responds to incentives and those in which it does not. 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]). SDT is clearly a special case of our model, and so implies both NIAC and NIAS. However the same signal structure must rationalize behavior in all decision problems. If attention 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_1^a and U_2^b, as long as the two are equal.[12]

[11] Note that these could be recorded as state dependent dollar prizes rather than direct utilities. Allowing for risk aversion rather than risk neutrality adds more notational complexity than warranted since results are unchanged in essentials.

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

Increasing attention in response 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]).

6.3 Experiment 2: Spillover Effects

Experiment 2 is designed to test whether the introduction of an "attention-inducing" action can spill over to increase the probability of choosing a previously available alternative. A theoretical example of this form is introduced in Matejka and McKay [2011] and described in table 2. It consists of two equally likely states (49 and 51 red balls). Decision problem 5 consists of two actions, a (which pays the same amount in both states) and b (which pays slightly more in state 2 and slightly less in state 1). Decision problems 6-8 add a further action c, which pays significantly more in state 1 and significantly less in state 2.

When action a and b are available there is little incentive to gather information, meaning that subjects may choose to remain uninformed and choose a. However, with c available also, it becomes more important to learn the true state, as c provides a high reward in state 1 but a low reward in state 2 - increasingly so for later decision problems. A rationally inattentive agent may therefore select a more informative attention strategy. If this learning suggests to the DM that state 2 is very likely, then it is optimal to choose action b.

This experiment allows us to differentiate between costly information processing and random utility models (RUMs) (McFadden [1974], Gul and Pesendorfer [2006]) which do not allow for flexible attention. RUMs take as given a probability measure over some family of utility functions. Prior to making a choice, one utility function gets drawn from this set according to the specified measure. The DM then chooses in order to maximize this utility function.[13]

A RUM could potentially explain an increase in accuracy as incentives increase in experiment 1, as this increases the value difference between the

[13] In the case of choice over lotteries, the family of utility functions can be over the lotteries themselves or, following Gul and Pesendorfer [2006], over the underlying prize space, with the utility of a lottery equal to its expectation according to the selected utility function.

two options. However, a general property of RUMs is monotonicity. Addition of a new action to the set of available choices cannot increase the probability that one of the pre-existing options will be chosen (Gul and Pesendorfer [2006], Luce and Suppes [1965]).

Monotonicity Axiom Given $A \in \mathcal{F}$, $a \in A$, $b \in F \backslash A$ and $\omega \in \Omega$,

$$Q(A)^a_\omega \geq Q(A \cup b)^a_\omega .$$

Monotonicity is violated by a model of costly inattention that exhibits information spillovers of the type described above: the introduction of action c increases the probability of choosing action b in state 2.

6.4 Experiment 3: The Intensive Margin

Table 3: Experiment 3

	Payoffs							
DP	U^a_1	U^a_2	U^a_3	U^a_4	U^b_1	U^b_2	U^b_3	U^b_4
9	1	0	10	0	0	1	0	10
10	10	0	1	0	0	10	0	1
11	1	0	1	0	0	1	0	1
12	10	0	10	0	0	10	0	10

In experiment 3 we vary the states in which information is valuable and measure the extent to which subjects focus their attention accordingly. All decision problems in this experiment involve four equally likely states comprising two identifiable groupings. States 1 and 2 are perceptually hard to distinguish from one another, being defined respectively by 29 and 31 red balls. States 3 and 4 are also hard to distinguish from another, being defined respectively by 69 and 71 red balls. There are four decision problems with two possible actions, still labelled a and b. The decision problems differ according to whether it is important to differentiate between states 1 and 2 (problem 9), states 3 and 4 (problem 10), neither (problem 11), or both (problem 12), as described in table 2.

7 Results

7.1 Attention is Limited and Flexible

Before implementing the NIAS and NIAC tests, we provide evidence that subjects are neither fully attentive nor completely inattentive. We also confirm the presence of the attentional flexibility that SDT and standard RUM models rule out.

The first point to observe is that the experiments produce choice data that is both stochastic and state dependent. Subjects gather some information prior to choice, but this information is incomplete. Using aggregate data from the simple two action cases of experiments 1 and 3 (in which there is a clear correct choice in each state), subjects made "mistakes", choosing the inferior action on 32% of all trials. In all three experiments, choice behavior is significantly different across states (Fisher's exact test, $p < 0.0001$). For example, in experiment 1, 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. For example, of the 46 subjects in experiment 1, 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. These results suggest that our subjects are absorbing some information about the state of the world, but are not fully informed when they make their choices.

Our data also rules out the fixed-signal SDT model which does not allow subjects to make better decisions as incentives increase in the symmetric case. As shown in figure 1b below, our aggregate data clearly exhibits such a change, with higher proportions of correct choices at higher incentive levels (rising from 62% in decision problem 1 to 77% in problem 4, significant at the 0.1% level while clustering at the individual). At the individual level, 54% of subjects exhibit significant changes in choice probabilities between decision problems at the 10% level.[14]

[14] All reported standard errors and statistical tests carried out using OLS regression of the choice of act on each trial on dummies associated with each decision problem. For aggregate

Experiment 2 provides evidence against RUMs with fixed information structures. Table 4 shows that the 44 subjects who took part in the experiment demonstrate clear violations of monotonicity. The introduction of action c increases the probability of choosing action b in state 2 from 23% in DP 5 to 39% in DP 8. Across all decision problems, the introduction of act c increases the choice of b in state 2 by an average of 12pp, significant at the 1% level. At the individual level, 51% of subjects show a significant violation of monotonicity of the type predicted by costly information acquisition theory at the 10% level.

Table 4

DP	$Q_1(b)$	$Q_2(b)$
5	17%	23%
6	15%	31%
7	12%	33%
8	13%	39%

7.2 NIAS and NIAC: Experiment 1

In the two state/two action set up of experiment 1, NIAS implies the existence of a cutoff posterior probability of state 1 that determines the optimal act. For posterior beliefs above 0.5, action a is optimal, while for lower posteriors, action b is optimal. This cutoff is shown in figure 1a together with the estimated posteriors associated with the choice of action a and action b at the aggregate level. This figure demonstrates that NIAS is satisfied in the aggregate data. At the individual level, for only 1 subject is there a statistically significant violation of NIAS (i.e. the estimated posterior is significantly lower than 0.5 when a is chosen or significantly higher than 0.5 when b is chosen at the 10% level). Moreover, as figure 2a shows, monetary losses due to NIAS violations are small.[15] As a benchmark, these losses are compared to those that would have been observed from a population of subjects choosing at random.[16] The

data we control for clustering at the subject level.

[15] Treating point estimates as each subject's true posterior beliefs

[16] The use of random benchmarks has been discussed by, for example, Beatty and Crawford [2011]. The precise procedure used to construct the random behavior is as follows: for each

observed distribution is significantly different from the simulated distribution at the 0.01% level (Kolmogorov-Smirnov test).

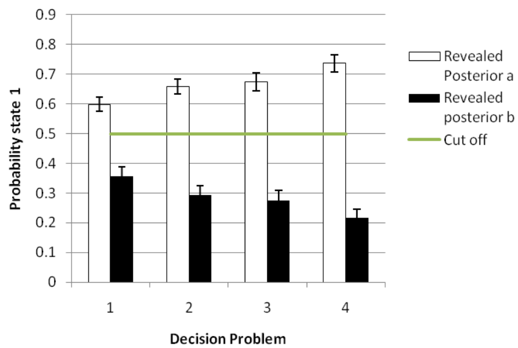


Figure 1a - NIAS Experiment 1

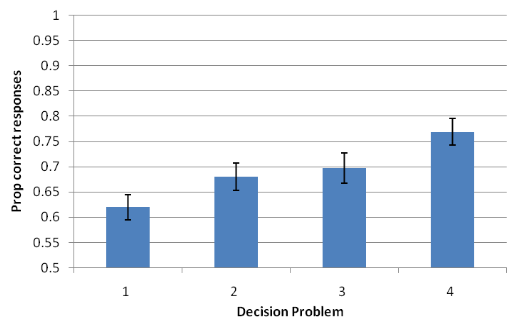


Figure 1b - NIAC Experiment 1

The NIAC condition in experiment 1 relates the change in incentives between decision problems to the change in τ_i, the probability that the correct decision is taken in state $i = 1, 2$ (action a in state 1, action b in state 2). For two state/two action problems of this type, NIAC implies the condition,

$$\Delta\tau_1\Delta(U_1^a - U_1^b) + \Delta\tau_2\Delta(U_2^b - U_2^a) \geq 0, \tag{5}$$

where Δx indicates the change in x between two decision problems. In experiment 1, equation 5 implies only that $\tau_1 + \tau_2$, the total probability of choosing the correct action, should be monotonic in rewards. Figure 1b shows that indeed the proportion of correct responses rises from 62% in decision problem 1 to 77% in problem 4. Differences between all pairs of decision problems are significant at the 1% level, apart from between problem 2 and 3, for which the difference is not significant.

At the individual level 83% of subjects show no significant violation of

decision problem and for each state, a random number is drawn for each available action. The probability of choosing each action from that state is then calculated as the value of the random number associated with that action divided by the sum of all random numbers.

the NIAC condition. Losses resulting from NIAC violations at the individual level are small, as shown in figure 2b. This figure plots the distribution of the maximal surplus possible by reassigning attention strategies to decision problems minus the surplus generated by the observed assignment for each individual.[17] The NIAC condition demands this number be zero. As a comparator, we show the distribution obtained from random choice. Again, the observed distribution is significantly different from the simulated distribution at the 0.01% level.

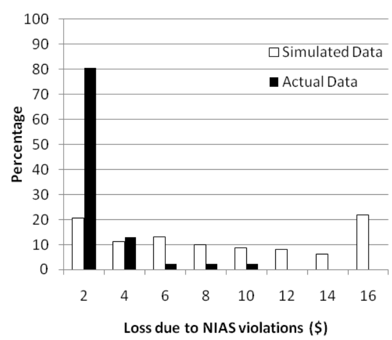
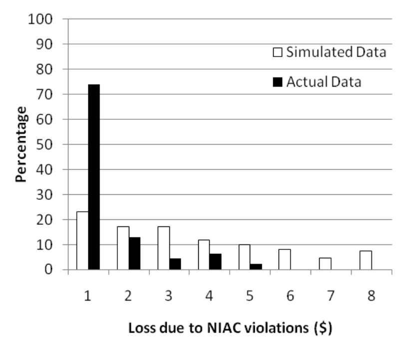


Fig 2a: NIAS losses
Experiment 1



Fib 2b: NIAC losses
Experiment 1

7.3 NIAS and NIAC: Experiment 2

For experiment 2, NIAS defines regions of acceptable posteriors for the choice of each action in each decision problem. Table 5 describes these regions, and the aggregate posteriors observed in the data.

[17] The actual surplus of a subject's attention strategy is calculated assuming no violations of NIAS.

Table 5: NIAS conditions for experiment 2						
DP	Range γ_1			Aggregate		
	b	a	c	γ_1^b	γ_1^a	γ_1^c
5	$[0, 50\%]$	$[50\%, 100\%]$	n/a	43	52	
6	$[0, 50\%]$	$[50\%, 65\%]$	$[65\%, 100\%]$	32	50	85
7	$[0, 50\%]$	$[50\%, 60\%]$	$[60\%, 100\%]$	27	51	82
8	$[0, 50\%]$	$[50\%, 57.5\%]$	$[57.5\%, 100\%]$	25	49	88

The aggregate data shows no significant violations of NIAS. At the individual level 91% of subjects show no significant violations of NIAS, and the cost of the resulting violations is small (Figure s1 in online appendix 3).

Applying bilateral NIAC to experiment 2 implies the following ranking on $Q_1(c) - Q_2(c)$[18],

$$Q_1^8(c^8) - Q_2^8(c^8) \geq Q_1^7(c^7) - Q_2^7(c^7) \geq Q_1^6(c^6) - Q_2^6(c^6).$$

In the aggregate the values of $Q_1(c) - Q_2(c)$ are 29%, 18% and 18% for decision problem 8, 7, and 6 respectively. DP 8 is significantly different from DP 7 and DP 6, though DP 6 and DP 7 are not significantly different from each other. At the individual level, 65% of subjects show no significant violations of NIAC, and again losses are small (figure s2 in the online appendix).

7.4 NIAS and NIAC: Experiment 3

The NIAS conditions for each decision problem in experiment 3 are,

$$U_1^a(\gamma_1^a - \gamma_2^a) + U_3^a(\gamma_3^a - \gamma_4^a) \geq 0$$

[18]If posterior beliefs when a is chosen in decision problem 5 made it preferable to choose c (if available), we additionally have the restriction,

$$Q_1^6(c) - Q_2^6(c) \geq Q_1^5(a) - Q_2(a).$$

However this is not the case is our aggregate data.

Table s2 in the supplemental material shows that this condition holds at the aggregate level. At the individual level, 92% of subjects show no significant violations of NIAS and, as shown in figure s1 in the online appendix, the losses amongst those that do not are again small.

With regard to the NIAC condition, the equivalent of condition 5 implies that subjects should make the right choice in a given state more often when the value of doing so is high.[19] More specifically, NIAC implies six inequalities based on binary comparisons of the 4 decision problems. Table 6 shows these inequalities, the average value of the left hand and right hand side variables in the aggregate data, and the probability associated with the test that these two are equal.

Table 6: NIAC conditions for experiment 3

Condition	LHS	RHS	P
$\tau_1^{10}+\tau_2^{10}+\tau_3^{9}+\tau_4^{9}\geq \tau_3^{10}+\tau_4^{10}+\tau_1^{9}+\tau_2^{9}$	72.8	64.9	0.01
$\tau_3^{9}+\tau_4^{9}\geq \tau_3^{11}+\tau_4^{11}$	68.2	63.3	0.38
$\tau_1^{10}+\tau_2^{10}\geq \tau_1^{11}+\tau_2^{11}$	77.3	66.9	0.02
$\tau_1^{12}+\tau_2^{12}\geq \tau_1^{9}+\tau_2^{9}$	74.8	63.7	0.02
$\tau_3^{12}+\tau_4^{12}\geq \tau_3^{10}+\tau_4^{10}$	69.1	66.3	0.33
$\tau_1^{12}+\tau_2^{12}+\tau_3^{12}+\tau_4^{12}\geq \tau_1^{11}+\tau_2^{11}+\tau_3^{11}+\tau_4^{11}$	72.0	65.1	0.10

In every case the inequality is satisfied by the point estimates in the aggregate data. In three of the cases the differences are statistically significant at the 5% level. At the individual level, 79% of subjects exhibit no significant failures of NIAC and, as table s2 in the appendix shows, the resulting losses are small. Overall, 75% of subjects exhibit no significant violation of either NIAS or NIAC.

[19] The precise condition is

$$\Delta\tau_1\Delta(U_1^a-U_1^b)+\Delta\tau_3\Delta(U_3^a-U_3^b)+\Delta\tau_2\Delta(U_2^b-U_2^a)+\Delta\tau_4\Delta(U_4^b-U_4^a)\geq 0.$$

8 Existing Literature

Many approaches have been taken to modelling costs and constraints on information acquisition in economic applications, including sequential search (McCall [1970]) selection of the variance of a normal signal (e.g. Verrecchia [1982]), and the binary choice to either be fully informed or not (Reis [2006]). More recently, Sims [2003] introduced the concept of rational inattention, in which the decision maker is free to choose any attention strategy they wish, with costs based on the Shannon mutual information between prior and posterior beliefs. 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.

A recent wave of literature shares our goal of capturing the observable implications of optimal acquisition of costly information. Matejka and McKay [2011] analyze the implications 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. Furthermore, unlike Matejka and McKay [2011] we provide necessary and sufficient conditions for our model, while unlike Ellis [2012] we provide conditions that are necessary and sufficient in finite data sets, making them applicable in practice.

A second decision theoretic approach to identifying optimal behavior in the face of costly attention is to examine choice over menus, in which attention costs are characterized by an aversion to contingent planning. de Oliveira *et al.* [2013] consider a model similar to ours in this setting.

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.

In the psychology literature, theories to which we are close in spirit are signal detection theory (Green and Swets [1966]) and categorization theory. A common feature is that the DM receives a signal and must choose the optimal action at each resulting posterior. These theories are connected to enormous experimental literatures in psychology that capture state dependent stochastic choice data. Unlike our model, signal detection theory generally fixes the attention strategy independent of incentives.

Despite the powerful psychological precedents, there is little experimental work on state dependent stochastic choice data within economics. One related paper is Cheremukhin *et al.* [2011], which uses a formulation similar to Matejka and McKay [2011] to estimate a rationally inattentive model of lottery choice. However they do not analyze state dependence in the resulting stochastic choice data.

9 Conclusions

We show that a general model of costly information acquisition is characterized by two simple and readily testable restrictions on state dependent stochastic choice data. We identify what can be recovered about information costs from such data. We provide experimental evidence that subjects do indeed adjust the information that they collect on the basis of the incentives inherent in their environment. Models that do not take this into account (such as signal detection theory and random utility models) fail to capture important aspects of the data.

We do not believe rational allocation of attention necessarily describes behavior in all circumstances. Indeed, one of the strengths of our approach is that it helps to identify when and how standard assumptions on information acquisition need to be relaxed. In this vein, we are currently exploring behavior of subjects facing important asymmetries in beliefs and in the cost of mistakes. In contrast, when the model does apply, one can test additional restrictions on the nature of costs. In this vein, we are currently exploring costs based on Shannon mutual information and various generalizations (see Caplin and Dean [2013]).

References

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

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.

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 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.

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.

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

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

R. Luce and P. Suppes. *Handbook of Mathematical Psychology Vol. III*, chapter Preference, utility, and subjective probability. Wiley, New York, 1965.

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.

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. CERGE-EI Working Papers wp442, The Center for Economic Research and Graduate Education - Economic Institute, Prague, June 2011.

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

Daniel McFadden. *Frontiers in Econometrics*, chapter Conditional logit analysis of qualitative choice behavior. Academic Press, New York, 1974.

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.

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.

Abraham Wald. *Sequential Analysis*. John Wiley and Sons, 1st edition, 1947.

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