Revealed Preference, Rational Inattention, and Costly Information Processing*

Andrew Caplin† and Mark Dean‡

December 2013- PRELIMINARY - DO NOT CIRCULATE

Abstract

We develop "revealed preference" tests for models of optimal information aquisition . The tests encompass rational inattention theory as well as 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. We find that subjects adjust the intensity and focus of their attention in response to incentives. Our tests provide quantitive confirmation that such adjustements are well-modeled as rationally responsive to costs.

1 Introduction

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

*An earlier version of this paper was circulated under the title "Rational Inattention and State Dependent Stochastic Choice". We thank Roland Benabou, Federico Echenique, Andrew Ellis, Daniel Martin, Filip Matejka, Alisdair McKay, Stephen Morris, Pietro Ortoleva, Daphna Shohamy, Severine Toussaert, Isabel Trevino, Laura Veldkamp and Michael Woodford for their constructive contributions. We also thank Samuel Brown for his exceptional research assistance.

†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

tance of information constraints has been increasingly recognized,[1] so an ever wider array of information gathering technology has been modelled : McCall [1970] considered the case of sequential search; Verrecchia [1982] the choice of variance of a normal signal; Sims [1998] an unrestricted choice of information structure with costs based on Shannon entropy. Many other technologies have been discussed in the literature.[2]

In this paper we introduce definitive behavioral conditions that cover a large class of costly information acquisition theories, including rational inattention, the selection of a signal from a pre-determined set, and sequential information acquisition.[3] We establish limits on what choice data can reveal about the costs of information. In an experimental implementation, we find that subjects adjust the intensity and focus of their attention in response to incentives. Our tests provide quantitative confirmation that such 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 unobservability of information acquisition costs motivates our approach. In the revealed preference spirit, our tests are necessary and sufficient for any arbitrary finite data set, and so can be readily applied in practice. The success or failure of these tests will identify the relative merits of specializing or amending the standard model.

Enriched choice data plays a central role in our tests. We utilize "state

[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]), buyers of second-hand cars focus their attention on the leftmost digit of the odometer (Lacetera *et al.* [2012]), while purchasers limit their attention to a relatively small number of websites when buying over the internet (Santos *et al.* [2012]).

[2] For example Reis [2006], van Nieuwerburgh and Veldkamp [2009], Woodford [2012], Gul *et al.* [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. Following up on this approach, Masatlioglu *et al.* [2012] and Manzini and Mariotti [2012] have mapped specific models search to their observable counterparts in choice data.

dependent" stochastic choice data, which described the decision maker's probability of choosing each available act in each (objective) state of the world.[4] While only recently introduced into revealed preference analysis (see Caplin and Martin [2011], henceforth CM13), it is standard in the econometric analysis of discrete choice. For example, Chetty *et al.* [2009] studies how choice distributions are impacted by an in principle observable state: the inclusion or exclusion of sales taxes in stated prices. He finds evidence of incomplete state awareness among buyers.[5] It is also easy to collect experimentally, as we describe below.

There are two conditions that render state dependent data consistent with optimal acquisition of costly information. First, a "no improving action switch" (NIAS) condition ensures that choices are optimal given what was learned about the state of the world, as in CM13. Second, a "no improving attention cycles" (NIAC) condition ensures that total utility cannot be raised by reassigning attention strategies across decision problems. While these conditions are obviously necessary, we show that they are also sufficient for the existence of a costly information representation.

We provide limits on what can be learned about information costs from state dependent stochastic choice. One cannot identify whether or not less informative strategies (in the sense of Blackwell) would have been more costly. One cannot identify whether or not it is feasible to mix attention strategies. One also cannot tell whether or not inattention is costless.[6] Yet there are readily computable limits on the costs of attention in any finite data set.

We implement our tests by experimentally eliciting state dependent stochastic choice data. Our experiments incentivize shifts in the intensity and the focus of attention. Our subjects are responsive to these shifts in incentives, for example discriminating more finely in their choices when learning is ex

[4]The key role of data enrichment has arised previously in our research 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.

ante more valuable due to the rich set of potentially available action choices. Our formal tests confirm that behavior is broadly in line while the costly information processing. Alternative theories in which learning is unresponsive to attentional incentives, are clearly rejected.[7]

Our paper is related to recent literature attempting to characterize 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. Manzini and Mariotti [2012] characterize a model of consideration sets using random choice. Ellis [2012] uses state dependent deterministic choice to study a model in which a decision make has a fixed set of signals to choose from. Masatlioglu *et al.* [2012] identify limited attention using standard choice data. de Oliveira *et al.* [2013] consider a more general model of attention, again using choice over menus as their data. Other recent work has considered the implications of unobserved information acquisition in dynamic (Dillenberger *et al.* [2012]) and strategic (Bergemann and Morris [2013], Penta [2012]) settings. More generally, our work fits into an ongoing resurgence in the use of revealed preference methods. Recent examples include Choi *et al.* [2007], Beatty and Crawford [2011] and Echenique *et al.* [2011].

Section 2 introduces the static model of costly information acquisition. Section 3 derives the testable implications for state dependent choice data. Section 4 provides the extension to sequential information acquisition. Section 5 establishes limits on identification. Section 6 details our experimental design, with results in section 7. Section 8 relates our work to the broader literature and outlines ongoing work. Section 9 concludes.

[7] In particular, we find monotonicity violations of the type suggested by Matejka and McKay [2011] that rule out standard random utility models.

2 A Model of Costly Information Acquisition

2.1 Decision Problems

We consider a decision making environment comprising a finite set of states of the world $\omega \in \Omega$ with cardinality M. We define $\Gamma = \Delta(\Omega)$ as the set of probability distributions over states. Given $\gamma \in \Gamma$, γ_ω denotes the probability of state ω. A decision maker (DM) has prior $\mu \in \Gamma$. As in CM13, we consider a DM with an expected utility function over a prize set. A decision problem consists of a finite set of available actions A from which the DM must choose. The state ω specifies precisely which prize corresponds to each such action. In the current analysis we treat the underlying utility function as known and suppress reference to the prize space. Instead we directly specify state dependent utility function $U : \Omega \times A \longrightarrow \mathbb{R}$ and let U_ω^a denote the utility of action a in state ω.[8] We let F be the set of all conceivable actions $a : \Omega \to \mathbb{R}^M$ with $\mathcal{F}$ comprising all non-empty finite subsets of F.

2.2 Behavior

Formally, the DM chooses an *attention strategy*. In the current section, we follow the rational inattention literature in treating such an attention strategy as a stochastic mapping from states of the world to subjective signals, the

[8] Throughout this paper, we assume that the DM's expected utility function and prior beliefs over states of the world are both known to the researcher - only attention strategies and costs are not directly observable. This assumption is in line with the focus of the paper, but is not central to our approach. By enriching the data set, we could recover beliefs and preferences from the choice data of the DM, and use these as a starting point for our representation. In order to recover utility, we could replace the "Savage style" acts 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 payoff are state independent). If we further add to our data set the choices of the DM over acts *before* the state of the world is determined (or at least in a situation in which they cannot exert any effort to determine that state) then we can also recover the DM's prior over objective states (again assuming expected utility maximization).

Alternatively, Caplin and Martin [2011] show how the conditions of the type presented in this paper could be extended to cover the case of unknown utility.

probabilities of which depend on the state of the world. We extend the model to sequential sampling in section 4.

Throughout the analysis we will be characterizing expected utility maximizers. Hence we identify each subjective signal with its associated posterior beliefs $\gamma \in \Gamma$, which is equivalent to the subjective information state the DM following the receipt of that signal. Having selected an attention strategy, the DM can condition choice of action only on these signals. For a particular prior, the set of feasible attention strategies is the set of stochastic mappings from objective information states to subjective signals that satisfy Bayes' law.[9]

Definition 1 *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 $\pi_\omega(\gamma) \equiv \pi(\omega)(\{\gamma\})$.

We use $\tilde{\Pi} \equiv \cup_\mu \Pi(\mu)$ to denote the set of attention strategies for all priors

Note that $\pi_\omega(\gamma)$ can be interpreted as the probability of signal γ conditional on state of the world ω.

There are costs of attention. An attention cost function maps attention strategies to the corresponding level of disutility. 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]), by allowing only partitional information structures (Ellis [2012]), or by allowing only normal signals (Verrecchia [1982]). To avoid triviality we assume that feasible attention strategies exist for all priors.

Definition 2 *An **attention cost function** is a mapping $K : \Gamma \times \tilde{\Pi} \to \bar{\mathbb{R}}$*

[9] Attention strategies are equivalent to temporal lotteries in models that capture preferences over the timing of the resolution of uncertainty (Kreps and Porteus [1978]).

with $K(\mu, \pi) \in \mathbb{R}$ for some $\pi \in \Pi(\mu)$ for all $\mu \in \Gamma$ and $K(\mu, \pi) = \infty$ for $\pi \notin \Pi(\mu)$. We let $\mathcal{K}$ denote the class of such functions.

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.

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 at 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:

$$\begin{aligned} G(A, \pi) &= \sum_{\gamma \in \Gamma(\pi)} \left[\sum_{\omega} \mu_\omega \pi_\omega(\gamma) \right] g(\gamma, A) \\ \text{where } g(\gamma, A) &= \max_{a \in A} \sum_{\omega} \gamma_\omega U_\omega^a. \end{aligned}$$

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,

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

2.3 State Dependent Stochastic Choice Data

The idealized data set that we use to test the model of rational inattention is state dependent stochastic choice data in a finite set of decision problems $D \subset \mathcal{F}$. Data of this general form is standard in psychometric research and can be readily gathered in the laboratory, as we demonstrate in section 6. To formalize, we define Q to be the set of mappings from Ω to probability distributions over F with finite support. Given $q \in Q$, we let q^a_ω 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, $F(q) = a \in F | q^a_\omega > 0$ for some $\omega \in \Omega$. For $A \in \mathcal{F}$, we define Q^A as all data sets with $F(q) \subset A$.

Definition 3 *A state dependent stochastic choice* ***data set*** (D, q) *comprises a finite set of decision problems* $D \subset \mathcal{F}$ *and a function* $q : D \to Q$*, with* $q(A) \in Q^A$*.*

A first requirement for existence of an attention cost function for which an optimal attention strategy generates all observed data is consistency of that strategy with the observed data for each observed decision problem A. We allow for mixed strategies $C : \Gamma(\pi) \to \Delta(A)$ in defining this form of consistency, with $C^a(\gamma)$ the probability of action $a \in A$ given $\gamma \in \Gamma(\pi)$. Following this we define the sought after representation.

Definition 4 *Attention strategy* $\pi \in \Pi$ *is* ***consistent with*** $A \in \mathcal{F}$ *and* $q \in Q^A$ *if there exists* $C : \Gamma(\pi) \to \Delta(A)$ *such that:*

1. *Final choices are optimal:*

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

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

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

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 rational inattention representation. The first establishes optimality of final choice given an attention strategy and applies to each decision problem separately. The second establishes optimality of the attention strategy and applies to the collection of decision problems. The key to our approach is the fact that state dependent stochastic choice data can be used to reveal much about a DM;s attention strategy.

3.1 Minimal Attention Strategies

If a DM is rationally inattentive, 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 $q \in Q$, define the* ***revealed posteriors*** $r(q) : F(q) \to \Gamma$ *by,*

$$[r(q)(a)]_\omega \equiv r_\omega^a(q) = \frac{\mu_\omega q_\omega^a}{\sum\limits_{\upsilon} \mu_\upsilon q_\upsilon^a}.$$

The revealed posterior $r_\omega^a(q)$ 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 define their attention strategy.[10] 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 state dependent stochastic

[10] As they would do if more informative information structures are more costly.

choice data. The probability of a 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 $q \in Q$, $\omega \in \Omega$, and $\gamma = r^a_\omega(q)$ for some $a \in F(q)$, define the* ***minimal attention strategy*** *$\bar{\pi}^q \in \Pi$ to satisfy,*

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

A key observation is that any attention strategy consistent with the data must be weakly more informative than the minimal attention strategy, in the sense of statistical sufficiency. 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 $|\Gamma(\rho)| \times |\Gamma(\pi)|$ stochastic 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 $q \in Q$, then it is sufficient for $\bar{\pi}^q$.*

Blackwell's theorem establishes the equivalence of the statistical notion of "more informative than" (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 is that there are no improving wholesale switches of action. Caplin and Martin [2011] show that this condition characterizes Bayesian behavior regardless of the rationality of the attention given. It specifies that, when one identifies in the data the revealed posterior associated with any chosen act, this action must be optimal at that posterior. The strategic analog is employed by Bergemann and Morris [2013] in characterizing Bayesian correlated equilibria.

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

$$\sum_{\omega} r_{\omega}^{a}(q) U_{\omega}^{a} \geq \sum_{\omega} r_{\omega}^{a}(q) U_{\omega}^{b},$$

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 the total gross utility can be increased by reassigning attention strategies across decision problems. The following example illustrates a violation of this condition.

Consider again the decision problem above 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 as follows (using the choice set A as an argument),

$$\begin{aligned} q_1^a(A) &= \frac{2}{3} = 1 - q_1^b(A); \\ q_2^a(A) &= \frac{1}{3} = 1 - q_2^b(A). \end{aligned}$$

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 the corresponding state dependent data set,

$$\begin{aligned} q_1^a(B) &= \frac{3}{4} = 1 - q_1^b(B); \\ q_2^a(B) &= \frac{1}{4} = 1 - q_2^b(B). \end{aligned}$$

The specified data looks problematic with respect to rational inattention. Act 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 rational inattention 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,

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

We conclude that gross benefit is maximized by the assignment of minimal attention strategies to decision problem 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 $A_J = A_1$,

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

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

The NIAC condition is analogous with acyclicity conditions that are standard in revealed preference theory (such as the Strict Axiom of Revealed Preference). Consider the two decision problem case described in equation 1. A further rearrangement of this condition implies that

$$G(A, \bar{\pi}^A) - G(A, \bar{\pi}^B) + G(B, \bar{\pi}^B) - G(B, \bar{\pi}^A) \geq 0$$

We can interpret $G(A, \bar{\pi}^A) - G(A, \bar{\pi}^B) < 0$ as $\bar{\pi}^B$ being "revealed more costly" than $\bar{\pi}^A$: $\bar{\pi}^B$ would have given higher gross value in decision problem A than would $\bar{\pi}^A$, so the fact that it was not chosen must mean that it is more costly. Thus, this condition can be interpreted as saying that if $\bar{\pi}^B$ has been revealed more costly than $\bar{\pi}^A$, we cannot also have that $\bar{\pi}^A$ is revealed more costly that $\bar{\pi}^B$. In fact, the condition says something more than that because, unlike in the standard revealed preference exercise, we have information about *how much* more costly $\bar{\pi}^B$ is than $\bar{\pi}^A$ in terms of expected utility. Therefore if we have data revealing that $\bar{\pi}^B$ is more costly than $\bar{\pi}^A$ by at least some amount x, we cannot also have information that implies that this difference must be less that x.

3.4 Characterization

NIAC and NIAS together are necessary and sufficient for (D, q) to have a rational inattention 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 rational inattention representation if and only if it satisfies NIAS and NIAC.*

4 Sequentially Rational Inattention

So far we have considered static models of costly attention. Yet 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 how learning evolved before choice took place is not directly observable. It turns that this model has no additional behavioral restrictions on the data: NIAS and NIAC are still necessary and sufficient for data to be consistent with optimal sequential attention.

4.1 Sequential Attention and Choice Strategies

We fix a time interval within which a decision is to be made. This decision period is sub-divided into a sequence of $T \geq 1$ sub-periods in each of which the DM decides how much attention to give to the decision at hand. The decision on when and how to pay attention and what to choose is assumed to be sequentially rational. We assume that there is no discount factor in operation during this decision period, so that the sequence of attention and action choice decisions is made to maximize the net undiscounted value of final prize utility less sequential attention costs. The sequence of attentional inputs and the actual decision time are both flexible, and it is assumed that neither is observed in the choice data.

The DM at the start of period 0 is endowed with prior beliefs $\mu \in \Gamma$. The first decision is whether or not to stop and choose at this time. If so, the stopping time is $\tau = 0$ and we define stopping set $S_0 = \mu$ and G_0 is empty. If not we define $G_0 = \mu$ and S_0 is 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 DM selects a deterministic stopping rule. We define $G_1 \subset \Gamma(\pi_1)$ to be all continuation posteriors, with $S_1 = \Gamma(\pi_1)/G_1$ the corresponding stopping set triggering a period 1 choice.

The process iterates from this point forward. We allow for history dependence by defining the continuation set G_t for $t \geq 0$ to comprise of sequences of posteriors $\gamma^t = (\gamma_1, .., \gamma_t) \in \Gamma^t$. 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 function.

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

The ensuing continuation set and attention strategy are correspondingly defined:

$$\begin{aligned} G_{t+1} &\subset \left\{\gamma^{t+1} \in \Gamma^{t+1} | \gamma^t \in G_t, \gamma_{t+1} \in \Gamma(\pi_t(\gamma_t))\right\} \equiv \Gamma^{t+1}(\pi^t); \\ S_{t+1} &\subset \Gamma^{t+1}(\pi^t)/G_{t+1}. \end{aligned}$$

The above fully specifies a sequential attention strategy. We let $\Sigma(\mu)$ be the set of such strategies, with generic element $\sigma \in \Sigma(\mu)$. For $t \leq \tau$ we let $\Sigma^t(\mu)$ comprise all pairs (π_s, G_s) for $1 \leq s \leq t$, with generic element $\sigma^t \in \Sigma^t(\mu)$. Note that a strategy induces a probability distribution over sequences of posteriors. Let $\rho^\sigma(\gamma^t)$ be the overall probability of posterior sequence γ^t given strategy σ, with $\rho^\sigma_\omega(\gamma^t)$ being the probability of this posterior in the state $\omega \in \Omega$,

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

4.2 Evaluating Strategies

We turn to the evaluation of a choice strategy σ. We assume that choices are made optimally given posteriors. Thus, when faced with decision problem A, for each $\gamma^t \in S_t$, the decision maker will receive utility $g(\gamma_t, A)$. 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 9 *Given $\mu \in \Gamma$ we define period attention cost functions $E : \Gamma^t \times \tilde{\Pi} \to \bar{\mathbb{R}}$ for $1 \leq t \leq T$ such that $E(\gamma^t, \pi)$ is the cost of using attention strategy π following sequence of posteriors γ^t. The cost of infeasible strategies with $\pi \notin \Pi(\gamma_t)$ is infinite. We let $\mathcal{E}$ denote the class of such functions.*

The above covers all standard 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 current posterior without changing in any way the ensuing analysis.[11]

We define a strategy σ as sequentially rational for decision problem A if it solves the following problem:

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

where $\gamma^0 \equiv \mu$.

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

4.3 SCI Representations

Our goal is to identify all data sets (D, q) that can be rationalized by $E \in \mathcal{E}$ as consistent with sequentially optimal behavior in the face of costly attention.

Definition 10 *Given decision problem $A \in \mathcal{F}$, $\sigma \in \Sigma(\mu)$ is* ***consistent with*** *$q \in Q$ if there exists a choice function $C : \Gamma \to \Delta(A)$ such that:*

1. *Final choices are optimal: given $1 \leq t \leq \tau$, $\gamma^t \in S_t$ such that $\gamma_t = \gamma$,*

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

$a \in A$ with $C^{\alpha}(\gamma) > 0$, and $b \in A$,

$$\sum_{\omega} \gamma_{\omega} U_{\omega}^{a} \geq \sum_{\omega} \gamma_{\omega} U_{\omega}^{b}.$$

2. The attention and choice functions match the data,

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

We can now define what it means for a data set to admits a sequential costly information representation.

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, the result holds because we can reduce the dynamic problem of sequential choices to a static problem of choice of strategy, due to time consistency.

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

5 Recoverability

In this section we establish limits on recoverability 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 recoverability by characterizing all cost functions that can rationalize a given data set.

5.1 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 (see also de Oliveira *et al.* [2013] and Yang [2011]).[12]

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

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

Not all restrictions on the form of the cost function can be so readily absorbed. We show in the Appendix that there are data sets satisfying NIAS and NIAC yet for which there exists no cost function that produces a rational inattention representation with a cost function that is strictly monotonic with the informativeness of the information structure (i.e. if π is sufficient for π' but π' is not sufficient for π, then $\bar{K}(\pi) > \bar{K}(\pi')$).

5.2 Mixture Feasibility

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

[12] 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.

Definition 12 *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. 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,

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

5.3 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 $\pi_\omega = 1$ for $\omega \in \Omega$.

Attentional cost function $K \in \mathcal{K}$ satisfies **normalization** if it is non-negative where real-valued, with $K(I) = 0$.

5.4 Theorem 2

Theorem 2 states that, whenever a rational inattention 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 rational inattention 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 some utility function if and only if they can be rationalized by a non-satiated, continuous, monotone, and concave utility function.

5.5 Recoverability

Theorem 1 tells us the conditions under which there exists an attentional cost function that will rationalize the SDSC data. We now provide conditions on the set of all such cost functions. We restrict ourselves to cost functions that satisfy weak monotonicity, so that we can treat minimal attention strategies as optimal. This means directly that, for all $A \in D$ and $\pi \in \Pi$,

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

The key question is what this condition implies for cost differences between between distinct minimal attention strategies in decision problems $A, B \in D$. To answer this we consider cycling these strategies among decision problems heading from A to B and also from B to A.

Consider first the direct switch from A to B. Since $\bar{\pi}^A$ was chosen in A and $\bar{\pi}^B$ was not,

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

Conversely, since $\bar{\pi}^B$ was chosen in B and $\bar{\pi}^A$ was not,

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

Potentially tighter bounds can be placed by considering all sequences of revealed attention strategy. Consider the corresponding inequalities in string $A_1 ... A_n \in D$ with $A_1 = A$ and $A_n = B$, where we use strict monotonicity to treat minimal attention strategies as optimal and also use them in representing maximum achievable utility,

$$\begin{aligned}
K(\bar{\pi}^{A_1}) - K(\bar{\pi}^{A_2}) &\leq G(A_1, \bar{\pi}^{A_1}) - G(A_1, \bar{\pi}^{A_2}); \\
K(\bar{\pi}^{A_2}) - K(\bar{\pi}^{A_3}) &\leq G(A_2, \bar{\pi}^{A_1}) - G(A_2, \bar{\pi}^{A_3}); \\
&\quad .... \\
K(\bar{\pi}^{A_{n-1}}) - K(\bar{\pi}^{A_n}) &\leq G(A_{n-1}, \bar{\pi}^{A_{n-1}}) - G(A_{n-1}, \bar{\pi}^{A_n}).
\end{aligned}$$

Summing these inequalities in light of weak monotonicity yields,

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

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

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

The above three conditions characterize all restrictions that can be placed on a weakly monotonic cost function that allows it to represent a data set satisfying NIAS and NIAC. Note first that NIAS and NIAC imply that the lower bound on $K(\bar{\pi}^A) - K(\bar{\pi}^B)$ cannot be higher than the upper bound, so that such a cost function exists. Note also that if one considers cost functions that satisfy inattention is free, one may be able to place absolute numerical restrictions on the cost function. Specifically, if there is a decision problem $A \in D$ in which attention is of no value (all acts have state independent payoffs), we know that the minimal attention strategy has cost zero. Applying all above inequalities from this decision problem bounds the level of costs. A related comment is that 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 on costs can be placed on models in which the data is generated by a finite set of possible attention strategies.

6 Experimental Design

6.1 Design

We introduce an experimental design that produces state dependent stochastic choice data. Subjects are shown a screen on which there are displayed 100 balls, some of which are red and some of which are blue. The state of the world is determined by the number of red balls on the screen. Prior to seeing the screen, subjects are informed of the probability distribution over states. They then choose among actions whose payoffs are state dependent. There is no external limit (such as a time constraint) on a subject's ability to collect information about the state of the world, nor any extrinsic cost to the subject of gathering information. Therefore the extent to which subjects fail to discern the true state of the world is due to their unwillingness to trade cognitive effort for monetary reward.

A decision problem is defined by the set of available actions, as it is in section 2.1. A subject faces each decision problem 50 times.[13] We estimate state dependent stochastic choice functions at the individual and aggregate

[13] 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.

level using the observed frequency of choosing each act 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 we would expect subjects to actively alter their attention in response to incentives. This allows us to meaningfully test our NIAS and NIAC conditions. It also allows us to show that models of flexible inattention capture important aspects of the data that missing from models that do not have this feature.

We conducted three experiments. Each experiment was performed by between 24 and 46 subjects recruited from the New York University student population.[14] Each subject answered 200 questions as well as 1 practice question. At the end of the 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 about 45 minutes to complete a session. Instructions are included in the supplemental material.

6.2 Experiment 1: Changes in Attentional Effort

Experiment 1 tests whether subjects increase overall attentional effort as incentives increase. It comprises of 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.[15] Table 1 describes the available actions in the four decision

[14] 46 subjects took part in experiment 1, 24 in experiment 2, and 45 in experiment 3. Each subject took part in only one experiment.

[15] 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.

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_3^a	U_4^a	U_1^b	U_2^b	U_3^b	U_4^b
5	1	0	10	0	0	1	0	10
6	10	0	1	0	0	10	0	1
7	1	0	1	0	0	1	0	1
8	10	0	10	0	0	10	0	10

Experiment 1 allows us to differentiate between a model in which attention responds to incentives, and one in which subjects make optimal choices on the basis of a fixed attention.[16] While more typical in psychology (for example signal detection theory (SDT) (Green and Swets [1966])), such models have also 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. This puts further restrictions of the data in this experiment. For any information structure, a is preferable in posteriors states such that $\gamma_1 > 0.5$, while b is preferable for posteriors such that $\gamma_1 < 0.5$, regardless of the value of x. Thus, if attention does not change as a function of incentives, neither should choice behavior vary across the decision problems in this experiment.[17]

6.3 Experiment 2: Focussing Attention

In experiment 2 we vary the states on which attention 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

[16] Equivalently, one could think of this as a model in which the choice of information structure is made before the decision problem is revealed.

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

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 5), states 3 and 4 (problem 6), neither (problem 7), or both (problem 8), as described in table 2.

6.4 Experiment 3: Informational Spillovers

Experiment 3 is designed to test whether the introduction of a new act to the choice set can alter the attention strategy in such a way as to increase the probability of choosing a previously available alternative. It is based on an example from Matejka and McKay [2011], and described in table 3. It consists of two equally likely states (49 and 51 red balls). Decision problem 9 consists of two actions, a (which pays the same amount in both states) and b (which pays slightly more is state 2 and slightly less in state 1). Decision problems 10-12 add a further action c, which pays significantly more in state 1 and significantly less in state 2.

Table 3: Experiment 3

	Payoffs					
DP	U_1^a	U_2^a	U_1^b	U_2^b	U_1^c	U_2^c
9	23	23	20	25	n/a	n/a
10	23	23	20	25	30	10
11	23	23	20	25	35	5
12	23	23	20	25	40	0

When act 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 more likely, then it is optimal to choose act b, producing a violation.

This experiment allows us to differentiate between costly information processing from random utility models (RUMs) (McFadden [1974], Loomes and Sugden [1995], 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.[18] Understanding the implications of RUMs in our setting is non-trivial, as they typically conditions out all observable states, as a result giving rise to state independent choice. One must therefore take a stand on what the DM knows about the true state of the world when they make their choice. The most flexible case which does not nest flexible inattention is one in which the DM receives a fixed signal about the state of the world, before a utility function is randomly drawn. This model nests as special cases RUMs in which the agent is uninformed (i.e. receives no signal) or fully informed.

A key property implied by the fixed-information RUM is monotonicity. This axiom states that the addition of a new act to the set of available choices cannot increase the probability that one of the pre-existing options will be chosen (Gul and Pesendorfer [2006] and Luce and Suppes [1965]).

Monotonicity Axiom Given $(\mu, A) \in \Gamma \times \mathcal{F}$, $h \in F \backslash A$ and $m \in \Omega$,

$$q_m^f(\mu, A) \geq q_m^f(\mu, A \cup h).$$

Monotonicity is violated by a model of costly inattention that exhibits information spillovers of the type described above: the introduction of act c increases the probability of choosing act b in state 2. Such evidence would therefore imply that RUM models that do not allow for flexible information acquisition are missing an important aspect of behavior.

[18] 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.

6.5 Attention is Limited and Flexible

Before implementing the NIAS and NIAC, we first provide evidence that subjects are neither fully attentive or completely inattentive. We also confirm the presence of the attentional flexibility that SDT and standard RUM models ignore.

The first point to observe is that the experiments produced 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 act cases of experiments 1 and 2 (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 (Fishers 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 choice.

Our data also rules out the fixed-signal SDT model. As discussed in section 6.2, this model does not allow for subjects to make better decisions as incentives increase in this symmetric case. As shown in figure 2b below, Our aggregate data clearly exhibits such a changes, 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.

Experiment 3 that provides evidence against a fixed-signal RUM. Table 4 shows that the 44 subjects that took part in experiment demonstrate clear violations of monotonicity. The introduction of action c increases the probability of choosing action b in state 2 from 23% to an average of 35% across decision problems 9-12, significant at the 1% level. At the individual level, 51% of subjects show a significant violation of monotonicity of the type predicted by rational inattention theory at the 10% level.

Table 4		
Decision problem	$q_1(b)$	$q_2(b)$
9	17%	23%
10	15%	31%
11	12%	33%
12	13%	39%

6.6 NIAS and NIAC Tests

6.6.1 Experiment 1

In the two state/two action set up of experiment 1, NIAS implies existence of a cutoff posterior probability of state 1 that determines the optimal act. For posteriors beliefs above 0.5, action a is optimal, while for lower posteriors, action b is optimal. This cutoff is shown in figure 2a together with the estimated posteriors associated with the choice of action a and action b at the aggregate level (treating all data as if it was generated by a single subject). 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 3a shows, monetary losses due to NIAS violations are small.[19] As a benchmark, these losses are compared to those that would have been observed from a population of subjects choosing at random.[20] The observed distribution is significantly different from the simulated distribution at the 0.01% level

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

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

(Kolmogorov-Smirnov test).

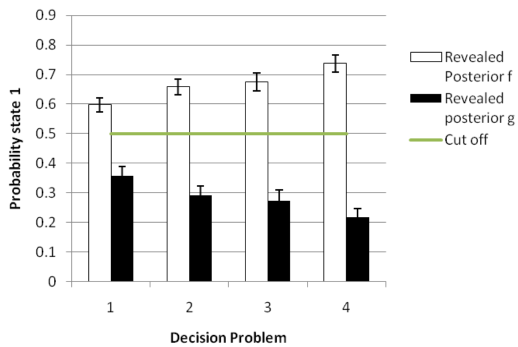


Figure 2a - NIAS Experiment 1

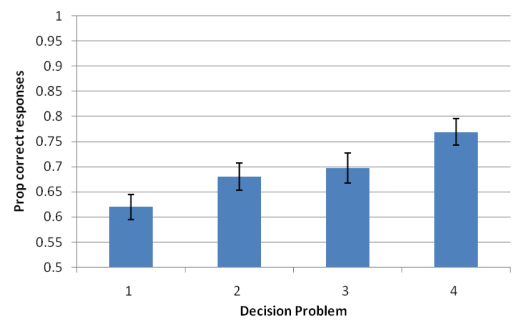
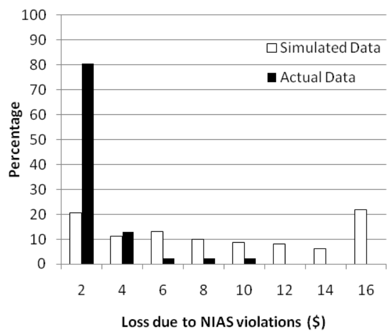


Figure 2b - NIAC Experiment 1

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. This expression has a natural interpretation. The first term is the change in the probability of choosing the correct action in state 1 multiplied by the change in the benefit of choosing the right action - i.e. the difference between the payoff of action a and b in that state. The second term is the change in the probability of choosing the right action in state 2 multiplied by the benefit of so doing. As $\Delta(U_1^a - U_1^b) = \Delta(U_2^b - U_2^a)$ for each pair of decision problems in experiment 2, equation 5 implies that $\tau_1 + \tau_2$, the total probability of choosing the right action, should be monotonic in rewards. Figure 2b 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.



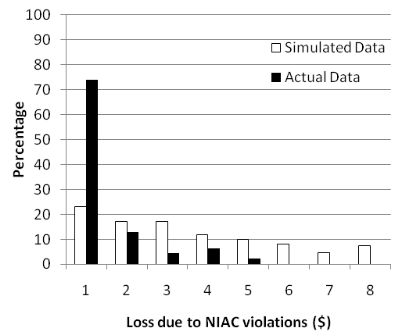


Fig 3a: NIAS losses
Experiment 1

Fib 3b: NIAC losses
Experiment 1

At the individual level 83% of subjects show no significant violation of the NIAC condition. Losses resulting from NIAC violations at the individual level are small, as shown in figure 3b. This figure plots the distribution of actual surplus minus the maximal surplus possible by reassigning attention strategies to decision problems for each individual.[21] The NIAC condition demands this number to 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

6.6.2 Experiment 2

Experiment 2 uses a four act set up to test whether subjects can focus their attention where it is most valuable. The NIAS conditions are,

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

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

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, the losses amongst those that do not are again small and significantly different from the random benchmark at the 0.01% level.

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.[22] More specifically, NIAC implies six inequalities based on binary comparisons of the 4 decision problems. Table 5 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 5: NIAC conditions for experiment 2			
Condition	LHS	RHS	P
$\tau_1^6+\tau_2^6+\tau_3^5+\tau_4^5 \geq \tau_3^6+\tau_4^6+\tau_1^5+\tau_2^5$	72.8	64.9	0.01
$\tau_3^5+\tau_4^5 \geq \tau_3^7+\tau_4^7$	68.2	63.3	0.38
$\tau_1^6+\tau_2^6 \geq \tau_1^7+\tau_2^7$	77.3	66.9	0.02
$\tau_1^8+\tau_2^8 \geq \tau_1^5+\tau_2^5$	74.8	63.7	0.02
$\tau_3^8+\tau_4^8 \geq \tau_3^6+\tau_4^6$	69.1	66.3	0.33
$\tau_1^8+\tau_2^8+\tau_3^8+\tau_4^8 \geq \tau_1^7+\tau_2^7+\tau_3^7+\tau_4^7$	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.

[22] The precise condition is

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

6.6.3 Experiment 3

For experiment 3, NIAS defines regions of acceptable posteriors for the choice of each act in each decision problem. Table 6 describes these regions, and the aggregate posteriors observed in the data

Table 6: NIAS conditions for experiment 3

DP	Range γ_1			Aggregate		
	b	a	c	γ_1^b	γ_1^a	γ_1^c
9	$[0, 40\%]$	$[50\%, 100\%]$	n/a	43	52	
10	$[0, 40\%]$	$[50\%, 65\%]$	$[65\%, 100\%]$	32	50	85
11	$[0, 40\%]$	$[50\%, 60\%]$	$[60\%, 100\%]$	27	51	82
12	$[0, 40\%]$	$[50\%, 57.5\%]$	$[57.5\%, 100\%]$	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 (table s1).

Applying bilateral NIAC to experiment 3 implies the following ranking on $q_1(c) - q_2(c)$,

$$q_1^{12}(c^{12}) - q_2^{12}(c^{12}) \geq q_1^{11}(c^{11}) - q_2^{11}(c^{11}) \geq q_1^{10}(c^{10}) - q_2^{10}(c^{10}).^{23}$$

In the aggregate data this ordering holds. The values of $q_1(c) - q_2(c)$ are 29%, 18% and 18% for decision problem 12, 11 and 10 respectively. DP 12 is significantly different from DP 11 and DP 10, though DP10 an DP11 are not significantly different from each other. At the individual level, 65% of subjects show no significant violations of NIAC.

[23] If it were the case that posterior beliefs when a is chosen in decision problem 9 are such it would be preferable to choose c^{10} (if available) we additionally have the restriction,

$$q_1^{10}(c^{10}) - q_2^{10}(c^{10}) \geq q_1^{9}(c^{9}) - q_2^{9}(c^{9}).$$

However this is not the case is our aggregate data.

7 Existing Literature

Much of the work on rational inattention in economics can be traced back to Sims [1998, 2003] which characterize the behavioral impact of constraints on information processing in linear quadratic control problems.[24] The rate of information flow is measured by Shannon mutual information. Sims [2003] shows that such a constraint generates behavior similar to that generated by an agent who observes the state of the world only noisily. However, the type of noise is determined endogenously, based on the incentives in the environment. Following this paper, mutual information constraints have been incorporated in an increasing number of economic settings, including consumption-savings problems (Sims [2006], Tutino [2008], and Mackowiak and Wiederholt [2010]), pricing problems (Mackowiak and Wiederholt [2009], Matejka [2010], Martin [2013]), monetary policy (Paciello and Wiederholt [2011]), and portfolio choice (Mondria [2010]).

In part, the focus on mutual information to measure the cost of information is justified by its central position in the information theory literature. The Shannon mutual information of two random variables is related to the expected length in bits of the optimally encoded signal needed to generate one from the other. It also has an axiomatic characterization which shows that information costs must be of this form if they are to obey certain intuitive properties (see for example Csiszár [2008]). Shannon mutual information costs also have interesting properties from an economic standpoint. As discussed in the text, Matejka and McKay [2011] demonstrate a strong relationship between mutual information based rational inattention and logit-style random choice. Cabrales *et al.* [2013] demonstrate a further interesting link between mutual information and economic behavior. They consider a "ruin averse" investor facing a class of no-arbitrage investment problems and show that the ranking of information structures based on willingness to pay is equivalent to that provided by mutual information.

[24] Although the study of costly information aquisition in economics goes back much further - for example Stigler [1961], Marschak [1971] and Milgrom [1981].

While much of the rational inattention literature has focussed on mutual information costs, a variety of other cost functions and constraints have been studied. Woodford [2012] points out that mutual information does not imply that less attention will be paid to rare events (as such attention is cheap in expectation), in violation of experimental results due to Shaw and Shaw [1977]. He therefore proposes an alternative measure in which the cost of an information structure is evaluated according to the related concept of the Shannon capacity. Gul *et al.* [2012] consider the behavior of households who are restricted to having "crude" consumption plans i.e. plans that are restricted to having at most n realizations. van Nieuwerburgh and Veldkamp [2009] consider a more general information cost function, based on the distance between prior and posterior variance. Saint-Paul [2011] considers the case in which DMs face Shannon cost, but can only choose discrete policy functions (essentially combining the approaches of Sims [2003] and Gul *et al.* [2012]). Reis [2006] considers the case of a binary information choice: in any given period either attention can be paid, and the state is fully revealed, or not, in which case no information is gathered. Even many of the articles that use mutual information costs effectively restrict the decision maker to choose Gaussian signals, implying additional constraints (see Sims [2006] for a discussion).

A key strength of our approach is that our model nests all of the above costs functions. The costs of allowable 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 rational inattention models currently in use.[25]

A recent wave of decision theoretic literature shares the goal of capturing the observable implications of inattention, both rational and otherwise. Closest in spirit to our work is Ellis [2012], who works with a data set similar to ours - state dependent choice functions. Ellis [2012] asks under what conditions can such data be rationalized by a model in which a DM has a set of available partitions on the state space, and selects the best partition for the

[25] Note that we consider only the instrumental value of information, not any intrinsic value that information might have as in Grant *et al.* [1998].

decision problem at hand. The characterization is based on the identification of cells in the partition with all objective states in which the same choice is made - an approach similar to that taken in our paper. This allows for the identification of the preferences revealed by choices.

There are three key differences between the theoretical section of our paper and Ellis [2012]. On the one hand, he places weaker requirements on the data: unlike our approach, the DM's utility function and prior beliefs are derived from behavior rather than directly observable. On the other hand he considers a more restrictive class of information restrictions: DMs effectively face a cost function which is zero for allowable partitions and infinity for all other information structures. This restriction rules out any stochasticity in choice, as well as many commonly used information cost functions (such as those based on Shannon mutual information). Moreover, the conditions of Ellis [2012] require the data to be observed from essentially all decision problems to be both sufficient and necessary. In contrast, our tests work on data collected from any arbitrary collection of decision problems.

A second decision theoretic approach to identifying rational attention is to examine choice over menus. Ergin and Sarver [2010] consider a model in which a DM makes choices *within* choice sets by optimally selecting a partition on (subjective and unobservable) states of the world, then choosing the best action conditional on that partition.[26] They characterize the implications of such a model for choices *between* choice sets. Costly contemplation is characterized by an aversion to contingent planning: an agent would prefer to find out which set they are choosing from and then choose from that set, rather that have to make contingent plans. Mihm and Ozbek [2012] extend this approach to the case in which there are observable states of the world, resulting in a representation similar to that considered in this paper.

Our work forms part of an ongoing project aimed at characterizing choice behavior when the internal information state of the agent is not directly observable. Block and Marschak [1960] early on stressed the difficulty in separating out theories of stochastic choice. van Zandt [1996] provided an explicit neg-

[26] Ortoleva [2013] provides an alternative model of 'thinking aversion'.

ative result in this regard, showing that any choice behavior is rationalizable in a model that allows for unobserved costly information acquisition if the state of the world is not observable. Caplin and Dean [2011] and Caplin *et al.* [2011] consider the case of sequential information search, using an extended data set to derive behavioral restrictions of search of this kind as well as of satisficing behavior. Gabaix *et al.* [2006] use Mouselab data to test a near-optimal model of sequential information search. Caplin and Martin [2011] introduce the NIAS condition to characterize subjective rationality in a single decision problem. Masatlioglu *et al.* [2012] characterize "revealed attention", using the identifying assumption that removing an unattended item from the choice set does not affect attention. **?** consider a dynamic problem in which the DM receives information in each period which is externally unobservable, characterizing the resulting preference over menus. Fudenberg and Strzalecki [2012] also consider a dynamic problem, characterizing dynamic stochastic choice rules that are consistent with rational inattention and Shannon mutual information costs.

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.[27] The chief distinction is that, unlike the rational inattention model, signal detection theory generally fixes the attention strategy independent of the incentive to learn implied by the decision making environment. These models are therefore a special case of our general formulation.

Despite the powerful psychological precedents, there is little experimental work on state dependent stochastic choice data within economics. There is so far as we know no work in either field that tests NIAC and NIAS directly. One related paper is Cheremukhin *et al.* [2011], which uses a formulation similar to Matejka and McKay [2011] to estimate a rationally inattentive model on subject's choice over lotteries. However they do not analyze the state

[27] We do not attempt to summarize the literature here - see Verghese [2003] for a review.

dependence in the resulting stochastic choice data.

8 Conclusions

As economists increasingly focus on attentional constraints, so the importance of rational inattention theory has grown. We characterize a general model of rational inattention which encompasses all models currently in the literature. The necessary and sufficient conditions are simple and readily testable. We find the model to do a qualitatively good job of explaining subject behavior in a simple experimental implementation. In contrast, traditional random utility models fail to capture important data features.

In addition to further investigating the comparison with random utility models, we are currently exploring the behavioral content of more structured models of attention costs, in particular the Shannon model.[28] We are also exploring the implications of the nature of attention costs for economic applications.

[28] 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. Robust predictions in games with incomplete information. *Econometrica*, 81(4):1251–1308, 07 2013.

Antonio Cabrales, Olivier Gossner, and Roberto Serrano. Entropy and the value of information for investors. *American Economic Review*, 103(1):360–77, February 2013.

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

Andrew Caplin and Mark Dean. Rational inattention, entropy, and choice: The posterior-based approach. Mimeo, Center for Experimental Social Science, New York University, 2013.

Andrew Caplin and Daniel Martin. A testable theory of imperfect perception. NBER Working Papers 17163, National Bureau of Economic Research, Inc, June 2011.

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

Syngjoo Choi, Raymond Fisman, Douglas Gale, and Shachar Kariv. Consistency and heterogeneity of individual behavior under uncertainty. *American Economic Review*, 97(5):1921–1938, December 2007.

Imre Csiszár. Axiomatic Characterizations of Information Measures. *Entropy*, 10:261–273, 2008.

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.

Federico Echenique, Sangmok Lee, and Matthew Shum. The money pump as a measure of revealed preference violations. *Journal of Political Economy*, 119(6):1201 – 1223, 2011.

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

Haluk Ergin and Todd Sarver. A Unique Costly Contemplation Representation. *Econometrica*, 78(4):1285–1339, 2010.

Drew Fudenberg and Tomasz Strzalecki. Recursive stochastic choice. Mimeo, Harvard University, 2012.

Xavier Gabaix, David Laibson, Guillermo Moloche, and Stephen Weinberg. Costly information acquisition: Experimental analysis of a boundedly rational model. *American Economic Review*, 96(4):1043–1068, September 2006.

Simon Grant, Atsushi Kajii, and Ben Polak. Intrinsic preference for information. *Journal of Economic Theory*, 83(2):233–259, December 1998.

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.

Faruk Gul, Wolfgang Pesendorfer, and Tomasz Strzalecki. Behavioral competitive equilibrium and extreme prices. Mimeo, Princeton University, 2012.

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

D.M. Kreps and E.L. Porteus. Temporal resolution of uncertainty and dynamic choice theory. *Econometrica*, 46(1):185–200, 1978.

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

Graham Loomes and Robert Sugden. Incorporating a stochastic element into decision theories. *European Economic Review*, 39(3-4):641–648, April 1995.

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.

Bartosz Mackowiak and Mirko Wiederholt. Optimal sticky prices under rational inattention. *American Economic Review*, 99(3):769–803, June 2009.

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

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

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

Jacob Marschak. Economics of information systems. *Journal of the American Statistical Association*, 66(333):192–219, March 1971.

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

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

Filip Matejka and Alisdair McKay. Rational inattention to discrete choices: A new foundation for the multinomial logit model. CERGE-EI Working Papers wp442, The Center for Economic Research and Graduate Education - Economic Institute, Prague, June 2011.

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

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

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

Maximilian Mihm and M. Kemal Ozbek. Decision making with rational inattention. Working paper, Social Science Research Network, 2012.

Paul R Milgrom. Rational expectations, information acquisition, and competitive bidding. *Econometrica*, 49(4):921–43, June 1981.

Jordi Mondria. Portfolio choice, attention allocation, and price comovement. *Journal of Economic Theory*, 145(5):1837–1864, September 2010.

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

Pietro Ortoleva. The price of flexibility: Towards a theory of thinking aversion. *J. Economic Theory*, 148(3):903–934, 2013.

Luigi Paciello and Mirko Wiederholt. Exogenous information, endogenous information and optimal monetary policy. Technical report, 2011.

Antonio Penta. Higher order uncertainty and information: Static and dynamic games. *Econometrica*, 80(2):631–660, 03 2012.

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.

Gilles Saint-Paul. A "quantized" approach to rational inattention. TSE Working Papers 10-144, Toulouse School of Economics (TSE), 2011.

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.

M. L. Shaw and P. Shaw. Optimal allocation of cognitive resources to spatial locations. *J Exp Psychol Hum Percept Perform*, 3(2):201–211, May 1977.

Christopher A. Sims. Stickiness. *Carnegie-Rochester Conference Series on Public Policy*, 49(1):317–356, December 1998.

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

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

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

Antonella Tutino. The rigidity of choice: Lifecycle savings with information-processing limits. Technical report, 2008.

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

Timothy van Zandt. Hidden information acquisition and static choice. *Theory and Decision*, 40:235–247, 1996.

Preeti Verghese. Visual search and attention: A signal detection theory approach. *Neuron*, 31(4):523–535, 2003.

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.

Ming Yang. Coordination with rational inattention. Technical report, 2011.

9 Appendix A: Proofs

9.1 Lemma 1

Lemma 1 If $\pi \in \Pi$ is consistent with $A \in \mathcal{F}$ and $q \in Q$, then it is sufficient for $\bar{\pi}^q$.

Proof. Let $\rho \in \Pi$ be an attention strategy that is consistent with $q \in Q$ in decision problem A. First, we list in order all distinct posteriors $\eta^i \in \Gamma(\rho)$ for $1 \leq i \leq |\Gamma(\rho)|$. Given that ρ is consistent with q, there exists a corresponding optimal choice strategy $C : \{1, ..., I\} \to \Delta(A)$, with $C^i(a)$ denoting the probability of choosing action $a \in F(q)$ with posterior η^i, such that the attention and choice functions match the data,

$$q^a_\omega = \sum_{i=1}^{I} \rho_\omega(\eta^i) C^i(a).$$

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

$$\bar{F}^j \equiv \{a \in F | r^a(q) = \gamma^j\}.$$

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

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

Note that this is indeed a $|\Gamma(\rho)| \times |\Gamma(\pi)|$ stochastic matrix $B \geq 0$ with $\sum_{j=1}^{J} b^{ij} = 1$ all i. Given $\gamma^j \in \Gamma(\pi)$ and $\omega \in \Omega$, note that,

$$\sum_{i=1}^{I} b^{ij} \rho_\omega(\eta^i) = \sum_{i=1}^{I} \rho_\omega(\eta^i) \sum_{a \in \bar{F}^j} C^i(a) = \sum_{a \in \bar{F}^j} q^a_\omega,$$

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

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

as required for sufficiency. ■

9.2 Theorem 1 and Corollary 1

Theorem 1 Data set (D, q) has a rational inattention representation if and only if it satisfies NIAS and NIAC.

Proof of Necessity. Necessity of NIAS follows much as in CM13. Fix $A \in D$, $\tilde{\pi}^A$ and $\tilde{C} : \Gamma(\tilde{\pi}^A) \to \Delta(A)$ in a rational inattention representation, and possible $a \in A$. By definition of a rational inattention representation,

$$\sum_{\gamma \in \Gamma(\tilde{\pi}^A)} \tilde{C}^a(\gamma) \left[\sum_{\omega \in \Omega} \gamma_\omega U^a_\omega \right] \geq \sum_{\gamma \in \Gamma(\tilde{\pi}^A)} \tilde{C}^a(\gamma) \left[\sum_{\omega} \gamma_\omega U^b_\omega \right] \text{ all } b \in A.$$

Substituting,

$$\gamma_\omega = \frac{\mu_\omega \tilde{\pi}_\omega(\gamma)}{\sum_{\upsilon} \mu_\upsilon \tilde{\pi}_\upsilon(\gamma)},$$

cancelling the common denominator $\sum_{\upsilon} \mu_\upsilon \tilde{\pi}_\upsilon(\gamma)$ in the inequality, substituting $q^a_\omega = \sum_{\gamma \in \Gamma(\tilde{\pi}^A)} \tilde{\pi}_\omega(\gamma) \tilde{C}^a(\gamma)$, and dividing all terms by $\sum_{\upsilon} \mu_\upsilon q^a_\upsilon$, we derive,

$$\sum_{\omega} r^a_\omega(q)\, U^a_\omega = \sum_{\omega} \left[\frac{\mu_\omega q^a_\omega}{\sum_{\upsilon} \mu_\upsilon q^a_\upsilon} \right] U^a_\omega \geq \sum_{\omega} \left[\frac{\mu_\omega q^a_\omega}{\sum_{\upsilon} \mu_\upsilon q^a_\upsilon} \right] U^b_\omega = \sum_{\omega} r^a_\omega(q)\, U^b_\omega,$$

establishing necessity of NIAS.

To confirm necessity of NIAC consider any sequence $A_1, A_2, ...A_J \in D$ with $A_J = A_1$ and corresponding attention strategy $\tilde{\pi}^j$ for $1 \leq j \leq J$. By optimality,

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

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

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

so that,

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

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

$$G(A_j, \tilde{\pi}^j) = G(A_j, \bar{\pi}^j) = \sum_{a \in A_j} \sum_{\omega} \mu_\omega q_\omega^a U_\omega^a.$$

Hence,

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

establishing NIAC.

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

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

We now allow for arbitrary matchings of attention strategies to decision problems. First, let b_{jl} denote the gross utility of decision problem j combined with minimal attention strategy l,

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

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

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

It is simple to see that the NIAC condition implies that the identify map $\omega^I(j) = j$ maximizes the sum over all matching functions $\omega \in \mathcal{M}$. Suppose to the contrary that there exists some alternative matching function that achieves a strictly higher sum, and denote this match $\omega^* \in \mathcal{M}$. In this case construct a first sub-cycle as follows: start with the lowest index j_1 such that $\omega^*(j_1) \neq j_1$.

Define $\omega^*(j_1) = j_2$ and now find $\omega(j_2)$, noting by construction that $\omega(j_2) \neq j_2$. Given that the domain is finite, this process will terminate after some $K \leq J$ steps with $\omega^*(j_K) = j_1$. If it is the case that $\omega^*(j) = j$ outside of the set $\cup_{k=1}^{K}$ j_k, then we know the increase in the value of the sum is associated only with this cycle, hence,

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

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

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

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

Standard duality theory implies that the optimal assignment $\omega^I(j) = j$ is associated with a system of prices on minimal attention strategies such that the increment in net payoff from any move of any decision problem is not more than the increment in the cost of the attention strategy.

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

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

or,

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

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

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

for all $b \in A$. That this choice rule and the corresponding minimal attention function match the data holds by construction. ■ ■

Corollary 1 Data set (D, q) has an SRI representation if and only if it satisfies NIAS and NIAC.

Proof. Sufficiency is implied by theorem 1 applied to the special case with $T = 1$. To prove necessity, we first construct a mapping from sequential to static attention strategies $\lambda : \Sigma(\mu) \longrightarrow \Pi(\mu)$, with $\lambda(\sigma) \in \Pi(\mu)$. Given $\sigma \in \Sigma(\mu)$ and $\gamma \in \Gamma$ we specify the corresponding state dependent probabilities as,

$$\lambda^{\sigma}_{\omega}(\gamma) = \sum_{\{\gamma^t \in S_t | \gamma_t = \gamma,\ 1 \le t \le \tau\}} \rho^{\sigma}_{\omega}(\gamma^t).$$

We also define a mapping from sequential to static attention cost functions, $\kappa : \mathcal{E} \longrightarrow \mathcal{K}$. Given $E \in \mathcal{E}$ and $\pi \in \Pi(\mu)$,

$$\kappa^E(\pi) = \begin{cases} \inf_{(\sigma \in \Sigma(\mu) | \lambda^{\sigma} = \pi\}} E(\sigma); \\ \infty \text{ if } \nexists \sigma \in \Sigma(\mu) \text{ s.t. } \lambda^{\sigma} = \pi. \end{cases}$$

Note that $\kappa^E : \Pi(\mu) \to \bar{\mathbb{R}}$ inherits regularity (existence of a real-valued strategy) from the corresponding assumption for $E : \Sigma(\mu) \to \bar{\mathbb{R}}$.

$$\rho^{\sigma}(\gamma^t) = \sum_{\omega} \mu_{\omega} \rho^{\sigma}_{\omega}(\gamma^t).$$

We show now that, given $(\tilde{E}, \tilde{\sigma})$ with $\tilde{E} \in \mathcal{E}$ and $\tilde{\sigma} : D \to \Sigma(\mu)$ that define an SCI representation, the pair $(\tilde{\kappa}(\tilde{E}), \tilde{\lambda})$ with $\tilde{\kappa} \equiv \tilde{\kappa}(\tilde{E}) \in \mathcal{K}$, $\tilde{E} \in \mathcal{E}$, and $\tilde{\lambda} : D \to \Pi(\mu)$ define a rational inattention representation, whereupon application of the necessity aspect of theorem 1 implies that the data satisfy NIAS and NIAC. Since $(\tilde{E}, \tilde{\sigma})$ define an SRI representation of (D, q), we know that for all $A \in D$, $\tilde{\sigma}(A)$ is consistent with $q(A)$ and satisfies $\tilde{\sigma}(A) \in \hat{\Sigma}(\tilde{E}, A)$. Hence there exists $\tilde{C} : \Gamma \to \Delta(A)$ such that, given $1 \le t \le \tau$, $\gamma^t \in S_t$ such that $\gamma_t = \gamma$, $a \in A$ with $\tilde{C}^{\alpha}(\gamma) > 0$, and $b \in A$,

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

In addition, the attention and choice functions match the data,

$$q^a_{\omega} = \sum_{t=0}^{\tau} \sum_{\gamma^t \in S_t} \tilde{\rho}_{\omega}(\gamma^t) \tilde{C}^a(\gamma_t),$$

where $\tilde{\rho}_{\omega}(\gamma^t) \equiv \rho^{\tilde{\sigma}(A)}_{\omega}(\gamma^t)$.

Using the same choice function $\tilde{C} : \Gamma \to \Delta(A)$, it is immediate that all choices are optimal in $\tilde{\lambda}^A \equiv \lambda\left[\tilde{\sigma}(A)\right]$ and that it rationalizes the data,

$$q_\omega^a = \sum_{t=0}^{\tau} \sum_{\gamma^t \in S_t} \tilde{\rho}_\omega(\gamma^t) \tilde{C}^a(\gamma_t) = \sum_{\gamma \in \Gamma} \sum_{\{\gamma^t \in S_t | \gamma_t = \gamma,\ 1 \leq t \leq \tau\}} \tilde{\rho}_\omega(\gamma^t) \tilde{C}^a(\gamma) = \sum_{\gamma \in \Gamma} \tilde{\lambda}_\omega^A(\gamma) \tilde{C}^a(\gamma).$$

Hence, for all $A \in D$, $\tilde{\lambda}^A \in \Pi(\mu)$ is consistent with $q(A)$. It remains only to confirm that $\tilde{\lambda}^A \in \hat{\Pi}(\kappa(\tilde{E}), A)$. Given that $(\tilde{E}, \tilde{\sigma})$ is an SRI representation, we know that, given $A \in D$,

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

Substitution shows that this implies that the corresponding property holds for $\tilde{\lambda}^A$ in relation to $\tilde{\kappa}$,

$$\tilde{\lambda}^A \in \arg \max_{\pi \in \Pi(\mu)} \sum_{\gamma \in \Gamma} \sum \mu_\omega \tilde{\lambda}_\omega^A(\gamma) g(\gamma, A) - \tilde{\kappa}(\pi),$$

completing the proof. ■

9.3 Theorem 2

Theorem 2 Data set (D, q) satisfies NIAS and NIAC if and only if it has a rational inattention representation with conditions K1 to K3 satisfied.

Proof. The proof of necessity is immediate from theorem 1. The proof of sufficiency proceeds in four steps, starting with a rational inattention representation $\left(\bar{K}, \bar{\pi}\right)$ of (D, q) of the form produced in theorem 1 based on satisfaction of the NIAS and NIAC conditions. A key feature of this function is that the function $\bar{K}$ is real-valued only on the minimal information strategies $\bar{\Pi} \equiv \{\bar{\pi}^A | A \in D\}$ associated with all corresponding decision problems, otherwise being infinite. The first step is the proof is to construct a larger domain $\mathring{\Pi} \supset \bar{\Pi}$ to satisfy three additional properties: to include the inattentive strategy, $I \in \mathring{\Pi}$; to be closed under mixtures so that $\pi, \eta \in \mathring{\Pi}$ and $\alpha \in (0, 1)$

implies $\alpha * \pi \oplus (1-\alpha) * \eta \in \mathring{\Pi}$; and to be "closed under garbling," so that if $\pi \in \mathring{\Pi}$ is sufficient for attention strategy $\rho \in \Pi$, then $\rho \in \mathring{\Pi}$. The second step is to define a new function $\mathring{K}$ that preserves the essential elements of $\bar{K}$ while being real-valued on the larger domain $\mathring{\Pi} \supset \bar{\Pi}$, and thereby to construct the full candidate cost function $\mathring{K} : \Pi \to \mathbb{R} \cup \infty$. The third step is to confirm that $\mathring{K} \in \mathcal{K}$ and that $\mathring{K}$ satisfies the required conditions K1 through K3. The final step is to confirm that $\left(\mathring{K}, \bar{\pi}\right)$ forms a rational inattention representation of (D, q).

We construct the domain $\mathring{\Pi}$ in two stages. First, we define all attention strategies for which some minimal attention strategy $\bar{\pi} \in \Pi$ is sufficient;

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

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

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

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

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

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

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

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

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

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

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

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

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

establishing that indeed $\rho \in \mathring{\Pi}$ since, with $\bar{\Pi}_S$ closed under garbling, $\pi^j \in \bar{\Pi}_S$ and ρ^j a garbling of π^j implies $\rho^j \in \bar{\Pi}_S$.

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

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

Note that $\bar{K}_S(\pi) = \bar{K}(\pi)$ all $\pi \in \bar{\Pi}$. To see this, consider $A, A' \in D$ with

$\bar{\pi}^{A'}$ sufficient for $\bar{\pi}^A$. By the Blackwell property, expected utility is at least as high using $\bar{\pi}^{A'}$ as using $\bar{\pi}^A$ for which it is sufficient,

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

At the same time, since $\left(\bar{K}, \bar{\pi}\right)$ is a rational attention representation of (D, q), we know that $\bar{\pi}^A \in \hat{\Pi}(K, A)$, so that,

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

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

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

The second stage in the construction is extend the domain of the cost function from $\bar{\Pi}_S$ to $\mathring{\Pi}$. As noted above, this set comprises all finite mixtures of elements of $\bar{\Pi}_S$,

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

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

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

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

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

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

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

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

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

confirming mixture feasibility.

We show also that $\mathring{K}$ satisfies K3, weak monotonicity in information. Consider $\pi, \eta \in \mathring{\Pi}$ with π sufficient for η. We know by Lemma 3 that we can find $J \in \mathbb{N}$, $\lambda \in S^{J-1}$, and corresponding elements $\{\pi^j\}_{j=1}^{J} \in \bar{\Pi}_S$ with fixed range

$\Gamma(\pi^j) = \Gamma(\pi)$ such that $\pi = \sum_{j=1}^{J} \lambda_j * \pi^j$ and such that,

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

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

$$\bar{K}_S(\pi^j) \geq \bar{K}_S(\eta^j).$$

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

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

confirming weak monotonicity.

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

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

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

and such that,

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

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

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

so that,

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

We conclude that there exists j such that,

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

Note that each $\pi^j \in \bar{\Pi}_S$ inherits its cost $\bar{K}_S(\pi^j)$ from an element $\bar{\pi}^j \in \bar{\Pi}$ that is the lowest cost minimal attention strategy according to $\bar{K}$ on set $\bar{\Pi}$ that is sufficient for π^j,

$$\bar{K}_S(\pi^j) = \bar{K}(\bar{\pi}^j),$$

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

$$G(A, \bar{\pi}^j) - \bar{K}(\bar{\pi}^j) \geq G(A, \pi^j) - \bar{K}_S(\pi^j) > G(A, \bar{\pi}^A) - \mathring{K}(\bar{\pi}^A).$$

To complete the proof it is sufficient to show that,

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

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

$$G(A, \bar{\pi}^j) - \bar{K}(\bar{\pi}^j) > G(A, \bar{\pi}^A) - \bar{K}(\bar{\pi}^A),$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

This follows since,

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

■

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

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

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

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

A first issue that we wish to avoid is limitless growth in the cardinality $J(n)$. The first key observation is that, by Charateodory's theorem, we can reduce the number of strictly positive weights in a convex combination $\pi = \sum_{j=1}^{J^*(n)} \lambda_j^*(n) * \pi^j(n)$ to have cardinality $J^*(n) \leq M + 1$. We confirm now that we can do this without raising the corresponding costs, $\sum_{j=1}^{J^*(n)} \lambda_j^*(n) \bar{K}_S(\pi^j(n))$.

Suppose that there is some integer n such that the original set of attention strategies has strictly higher cardinality $J(n) > M+1$. Suppose further that the first selection of $J^1(n) \leq M+1$ such posteriors for which there exists a strictly positive probability weights $\delta_j^1(n)$ such that $\pi = \sum_{j=1}^{J^1(n)} \delta_j^1(n) * \pi^j(n)$ has higher such costs (note WLOG that we are treating these as the first $J^1(n)$ attention strategies in the original list). It is convenient to define $\delta_j^1(n) = 0$ for $J^1(n)+1 \leq j \leq J(n)$ so that we can express this inequality in the simplest terms,

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

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

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

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

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

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

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

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

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

$$\left|\Gamma(\pi^j(n))\right| = K^j$$

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

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

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

which is possible because the sequence of posteriors lives in a compact set, and so have a convergent subsequence.

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

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

Again, this is possible because the state dependent probabilities and weights live in a compact set.

The final selection of a subsequence ensures that, given $1 \leq j \leq J$, each

$\pi^j(n) \in \bar{\Pi}_S$ has its value defined by precisely the same minimal attention strategy $\bar{\pi}^j \in \bar{\Pi}$ as the least expensive among those that were sufficient for it and hence whose cost it was assigned in function $\bar{K}_S$. Technically, for each $1 \leq j \leq J$,

$$\bar{K}_S(\pi^j(n)) = \bar{K}(\bar{\pi}^j),$$

for $1 \leq n \leq \infty$: this is possible because the data set and hence the number of minimal attention strategies is finite.

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

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

with state dependent probabilities,

$$\left[\pi^j(L)\right]_\omega \left(\gamma^{jk}(L)\right) = \pi_\omega^{jk}(L).$$

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

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

We know from the construction that, for each n,

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

Hence the result is established provided only,

$$\bar{K}_S(\pi^j(L)) \leq \bar{K}(\bar{\pi}^j),$$

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

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

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

9.4 No Strong Blackwell

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

$$(U_1^a, U_2^a) = (10, 0)\,; (U_1^b, U_2^b) = (0, 10)\,; (U_1^c, U_2^c) = (7.5, 7.5)\,.$$

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

$$q_1^a = q_2^b = \frac{3}{4} = 1 - q_1^b = 1 - q_2^a.$$

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

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

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