# Experimental Tests of Rational Inattention*

Mark Dean† and Nathaniel Neligh‡

June 2017

**Abstract**

We use laboratory experiments to test models of 'rational inattention', in which people acquire information to maximize utility from subsequent choices net of information costs. We show that subjects adjust their attention in response to changes in incentives a manner which is broadly in line with the rational inattention model but which violates models such as random utility in which attention is fixed. However, our results are not consistent with information costs based on Shannon entropy, as is often assumed in applied work. We find more support for a class of 'posterior separable' cost functions which generalize the Shannon model.

# 1 Introduction

Attention is a scarce resource. The impact of attentional limits has been identified in many important economic settings.[1] The importance of limits on attention has lead to a wide range

*We thank in particular Andrew Caplin, who was instrumental in the development of this research agenda and many of the ideas embodied in this paper, and Stephen Morris and Isabel Trevino who were involved in the design of Experiment 4 as part of a separate project (see Dean *et al.* [2016]). Also John Leahy, Daniel Martin, Filip Matejka, Pietro Ortoleva and Michael Woodford for their constructive contributions, as well as the members of the Cognition and Decision Lab at Columbia University and numerous seminar participants. We acknowledge the Junior Faculty Summer Research Support Grant from Columbia University which provided funding for the project. Research funds were also provided by Princeton University and New York University. This paper reports on experiments similar to those in Caplin and Dean [2013] and Caplin and Dean [2014] and subsumes those parts of that paper that are common.

†Department of Economics, Columbia University. Email: mark.dean@columbia.edu

‡Department of Economics, Columbia University. Email: nln2110@columbia.edu

[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]). Purchasers limit their attention to a relatively small number of websites when buying over the internet (Santos *et al.* [2012]).

of modeling approaches aimed at understanding behavior under such constraints. Particularly influential are models of 'rational inattention'[2] which assume that people choose the information they attend to in order to maximize the expected utility of subsequent choices net of informational costs.[3]

The widespread use of the rational inattention model leads to a number of natural research questions. First, do people in fact actively adjust their attention in response to incentives? Second, do they do so in line with the predictions of the rational inattention model? Third, what do the costs of attention look like? Fourth, how much heterogeneity in information costs is there in the population?

In this paper we use a sequence of four laboratory experiment to provide answers to these questions. The basic set up is a simple information acquisition task in which subjects are presented with a number of balls on the screen which can either be red or blue. They must then choose between different actions, the payoff of which depend on the fraction of balls which are red (which we will call the 'state of the world'). The prior probability of each state is known to the subject. There is no time limit or extrinsic cost of information in the experiment, so if subjects face no intrinsic cost of information acquisition the experiment would be trivial: they would simply ascertain the number of red balls on the screen and choose the best action given this state. As we shall see, subjects in general do not behave in this way.

Within this setup the four experiments vary different features of the decision problem, including the range of available actions, the value of correct choice and the prior probability of possible states. By repeatedly exposing the subject to each decision problem we can collect 'state dependent stochastic choice' (SDSC) data, or the probability that each action is chosen in each state. Such data is particularly useful for testing models of rational inattention, and learning about attention cost (see Caplin and Dean [2015]).

In order to answer the four questions above, we rely on several recent theoretical advances. Caplin and Martin [2015b] and Caplin and Dean [2015] provide necessary and sufficient conditions for SDSC to be consistent with a general model of rational inattention which is agnostic about information costs (henceforth the general model). The No Improving Action Switches (NIAS) condition ensures that choice of action is optimal given the information

[2] We will use the term 'rational attention' to describe this entire model class, while recognizing that others use this term to refer to the specific case when costs are based on the Shannon mutual information between prior and posterior beliefs. Our definition covers almost all models of costly information acqusition, as we discuss in section 2.2. We refer to the latter as the 'Shannon model'.

[3] Recent examples include Sims [2003], van Nieuwerburgh and Veldkamp [2009], Hellwig *et al.* [2012], Matejka and McKay [2015] and Caplin and Dean [2015].

gathered, while the No Improving Attention Cycle (NIAC) condition ensures that the allocation of attention across decision problems can be rationalized by some cost function. Using this 'revealed cost' approach, bounds can be placed on the costs associated with different information structures.

Complementary to this agnostic approach to attention costs, we also ask whether our data is in line with specific cost functions. Here we focus on costs which are linear in Shannon mutual information (henceforth the Shannon model), which was introduced to the economics literature by Sims [2003], and has proved extremely popular in subsequent empirical and theoretical work.[4] The use of mutual information costs has been justified on information theoretic grounds, related as they are to the average number of signals needed to generate a given set of posterior beliefs (see Sims [2003]). Compared to the general model, the Shannon model is extremely restrictive, with only a single parameter related to the marginal cost of information. Recent works by Matejka and McKay [2015] and Caplin *et al.* [2017][5] have highlighted a number of behavioral regularities implied by the Shannon model: The Locally Invariant Posteriors (LIP) and Invariant Likelihood Ratio (ILR) conditions restrict how optimal information acquisition responds to changes in prior beliefs and incentives respectively. The Shannon model also implies that behavior is invariant to the addition or subtraction of states which are identical in terms of the payoffs of all available actions - a property labeled Invariance Under Compression (IUC) by Caplin *et al.* [2017]. This property means the model is incommensurate with any notion of 'perceptual distance', by which some states are easier to differentiate than others. Because the Shannon model is so restrictive, we also consider an intermediate class of 'posterior separable' models, introduced by Caplin and Dean [2013], which retain the LIP feature of the Shannon models, but relaxes other implications. Models in this class have recently been used for several economic applications (see for example Clark [2016] and Morris and Strack [2017]).

Experiment 1 is designed to distinguish rational inattention from models in which attention is unresponsive but choice is stochastic - including random utility (Block and Marschak [1960]) and signal detection theory (Green and Swets [1966]). The key observation is that these alternative models imply the property of Monotonicity: adding alternatives to the choice set cannot increase the probability of choosing previously available options. This property is not implied by models of rational inattention, and Matejka and McKay [2015] describe a scenario in which informational spillovers could lead a rationally inattentive decision maker to violate monotonicity. We provide an experimental implementation and find

[4] See for example the application of the model to investment decisions (e.g Mondria [2010]), global games (Yang [2015]), and pricing decisions (Mackowiak and Wiederholt [2009], Matějka [2016], Martin [2017]).

[5] See also Caplin and Dean [2013].

that monotonicity is strongly rejected by the data in a manner broadly consistent with rational inattention (as characterized by the NIAS and NIAC conditions), indicating the need for models in this class.

Experiment 2 examines the response of information acquisition to incentives using a simple two state/two action design in which we vary the benefit from choosing the correct action. This provides a simple test of the NIAC and NIAS conditions, in an admittedly undemanding setting. The aggregate data strongly supports both conditions, and at the individual level 81% of subjects exhibit no significant violation of either. This implies that there is a notion of information costs which rationalizes the behavior of the majority of subjects. We find a high degree of heterogeneity in these costs across subjects. We also ask whether the responsiveness of our subjects to changes in incentives is consistent with the Shannon model, and find that the answer is largely negative. Both in the aggregate and at the individual level we find that subjects typically respond less strongly than predicted by Shannon costs. Behavior is better matched by a simple, two parameter extension of the Shannon model in the posterior separable class.

Experiment 3 looks at the reaction of behavior to changes in prior beliefs, again in a simple two state/two action setting. This provides us both with a more sophisticated test of the NIAS condition, and a test of the LIP condition which characterizes the class of posterior separable model. The former condition is largely satisfied, both at the aggregate and the individual level - in particular we find little evidence of base rate neglect in our subject's choices. The evidence for the LIP condition is more mixed, but some aspects of the data do agree with this more stringent condition.

Our final experiment tests IUC using a design with multiple states but only two acts to choose from, the payoff of which is the same in many different states. The Shannon model implies that behavior should be the same in all states in which both actions pay the same amount. This is incommensurate with the idea of a perceptual distance, by which some states are easier to differentiate between than others. We show that, in our baseline experimental set up which has a natural notion of perceptual distance, IUC performs poorly. However, in an alternative setting based on letter recognition, it provides a reasonable approximation of behavior.

Our paper provides an empirical counterpart to the large recent theoretical literature on limited attention, which includes both the rational inattention models discussed above and other approaches in which attention is not modelled as the result of an optimizing process (for example Masatlioglu *et al.* [2012], Manzini and Mariotti [2014], Lleras *et al.* [2017]). To our knowledge there is surprisingly little experimental work testing such models. Notable exceptions include Gabaix *et al.* [2006], Caplin *et al.* [2011], Taubinsky [2013] and Khaw *et al.* [2016]. These papers are designed to test models which are very different to those we consider here, and as such make use of very different data. Pinkovskiy [2009] and Cheremukhin *et al.* [2015] fit the Shannon model using data on stochastic choice between lotteries, but do not test the sharp behavioral predictions from that model as we do here. Bartoš *et al.* [2016] report the results of a field experiment which supports rationally inattentive behavior in labor and housing markets. More broadly, our work fits in to a recent move to use richer data to understand the process of information acquisition (for example Krajbich *et al.* [2010], Brocas *et al.* [2014], Polonio *et al.* [2015], and Caplin and Martin [2015a]). In contrast to the relatively small literature in economics, there is a huge literature in psychology that examines behavior in perceptual tasks which are similar to some of our experiments, though the data is analyzed in very different ways (for example see Ratcliff *et al.* [2016] for a recent review, and Krajbich *et al.* [2011] for an application to economic decision making). We discuss our relationship to these papers in section 6.

The paper is organized as follows. Section 2 describes the theory underlying our experiments. Section 3 describes the experimental design in detail. Section 4 provides results, section 5 provides additional discussion, and section 6 describes the related literature.

# 2 Theory

## 2.1 Set-Up and Data

For our discussion of the testable implications of the rational inattention model we use the set up and notation of Caplin and Dean [2015].

We consider a decision maker (DM) who chooses among actions, the outcomes of which depend on which of a finite number of states of the world $\omega \in \Omega$ occurs. The utility of action $a$ in state of the world $\omega$ is denoted by $u(a, \omega)$.

A decision problem is defined by a set of available actions $A$ and a prior over states of the world $\mu \in \Delta(\Omega)$, both of which we assume can be chosen by the experimenter. The data observed from a particular decision problem is a *state dependent stochastic choice (SDSC)* function, which describes the probability of choosing each available action in each state of the world. For a decision problem $(\mu, A)$ we use $P_{(\mu,A)}$ to refer to the associated SDSC function, with $P_{(\mu,A)}(a|\omega)$ the probability that action $a \in A$ was chosen in state $\omega \in \Omega$ (where it will not cause confusion, we will suppress the subscript on $P$). Note that a SDSC function also implies a posterior probability distribution over states, $\gamma^a$, associated with each action $a \in A$ which is chosen with positive probability. By Bayes' rule we have

$$\gamma^a(\omega) = P(\omega|a) = \frac{\mu(\omega)P(a|\omega)}{\sum_{\omega' \in \Omega} \mu(\omega')P(a|\omega')}. \tag{1}$$

This construct will be useful in testing the various theories we discuss below.

## 2.2 The Rational Inattention Model

The rational inattention model assumes that the DM can gather information about the state of the world prior to choosing an action. Importantly, they can choose what information to gather conditional on the decision problem they are facing. The DM must trade off the costs of information acquisition against the benefits of better subsequent choices. The rational inattention model assumes that the DM solves this trade off optimally.

In each decision problem, the DM chooses an *information structure*: a stochastic mapping from objective states of the world to a set of subjective signals. While this formalization sounds somewhat abstract, its subsumes the vast majority of models of optimal information acquisition that have been proposed (see Caplin and Dean [2015]). Note that we assume that the subject's choice of information structure is not observed, and so has to be inferred from their choice data.

Having selected an information structure, the DM can condition choice of action only on these signals. For notational convenience we identify each signal with its associated posterior beliefs $\gamma \in \Gamma$, which is equivalent to the subjective information state of the DM following the receipt of that signal. Feasible information structures satisfy Bayes' rule, so for any prior $\mu$ the set of possible structures $\Pi(\mu)$ comprises all mappings $\pi : \Omega \rightarrow \Delta(\Gamma)$ that have finite support $\Gamma(\pi) \subset \Gamma$ and that satisfy Bayes' rule, meaning that for all $\omega \in \Omega$ and $\gamma \in \Gamma(\pi)$,

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

where $\pi(\gamma|\omega)$ is the probability of signal $\gamma$ given state $\omega$ and $\gamma(\omega)$ is the probability of state $\omega$ conditional on receiving signal $\gamma$.

We assume that there is a cost associated with the use of each information structure, with $K(\mu, \pi)$ denoting the cost of information structure $\pi$ given prior $\mu$. We define $G$ as the gross payoff of using a particular information structure in a particular decision problem. This is calculated assuming that actions are chosen optimally following each signal,

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

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

The rational inattention model assumes that the decision maker choose actions in order to maximize utility given information, and chooses information structures to maximize utility net of costs, i.e.

$$G(\mu, A, \pi) - K(\mu, \pi)$$

We do not a priori rule out the possibility that the cost of some information structures is infinite, meaning that this formalization can cope with models in which the DM is restricted to choosing certain types of information structure (for example normal signals).[6] Note the cost function $K$ is essentially a high-dimentional free parameter in this model.

Caplin and Dean [2015] provide necessary and sufficient conditions on SDSC data such that there exists some cost function which rationalizes the observed pattern of behavior. We call this the *general model* of rational inattention. The *No Improving Action Switches* (NIAS) condition, introduced by Caplin and Martin [2015b], ensures that choices are consistent with efficient use of whatever information the DM has. It states that, for any action $a$ which is chosen with positive probability, it must be that $a$ maximizes expected utility given $\gamma^a$ - the posterior distribution associated with that act. The *No Improving Attention Cycles* (NIAC) condition ensures that choice of information itself is rationalizable according to some underlying cost function. It relies on the concept of a *revealed information structure,* which can be recovered from the data by assuming that each chosen action is associated with exactly one signal.[7] Essentially, NIAC states that the total gross value of information (measured by $G$) in a collection of decision problems cannot be increased by switching revealed information structures between those problems.

In the interests of brevity, we do not provide a formal definition of NIAS or NIAC here

[6]Though we do rule out costs being infinte for all information costs so that optimization is well defined.

[7]Note that we do not require that this is true in the underlying model. Caplin and Dean [2015] show that constructing a revealed information structure in this manner is enough to test all models in the rational inattention class.

(we refer the interested reader to Caplin and Dean [2015]). Instead we will describe in section 3 how these conditions apply to our specific experiments. However, we highlight one important structural feature: because the general model allows costs to be indexed by priors in an arbitrary way, it puts no restriction on behavior as prior changes, only as the set of available actions change.

We emphasize that the flexibility in the choice of the function $K$ means that general model includes as special cases almost all models of optimal costly information acquisition that have been discussed in the literature, including those in which agents can either pay to receive information or not,[8] those that apply additive normal noise to the agent's information and then allows them to pay a cost to decrease the variance of that noise,[9] or those in which the decision maker chooses a partition structure on the state space.[10] See Caplin and Dean [2015] for a discussion.

### 2.2.1 Rational Inattention vs Other Models of Stochastic Choice

Rational inattention is, of course, not the only model which allows for stochasticity in choice. Two highly influential alternatives are the random utility model (Block and Marschak [1960], McFadden [1974], Gul and Pesendorfer [2006]) and Signal Detection Theory (Green and Swets [1966]). Here we describe how these can be differentiated from rational inattention.

The random utility model (RUM) assumes that people have many possible utility functions which may govern their choice. On any given trial one of these utility functions is selected according to some probability distribution, and the DM will choose in order to maximize that function. Stochasticity therefore derives from changes in preferences, rather than noise in the perception of the state of the world.

Typically the RUM has not been applied to situations in which there is an objective, observable state of the world, and there are many possible ways that the model could be adapted to such a situation.[11] However, as long as we maintain the assumption that the DM does not actively change their choice of information in response to the decision problem, all variants of the RUM will imply the property of Monotonicity. This states that adding new alternatives to the choice set cannot increase the probability of an existing alternative being chosen:

[8] For example Grossman and Stiglitz [1980], Barlevy and Veronesi [2000] and Reis [2006].

[9] For example Verrecchia [1982] and Hellwig *et al.* [2012] .

[10] For example Robson [2001] and Ellis [2013].

[11] For example, the DM could be fully informed about the underlying state, have no information about the state, or receive a noisy signal regarding the state.

**Definition 1** *A SDSC satisfies Monotonicity if, for every $\mu \in \Delta(\Omega)$, $A \subset B$, $\omega \in \Omega$ and $a \in A$*

$$P_{(\mu,A)}(a|\omega) \geq P_{(\mu,B)}(a|\omega)$$

That Monotonicity is a necessary property of data generated by random utility models is intuitively obvious: Adding new alternatives to a set $A$ can only (weakly) reduce the set of utility functions for which any $a \in A$ is optimal. However, Monotonicity is *not* implied by rational inattention models, as illustrated by Matejka and McKay [2015]. The introduction of a new act can increase the incentives to information acquisition, which may in turn lead the DM to learn that an existing act was of high value. Matejka and McKay [2015] demonstrate that the Shannon model will always generate a violation of Monotonicity across some decision problems. We make use of this insight in Experiment 1.

Signal Detection Theory (SDT) is popular model in the psychological literature on perception and choice. Essentially it assumes that people receive a noisy signal about the state of the world, then choose actions optimally given subsequent beliefs. As such, it is a special case of the general model in which the costs of all but one information structure are infinite. A subject behaving according to SDT will therefore satisfy NIAC and NIAS. However, they will also satisfy Monotonicity: as information selection cannot adjust, the only way that adding a new option can affect choice is by being chosen instead of one of the existing options upon the receipt of some signal. Thus a violation of Monotonicity rules out SDT as well as random utility.

## 2.3 The Shannon Model

The general model is completely agnostic about the form of information costs.[12] However, for many applied purposes, specific cost functions are assumed. One of the most popular approaches is to base costs on the Shannon mutual information between states and signals. Introduced to the economics literature by Sims [2003], Shannon costs can be justified on axiomatic or information theoretic grounds, and have been widely applied in the subsequent literature.

Mutual information costs have the following form

$$K_s(\mu, \pi) = -\kappa \left[ \sum_{\gamma \in \Gamma(\pi)} \pi(\gamma) H(\gamma) - H(\mu) \right]$$

[12] Other than assuming that they are additively separable from the utility gained from choice of action.

where $\pi(\gamma) = \sum_{\omega\in\Omega} \mu(\omega)\pi(\gamma|\omega)$ is the unconditional probability of signal $\gamma$ and $H(\gamma) = \sum_{\omega\in} -\gamma(\omega) \ln \gamma(\omega)$ is the Shannon entropy of distribution $\gamma$.[13] Mutual information can therefore be seen as the expected reduction in entropy due to the observation of signals from the information structure.

Clearly, the Shannon model puts much more structure on information costs than the general model as it has essentially one degree of freedom: the marginal cost of mutual information governed by $\kappa$. This in turn means that the Shannon model puts much tighter restrictions on behavior than the general model. These restrictions have been discussed in several recent papers (particularly Caplin and Dean [2013], Matejka and McKay [2015] and Caplin *et al.* [2017]). In this paper we shall concern ourselves with three implications of the Shannon model: Invariant Likelihood Ratio, Locally Invariant Posteriors and Invariance Under Compression.

The Invariant Likelihood Ratio (ILR) property (Caplin and Dean [2013]) states that for any two chosen actions, the posterior probabilities about a particular state conditional on those actions depend only on the relative payoffs of those actions and information costs

$$\frac{\gamma^a(\omega)}{\gamma^b(\omega)} = \frac{\exp(u(a,\omega)/\kappa)}{\exp(u(b,\omega)/\kappa)}$$

As we shall see in the discussion of experiment 2 below, this puts tight restrictions on the way in which information acquisition can change with the rewards for doing so.

The Locally Invariant Posterior (LIP) property states that local changes in prior beliefs do not lead to changes in optimal posterior beliefs.[14] Specifically, if, for some decision problem $(\mu, A)$, the associated SDSC reveals some set of posteriors $\{\gamma^a\}_{a\in A}$, and we change the prior to some $\mu'$ in such a way that these posteriors are still feasible (i.e. $\mu'$ is in the convex hull of $\{\gamma^a\}_{a\in A}$), the LIP property states that precisely these posteriors should also be used in the decision problem $(\mu', A)$. We will test this proposition in experiment 3.

ILR shows that, according to the Shannon model, posterior beliefs depend only of the payoffs of actions in a particular state, not on any other features of the state. This implies that behavior should not be affected by adding or subtracting states which are identical in payoff terms for all acts. Caplin *et al.* [2017] show that this 'Invariance Under Compression' property fully characterizes the Shannon model within a the broader class of posterior separable models described below. Behaviorally, one implication of this is that the Shan-

[13] Recall that we identify a signal with its resulting posterior ditribution.

[14] Again, see Caplin and Dean [2013] and Caplin *et al.* [2017] for further details.

non model lacks any notion of 'perceptual distance': that some states might be harder to differentiate than others. We test this implication in experiment 4.

## 2.4 Posterior Separable Information Costs

So far we have considered only the general model, with completely unconstrained costs, and the Shannon model, in which costs are completely pinned down. An intermediate class of models, introduced in Caplin and Dean [2013], is defined by 'posterior separable' information costs. These models keep the structure of Shannon mutual information, but relax the precise functional form. Specifically, posterior separable cost functions are those that can be written as

$$K_T(\mu,\pi) = \left[\sum_{\gamma\in\Gamma(\pi)} \pi(\gamma)T(\gamma) - T(\mu)\right]$$

for some strictly convex function $T$.[15] Posterior separable cost functions satisfy LIP, but not necessarily ILR or IUC. They can therefore allow for different elasticities of attention with respect to incentives as well as the possibility of different perceptual differences between states.

One type of posterior separable cost function that we will fit to the experimental data is given by the parameterized class $T_{\{\rho,\kappa\}} \in \mathcal{K}_{PS}$:

$$T_{\{\rho,\kappa\}}(\gamma) = \begin{cases} -\kappa\left(\sum_{m=1}^{M}\gamma_m\left[\frac{\gamma_m^{1-\rho}}{(\rho-1)(\rho-2)}\right]\right) & \text{if } \rho\neq 1 \text{ and } \rho\neq 2; \\ -\kappa\left(\sum_{m=1}^{M}\gamma_m\ln\gamma_m\right) & \text{if } \rho = 1. \\ -\kappa\left(\sum_{m=1}^{M}\gamma_m\frac{\ln\gamma_m}{\gamma_m}\right) & \text{if } \rho = 2. \end{cases},$$

were $\gamma_m$ is the posterior probability of state $m$. If there are only two states of the world, derivatives with respect to $\gamma$ obey,

$$\begin{aligned} \frac{\partial T_{\{\rho,\kappa\}}(\gamma)}{\partial\gamma_1} &= \begin{cases} \kappa\left(\frac{\gamma_1^{1-\rho}-(1-\gamma_1)^{1-\rho}}{(\rho-1)}\right) & \text{if } \rho\neq 1; \\ \kappa\left(\ln\gamma_1-\ln(1-\gamma_1)\right) & \text{if } \rho = 1. \end{cases} \\ \frac{\partial^2 T_{\{\rho,\kappa\}}(\gamma)}{\partial\left(\gamma_1\right)^2} &= \kappa\left(\gamma_1^{-\rho}+(1-\gamma_1)^{-\rho}\right) \end{aligned}$$

[15] In Caplin *et al.* [2016] this class of models is refered to as 'uniformly posterior separable' to differentiate them from a broader class of models in which the function $T$ is allowed to vary with the prior.

We use this class of functions because they provide a simple and easy to estimate way of generalizing the Shannon model to allow for extra flexibility in the response of attention to incentives, similar to the way that constant relative risk aversion generalizes log utility: note that the second derivative of these costs functions is continuous in $\rho$, with the Shannon entropy cost function fitting smoothly into the parametric class at $\rho = 1$.

# 3 Experimental Design

## 3.1 Set Up

We now introduce our experimental design which produces state dependent stochastic choice data for each subject.

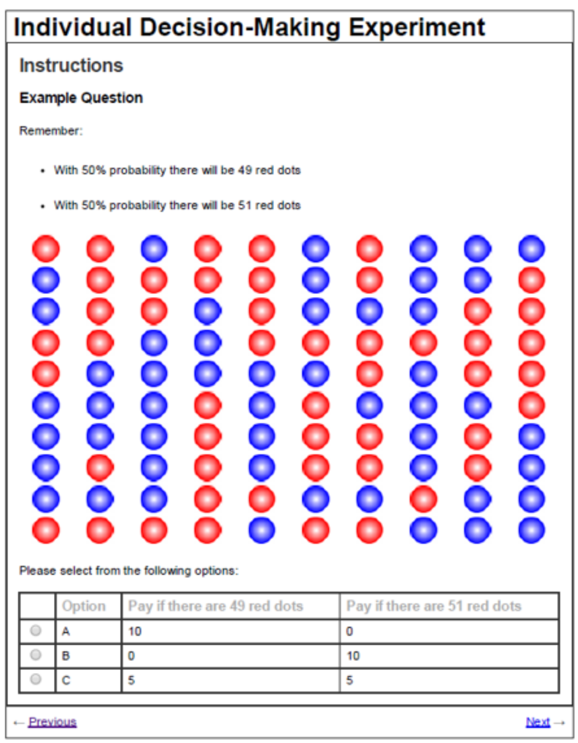



Figure 1: Typical Screenshot

In a typical question in the experiment, a subject is 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 such states. Having seen the screen, they choose from a number of different actions whose payoffs are state dependent. As in the theory, a decision problem is defined by this prior distribution and the set of available actions. Figure 1 shows a typical screenshot from the experiment.

A subject faces each decision problem multiple times, allowing us to approximate their state dependent stochastic choice function. In any given experiment, the subject faces several different problems. All occurrences of the same problem are grouped, but the order of the problems is block-randomized. At the end of the experiment, one decision problem is selected at random for payment.

There are several things to note about our experimental design. First there is no external limit (such as a time constraint) on a subject's ability to collect information about the state of the world. If they so wished, subjects could determine the state with a very high level of precision in each question by precisely counting the number of red balls - a very small number of subjects do just this. We are therefore not studying hard limits to a subject's perceptual ability to determine the state, as is traditional in many psychology experiments (see section 6 for a discussion). At the same time, there is no 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 and time for better information, and so higher payoffs.[16]

Second, in order to estimate the state dependent stochastic choice function we treat the multiple times that a subject faces the same decision making environment as multiple independent repetitions of the same decision problem. 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. We discuss this assumption further in section 4.1.

Third, in experiments where it is important, we pay subjects in 'probability points' rather than money - i.e. subjects are paid in points which increase the probability of winning a monetary prize. We do so in order to get round the problem that utility itself is not directly observable. This is not a problem if utility is linearly related to the quantity of whatever we use to pay subjects. Expected utility theory implies that utility is linear in probability points but not monetary amounts.

[16]Subjects had a fixed number of tasks to complete during the course of the experiment. They were told that when they had completed the experiment they had to stay in the lab until all subjects had finished the experiment.

An example of the experimental instructions can be found in the online appendix.

## 3.2 Experiment 1: Testing for Responsive Attention

Experiment 1 is designed to elicit violations of Monotonicity, which therefore also violate the predictions of the RUM and SDT. Based on a thought experiment discussed in Matejka and McKay [2015], the design requires subjects to take part in two decision problems (DP) described in table 1 below. Payment was in probability points with a prize of $20. Each subject faces 75 repetitions of each DP.

Table 1: Experiment 1

| | Payoffs | | | | | |
|---|---|---|---|---|---|---|
| DP | $U(a,1)$ | $U(a,2)$ | $U(b,1)$ | $U(b,2)$ | $U(c,1)$ | $U(c,2)$ |
| 1 | 50 | 50 | $b_1$ | $b_2$ | n/a | n/a |
| 2 | 50 | 50 | $b_1$ | $b_2$ | 100 | 0 |

The structure of the two DP is as follows. There are two equally likely states of the world - 1 and 2 (represented by 49 and 51 red balls respectively). In DP 1, the subject has the choice between the sure-thing option $a$, which pays 50 probability points, and an option $b$ which pays off less than $a$ in state 1, but more in state 2 (i.e. $b_1 < 50 < b_2$). However, $b_1$ and $b_2$ are chosen to be relatively close to 50. We used 4 different values for $b_1$ and $b_2$ as described in table 2.

Table 2: Treatments for Exp. 1

| Treatment | Payoffs | |
|---|---|---|
| | $b_1$ | $b_2$ |
| 1 | 40 | 55 |
| 2 | 40 | 52 |
| 3 | 30 | 55 |
| 4 | 30 | 52 |

The incentive for gathering information in DP 1 is low. The subject can simply choose $a$ and guarantee that they will receive 50 points. If they try to determine the state then half the time they will find out that it is highly likely to be 1, in which case $a$ is better than $b$. Even if they do find out that the state is highly likely to be 2 the additional payoff over simply choosing $a$ is low. Thus, for many information cost functions, the optimal strategy for DP 1 will be to remain uninformed and select $a$.

In DP 2, the option $c$ is added. This increases the value of information acquisition, as $c$ pays a high number of points in state 1 and a low number in state 2. Thus, the addition of $c$ may lead subjects to identify the true state with a high degree of accuracy. However, having done so, half the time they will determine that the state is in fact 2, in which case $b$ is the best option. Thus, there is potentially a 'spillover' effect of adding $c$ to the choice set which is to increase the probability of selecting $b$. It is this violation of monotonicity we will look for in the data. Matejka and McKay [2015] show that, for a DM with Shannon costs, such violations are guaranteed for some parameterization of this class of decision problem.

Experiment 1 also provides a first test for the NIAS and NIAC conditions which characterize the general model. Recall that the NIAS condition requires that subjects make optimal choices given their revealed posteriors. So, for example, at the posteriors $\gamma^a$ from DP 1 it must be the case that the expected utility of $a$ is higher than that of $b$ This implies that

$$\gamma_1^a(1)50 + \gamma_1^a(2)50 \geq \gamma_1^a(1)b_1 + \gamma_1^a(2)b_2 \Rightarrow$$

$$\frac{\mu(1)P_1(a|1)}{P_1(a)}50 + \frac{\mu(2)P_1(a|2)}{P_1(a)}50 \geq \frac{\mu(1)P_1(a|1)}{P_1(a)}b_1 + \frac{\mu(2)P_1(a|2)}{P_1(a)}b_2 \Rightarrow$$

$$P_1(a|1)(50 - b_1) + P_1(a|2)(50 - b_2) \geq 0$$

where $P_i$ is the SDSC data generated by DP $i$.

In DP 1 the NIAS conditions imply two comparisons which can be collapsed into a single inequality. In DP 2 there are six inequalities to check, not all of which will necessarily bind. The derivation of these conditions is described in appendix A1 and summarized in table 3.

Table 3: NIAS tests for Experiment 1

| DP | Comparison | Condition |
|---|---|---|
| 1 | N/A | $P_1(a\|1)(50 - b_1) + P_1(a\|2)(50 - b_2) - (100 - (b_1 + b_2)) \geq 0$ |
| 2 | a vs b | $P_2(a\|1)(50 - b_1) + P_2(a\|2)(50 - b_2) \geq 0$ |
| 2 | a vs c | $P_2(a\|2) - P_2(a\|1) \geq 0$ |
| 2 | b vs a | $P_2(b\|1)(b_1 - 50) + P_2(b\|2)(b_2 - 50) \geq 0$ |
| 2 | b vs c | $P_2(b\|1)(b_1 - 100) + P_2(b\|2)b_2 \geq 0$ |
| 2 | c vs a | $P_2(c\|1) - P_2(c\|2) \geq 0$ |
| 2 | c vs b | $P_2(c\|1)(100 - b_1) - P_2(c\|2)b_2 \geq 0$ |

NIAC is the condition that guarantees that behavior can be rationalized by an underlying cost function. It states that total gross surplus (measured by the function $G$) cannot be increased by switching the assignment of information structures from those revealed in the data: i.e. using the information structure revealed in DP1 in DP2 and visa versa (see Caplin and Dean [2015] for further details). In appendix 1 we show that (assuming NIAS holds) this boils down to the condition that

$$P_2(c|\omega_1) - P_2(c|\omega_2) - (P_1(a|\omega_1) - P_1(a|\omega_2)) \geq 0.$$

This essentially implies that the DM be (weakly) more informed when choosing $c$ in DP 2 than when choosing $a$ in DP 1 - a condition which makes intuitive sense: the rewards to information processing are higher in DP 2 than in DP 1. Notice that these conditions also imply if the DM chooses to be uninformed in DP 2 - meaning that $b$ is never chosen and $P_2(c|\omega_1) = P_2(c|\omega_2)$, then $b$ also cannot be chosen in DP 1.

Together, NIAS and NIAC are necessary and sufficient for the data in experiment 1 to be consistent with the general model.

### 3.3 Experiment 2: Changing Incentives

Our second experiment is designed to examine how subjects change their attention as incentives change. We do so using the simplest possible design: decision problems consist of two actions and two equally likely states, with the reward for choosing the 'correct' state varying between problems. Table 4 shows the four DPs that were administered in experiment 2. Payoffs were in probability points for a prize of $40, with subjects facing 50 repetitions of each DP. Again, states 1 and 2 were represented by 49 and 51 red balls respectively.

Table 4: Experiment 2

| | Payoffs | | | |
|---|---|---|---|---|
| DP | $U(a,1)$ | $U(a,2)$ | $U(b,1)$ | $U(b,2)$ |
| 3 | 5 | 0 | 0 | 5 |
| 4 | 40 | 0 | 0 | 40 |
| 5 | 70 | 0 | 0 | 70 |
| 6 | 95 | 0 | 0 | 95 |

The primary aim of this experiment is to provide estimates of the cost function associated with information acquisition. However, in order for this to be meaningful it must be the case that behavior is rationalizable with some underlying informational cost function. We therefore begin by testing the NIAS and NIAC conditions which are necessary and sufficient for such a cost function to exist. In this setting these conditions take on a particularly simple form. NIAS - which guarantees that subjects are using the information they have efficiently - requires that

$$P_i(a|1) \geq P_i(a|2) \text{ for } i \in \{5, 6, 7, 8\}.$$

This condition simply states that the subject must be more likely to choose the action $a$ in state 1 (when it pays off a positive amount) than in state 2 (when it does not). If and only if this condition holds then $a$ (resp. $b$) is the optimal choice of action given the posterior probabilities over states when $a$ ($b$) was chosen. See Caplin and Dean [2015] section E for the derivation of the NIAS and NIAC conditions for experiments 2 and 3.

NIAC is the condition which ensures that behavior is consistent with some underlying cost function. In this setting it is equivalent to requiring that subjects become no less accurate as incentives increase - i.e.

$$\begin{aligned} & P_8(a|1) + P_8(b|2) \\ \geq \; & P_7(a|1) + P_7(b|2) \\ \geq \; & P_6(a|1) + P_6(b|2) \\ \geq \; & P_5(a|1) + P_5(b|2) \end{aligned}$$

This condition guarantees that gross payoff cannot be increased by reallocation information structures across decision problems.

Having established that some rationalizing cost function exists, we can consider what it looks like. Of particular interest is whether behavior is consistent with Shannon costs. In order to determine this, we can make use of the ILR condition above. Assuming that utility is linear in probability points, this implies that

$$\begin{aligned} \kappa &= \frac{\ln(\gamma_5^a(1)) - \ln(\gamma_5^b(1))}{5} \\ &= \frac{\ln(\gamma_6^a(1)) - \ln(\gamma_6^b(1))}{40} \\ &= \frac{\ln(\gamma_7^a(1)) - \ln(\gamma_7^b(1))}{70} \\ &= \frac{\ln(\gamma_8^a(1)) - \ln(\gamma_8^b(1))}{95} \end{aligned}$$

Where $\gamma_j^a(1)$ is the posterior probability of state 1 in decision problem $j$ following the choice of action $a$ (recall that these posteriors can be directly observed in the SDSC data). Moreover, the symmetry of the Shannon model implies that $\gamma_j^a(1) = \gamma_j^b(2)$.

Thus, while the general model implies only that the probability of making the correct choice is non-decreasing in reward, the Shannon model implies a very specific rate at which subjects must improve. Effectively, behavior in a single decision problem pins down the model's one free parameter, $\kappa$, which then dictates behavior in all other decision problems. The class of models introduced in section 2.4 relax this constrain somewhat - allowing for two parameters rather than one. We can use the data from experiment 2 to fit this class of models in order to determine if provides an improvement over the Shannon assumption.

## 3.4 Experiment 3: Changing Priors

The third experiment studies the impact of changing prior probabilities. Again we use the simplest possible setting with two states (47 and 53 red balls respectively)[17] and two acts. Again there are 4 decision problems, each of which is repeated 50 times. Because this experiment made use of only two payoff levels, payment was made in cash, rather than probability points. Table 5 describes the 4 decision problems with payoffs denominated in US Dollars

Table 5: Experiment 3

| | | Payoffs | | | |
|---|---|---|---|---|---|
| DP | $\mu(1)$ | $U(a,1)$ | $U(a,2)$ | $U(b,1)$ | $U(b,2)$ |
| 7 | 0.50 | 10 | 0 | 0 | 10 |
| 8 | 0.60 | 10 | 0 | 0 | 10 |
| 9 | 0.75 | 10 | 0 | 0 | 10 |
| 10 | 0.85 | 10 | 0 | 0 | 10 |

Each DP has two acts which pay off \$10 in their correct state. The only thing that changes between the decision problems is the prior probability of state 1, which increases from 0.5 in DP 7 to 0.85 in DP 10.

The general model has only a limited amount to say about behavior in experiment 3. NIAC has no bite, as the general model puts no constraint on how information costs change with changes in prior beliefs. However, NIAS must still hold - subjects must still use whatever

[17] We used a somewhat easier setting for this experiment (relative to experiment 2) in order to ensure that most subjects collected some information in the baseline DP 7.

information they have optimally. For this experiment the NIAS condition implies

$$P_i(a|1) \geq \frac{2\mu(1)-1}{\mu(1)} + \frac{1-\mu(1)}{\mu(1)} P_i(a|2)$$

Again this condition derives from the necessity that each action be the best option at the distribution over states conditional on it being chosen.

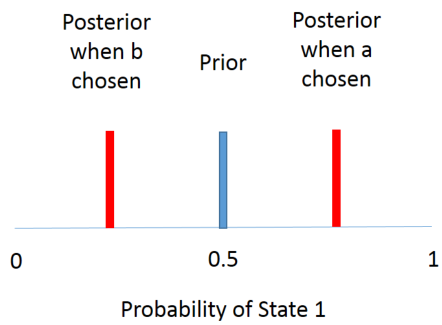
A natural alternative model is one of base rate neglect (see for example Tversky and Kahneman [1974]), in which subjects ignore changes in prior probabilities when assessing alternatives. A DM who ignored the impact of changing priors on their posterior would be in danger of violating NIAS as $\mu(1)$ increases.

In contrast, the posterior separable class of models puts a lot of structure on behavior as priors change, as captured by the LIP condition. Figure 2 demonstrates the testable implications of the LIP condition for experiment 3. First, one observes the posterior beliefs associated with the choice of $a$ and $b$ in DP 7, when the $\mu(1) = 0.5$ (panel a). Then, as the prior probability of state 1 increases, there are only two possible responses. First, if the prior remains inside the convex hull of the posteriors used at $\mu(1) = 0.5$ the subject must use precisely the same posteriors (panel b).[18] Second, if the prior moves outside the convex hull of the posteriors used at $\mu(1) = 0.5$ the subject should learn nothing, and choose option

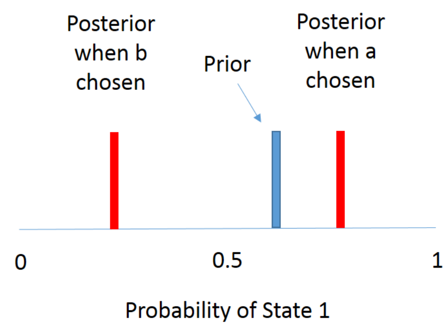
[18] Of course, for Bayes' rule to hold, it must also be the case that the unconditional probability of choosing option $a$ increases.

$a$ in all trials (panel c).

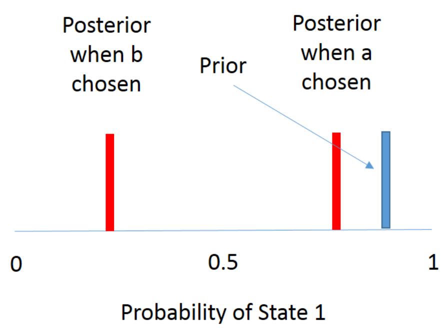
Figure 2: Locally Invariant Posteriors




Panel a: Posteriors used when $\mu(1) = 0.5$




Panel b: Same posteriors are used if they remain feasible




Panel c: No learning takes place if posteriors are infeasible

## 3.5 Experiment 4: Invariance Under Compression

Our final experiment is designed to test the property of IUC which is inherent in the Shannon model.[19] Consider the set up described in table 6. There are $N$ equally likely states of the world and two actions, $a$ and $b$. Action $a$ pays off \$10 in states of the world $\{1, ... \frac{N}{2}\}$ and

[19] This experimental design was developed a part of a distinct project on information acquisition in global games. See Dean *et al.* [2016].

zero otherwise, while action $b$ pays off \$10 in states $\{\frac{N}{1}+1, N\}$ and zero otherwise.

| Table 6: Experiment 4 | | |
|---|---|---|
| | Payoffs | |
| States | $1, ...\frac{N}{2}$ | $\frac{N}{2}+1, .., N$ |
| $a$ | 10 | 0 |
| $b$ | 0 | 10 |

The predictions of the Shannon model in this environment can be readily determined from the ILR condition, which shows that posterior beliefs following the choice of each act depend only on the relative payoff the available acts in that state. This implies immediately that behavior should be equivalent in all states between 1 and $\frac{N}{2}$ and in all states between $\frac{N}{2}+1$ and $N$. This is a manifestation of the IUC condition.

This illustrates an important behavioral feature of the Shannon model, which is that it is symmetric. A permutation of prior beliefs and payoffs across states should lead to the equivalent permutation of SDSC data, so behavior is essentially invariant to the labelling of states. The model lacks any sense of 'perceptual distance', by which some states are harder to distinguish that others. According to the Shannon model, subjects are no more likely to make mistakes in states that are close to the threshold of $\frac{N}{2}$ than those that are far away.

Whether or not this is a reasonable assumption is likely to depend on the task at hand. We therefore test this implication in two different decision making environments which differ in the extent to which there is a natural perceptual distance between states. The first (the 'Balls' treatment) makes use of the same interface as experiments 1-3: states are represented by the number of red balls centered around 50. Subjects in this experiment faced four different DPs, each of which was repeated 50 times. DPs varied in the number of possible states - from 8 to 20 (so, for example, in the 8 state treatment there could be between 47 and 54 red balls, while in the 20 state treatment there could be between 41 and 60 red balls).

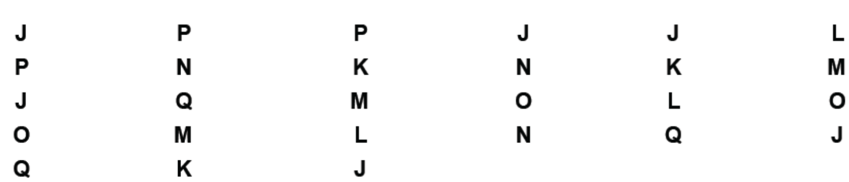



Figure 3: Alternative perceptual task

The second environment (the 'Letters' treatment) makes use of a different perceptual task. Subjects are shown a grid of letters, of the type displayed in figure 3. One of these letters appears more often than the others. We refer to this as the 'state' letter. The position of the state letter in the alphabet determines the state: which act pays off depends on whether the state letter is before or after the letter $\not{N}$. Again, subjects in this treatment faced 50 repetitions of 4 decision problems which varied the number of state letters in each grid (non-state letters were always repeated 3 times) and the number of possible states (i.e. the number of possible letters). Table 7 summarizes the various decision problems that go to make up experiment 4. Subjects either faced DPs 11-14 (i.e. the balls treatment) or 15-17 (the letters treatment).

Table 7: Treatments for Experiment 4

| DP | Stimuli | # States | # Letter Repetitions |
|---|---|---|---|
| 11 | Balls | 8 | N/A |
| 12 | Balls | 12 | N/A |
| 13 | Balls | 16 | N/A |
| 14 | Balls | 20 | N/A |
| 15 | Letters | 8 | 4 |
| 16 | Letters | 12 | 6 |
| 17 | Letters | 16 | 7 |
| 18 | Letters | 20 | 10 |

# 4 Implementation and Results

Subjects were recruited from the New York University and Columbia University student populations.[20] At the end of each session, one question was selected at random for payment, the result of which was added to the show up fee of $10. Subjects usually took between 45 minutes and 1.5 hours to complete a session, depending on the experiment. Instructions are included in the online appendix.

[20] Using the Center for Experimental Social Science subject pool at NYU and the Columbia Experimental Laboratory in the Social Science subject pool at Columbia.

## 4.1 Matching Theory to Data

The theoretical implications above are couched in terms of the population distribution of SDSC - i.e. the true probability of a given subject choosing each possible alternative in each state of the world. Of course this is not what we observe in our experiment for two reasons. First, we are only able to make inferences based on estimates of these underlying parameters from finite samples. Second, in order to generate these samples we will need to aggregate over decision problems and/or individuals.

Taking the second problem first - we make use of two types of aggregation in the following results. First, because we make each subject repeat the same decision problem numerous times, we can estimate SDSC data at the subject level. Second, we can aggregate over subjects who have faced the same decision problem which gives us more observations and so more power. In principle, both of these might be problematic. In the former case, while each repetition of the decision problem is the same if states are defined by number of red balls, the actual configuration of red and blue balls vary from trial to trial in order to prevent learning. It could be that some configurations are easier to understand than others (in extremis, all the red balls could appear on the left of the screen while all the blue ones appear on the right). Aggregating across individuals may also cause problems, because different individuals may have different costs of attention.

For most of the tests that we perform this aggregation does not present a problem. For example, in experiment 1 we look for violations of Monotonicity by studying whether the probability of choosing $b$ increases when $c$ is introduced to the action set. Consider a DM for whom monotonicity holds conditional on the difficulty of the problem, as represented by the configuration of dots on the screen. This means that, when sampling from different configurations, the distribution of probabilities of $b$ being selected when $c$ was not available should stochastically dominate that when $c$ is available, and so Monotonicity should hold in expectation. Similarly, aggregating across subjects for whom Monotonicity holds should lead to monotonic data.

Two case in which aggregation may cause problems are (i) when we are estimating the rate at which accuracy responds to incentives, for example when comparing the data to the Shannon model in experiment 2, and (ii) when testing the LIP condition in experiment 3. We discuss this issue in more depth in section 5 below.

Once we have done the aggregation, we are still faced with the fact that we only observe estimates based on finite samples, and so we can only make probabilistic statements about whether a given condition holds for the underlying data generating process. Broadly speaking, there are two possible types of test we can perform: we can either look for evidence that an axiom is violated, or that it holds. Take again the example of Monotonicity, which states that $P_{\{a,b\}}(b|2) \geq P_{\{a,b,c\}}(b|2)$. On the one hand, we could ask whether one can reject the hypothesis that $P_{\{a,b\}}(b|2) < P_{\{a,b,c\}}(b|2)$. On the other, one could try to reject the hypothesis that $P_{\{a,b\}}(b|2) \geq P_{\{a,b,c\}}(b|2)$. In the former case, a rejection of the hypothesis would provide convincing evidence that the axiom holds. In the second, it would provide convincing evidence that the axiom is violated. The difference between the two tests is whether the axiom is given the 'benefit of the doubt', in terms of data which is not statistically distinguishable from $P_{\{a,b\}}(b|2) = P_{\{a,b,c\}}(b|2)$. Note that the probability of observing such data should fall as more data is collected, and so power increases. Typically we will use the former approach for data aggregated across subjects, where we have enough observations to provide powerful tests, and the latter for individual level data where we have less power.

Note that the null hypotheses above are defined in terms of inequalities. This is typically the case for the tests we employ. When testing against a null hypothesis which encompasses an entire region of the parameter space, there are two possible approaches. The Bayesian approach is to assign some prior to the parameter space and then update it using the data. The null is rejected if 95% of the posterior weight falls outside the null region. The frequentist approach simply treats the null hypothesis as a single point hypothesis placed at the location in the null region which is the most favorable to the null hypothesis. In this paper we will use this approach - so, in the case of Monotonicity, we will use a two sided test against the null of $P_{\{a,b\}}(b|2) = P_{\{a,b,c\}}(b|2)$, regardless of whether we are taking as the null that the axiom holds or that it is violated.

When aggregate data is used, standard errors are corrected for clustering at the subject level.

A further potential issue is the fact that attention costs may vary over the course of the experiment due to fatigue or learning effects. We discuss this issue in section 5.1.

## 4.2 Experiment 1: Testing for Responsive Attention

Table 8 summarizes the results of the Monotonicity tests from experiment 1. The first panel reports the probability of choosing action $b$ in state 1 with and without $c$ available, and the results of a statistical test to determine whether the latter is higher that the former. The second panel repeats the exercise for $P(b|2)$. The final column reports the fraction of subjects who show a significant violation of monotonicity. 28 subjects took part in this experiment, evenly divided across the 4 treatments.

Table 8: Results of Experiment 1[21]

| | | $P(b\|1)$ | | | $P(b\|2)$ | | | |
|---|---|---|---|---|---|---|---|---|
| Treat | N | $\{a,b\}$ | $\{a,b,c\}$ | Prob | $\{a,b\}$ | $\{a,b,c\}$ | Prob | % Subjects |
| 1 | 7 | 2.9 | 6.8 | 0.52 | 50.6 | 59.8 | 0.54 | 29 |
| 2 | 7 | 5.7 | 14.7 | 0.29 | 21.1 | 63.1 | 0.05 | 43 |
| 3 | 7 | 9.5 | 5.0 | 0.35 | 21.4 | 46.6 | 0.06 | 29 |
| 4 | 7 | 1.1 | 0.8 | 0.76 | 19.9 | 51.7 | 0.02 | 57 |
| Total | 28 | 4.8 | 6.6 | 0.52 | 28.3 | 55.6 | <0.01 | 39 |

Aggregating across individuals and treatments (final row), we find a significant violation of monotonicity in the direction predicted by models of rational inattention. The probability of choosing $b$ in state 2 increases from 28.3% to 55.6% following the introduction of $c$, significant at the 1% level. The increase in the choice of $b$ in state 1 is small and insignificant. At the individual level, 39% of subjects show a significant violation of monotonicity. Disaggregating by treatment, we see that the point estimate of $P(b|2)$ increases with the introduction of $c$ in all treatments, significantly so (at the 10% level) in treatments 2-4.

Table 9 reports the results of the NIAS tests for experiment 1 using aggregate data.[22] The first column reports the mean value for the LHS of the tests described in table 3. Recall that the NIAS condition requires each of these to be positive. The second column reports the probability associated with a test of the hypothesis that this value is equal to zero. Five of the seven tests provide strong evidence in favor of NIAS with point estimates significantly greater than zero. The two remaining tests have estimates which are not significantly different from

[21] P values from a logit regression of the choice of option $b$ on dummies representing whether or not $c$ was present and whether the state was 1 or 2. Standard errors clustered at the individual level.

[22] Estimate for the first row generated by constructing, for each choice and each individual

$$\begin{aligned}&\frac{1(choose\_a).1(state\_1)}{P(1)}(50-b1)\\&+\frac{1(choose\_a).1(state\_2)}{P(2)}(50-b2)\\&-100+(b_1+b_2)\end{aligned}$$

where $1(choose\_a)$ is a dummy which takes the value 1 if $a$ is chosen, $1(state\_i)$ is a dummy which takes the value 1 if the state is $i$ and $P(i)$ is the empirical frequencey of state $i$. Averaging over these values provides an estimate of the LHS of the first NIAS test described in table 3. P values were found using bootstrapping with standard errors clustered at the individual level.

Data for other rows constructed using the same method.

zero. In the comparison between $a$ and $c$ in DP 2 the point estimate is actually negative - though not significantly so. This implies that people were choosing $a$ when in fact it would have provided (marginally) higher expected utility to choose $c$. One possible explanation for this is a form of 'certainty bias' for probability points: subjects may have liked the fact that $a$ provides a 'sure thing' of 50 points, while $c$ is 'risky'

Table 9: NIAS Tests for Experiment 1

| | Aggregate Data | |
|---|---|---|
| Test | Est. | P |
| NIAS DP 1 | 0.30 | 0.41 |
| NIAS DP 2 a vs b | 5.46 | 0.00 |
| NIAS DP 2 a vs c | -0.02 | 0.31 |
| NIAS DP 2 b vs a | 1.07 | 0.06 |
| NIAS DP 2 b vs c | 25.57 | 0.00 |
| NIAS DP 2 c vs a | 0.47 | 0.00 |
| NIAS DP 2 c vs b | 30.66 | 0.00 |

The NIAC condition requires that $(P_2(c|\omega_1) - P_2(c|\omega_2)) - (P_1(a|\omega_1) - P_1(a|\omega_2))$ is greater than zero. In the aggregate data the expected value of this expression is 0.234, significantly different from 0 at the 5% level.[23]

At the individual level we observe only a small number of significant violations of NIAS or NIAC. Of the 28 tests of NIAS in DP 1 we find 3 violations. In DP 2 of the 168 tests we find 6 violations. For NIAC, we find 2 significant violations in 28 tests.

## 4.3 Experiment 2: Changing Incentives

We next report the results from experiment 2 in which we examine subjects' responses as we change incentives. 52 subjects took part in this experiment.

We begin by testing the NIAS and NIAC conditions which are necessary and sufficient for the general model. Table 10 reports the results of the NIAS test - which requires that the probability of choosing $a$ in state 1 must be higher than in state 2 - using aggregate data. It shows the probability of choosing $a$ in each state for each decision problem, and the probability of failing to reject the null that NIAS is violated. The aggregate data firmly

[23] Point estimates and standard errors calculated as in the NIAS tests above.

supports the NIAS condition.

| Table 10: NIAS Test from Experiment 2[24] | | | |
|---|---|---|---|
| DP | $P_j(a\|1)$ | $P_j(a\|2)$ | Prob |
| 3 | 0.74 | 0.40 | 0.00 |
| 4 | 0.76 | 0.34 | 0.00 |
| 5 | 0.78 | 0.33 | 0.00 |
| 6 | 0.78 | 0.28 | 0.00 |

Figure 4 shows the probability of choosing the 'correct' act in each DP, averaging across all subjects, which allows us to test the NIAC condition which states that this probability should be non-decreasing in the reward level. The point estimates obey this pattern, with accuracy increasing from 67% at the 5 probability point level to 75% at the 95 probability point level. Most of the differences between DPs are significant at the 10% level.[25]

[24] Results of a logistic regression of choice of action a on a dummy for state 1. P value reported is that associated with the state 1 dummy. Standard errors clustered at the subject level.

[25] Estimates and standard errors produced using a logit regression of correct choice on treatement, with standard errors clustered at the individual level. The success rate at the 5 probabilty point level is significantly different from that of 95 prob. point level at <0.1%, and different from the 40 and 70 levels at 10%. The 40 probability point level is significantly different from the 95 level at 10%. The 70 probability point level is not significantly different from either the 40 or the 95 level.

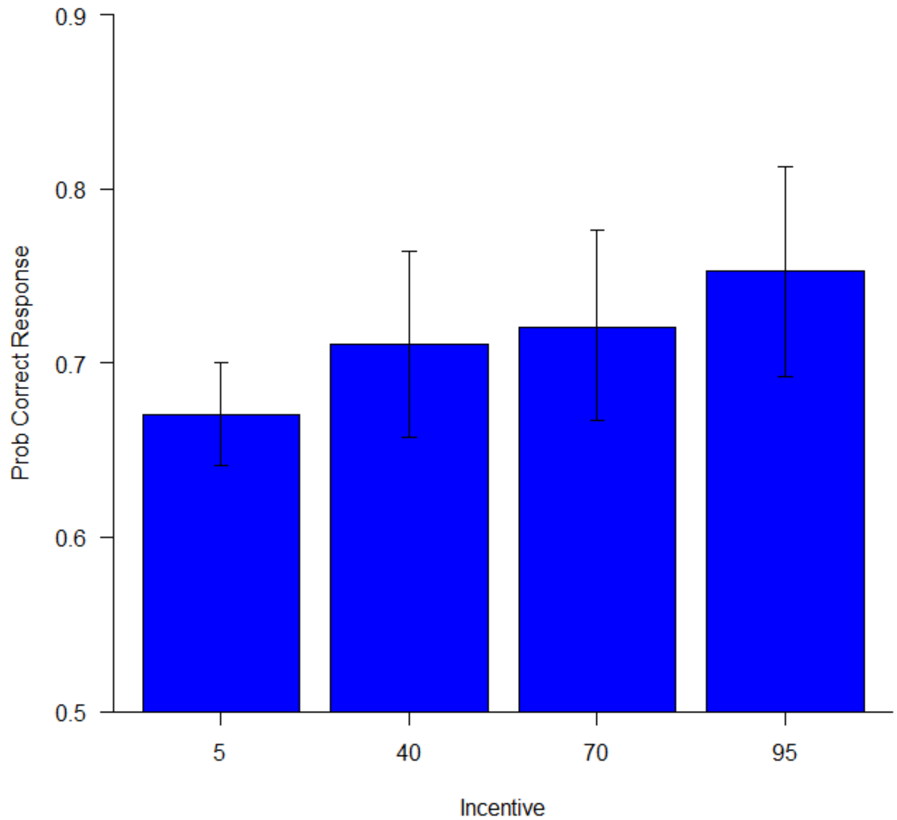



Figure 4: Probability of Correct Response by Decision Problem

Table 11 reports the individual level data, and in particular the fraction of subjects who exhibit significant violations of the NIAS condition, the NIAC condition, both or neither. 81% of subjects show no significant violations of either condition.[26]

[26]We checked the NIAC condition and the NIAS conditions separately for each individual. The NIAS condition was tested by simply estimating a robust OLS model regressing probability of choice of action $a$ on state. If the coefficient was significantly negative that is considered a significant violation of NIAS.

NIAC was checked by estimating a GLM regression. In this model a dummy for correct response was regressed against dummies for the three higher incentive levels. We then preformed an F-test of the joint restrictions that (i) the dummy on to 40 probability point level was greater than or equal to 0, (ii) that the dummy on the 70 point level was greater than or equal to that on the 40 point level and (iii) that the dummy on the 95 level was greater than equal to that on the 70% level. Subjects were categorized as violating NIAC if these restrictions were jointly rejected.

| Table 11: Individual Level Data from Experiment 2 | |
|---|---|
| Violate | % |
| NIAS Only | 2 |
| NIAC Only | 17 |
| Both | 0 |
| Neither | 81 |

Table 11 implies that most of our subjects do not have significant violations of the NIAS and NIAC conditions and therefore act as if they maximize relative to some underlying cost function. Figure 5 gives some idea of the heterogeneity of those costs across subjects. It graphs the probability of choosing the correct action at the 5 point level vs the 95 point level for each subject. The fact that most points fall above the 45 degree line is the defining feature of rational inattention. However, within this constraint there is still a great deal of variation. Our data set includes 'high fixed cost, high marginal cost' individuals who gather little information regardless of reward: their accuracy is near 50% for the low and high reward levels. It includes 'low fixed cost' subjects who have accuracy close to 100%, even in the low reward decision problem. Finally there are 'high fixed cost, low marginal cost' subjects, who actively adjust their accuracy as a function of reward.

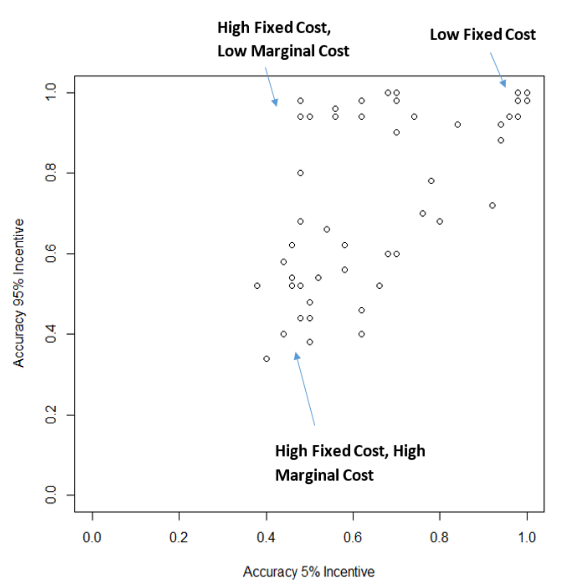



Figure 5: Individual Accuracy 5 Point vs 95 Point Reward

We next examine the extent to which subjects behave as if their costs are in line with the Shannon model. Figure 6 shows the estimated costs $\kappa$ from each decision problem and in each state using aggregate data. The Shannon model would predict that these should be equal. As we can see this is not the case: estimated costs are increasing in reward level - the estimated costs are significantly different at the 0.01% level between the 5 and 95 point reward levels. The fact that estimated costs are increasing implies that subjects are not increasing their accuracy fast enough in response to changing incentives relative to the predictions of the Shannon model.[27]

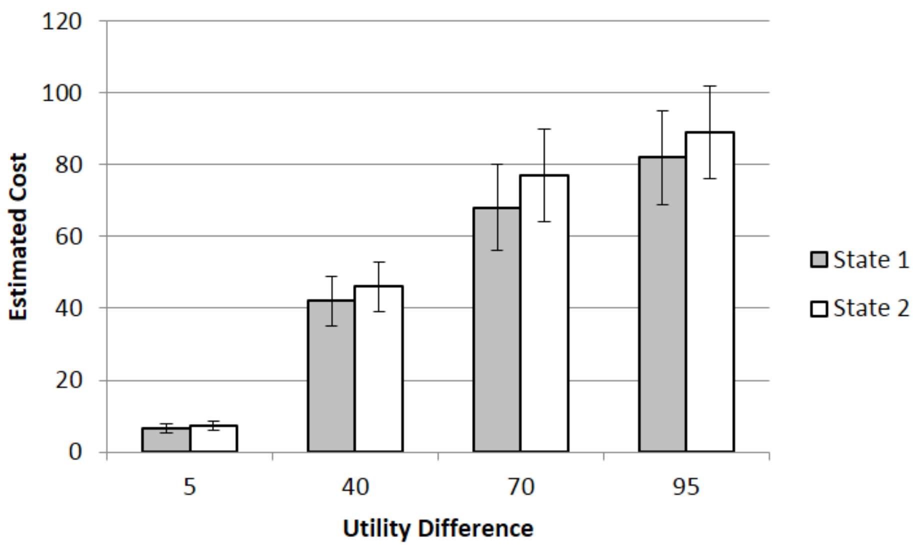



Figure 6: Estimated Costs

We next examine whether the class of cost functions introduced in section 2.4 does a better job of fitting the data. Figure 7 shows actual vs predicted accuracy at each reward level for the Shannon model and the model defined by the posterior separable cost function $T_{\{\rho,\kappa\}}$.[28] Since the Shannon model is nested within the class of $T_{\{\rho,\kappa\}}$ functions, this broader

[27] Cost estimates and standard errors based on an OLS regression of response on treatment and state dummies, which were then converted into cost estimates using the method discussed in section 3.3. Standard errors are clustered at the individual level.

[28] Models are fitted to the data by maximum likelihood. The error bounds were found through bootstrap resampling of participants and reestimation of the parameters of the models. The error bars on the Shannon model are very small for the low incentive level, because the model essentially pins down the intercept while leaving the slope free. Therefore, a wide range of prections for the incentive level ninety-five will come from a set of parameters which only produce a very narrow range of predictions at incentive level five. The data for the higher incentive levels are then very precisely pinning down the predictions at low incentive levels

class must weakly provided a better fit of the data. However, criteria that punish models for having additional parameters still suggest rejecting Shannon in favor of the broader parametrized class. For example the Akaike Information Criterion (AIC) is lower for the $T_{\{\rho,\kappa\}}$ model than for the Shannon model.[29]

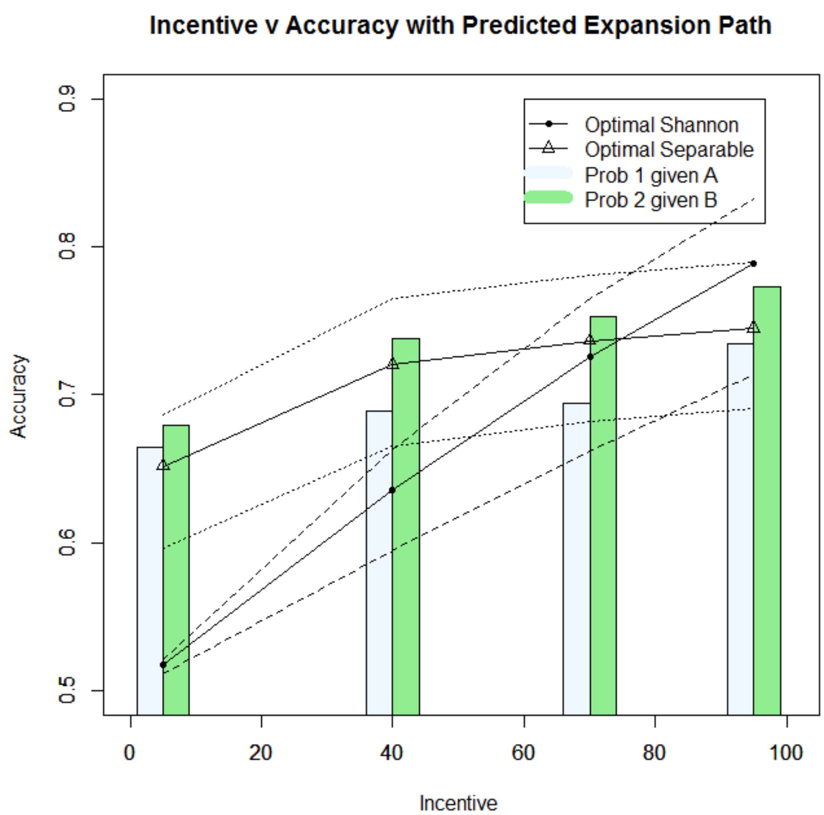



Figure 7

At the individual level we also see significant violations of the Shannon model. Figure 8 shows a scatter plot of the predicted vs actual accuracy for each subject in the 70 point DP, where the prediction is made using the Shannon model and the accuracy displayed at the 40 point level.[30] The scatter plot shows more subjects below the 45 degree line (i.e. are less accurate than predicted) than above (more accurate than predicted).[31] For each

even if the low incentive level data does not match those predictions well.

[29] The AIC is 12750.67 for the Shannon model and 12424.41 for the $T_{\{\rho,\kappa\}}$ model.

[30] We use these two reward levels to illustrate our findings because the predictions derived from more extreme comparisons typically cluster at the extremes, making the associated graph hard to interpret.

[31] For the analysis described in this paragraph we drop observations in which the point estimate for accuracy

subject and pair of reward levels we can test for significant violations of the Shannon model which indicate ‘too slow’ adjustment relative to the Shannon model (i.e. the accuracy at the higher reward is lower than it should be given the accuracy at the lower reward level), and for violations which indicate ‘too fast’ adjustment (accuracy at the higher reward level is higher than it should be).[32] Of the 221 possible comparisons, we find 66 violations of the ‘too slow’ variety and 8 of the ‘too fast’ variety. 21 subjects exhibit ‘too slow’ violations only, 4 exhibit ‘too fast’ violations’ only, 2 have examples of both and 21 examples of neither.

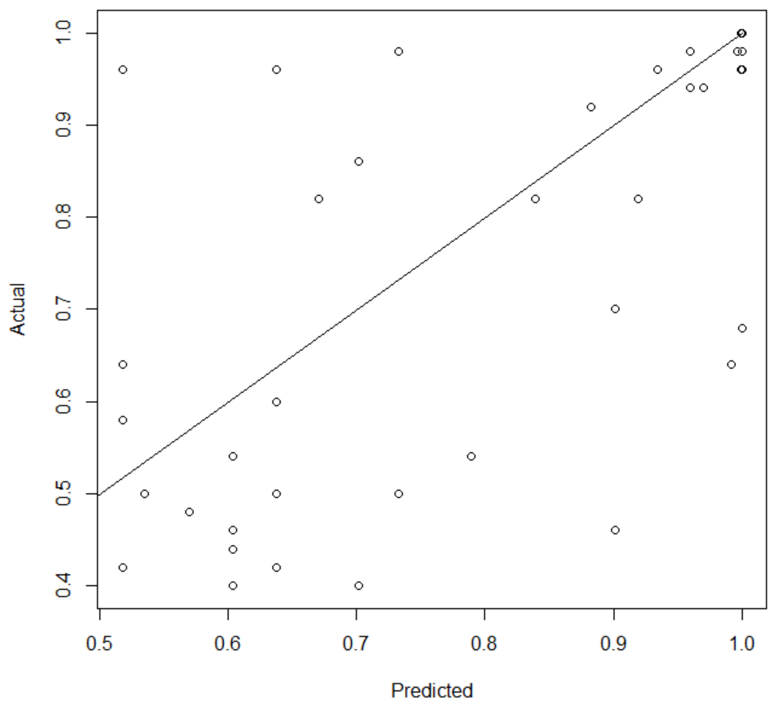



Figure 8: Predicted vs actual accuracy in the 70% payoff treatment

at the lower reward level is below 50%, as this does not allow us to recover the cost parameter of the Shannon model and so make predictions for the higher cost level.

[32]For each person and each incentive level pair we regress correctness on incentive level and a dummy for the higher incentive level with no constant using a logit regression. Note that a logit regression of correctness on incentive level with no constant is equivalent to fitting a Shannon model in this case. Significant coefficients on the high incentive dummy mean significant violations of Shannon. Positive coefficients mean that accuracy is responding too fast while negative coefficients mean it is responding too slow.

It could be that the violations of Shannon we observe are driven by those subjects that do not satisfy the conditions of the general model - i.e. violate NIAS or NIAC. In order to explore this possibility we repeat our analysis dropping such subjects and report the results in Appendix A2. We still find widespread and systematic violations of the Shannon model when focusing only on subjects whose behavior is rationalizable using some cost function.

## 4.4 Experiment 3: Changing Priors

We first examine the extent to which the 54 subjects in experiment 3 obeyed NIAS. Table 12 shows the aggregate probability of choosing act $a$ in state 2, the resulting constraint on the probability of choosing $a$ in state 1, and the related probability in the data. The final column shows the probability of failing to reject the null hypothesis that NIAS is violated in the aggregate data. Broadly speaking, NIAS holds at the aggregate level: the point estimates for $P(a|1)$ are at or above the constraint for all decision problems, significantly so for decision problems 7-9.

Table 12: NIAS Test[33]

| DP | $P_j(a\|2)$ | Constraint on $P_j(a\|1)$ | $P_j(a\|1)$ | Prob |
|---|---|---|---|---|
| 7 | 0.29 | 0.29 | 0.77 | 0.000 |
| 8 | 0.38 | 0.39 | 0.88 | 0.000 |
| 9 | 0.40 | 0.80 | 0.90 | 0.045 |
| 10 | 0.51 | 0.91 | 0.91 | 0.538 |

This pattern is repeated at the individual level, where we see only a small number of subjects exhibiting significant violations of NIAS, as reported in table 13. Thus, we see little evidence of base rate neglect in this data.

Table 13: Individual Level NIAS violations[34]

| Decision problem | 7 | 8 | 9 | 10 |
|---|---|---|---|---|
| Prior | 0.5 | 0.6 | 0.75 | 0.85 |
| % Significant Violations | 0 | 2 | 2 | 11 |

We next study the degree to which our data supports the predictions of the posterior separable model in the form of the LIP condition. In order to do so, we first divide subjects

[33]Tests based on an OLS regression of choice of action on state for each treatment to obtain estimates of $P_j(a|2)$ and $P_j(a|1)$. Standard errors clustered at the individual level. These estimates then used in a test of the linear restriction implied by the NIAS model.

[34]Tests based on the same method reported for the aggregate data in table 13.

based on the estimated posteriors in DP 7, in which both states are equally likely. The important distinction is where the posterior associated with the choice of action $a$ falls relative to the priors for DPs 8-10. Table 14 shows this categorization.

Table 14: Categorization Based on Posteriors from DP 7

| Posterior Range | N | % |
|---|---|---|
| [0.5,0.6) | 14 | 25 |
| [0.6,0.75) | 12 | 22 |
| [0.75,0.85) | 12 | 22 |
| [0.85,1] | 16 | 29 |

The first prediction of the posterior separable model is that, in DP $i$ with prior $\mu_i(1)$, subjects who use a posterior $\gamma_7^a(1) < \mu_i(1)$ should exclusively choose action $a$, while those with $\gamma_7^a(1) > \mu_i(1)$ should choose both $a$ and $b$, where $\gamma_7^a$ refers to the posteriors revealed in DP 7 given the choice of $a$. Table 15 tests this 'no learning' prediction. The top panel divides subjects into those who have a threshold (i.e. posterior belief from DP 7) above $\mu_i(1)$, and those for whose threshold is below $\mu_i(1)$ for $\mu_8(1) = 0.6$, $\mu_9 = 0.75$ and $\mu_{10} = 0.85$. For each of these decision problems, and each of these groups, it then reports the fraction of subjects who exclusively choose $a$. The second panel repeats the exercise but allows for some mistakes on the part of the subject by replacing the condition 'never choose $a$' with the condition 'choose $a$ less than 3 times' (out of 50) in the DP.

Table 15: Testing the 'no learning' prediction

| | | $\mu(1)$ | | |
|---|---|---|---|---|
| | | DP8 | DP9 | DP10 |
| | | 0.6 | 0.75 | 0.85 |
| Never choose $b$ | $\gamma_7^a(1) < \mu_i(1)$ | 35% | 27% | 29% |
| | $\gamma_7^a(1) \geq \mu_i(1)$ | 0% | 7% | 13% |
| Choose $b< 3$ | $\gamma_7^a(1) < \mu_i(1)$ | 50% | 27% | 37% |
| | $\gamma_7^a(1) \geq \mu_i(1)$ | 3% | 7% | 25% |

Table 15 shows that, while it does not perfectly match our data, the 'no learning' prediction does produce the correct comparative statics. In each DP, around 30% of the subject who should exclusively choose $a$ do so, higher than the equivalent fraction for those who should be choosing both $a$ and $b$. These differences are significant at the 5% level for DP 8 and 9, but not for DP 10.

The second part of the LIP condition states that, in each DP, subjects who are still actively gathering information[35] should use the same posteriors as they did in DP 7. Figure 9 tests this hypothesis. Panel a focusses on DP 8. It reports data exclusively on subjects who should be choosing both $a$ and $b$ in this DP according to the posterior separable model (i.e. those for whom $\gamma_7^a > 0.6$). It shows the estimated posteriors associated with the choice of action $a$ and $b$ in DP 7 and DP 8 aggregating across all such subjects. The LIP prediction is that these posteriors should be the same. Panels b and c repeat this analysis for DPs 9 and 10. Figure 9 shows that LIP holds relatively well when comparing the 0.5 and 0.6 posteriors: neither the posterior associated with the choice of $a$ nor the one associated with $b$ is significantly different across the two decision problems. However, LIP starts to break down as the prior becomes more skewed: The probability of state 2 given the choice of $b$ (i.e. $\gamma_j^b(2)$) is significantly lower when the prior is 0.75 or 0.85 than when it is 0.5 (P<0.01)).[36]

[35] i.e. those for whom their posterior beliefs from DP 7 fall above the prior in that DP.

[36] Tests based on OLS regressions with standard errors clustered at the individual level.

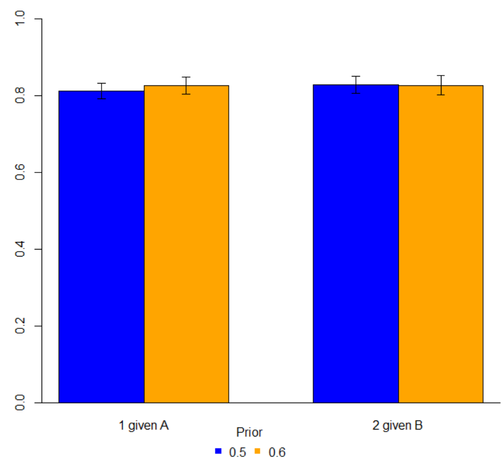



Figure 9a: Subjects with threshold above 0.6

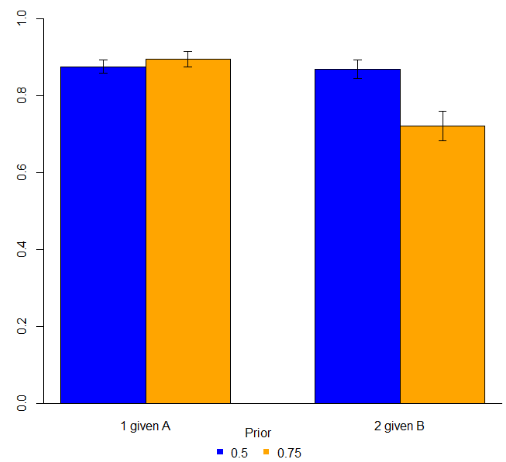



Figure 9b: Subjects with thresholds above 0.75

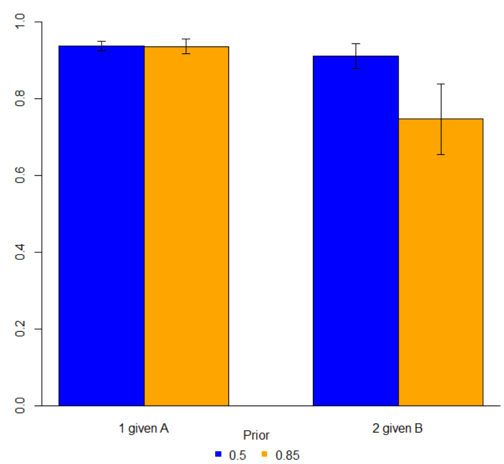



Figure 9c: Subjects with threshold above 0.85

## 4.5 Experiment 4: Symmetry

23 subjects took part in the 'Balls' treatment and 24 in the 'Letters' treatment for experiment 4. The results are summarized in figures 10 and 11, which show the probability of choosing

the correct action as a function of the state for each DP and each treatment.

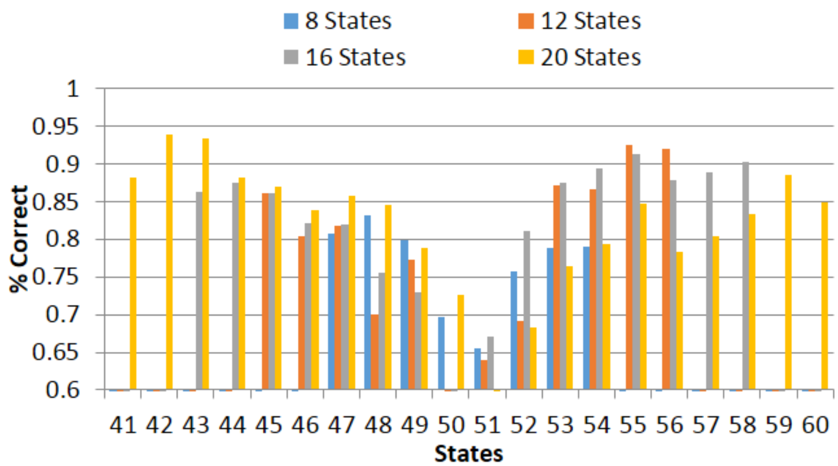



Figure 10: Balls Treatment

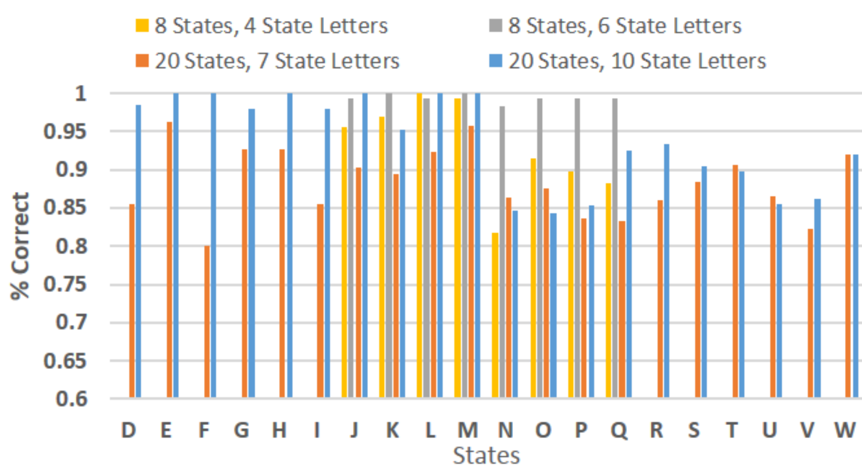



Figure 11: Letters Treatment

These figures show clear and systematic violations of symmetry in the 'Balls' treatment but not in the 'Letters' treatment. Figure 10 shows that, for DPs 11-14 subjects were more likely to make mistakes in states near the threshold of 50. This observation is confirmed by regression analysis, which finds a significant and positive correlation between distance from threshold and probability correct for each DP. No such relationship is observed in the letters treatment.[37] While there are variations in the probability correct across states,[38] these are not significantly correlated with distance from the threshold of N for any decision problem or in aggregate.

# 5 Discussion

Our overall conclusions from this set of experiments are (1) that experimental subjects clearly adapt their attention strategy in response to incentives; (2) that they do so broadly in line with the general model of rational inattention, at least in the aggregate data and in the simple environments we consider; (3) that the Shannon model has some significant difficulties in explaining our data, both in terms of the relationship it predicts between changing rewards and information gathering, and its unrealistic symmetry properties; and (4) the broader class of PS models improves on Shannon by dropping the symmetry property and allowing for a better fit of the relationship between incentives and information gathering; (5) the LIP condition which characterizes such models is violated but also appears to have some empirical bite.

## 5.1 Aggregation and Order Effects

In this section we discuss some of the issues which could effect these conclusions. In particular, could aggregation and order effects be responsible for some of the results we find, and so be the reason we have rejected some models? As discussed in section 4.1, there are two forms of aggregation which might be problematic: across individuals with different cost functions, and across decision problems with different degrees of difficulty. Of the two, we expect the former to be the primary source of variability. Given the large number of balls on the display, the law of large numbers means that we do not expect significant variation in costs across repetitions. For example, difficulty may be related to the degree to which balls

[37] Results from an OLS regression. A distance measure was constructed measuring the absolute distance between the state and the threshold. In the Balls treatment this is equal to the difference between the number of balls on the screen and 50. For the letters treatment this is the number of letters between the state letter and N. Choice is then regressed on distance, which action is correct, and DP separately for balls and letters treatments, aggregating across decision problem. Standard errors clustered at the subject level. The estimated coefficient on distance is 0.032 (P<0.001) in the balls treatment and 0.001 (P=0.694) in the letters treatment.

[38] The main difference in accuracy across states is due to subjects on average being more accurate in states below N than above. The reason for this is that subjects who always gave the same response overwhelmingly chose action $a$, rather than $b$.

clustering by color, the variance of which will be low when the number of balls is large.

Of the tests that we report, those that are potentially affected by aggregation issues are the test of the ILR condition in experiment 2 and the test of LIP in experiment 3. In both cases we report individual level as well as aggregate results. In the former case it is true that variability in information costs could lead to violations of the predictions of the Shannon model in the direction we observe. Data generated by aggregating across different cost levels in each decision problem would respond more slowly to incentives than the Shannon model would predict under the assumption of no cost variation. However, if such variations cause the model to fail at the individual level in an experimental situation where we believe costs to be relatively stable, they are likely to cause trouble for other applications as well. It is hard to think of an application of the model which does not require some aggregation of data.

In the case of the LIP condition, variability in costs would also bias the test towards a rejection of the 'no learning' condition: for example a subject who faced a particularly low cost realization for (say) $\mu(1) = 0.6$ might seek information and choose action $b$, even if they would choose to be uninformed at average information costs. Thus the success rate we find should be treated as a lower bound.

A further question is whether we find evidence of order effects in our data - i.e. evidence that subject's performance changes through the experiment due to, for example, learning effects or fatigue. Our design randomizes the order in which subjects face decision problems, which has two advantages. First, we can estimate the impact of order on performance, and second, such effects should wash out in the aggregate data. Order effects are of most interest in experiments 2 and 3, in which they could have a substantial effect on our conclusions. Appendix A3 reports the result of regressions of accuracy (i.e. the probability of picking the rewarding action) on order (i.e. in which block the question occurred between 1 and 4) while controlling for the type of question and clustering standard errors at the subject level. We find significant order effects in experiment 2 but not in experiment 3. In experiment 2 subjects were more accurate in the first block. No other differences were significant. These order effects could lead to some of the individual level violations of the NIAC and ILR conditions we observe. To the extent that there are undetected order effects in experiment 3, they may explain some of the failures of the 'no learning' prediction discussed in Table 16 However, they should not lead to the significant violations of LIP we see in figure 9.

We note that we do not necessarily see the potential presence of order effects as problem with our experimental design. Rather, it tells us when and how models of rational inattention can be applied. If people exhibit significant fluctuations in the cost of effort due to, for example, fatigue, this means that the general model may work well in making aggregate predictions but be less effective in making predictions at an individual level, unless these fluctuations can be controlled for.

# 6 Related Literature

Many papers have established the importance of attention limits in economically interesting contexts, including consumer choice,[39] financial markets,[40] and voting behavior.[41] There have, however, been far fewer papers that have attempted to test models of inattention. In the experimental literature, Caplin *et al.* [2011] and Geng [2016] test models of sequential search in the 'satisficing' tradition of Simon [1955]. While these papers find evidence of satisficing in the context of choice amongst a large numbers of easily understood alternatives, such models are clearly not suitable for understanding behavior when faced with a small number of difficult to understand alternatives, as we examine in this paper. Indeed, as satisficing behavior can be optimal given a particular information cost function (see Caplin *et al.* [2011]), the satisficing model can be seen as a special case of the models studied here.

Gabaix *et al.* [2006] test a dynamic model of information acquisition in which agents are partially myopic, and so not fully rational, which they label a model of 'directed cognition'. Unlike out paper, search costs are imposed explicitly either through financial costs or time limits. Instead, our aim is to learn about the intrinsic costs to information acquisition that decision makers face. Gabaix *et al.* [2006] also make use of a very different data set, looking at the sequence in which data is collected using Mouselab,[42] rather than the resulting pattern of stochastic choice. The optimal sequence of data acquisition in their set up cannot be readily determined, meaning that it is hard to tell whether their directed cognition model describes the data better than a fully rational alternative.[43] More recent work (Taubinsky [2013], Goecke *et al.* [2013], Khaw *et al.* [2016]) has also focussed on the dynamics of information

[39] Chetty *et al.* [2009], Hossain and Morgan [2006], Allcott and Taubinsky. [2015], Lacetera *et al.* [2012], Pope [2009], Santos *et al.* [2012].

[40] DellaVigna and Pollet [2007], Huberman [2015], Malmendier and Shanthikumar [2007], Bernard and Thomas [1989], Hirshleifer *et al.* [2009].

[41] Shue and Luttmer [2009], Ho and Imai [2008].

[42] An earlier literature used tools such as Mouselab and eye tracking to document what information individuals gather during the process of choice - see Payne *et al.* [1993], and Brocas *et al.* [2014] for a more recent application of these methods to choice in strategic settings. These papers have not genrally used the data to compare behavior to rational benchmarks.

[43] Though see Sanjurjo [2017].

acquisition.

A third set of papers (Pinkovskiy [2009] and Cheremukhin *et al.* [2015]), estimate the Shannon model on experimental data sets in which people make binary choices between gambles. These papers make use of standard stochastic choice data - modeling inconsistent choices as mistakes caused by lack of information - and not the SDSC data we introduce in this paper. While they typically find the Shannon model does well relative to other, non-rational models of stochasticity, they do not focus on the specific features that characterize this model within the general rational inattention class, such as ILR and LIP. For example, Cheremukhin *et al.* [2015] reports that accuracy increases with incentives - effectively a test of NIAC, which is a property of all models of rational inattention. There is no test of the specific properties which characterize the Shannon model.

In contrast to the relatively small amount of work in economics, there is a huge literature in psychology which has used SDSC data in order to understand the processes underlying perception and choice. Many of these studies are used to test the implications of the sequential sampling class of models, in which agents gain information over time, allowing them to arrive at their final decision.[44] Other work has focussed on testing the SDT paradigm introduced in section 2.2.1. See Yu [2014] and Ratcliff *et al.* [2016] for recent reviews, and Krajbich *et al.* [2011] for a discussion of the application of sequential sampling models to economic choice. Some of these studies are similar the design of experiments 2 and 3 in this paper - varying the reward level and prior beliefs in a choice between two uncertain alternatives. Typically these studies focus on subject's ability to successfully complete perceptual tasks.[45] and have design elements that make them unsuitable for our purpose - for example a lack of explicit incentives (e.g. van Ravenzwaaij *et al.* [2012] study the effect of changing priors in an unincentivized task) or a focus on a specific clinical population (for example Reddy *et al.* [2015] look at the response to incentives in schizophrenic subjects). To our knowledge, none of these studies perform the specific tests of the various classes of rational inattention model that we describe here. Neither does the literature include an equivalent of our experiments 1 and 4.

[44] See for example Ratcliff and McKoon [2008] for an introduction to this class of models.

[45] Probably most popular are dot motion tasks (Britten *et al.* [1992]), in which participants are shown screens with numerous moving dots and are asked to determine the overall direction of motion of the group. Ratcliff *et al.* [2016] reviews several studies of this type. Another common perceptual task is the lexical differentiation task (e.g. Zandt *et al.* [2000]) in which participants are asked to differentiate between letters or words based on some given rule. The last common experimental approach is static geometric estimate (e.g. Ratcliff and Smith [2004]). In these studies, participants are asked to categorize static images based on some visual characteristic such as distance, length, or orientation. It is this static geometric discrimination task that the experiments in this study most closely resembles, although to our knowledge no psychology study has used our precise perceptual task.

# References


Hunt Allcott and Dmitry Taubinsky. Evaluating behaviorally motivated policy: Experimentalevidence from the lightbulb market. *American Economic Review*, 105(8):2051–2538, 2015.

Gadi Barlevy and Pietro Veronesi. Information acquisition in finnancial markets. *Review of Economic Studies*, 67(1):79–90, 2000.

Vojtěch Bartoš, Michal Bauer, Julie Chytilová, and Filip Matějka. Attention discrimination: Theory and field experiments with monitoring information acquisition. *American Economic Review*, 106(6):1437–1475, 2016.

Victor Bernard and Jacob Thomas. Post-earnings-announcement drift delayed price response or risk premium. *Journal of Accounting Research*, 27:1–36, 1989.

Henry David Block and Jacob Marschak. *Contributions to Probability and Statistics*, volume 2, chapter Random orderings and stochastic theories of responses, pages 97–132. Stanford University Press, 1960.

Kenneth H Britten, Michael N Shadlen, William T Newsome, and J Anthony Movshon. The analysis of visual motion: a comparison of neuronal and psychophysical performance. *Journal of Neuroscience*, 12(12):4745–4765, 1992.

Isabelle Brocas, Juan D Carrillo, Stephanie W Wang, and Colin F Camerer. Imperfect choice or imperfect attention? understanding strategic thinking in private information games. *The Review of Economic Studies*, page rdu001, 2014.

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

Andrew Caplin and Mark Dean. Revealed preference, rational inattention, and costly information acquisition. *NBER Working Papers*, (19876), Jan 2014.

Andrew Caplin and Mark Dean. Revealed preference, rational inattention, and costly information acquisition. *The American Economic Review*, 105(7):2183–2203, 2015.

Andrew Caplin and Daniel Martin. The dual-process drift diffusion model: Evidence from response times. *Economic Inquiry*, 2015.

Andrew Caplin and Daniel Martin. A testable theory of imperfect perception. *The Economic Journal*, 125(582):184âĂŞ202, Jul 2015.

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

Andrew Caplin, Mark Dean, and John Leahy. Rational inattention, optimal consideration sets and stochastic choice. 2016.

Andrew Caplin, Mark Dean, and John Leahy. Rationally inattentive behavior: Characterizing and generalizing shannon entropy. 2017. Unpublished.

Anton Cheremukhin, Anna Popova, and Antonella Tutino. A theory of discrete choice with information costs. *Journal of Economic Behavior & Organization*, 113:34–50, 2015.

Raj Chetty, Adam Looney, and Kory Kroft. Salience and taxation: Theory and evidence. *American Economic Review*, 99(4):1145–1177, 2009.

Aubrey Clark. Contracts for information acquisition. 2016.

Mark Dean, Stephen Morris, and Isabel Trevino. Endogenous information structures and play in global games. Mimeo, 2016.

Stefano DellaVigna and Joshua Pollet. Demographics and industry returns. *American Economic Review*, 97(5):1667–1702, 2007.

Andrew Ellis. Foundations for optimal inattention. 2013.

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

Sen Geng. Decision time, consideration time, and status quo bias. *Economic Inquiry*, 54(1):433–449, 2016.

Henry Goecke, Wolfgang J Luhan, and Michael WM Roos. Rational inattentiveness in a forecasting experiment. *Journal of Economic Behavior & Organization*, 94:80–89, 2013.

David Green and John Swets. *Signal Detection Theory and Psychophysics*. Wiley, New York, 1966.

Sanford Grossman and Joseph Stiglitz. On the impossibility of informationally effcient markets. *American Economic Review*, 70(3):393–408, 1980.

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

Christian Hellwig, Sebastian Kohls, and Laura Veldkamp. Information choice technologies. *The American Economic Review*, 102(3):35–40, 2012.

David Hirshleifer, Sonya Seongyeon Lim, and Siew Hong Teoh. Driven to distraction: Extraneous events and underreaction to earnings news. *Journal of Finance*, 64(5):2289–2325, 2009.

Daniel Ho and Kosuke Imai. Estimating causal effects of ballot order from a randomized natural experiment: The california alphabet lottery, 1978-2002s. *Public Opinion Quarterly*, 72(2):216–240, 2008.

Tanjim Hossain and John Morgan. ...plus shipping and handling: Revenue (non) equivalence in field experiments on ebay. *The B.E. Journal of Economic Analysis & Policy*, 6(2):1–30, 2006.

Gur Huberman. Familiarity breeds investment. *Review of Financial Studies*, 14(3):659–680, Jun 2015.

Mel Win Khaw, Luminita Stevens, and Michael Woodford. Discrete adjustment to a changing environment: Experimental evidence. Technical report, National Bureau of Economic Research, 2016.

Ian Krajbich, Carrie Armel, and Antonio Rangel. Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, 13(10):1292–1298, 2010.

I Krajbich, D Lu, C Camerer, and A Rangel. The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in Psychology*, 3:193–193, 2011.

Nicola Lacetera, Devin Pope, and Justin Sydnor. Heuristic thinking and limited attention in the car market. *American Economic Review*, 102(5):2206–2236, 2012.

Juan Sebastian Lleras, Yusufcan Masatlioglu, Daisuke Nakajima, and Erkut Y Ozbay. When more is less: Limited consideration. *Journal of Economic Theory*, 170:70–85, 2017.

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

Ulrike Malmendier and Devin Shanthikumar. Are small investors naive about incentives? *Journal of Financial Economics*, 85(2):457–489, Aug 2007.

Paola Manzini and Marco Mariotti. Stochastic choice and consideration sets. *Econometrica*, 82(3):1153–1176, 2014.

Daniel Martin. Strategic pricing with rational inattention to quality. Mimeo, New York University, 2017.

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

Filip Matejka and Alisdair McKay. Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–98, 2015.

Filip Matějka. Rationally inattentive seller: Sales and discrete pricing. *The Review of Economic Studies*, 83(3):1125–1155, 2016.

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

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

Stephen Morris and Philipp Strack. The wald problem and the equivalence of sequential sampling and static information costs. 2017.

John W Payne, James R Bettman, and Eric J Johnson. *The Adaptive Decision Maker*. Cambridge University Press, 1993.

Maxim Pinkovskiy. Rational inattention and choice under risk: Explaining violations of expected utility through a shannon entropy formulation of the costs of rationality. *Atlantic Economic Journal*, 37(1):99–112, March 2009.

Luca Polonio, Sibilla Di Guida, and Giorgio Coricelli. Strategic sophistication and attention in games: an eye-tracking study. *Games and Economic Behavior*, 94:80–96, 2015.

Devin Pope. Reacting to rankings: Evidence from "america's best hospitals". *Journal of Health Economics*, 28(6):1154–1165, 2009.

Roger Ratcliff and Gail McKoon. The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, 20(4):873–922, April 2008.

Roger Ratcliff and Philip Smith. A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, 111(2):333–367, 2004.

Roger Ratcliff, Philip Smith, Scott Brown, and Gail McKoon. Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, 20(4):260–281, April 2016.

L Felice Reddy, William P. Horan, Deanna M. Barch, Robert W. Buchanan, Eduardo Dunayevich, James M. Gold, Naomi Lyons, Stephen R. Marder, Michael T. Treadway, Jonathan K. Wynn, Jared W. Young, and Michael F. Green. Effort-based decision-making paradigms for clinical trials in schizophrenia: Part 1-psychometric characteristics of 5 paradigms. *Schizophrenia Bulletin*, 41(5):1045–1054, September 2015.

Ricardo Reis. Inattentive producers. *Review of Economic Studies*, 73(3):793–821, 2006.

Arthur J Robson. The biological basis of economic behavior. *Journal of Economic Literature*, 39(1):11–33, 2001.

Adam Sanjurjo. Search with multiple attributes: theory and empirics. *Games and Economic Behavior*, 2017.

Babur De Los Santos, Ali Hortaçsu, and Matthijs Wildenbeest. Testing models of consumer search using data on web browsing and purchasing behavior. *American Economic Review*, 102(6):2955–2980, Oct 2012.

Kelly Shue and Erzo Luttmer. Who misvotes? the effect of differential cognition costs on election outcomes. *American Economic Journal: Economic Policy*, 1(1):229–257, 2009.

Herbert Simon. A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1):99–118, Feb 1955.

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

Dmitry Taubinsky. From intentions to actions: A model and experimental evidence of inattentive choice. *Working Paper*, 2013.

Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131, 1974.

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

Don van Ravenzwaaij, Martin Mulder, Francis Tuerlickx, and Eric-Jan Wagenmakers. Do the dynamics of prior information depend on task context? an analysis of optimal performance and an empirical test. *Frontiers in Psychology*, 3(132), May 2012.

Robert Verrecchia. Information acquisition in a noisy rational expectations economy. *Econometrica*, 50(6):1415–1430, 1982.

Ming Yang. Coordination with flexible information acquisition. *Journal of Economic Theory*, 158:721–738, 2015.

Angela Yu. . *Encyclopedia of Computational Neuroscience*, chapter Decision Making Tasks. Springer, 2014.

Trisha Van Zandt, Hans Colonius, and Robert Proctor. A comparison of two response time models applies to perceptual matching. *Psychometric Bulletin & Review*, 7(2):208–256, 2000.

# 7 Appendix

## 7.1 Appendix A1: NIAS and NIAC for Experiment 1

NIAS demands that, for each action $a \in A$ chosen with positive probability

$$\sum_{\omega \in \Omega} \mu(\omega) P(a|\omega) \left(u(a, \omega) - u(a', \omega)\right) \geq 0$$

for every other available alternative $a' \in A$.

For notational convenience, we will use $P$ to denote the SDSC data arising from the decision problem $\{a.b\}$ and $\hat{P}$ for that arising from $\{a, b, c\}$.

Taking the former DP first, the comparison of $a$ to $b$ requires

$$P(a|\omega_1)(50 - b_1) + P(a|\omega_2)(50 - b_2) \geq 0$$

while the comparison of $b$ to $a$ requires

$$\left(1 - P(a|\omega_1)\right)(b_1 - 50) + \left(1 - P(a|\omega_2)\right)(b_2 - 50)) \geq 0$$

or

$$P(a|\omega_1)(50 - b_1) + P(a|\omega_2)(50 - b_2) \geq 100 - (b_1 + b_2)$$

As, in all our treatments, $b_1 + b_2 < 100$ it is only the latter condition that binds.

In the DP in which the DM chooses from $\{a, b, c\}$ the comparison of $a$ to $b$ again requires

$$\hat{P}(a|\omega_1)(50 - b_1) + \hat{P}(a|\omega_2)(50 - b_2) \geq 0$$

while the comparison of $a$ to $c$ demands

$$\begin{aligned} \hat{P}(a|\omega_1)(50 - 100) + \hat{P}(a|\omega_2)(50) &\geq 0 \Rightarrow \\ 50\left(\hat{P}(a|\omega_2) - \hat{P}(a|\omega_1)\right) &\geq 0 \\ &\Rightarrow \hat{P}(a|\omega_2) \geq \hat{P}(a|\omega_1) \end{aligned}$$

The comparison of $b$ to $a$ gives

$$\hat{P}(b|\omega_1)(b_1 - 50) + \hat{P}(b|\omega_2)(b_2 - 50) \geq 0$$

And that of $b$ to $c$

$$\hat{P}(b|\omega_1)(b_1 - 100) + \hat{P}(b|\omega_2)b_2 \geq 0$$

The comparison of $c$ to $a$ gives

$$\begin{aligned} \hat{P}(c|\omega_1)\,(100 - 50) + \hat{P}(c|\omega_2)\,(-50) &\geq 0 \Rightarrow \\ 50\left(\hat{P}(c|\omega_1) - \hat{P}(c|\omega_2)\right) &\geq 0 \\ &\Rightarrow \hat{P}(c|\omega_1) \geq \hat{P}(c|\omega_2) \end{aligned}$$

While the comparison of $c$ to $b$ gives

$$\hat{P}(c|\omega_1)(100 - b_1) - \hat{P}(c|\omega_2)b_2 \geq 0$$

Note that not all of these constraints will hold simultaneously.

NIAC requires that the total surplus generated from the observed matching of information structures to decision problems is greater than that generated by switching revealed information structures across decision problems

$$\begin{aligned} & G(\mu, \{a, b\}, \pi) + G(\mu, \{a, b, c\}, \hat{\pi}) \\ \geq\ & G(\mu, \{a, b\}, \hat{\pi}) + G(\mu, \{a, b, c\}, \pi) \end{aligned} \tag{2}$$

where $\pi$ is the revealed posterior from data $P$ generated from choice set $\{a, b\}$ and $\hat{\pi}$ is the revealed information structure from data set $\hat{P}$ generated from choice set $\{a, b, c\}$. See Caplin and Dean [2015] for a formal definition of the revealed information structure, but essentially it assumes that the DM used an information structure which consists of the posteriors described in equation 1 for each chosen act, with the probability of receiving that posterior given by the (unconditional) probability of choosing the associated act.

Assuming NIAS holds, we can calculate $G(\mu, \{a, b\}, \pi)$ directly from the data: these are just the gross utilities derived from SDSC observed in each DP, so

$$\begin{aligned} G(\mu, \{a,b\}, \pi) &= (P(a \cap \omega_1) + P(a \cap \omega_2))\, 50 + P(b \cap \omega_1)b_1 + P(b \cap \omega_2)b_2 \\ &= 0.5\left[(P(a|\omega_1) + P(a|\omega_2))\, 50 + P(b|\omega_1)b_1 + P(b|\omega_2)b_2\right] \end{aligned}$$

where we have used the fact that $\mu(1) = \mu(2) = 0.5$. Similarly for $G(\mu, \{a,b,c\}, \hat{\pi})$ we have

$$G(\mu, \{a,b,c\}, \hat{\pi}) = 0.5\left[\left(\hat{P}(a|\omega_1) + \hat{P}(a|\omega_2)\right) 50 + \hat{P}(b|\omega_1)b_1 + \hat{P}(b|\omega_2)b_2 + \hat{P}(c|\omega_1)100\right]$$

Recall that $G(\mu, \{a,b\}, \hat{\pi})$ is the hypothetical utility generated from using information structure $\hat{\pi}$ in DP $\{a,b\}$. This means that we have to calculate the optimal action to take from the posteriors $\hat{\gamma}^a$, $\hat{\gamma}^b$ and $\hat{\gamma}^c$ associated with acts $a$ $b$ and $c$ in the DP in which $\hat{\pi}$ is observed. Note that, assuming NIAS hold, it must be the case that $a$ is still optimal from $\hat{\gamma}^a$ and $b$ is still optimal from $\hat{\gamma}^b$ in the new problem. The question is therefore only whether the DM should choose $a$ or $b$ from $\hat{\gamma}^c$. Note, however, that NIAS implies that

$$\begin{aligned} \hat{\gamma}^c(\omega_1)100 &\geq \hat{\gamma}^c(\omega_1)50 + (1 - \hat{\gamma}^c(\omega_1))50 \\ &\Rightarrow \hat{\gamma}^c(\omega_1) \geq \frac{1}{2} \end{aligned}$$

which in turn implies that it is optimal to choose $a$ rather than $b$ from this posterior. We therefore have

$$\begin{aligned} G(\mu, \{a,b\}, \hat{\pi}) = &\left(\hat{P}(a|\omega_1) + \hat{P}(a|\omega_2) + \hat{P}(c|\omega_1) + \hat{P}(c|\omega_2)\right) 50 \\ &+\hat{P}(b|\omega_1)b_1 + \hat{P}(b|\omega_2)b_2 \end{aligned}$$

Similarly, in order to calculate $G(\mu, \{a,b,c\}, \pi)$ we need to figure out the optimal choice of action from $\gamma^a$ and $\gamma^b$ associated with the choice of $a$ and $b$ in $\{a,b,c\}$. Again from NIAS it is obvious that it must be the case that $\gamma^b(\omega_1) \leq \frac{1}{2}$, and so it cannot be optimal to choose $c$ from this posterior. NIAS also implies that it must be better to choose $b$ rather than $a$ from this posterior. Further, note that by Bayes rule we have

$$P(a)\gamma^a(\omega_1) + (1 - P(a))\gamma^b(\omega_1) = \frac{1}{2}$$

Thus, as $\gamma^b(\omega_1) \leq \frac{1}{2}$ it must be the case that $\gamma^a(\omega_1) \geq \frac{1}{2}$, meaning that $c$ is weakly optimal from this posterior. This means that

$$G(\mu, \{a, b, c\}, \pi) = P(b|\omega_1)b_1 + P(b|\omega_2)b_2 + P(a|\omega_1)100$$

Plugging these into inequality 2 and cancelling gives

$$\begin{aligned} & (P(a|\omega_1) + P(a|\omega_2))\, 50 + \hat{P}(c|\omega_1)100 \\ \geq \ & \left(\hat{P}(c|\omega_1) + \hat{P}(c|\omega_2)\right) 50 + P(a|\omega_1)100 \end{aligned}$$

or

$$\hat{P}(c|\omega_1) - \hat{P}(c|\omega_2) \geq P(a|\omega_1) - P(a|\omega_2)$$

This expression has a natural interpretation when one notes that NIAS implies that $\hat{P}(c|\omega_1) \geq \hat{P}(c|\omega_2)$ and $P(a|\omega_1) \geq P(a|\omega_2)$: it implies that the DM has to be more informed when choosing $c$ in DP $\{a, b, c\}$ than when choosing $a$ in DP $(a, b\}$. In particular, if the DM chooses to gather no information in the former problem, meaning that $\hat{P}(c|\omega_1) = \hat{P}(c|\omega_2)$, it must also be the case that $P(a|\omega_1) = P(a|\omega_2)$, and so the DM is uninformed in the first problem. NIAS in turn implies that in such cases $a$ must be chosen exclusively in $\{a, b\}$.

## 7.2 Appendix A2: Shannon without Subjects who Violate NIAS or NIAC

In this appendix we rerun the analysis testing the Shannon model using the data from experiment 2 while excluding those subjects who exhibit significant violations on NIAS and NIAC. We will refer to the remainder as 'consistent' subjects.

Figure A2.1 shows estimated costs $\kappa$ using aggregate data , replicating the analysis of figure 6. Again, we see that costs are significantly higher at the 95 point level than at the 5 point level, indicating that adjustment is again too slow relative to the Shannon model

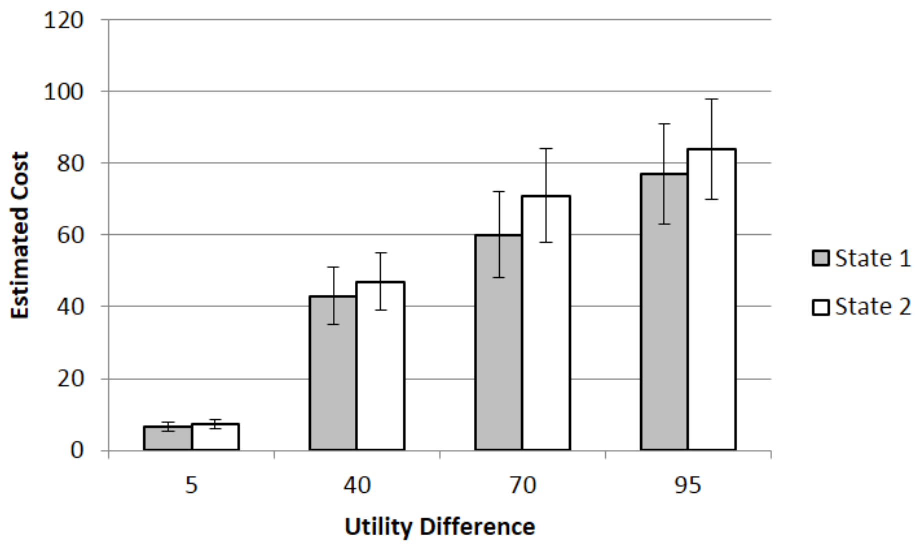



Figure A2.1: Estimated Costs - Consistent Subjects Only

Figure A2.2 replicates the analysis of figure 7 using only consistent subjects. Again we see that the models from the broader $T_{\{\rho,\kappa\}}$ class outperform the Shannon model, with an

AIC of 9852 vs 10123 for the Shannon model

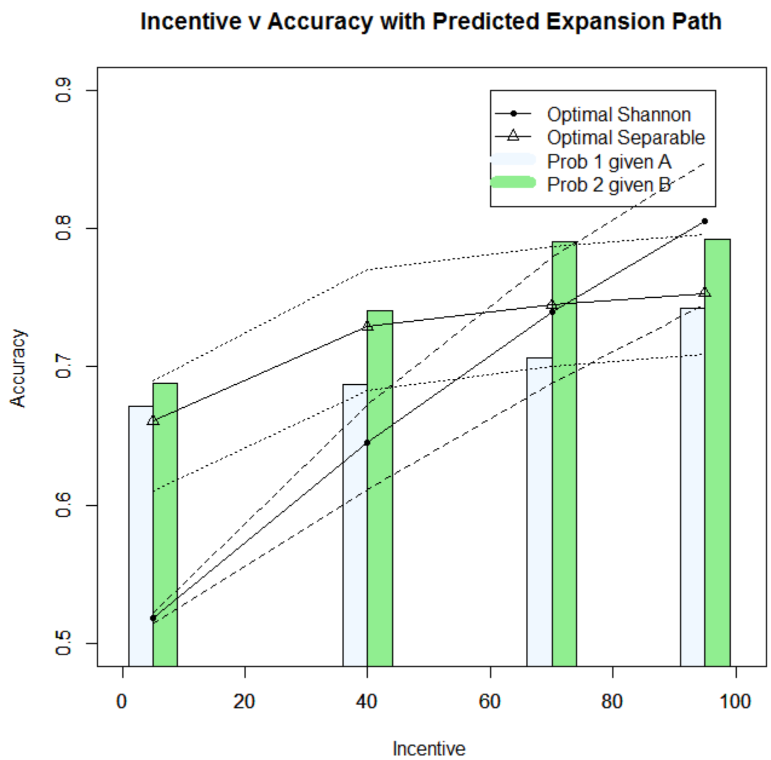



Figure A2.2: Consistent subjects only

Figure A2.3 replicates the individual level analysis of figure 8. As with the equivalent analysis in section 4.3, we drop observations in which accuracy at the lower reward level is below 50%. Of the 178 possible comparisons, we find 42 violations of the ‘too slow’ variety and 5 of the ‘too fast’ variety. 15 of subjects exhibit ‘too slow’ violations only, 3 exhibit ‘too fast violations' only and 21 have examples of neither.

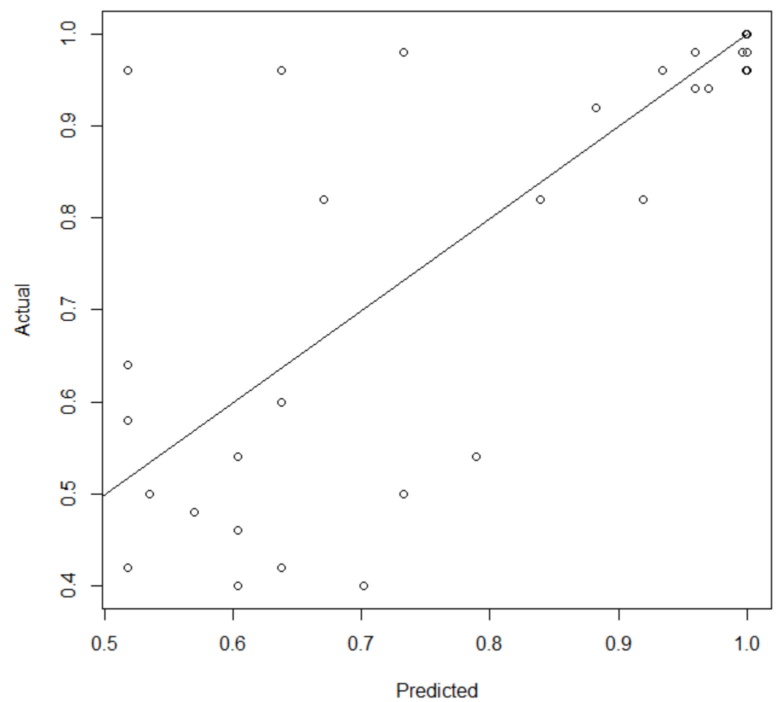



Figure A2.3: Predicted vs actual accuracy in the 70% payoff treatment

## 7.3 Appendix A3: Order Effects

Tables A3.1 and A3.2 report the result of regressions of accuracy (i.e. the probability of picking the rewarding action) on order (i.e. in which block the question occurred, between 1 and 4) controlling for the type of question and clustering standard errors at the subject level for experiments 2 and 3. In both cases the excluded category is block 1 - i.e. the first set of questions answered. The lower and upper CI refer to the upper and lower bounds to the 95% confidence interval, while Prob refers to the probability of rejecting the null hypothesis that the coefficient is equal to zero.

| Table A3.1: Order Effects - Experiment 2 | | | | |
|---|---|---|---|---|
| Block | Coefficient | Lower CI | Upper CI | Prob |
| 2 | -0.05 | -0.09 | -0.00 | 0.04 |
| 3 | -0.07 | -0.11 | -0.03 | 0.00 |
| 4 | -0.06 | -0.11 | -0.02 | 0.03 |

| Table A3.2: Order Effects - Experiment 3 | | | | |
|---|---|---|---|---|
| Block | Coefficient | Lower CI | Upper CI | Prob |
| 2 | -0.01 | -0.05 | 0.04 | 0.81 |
| 3 | -0.02 | -0.07 | 0.02 | 0.34 |
| 4 | -0.02 | -0.08 | 0.02 | 0.19 |