Experimental Tests of Rational Inattention*

Mark Dean† and Nathaniel Neligh‡

June 2022

Abstract

We use laboratory experiments to test models of rational inattention, in which people acquire information to maximize utility net of information costs. We show that subjects adjust their attention in response to changes in incentives in line with the rational inattention model. However, our results are qualitatively inconsistent with information costs that are linear in Shannon entropy, as is often assumed in applied work. Our data is best fit by a generalization of the Shannon model which allows for a more flexible response to incentives and for some states of the world to be harder to distinguish than others.

1 Introduction

It is now well established that economic actors often do not use all relevant information when making choices, meaning that they make 'mistakes' relative to a full information benchmark: shoppers may buy unnecessarily expensive products due to their failure to notice whether or

*We thank in particular Andrew Caplin for his extensive input, 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]). We also thank Sandro Ambuehl, Alessandra Casella, David Freeman, Judd Kessler, 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 This paper reports on experiments similar to those in the unpublished working papers Caplin and Dean [2013] and Caplin and Dean [2014] and subsumes those parts of that paper that are common. None of the results presented in this paper have been published previously.

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

‡Haslam College of Business, University of Tennessee, nneligh@utk.edu

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]). This observation has far reaching consequences for economic modelling from both a positive and normative perspective, breaking as it does the link between choice and revealed preference.

The importance of informational limits and choice mistakes has lead to the development of a wide variety of models that attempt to capture these constraints. The random utility model (RUM) provides one way to introduce limits on a decision maker (DM)'s ability to perceive differences between alternatives.[1] In psychology, signal detection theory (SDT) provides a model in which a DM responds optimally to imperfectly observed information about the world (Green and Swets [1966]). Marketing and, more recently, economics have made use of models of consideration sets, in which the DM considers only a subset of available alternatives (Hauser and Wernerfelt [1990], Manzini and Mariotti [2014]).

Recently, the concept of rational inattention has proved influential in modelling behavior when attention is limited.[2] Rationally inattentive behavior is defined by two assumptions: Choice is optimal conditional on the information received; and the DM chooses what information to gather in order to maximize the utility of subsequent choice net of costs. Within this framework, subclasses of models can be defined by the nature of costs. Particularly popular are costs based on Shannon mutual information, which measures the expected change in entropy between prior and posterior beliefs. Within this class, recent focus has been given to costs which are linear in mutual information (Matejka and McKay [2015]), though other alternatives have been considered, including hard constrains on, or nonlinear functions of mutual information (for example Sims [2003], Paciello and Wiederholt [2014]). Other authors have considered 'uniformly posterior separable' cost functions, which generalize Shannon costs to the expected change of any convex function of beliefs (Caplin and Dean [2013], Gentzkow and Kamenica [2014], Caplin *et al.* [2022]).

In this paper we use a sequence of laboratory experiments to examine the empirical validity of the rational inattention model in a simple perceptual environment. We begin by using non-parametric tests that differentiate between four nested classes of model. First, between those in which the DM adjusts their information acquisition in response to the

[1] This interpretation is made explicit in Block and Marschak [1960] - see Caplin [2016].

[2] We use the term 'rational attention' to describe any model in which information is chosen to maximize expected utility net of some additive cost term, while recognizing that others use it to refer to the specific case when costs are based on the Shannon mutual information between prior and posterior beliefs. We refer to the latter as the 'Shannon model'.

decision problem they are facing (as in the rational inattention model) and those in which they do not (including RUM, SDT and most models of consideration sets). Second, between models in which information choice can be rationalized as optimal relative to some underlying cost function (which we call the general model) and those in which it cannot. Third, between models that assume costs are uniformly posterior separable and those that do not, and finally between those that assume costs are linear in Shannon mutual information (which we call the Shannon model) and those that do not. Having established that our subjects are broadly consistent with the general model we then use structural estimation techniques to establish which of a set of popular parametric forms best describes their behavior.

Our main experimental environment 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 depends on the fraction of balls which are red (which we 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 environment our experiments examine the impact of changing four key features of the decision problem. Experiment 1 varies the set of available options. Experiment 2 changes the incentives for making the correct choice. Experiment 3 changes prior beliefs, while experiment 4 changes the number of states while keeping the number of options fixed. In combination with recent theoretical work (Matejka and McKay [2015], Caplin and Dean [2015], Caplin *et al.* [2022]), these four experiments provide enough data to distinguish between the model classes described above. They also provide a rich data set on which to perform structural estimations.

Our analysis establishes five key results. First, experiment 1 shows that adding new alternatives to a choice set can *increase* the likelihood of existing alternatives being chosen. Such behavior is inconsistent with models of fixed attention, as well as RUM, but is consistent with models of rational inattention in which subjects internalize the informational spillovers that arise from the addition of new objects to the choice set (Matejka and McKay [2015]).

Second, across all four experiments, behavior generally satisfies the No Improving Action Switches (NIAS) and No Improving Attention Cycle (NIAC) conditions of Caplin and Dean [2015], meaning that subjects are consistent with the general model of rational inattention.

While the experimental environments we study are deliberately simple, there are well know behavioral biases that could lead to violations of one or other of the relevant conditions, including base rate neglect (Kahneman and Tversky [1982]) and the possibility that people may perform worse when faced with higher stakes (Ariely *et al.* [2009]). The fact that behavior is consistent with the general model also justifies further analysis to determine which type of cost function best describes our subjects' behavior.

Third, experiment 3 provides mixed evidence in support of the key prediction of uniformly posterior separable cost functions: the Locally Invariant Posteriors (LIP) condition (Caplin *et al.* [2022]). This states that, as prior beliefs change, posterior beliefs should not if they remain feasible. In five out of six tests this is the case in our data. However, a joint test that all conditions hold simultaneously is rejected.

Fourth, we find that our subjects are inconsistent with two important predictions of the Shannon model. In experiment 2 we show that our subjects are less responsive to incentives than the Shannon model would predict, violating the Invariant Likelihood Ration (ILR) property (Caplin and Dean [2013]). In experiment 4 we show that subjects do not behave identically in payoff-identical states in an environment in which there is a natural notion of a perceptual distance. This contradicts the Invariance Under Compression (IUC) condition which isolates the Shannon model in the uniformly posterior separable class (Caplin *et al.* [2022]).

Finally we show that parametric generalizations of the Shannon model which address these issues provide a qualitatively better fit to our data that the Shannon model itself. The best fitting models allow for a more flexible response to incentives (for example by replacing the assumption of Shannon entropy with an alternative two parameter version such as generalized entropy). They also allows for the fact that some states may be harder to differentiate between than others, using the 'Neighborhood' structure of Hébert and Woodford [2021].

These five findings put significant constraints on the type of inattention model that can explain stochastic choice in our setting. One obvious question is the extent to which these results generalize to other informational environments. To provide a partial answer, we run a second experiment in which the subjects can identify the state by solving equations, a task that has been used previously by Ambuehl *et al.* [2020]. As we describe in section 6, the overall picture looks similar in this new environments: subjects actively adjust their attention, are broadly in line with the NIAS and NIAC conditions, but exhibit violations of symmetry and are less responsive to incentives than the Shannon model would predict. The main differences we find from the original experiment are (i) subjects are even less responsive to changes in incentives and (ii) we find less support for the LIP condition.

We argue that these results have important implications beyond the specific experimental tasks we used. In our main experiment subjects have to approximate a numeric quantity. Research in psychology describes how common such tasks are in everyday life (see for example Gandini *et al.* [2008]). Moreover, our task shares important features with economically important perceptual environments. For example one of the key violations of Shannon model we observe is based on the fact that information is presented in such a way that there is a natural metric on the state space, with closer states harder to distinguish that those further apart. Moreover, the subjects do not have access to any obvious technology that would allow them to collect information in a way that does not have this property. The Shannon model is often applied to economic environments that arguably also have this feature, such as the perception of wealth (Sims [2003]) or prices (Matějka [2015]).

Our findings also have practical implications for information gathering environments to which they generalize The fact that we find violations of monotonicity is a potential issue for the large literature that makes use of the RUM, particularly when it is used to capture stochasticity in perception rather than tastes. The fact that we find limited support for the Shannon model is also potentially problematic for the 'generalized logit' model (Matejka and McKay [2015]) which shares many features of the logit RUM. The importance of perceptual distance in information acquisition has implications for coordination in global games (Yang [2015], Morris and Yang [2021]), and the welfare implications of rational inattention (Angeletos and Sastry [2019]). Our finding that locally invariant posteriors holds in some situations is useful as this assumption greatly simplifies the analysis of behavior as prior beliefs change (see for example Martin [2017]).

We stress that we do not believe that our results are universal. Indeed, we would expect that changing some features of the informational environment would have a radical effect on behavior. A time limit might lead subject to act as if they had a hard informational constraint. Providing subjects with more sophisticated information gathering technologies might mean they could ignore any 'perceptual distance' between states if it is payoff irrelevant, and so behave more in line with the Shannon model. Other environments may have no natural perceptual distance at all. We believe that understanding the role of these environmental factors is crucial for developing theories of attention and information acquisition. Much like utility maximization or Nash equilibrium, one of the potential strengths of rational inattention is its portability - it can potentially make predictions in *any* information gathering task. Similarly, due to its basis in optimal coding theory, the Shannon model might stake a claim at being widely applicable across many domains. Understanding the limits of this generality - i.e. the situations in which the general and Shannon models work well and where they do not - is an important step in the development of broadly applicable models of information acquisition. We draw the analogy with experimental game theory, where an understanding of which simple, stylized settings generate behavior consistent with Nash equilibrium has been useful in developing new models, such as level K (Nagel [1995]) and Quantal Response Equilibrium (McKelvey and Palfrey [1995])

To the best of our knowledge ours is the first paper to use experimental data to implement the tests of rationally inattentive behavior that have been uncovered by the recent theoretical literature. Overall, there is surprisingly little experimental work in economics testing models of inattention. Notable exceptions include Gabaix *et al.* [2006], Caplin *et al.* [2011], Taubinsky [2013], Khaw *et al.* [2017] and Ambuehl and Li [2018]. 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. Dewan and Neligh [2020] and Caplin *et al.* [2020] study a variety of properties of information costs using tasks similar to our experiment 2, but do not replicate our other experiments. Contemporaneous to this paper, Ambuehl *et al.* [2020] test two implications of the Shannon model in a market setting - finding support for both - but do not test the two implications of Shannon we find violated in our study. Using a subset of the data from our study, Denti [2020] finds mixed evidence in support for uniformly posterior separable models. Pinkovskiy [2009] and Cheremukhin *et al.* [2015] fit generalizations of 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, but which is not designed to test the necessary and sufficient conditions of rational inattention as we do in this paper. 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 (for example see Ratcliff *et al.* [2016] for a recent review, and Krajbich *et al.* [2011] for an application to economic decision making). These studies differ from ours in many ways including the nature of the decision problem, incentivization, type of task and way in which the data is analyzed. We discuss our relationship to these literatures in section 7.

The paper is organized as follows. Section 2 describes the theory underlying our experiments. Section 3 describes the design of our main experiment in detail. Section 4 provides results of the qualitative tests of the RI and Shannon models, and section 5 describes our estimation results. Section 6 reports the result of our follow up experiment, section 7 describes the related literature and section 8 concludes.

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 ω 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 (μ, 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 conditional probability distribution over states, γ^a, associated with each action $a \in A$ which is chosen with positive probability. By Bayes' rule we have

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

These constructs, which we term 'revealed posteriors', 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.

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 those signals. For notational convenience we identify each signal with its associated posterior beliefs $\gamma \in \Gamma$. Feasible information structures satisfy Bayes' rule, so for any prior μ the set of possible structures $\Pi(\mu)$ comprises all mappings $\pi : \Omega \to \Delta(\Gamma)$ 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 γ given state ω and $\gamma(\omega)$ is the probability of state ω conditional on receiving signal γ. Note that γ is distinct from γ^a. The former represents the decision maker's beliefs after the receipt of a signal and as such is not observable to the experimenter, while the latter represents state probabilities conditional on the choice of action a, and so can be estimated from SDSC data.

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 π given prior μ. 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.

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

The rational inattention model assumes that the DM chooses information structures to maximize utility net of costs[3]

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

Our assumptions on the data mean that G is observable but K is not. We use the

[3]Here we impose separability between the costs and benefits of information. One could consider models which relax this assumption along the lines of Chambers *et al.* [2020], but as we shall see the separable model will do a good job of fitting our data.

convention of describing this as a *general model* of rational inattention. Other authors have used rational inattention to refer to the case in which costs are based on mutual information. We refer to this as the Shannon model, as discussed below.

Caplin and Dean [2015] provide necessary and sufficient conditions on SDSC data such that there exists some cost function which rationalizes the general model. 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 γ^a - the posterior distribution associated with that act. The NIAS condition holds for any model in which information is used optimally - regardless of how this information is selected - and so is not specific to the case of rational inattention.

The *No Improving Attention Cycles* (NIAC) condition ensures that choice of information structure itself is rationalizable according to some underlying cost function. It relies on the concept of a *revealed information structure*. Caplin and Dean [2015] provide a formal definition, but essentially the revealed information structure 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.[4] NIAC then states that the total gross payoff (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 (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.

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. In particular, because we do not a priori rule out the possibility that the cost of some information structures is infinite, this formalization can cope with models in which the DM is restricted to choosing from certain types of information structure, such as those consisting of normal signals,[5] or in which information is free up to a hard capacity constraint.[6] The only substantive assumption is that the objective function

[4] 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.

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

[6] For example Sims [2003].

is additively separable between gross utility and costs (see Chambers *et al.* [2020] for a discussion of non-separable models).

2.2.1 Rational Inattention vs Other Models of Stochastic Choice

Rational inattention is 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. Changes in utility can be interpreted as changes in utility or in perception, as made explicit in Block and Marschak [1960]. 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.[7] 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:

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 acquire information, which may in turn lead the DM to learn that an existing act was of high value. 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

[7] 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.

of the world, then choose actions optimally given subsequent beliefs. As such, it is a special case of the general model with the added assumption that 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 Shannon and Posterior Separable Models

The general model is almost completely agnostic about the form of information costs. 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. Shannon costs can be justified on axiomatic or information theoretic grounds (see for example Matejka and McKay [2015]), and have been widely applied in the subsequent literature.

Mutual information costs have the following form

$$K_s(\mu, \pi) = \kappa \left[\sum_{\gamma \in \Gamma(\pi)} \pi(\gamma) \left[-H(\gamma) \right] - \left[-H(\mu) \right] \right] \tag{2}$$

where $\pi(\gamma) = \sum_{\omega \in \Omega} \mu(\omega) \pi(\gamma|\omega)$ is the unconditional probability of signal γ and $H(\gamma) = \sum_{\omega \in} -\gamma(\omega) \ln \gamma(\omega)$ is the Shannon entropy of distribution γ.[8] Mutual information can therefore be seen as the expected reduction in entropy due to the observation of signals from the information structure.

We focus on the case in which costs are linear in Shannon mutual information, which we refer to as the Shannon model. In section 5 we discuss costs which are nonlinear transforms of mutual information. An alternative model is one in which subjects have a fixed mutual information constraint, with zero costs up to this constraint and infinite beyond it (e.g. Sims [2003]). This model has the implication that subjects cannot gather more information as incentives increase. As this is strongly rejected by the results of experiment 2 below, we do not focus on this case.[9]

[8] Recall that we identify a signal with its resulting posterior distribution.

[9] We also focus on the 'unrestricted' version of the Shannon model, in which the DM is free to choose any information structure they wish. A possible modification is to restrict the DM to learn about certain events

Clearly, the Shannon model puts much more structure on information costs than the general model, which in turn means that it puts much tighter restrictions on behavior. These restrictions have been discussed in several recent papers (particularly Caplin and Dean [2013], Matejka and McKay [2015] and Caplin *et al.* [2022]). Partly because of it's restrictive nature, researchers have also considered a generalization of the Shannon cost function to *uniformly posterior separable* costs. These keep the functional form (and so some of the analytical tractability) of the Shannon function, but replace $-H$ in equation (2) with any arbitrary convex function of posterior beliefs (see for example Caplin and Dean [2013] and Gentzkow and Kamenica [2014]). Morris and Strack [2017], and Hébert and Woodford [2019] show that cost functions of this class are consistent with models of optimal sequential learning.[10]

Caplin *et al.* [2022] show that a key defining characteristic of uniformly posterior separable costs is the Locally Invariant Posteriors (LIP) condition. This states that local changes in prior beliefs do not lead to changes in optimal posterior beliefs. Specifically, if, for some decision problem (μ, A), the associated SDSC reveals some set of posteriors $\{\gamma^a\}_{a \in A}$, and we change the prior to some μ' in such a way that these posteriors are still feasible (i.e. μ' 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 (μ', A). We will test this proposition in experiment 3.

We also consider two properties that are specific to the Shannon model. The Invariant Likelihood Ratio (ILR) property (Caplin and Dean [2013]) states that for any two chosen actions, the posterior probabilities of 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 ILR condition also implies that posterior beliefs depend only on 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

independently - for example Mackowiak and Wiederholt [2009] require that firms have to receive distinct signals regarding aggregate and idiosyncratic shocks. The question of whether a model of this type, for example requiring the subject to learn separately about the color of each ball, could explain our data is an interesting avenue for future research.

[10] Caplin *et al.* [2022] also differentiate between posterior separable cost functions, for which costs are allowed to change arbitrarily with prior beliefs, and *uniformly* posterior separable costs, for which the basic form of the cost function remains unchanged if the support of the prior remains unchanged.

for all acts. Caplin *et al.* [2022] show that this 'Invariance Under Compression' (IUC) property fully characterizes the Shannon model within the uniformly posterior separable class. Behaviorally, one implication of this property is that the Shannon model lacks any notion of 'perceptual distance': that some states might be harder to differentiate than others. We test this implication in experiment 4.[11]

3 Experimental Design

3.1 Set Up

We now introduce the main experimental design we use to produce state dependent stochastic choice data for each subject. 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 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 (DP) is defined by the prior distribution and the set of available actions. No feedback was given to the subject about the true state after each

[11] The ILR property does not imply IUC, even under the assumption that costs are posterior separable. This is because the ILR property only restricts behavior across acts which are chosen with positive probability. The IUC condition also puts restrictions on which acts are chosen with positive probability.

choice. Figure 1 shows a typical screenshot from the experiment.

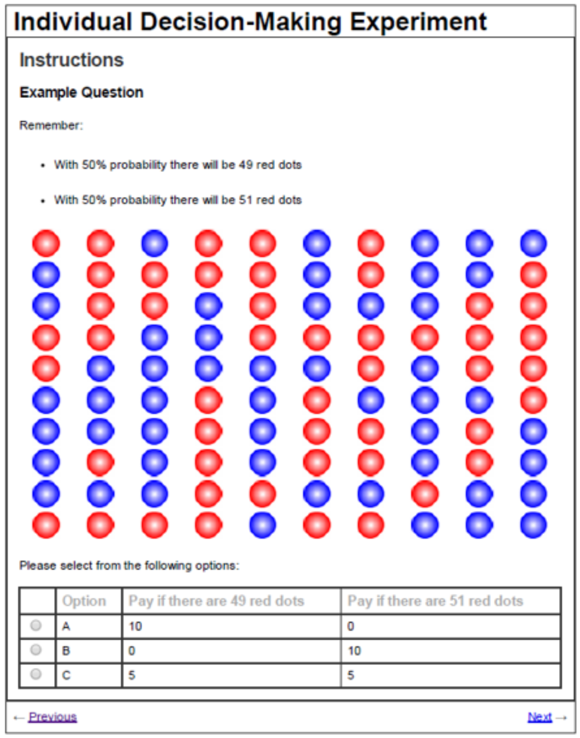


Figure 1: A typical decision problem

Each experiment consists of a small number of decision problems (between 2 and 4). A subject faced many repetitions of each decision problem (between 50-75 questions for each). The order in which subjects faced decision problems was randomized, but all repetitions of the same decision problem were grouped together (so, for example, in experiment 1 the subject would face either 75 repetitions of DP 1 then 75 repetitions of DP 2 or visa versa). At the end of the experiment, one decision problem was selected at random for payment.

There are several things to note about our experimental design. First there is no externally imposed 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 7 for a discussion). At the same time, there is no explicit 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.[12]

Second, in order to estimate the state dependent stochastic choice function we treat the multiple times that a subject faced the same decision making environment as multiple independent repetitions of the same decision problem. To prevent subjects from learning to recognize patterns, we randomized 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 4.6.

Third, in experiments where it is important, we paid subjects in 'probability points' rather than money - i.e. subjects were paid in points which increased the probability of winning a monetary prize. We do so in order to get round the problem that utility 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.

Fourth, we collected only choice data (not, for example, elicited beliefs) in a setting where subjects must gather their own information. One alternative design would be to ask subjects to choose between information structures directly. While such an experiment would be complementary, we believe there to be an advantage to understanding what subjects pay attention to when faced with the intrinsic costs of gathering and processing information, rather than when choosing from an extrinsically imposed menu of information structures. A second alternative design would be to have measured beliefs directly at the time of choice. Again we see an advantage in recovering implied beliefs from choice: it might be that subjects do not have direct access to the beliefs underlying their decisions, or find it hard to articulate them. Moreover, the theoretical work discussed in the previous section shows that SDSC provides a rich data set to test models of inattention without the need for stated beliefs: the revealed posteriors γ^a are sufficient.

A copy of the experimental instructions can be found in appendix A0.

[12]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. This reduced the level and heterogeneity of the opportunity cost of time for the subjects, meaning the primary cost they faced was the cognitive cost of gathering information. As such, our information cost estimates are likely to be lower bounds on the costs we would find if subjects had a free choice of what to do after they had completed the experiment.

3.2 Experiment 1: Testing for Responsive Attention

Experiment 1 is designed to test the Monotonicity axiom. Finding evidence of systematic violations of this condition has important ramifications, as it is a key implication of two important model classes: RUM in economics and SDT in psychology.

Based on a thought experiment discussed in Matejka and McKay [2015], the design requires subjects to take part in two decision problems described in table 1 below. Payment was in probability points with a prize of \$20. Each subject faced 75 repetitions of two decision problems..

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

Payment for each action (a,b and c) in each state (1 and 2) in probability points.

The structure of the two DPs is as follows. There are two equally likely states - 1 and 2 (defined as 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 less than a in state 1, but more in state 2 (i.e. $b_1 < 50 < b_2$). 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.[13]

[13] We use multiple values in order to explore the paramater space somewhat - a priori we did not know the values of b_1 and b_2 that would generate violations of monotonicity.

Table 2: Treatments for Experiment 1		
Treatment	Payoffs	
	b_1	b_2
1	40	55
2	40	52
3	30	55
4	30	52
Payment for action b in probability points across 4 treatments.		

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 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. In the interests of brevity, we relegate a formal derivation of these tests to Appendix A1, but we note here that the NIAS condition is relatively demanding, consisting of 7 inequalities (one from DP 1 and six from the three pairwise comparisons in DP 2). The NIAC condition effectively states that the net probability of state 1 when choosing a in DP 1 can be no higher than than the net probability of the same state when choosing c in DP 2.

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 3 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 four decision problems. Again, states 1 and 2 were represented by 49 and 51 red balls respectively.

Table 3: 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

Payoff for each action (a and b) in each state (1 and 2) in probability points.

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 cost. 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 requires that

$$P_i(a|1) \geq P_i(a|2) \text{ for } i \in \{3, 4, 5, 6\}.$$

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).[14]

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

[14] See Caplin and Dean [2015] section E for the derivation of the NIAS and NIAC conditions for experiments 2 and 3.

as incentives increase - i.e.

$$\begin{aligned} & P_6(a|1) + P_6(b|2) \geq P_5(a|1) + P_5(b|2) \\ \geq \;& P_4(a|1) + P_4(b|2) \geq P_3(a|1) + P_3(b|2) \end{aligned}$$

We emphasize that we see these tests of NIAS and NIAC as an important precursor to identifying what information costs look like. While one could think of hypotheses that would lead to violations of this condition - for example the problem of choking under high stakes, as described in Ariely *et al.* [2009] - we do not necessarily see these as particularly likely a priori.

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. Assuming that utility is linear in probability points, this implies that

$$\begin{aligned} \kappa &= \frac{5}{\ln(\gamma_3^a(1)) - \ln(\gamma_3^b(1))} = \frac{40}{\ln(\gamma_4^a(1)) - \ln(\gamma_4^b(1))} \\ &= \frac{70}{\ln(\gamma_5^a(1)) - \ln(\gamma_5^b(1))} = \frac{95}{\ln(\gamma_6^a(1)) - \ln(\gamma_6^b(1))}, \end{aligned} \tag{3}$$

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 inferred from 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, κ, which then dictates behavior in all other decision problems.

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)[15] and two acts. Again there are four decision problems, each of which is repeated 50 times. Because this

[15] 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.

experiment made use of only two payoff levels, payment was made in cash, rather than probability points. Table 4 describes the 4 decision problems with payoffs denominated in US Dollars.

Table 4: Experiment 3

DP	$\mu(1)$	Payoffs $U(a,1)$	$U(a,2)$	$U(b,1)$	$U(b,2)$
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

Prior probability of state 1 ($\mu(1)$) and payoff for each action (a and b) in each state (1 and 2) in USD

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

$$P_i(a|1) \geq \frac{2\mu(1)-1}{\mu(1)} + \frac{1-\mu(1)}{\mu(1)} P_i(a|2) \tag{4}$$

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. 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, uniformly posterior separable models puts a lot of structure on behavior in Experiment 3, as captured by the LIP condition. First, one observes the posterior beliefs associated with the choice of a and b in DP 7, when $\mu(1) = 0.5$. Then, as the prior probability of state 1 increases, there are only two possible responses. 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. 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 a in all questions.

This experimental design is based on directly informing subjects of the prior probabilities. It is therefore a joint test of the NIAS and LIP conditions and the assumption that priors are fully internalized. While our results show that subjects do change their behavior across treatments, we cannot rule out the possibility that some subjects are not fully aware of the change in priors. An alternative design in which prior beliefs are measured, rather than assumed, is an interesting avenue for future research.[16]

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

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. Note that, as we shall see in Section 5, one can construct cost functions such that ILR holds in experiment 4 but not experiment 2 and visa versa.

We test this implication using our experimental interface: 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).[18]

[16] We thank Sandro Ambuehl for this suggestion.

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

[18] A previous version of the paper reported the results of another treatment in which the state of the world was determined by by the number of letters on the screen. We omit these results for brevity.

4 Implementation and Results

Subjects were recruited from the New York University and Columbia University student populations.[19] 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 appendix A0.

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 repetitions of the same decision problem and/or individuals.

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. We relegate a discussion of the problems that aggregation causes to section 4.6, noting here only that most of our tests are robust to this issue.

Because we observe estimates of the SDSC function based on finite samples 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 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

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

observing such data should fall as more data is collected, and so power increases.[20] 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.

It is important to note that this approach means that, at least for our aggregate results, a lack of power would make it more likely that we would reject a particular model: success is only declared if the point estimate of a parameter is of the right sign *and* is significantly different from the boundary.

The null hypothesis above is 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 derive our p-values by using 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.

In situations where we are interested in the precise value of conditional action probabilities or some transformation of those probabilities, OLS regressions (i.e. a linear probability model) were employed, as the OLS coefficients provide unbiased estimates of these quantities.[21] When we are only concerned with differences in probabilities, we employ logistic regressions for its better properties when probabilities are extreme. When aggregate data is used, standard errors are corrected for clustering at the subject level.

4.2 Experiment 1: Testing for Responsive Attention

28 subjects took part in this experiment, evenly divided across the 4 treatments.[22]

Table 5 summarizes the results of the Monotonicity test. The first panel reports $P(b|1)$: the probability of choosing action b in state 1 with and without c available - aggregated across

[20] Assuming that the values are not equal in the population.

[21] This is true as long as the linear probability model cannot predict probabilities greater than zero or less than one, as is the case inour implementations.

[22] One session of 12 subjects on 11th May 2016 and one session of 16 subjects on 27th September 2016, both run at the CELSS laboratory at Columbia University.

all subjects. In order to establish whether the introduction of c increases the probability of b being chosen in this state we run a logit regression of the form

$$1(\text{choice}_{i,j} = b) = \beta_0 + \beta_1 1(c \text{ present}_{i,j}) + \varepsilon_{i,j} \tag{5}$$

where 1 is the indicator function, $\text{choice}_{i,j}$ is the choice in round j of subject i and $\varepsilon_{i,j}$ is the logistic error term, clustered at the subject level. For the first panel, this regression is performed on all observations in which the state is 1 and the treatment is as labelled in the first column. The column 'prob' reports the probability associated with the null hypothesis that $\beta_1 = 0$. The second panel repeats the exercise for $P(b|2)$, with regression (5) run on all observations in which the state is 2. The final column reports the fraction of subjects who show a significant violation of Monotonicity at the 5% level - i.e. regression (5) was run on each subject and each observation in which the state is 2, with the subject counted as violating monotonicity if $\beta_1 > 0$ and the hypothesis that $\beta_1 = 0$ can be rejected at the 5% level.

Table 5: Results of Experiment 1

		$P(b\|1)$			$P(b\|2)$			
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

Panel 1 reports fraction of observations in which b was chosen when the state was 1 and options a and b were available (column 1) or when a, b and c were available (column 2). Column 3 reports the probability associated with the null hypothesis that the probabilities in column 1 and 2 are equal. Panel 2 repeats the exercise for observations in which the state is 2. The final column reports the fraction of subjects who choose b significantly more in state 2 when c is available.

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.

Results for NIAS and NIAC tests are reported in appendix A1. Broadly speaking behavior is in line with the general model of rational inattention. In the aggregate data, 5 of the 7 point estimates for the NIAS tests are of the right sign and significantly different from the boundary. This is also the case for the NIAC test. The other two NIAS tests are not significantly different from the boundary. At the individual level, of the 196 NIAS tests we found 9 significant violations (5%), and 2 significant violations of the 28 NIAC tests (7%).

4.3 Experiment 2: Changing Incentives

We next report the results from experiment 2 in which we examine how subjects' responses change with incentives. 52 subjects took part in this experiment.[23] Table 6 reports the aggregate probabilities of choosing action a in each state and in each decision problem, while table A2.1 in appendix 2 reports the results of OLS and logit regression of the choice of a on state and decision problem dummies, i.e.

$$1(\text{choice}_{i,j} = a) = \beta_0 + \beta_{2,3} 1(\text{state=2} \cap \text{DP=3}) + \sum_{k=1}^{2} \sum_{m=4}^{6} \beta_{k,m} 1(\text{state=k} \cap \text{DP=m}) + \varepsilon_{i,j} \quad (6)$$

with errors clustered at the subject level. These regressions are used to perform the statistical tests in the following analysis.

We begin by testing the NIAS and NIAC conditions. Table 6 reports the results of the test of NIAS - 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 p-value for the null that NIAS is violated (i.e. that $\beta_{2,3} = 0$ for DP 3, and $\beta_{1,m} = \beta_{2.m}$ for DP 5,...,6. The aggregate data firmly supports the NIAS

[23] Three sessions of 22, 16 and 14 subjects taking place on 5th Dec 2016, 15th December 2016 and 20th Jan 2017 at the CELSS laboratory at Columbia University.

condition.

Table 6: NIAS Test from Experiment 2			
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

Probability that action a is chosen in state 1 and 2 for each DP - aggregate data. 'Prob' reports the probability associated with the test that $P_j(a|1) = P_j(a|2)$.

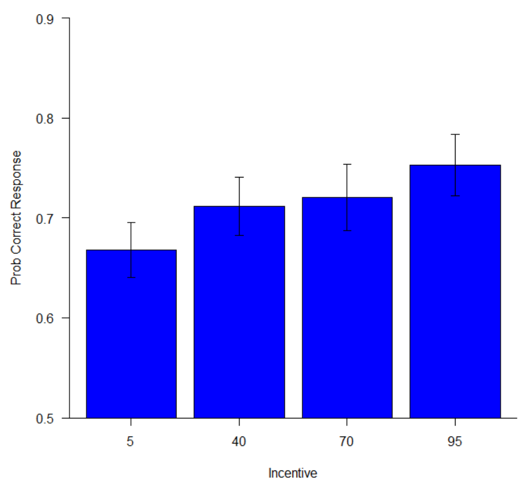


Figure 2: Probability of correct response by decision problem - aggregate data from experiment 2. Bars show standard errors, clustered at the individual level.

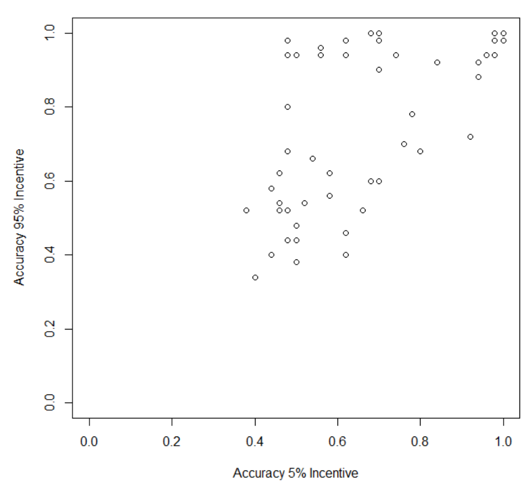


Figure 3: Scatter plot showing the sample probability of choosing the correct option of each subject in experiment 2 at the 5 point vs 95 point reward level.

Figure 2 shows the probability of choosing the 'correct' act in each DP, averaging across all subjects. This allows us to test the NIAC condition which states that this probability should be non-decreasing in the reward level. The point estimates from the aggregate data obey this pattern, with accuracy increasing from 67% at the 5 probability point payment level to 75% at the 95 probability point payment level. Table A2.2 in appendix A2 reports the results of two sided tests of equality of accuracy between each incentive level, based on the coefficients from regression specified in equation (6). Using the OLS regressions, the success rate at 5 probability points is significantly different from that at 95 points at <0.1%, and different from 40 and 70 points at 10%. Behavior at 40 probability points is significantly different from 95 points at 10%. Behavior at the 70 probability points level is not significantly different from either 40 or the 95 point level. The results from the logit regressions are similar.

Table 7 reports the fraction of subjects who exhibit significant violations of the NIAS condition, the NIAC condition, both or neither, using individual level regressions equivalent to equation (6)[24]. 81% of subjects show no significant violations of either condition.

Table 7: Individual Level Data from Experiment 2

Violate	Data
NIAS Only	2
NIAC Only	17
Both	0
Neither	81

Percentage of subjects who violate NIAS, NIAC, both or neither

Table 7 implies that most of our subjects do not have significant violations of the NIAS and NIAC conditions and therefore act as if they maximize payoffs net of some underlying attention cost function. Figure 3 gives some idea of the heterogeneity of those costs across subjects. It graphs the probability of choosing the correct action at the 5 point reward vs the 95 point reward 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 individuals who gather little information regardless of reward: their accuracy is near 50% for the low and high reward levels. It also includes subjects who have accuracy close to 100%, even in the low reward decision problem.

[24] A subjects was considered to violate NIAS if the relevant paramater tests indicated that $P_j(a|1) < P_j(a|2)$ and the hypothesis of equality could be rejected at the 5% level in the logit regression.

NIAC was checked by first using an F-test of the joint restrictions that (i) accuracy at the 40 percentage point (pp) level was equal to that at the 5 pp level, (ii) accuracy at the 70 pp level was equal to that on 40 pp level and (iii) accuracy at the 95 pp level equal to that at the 70 pp level. Subjects were categorized as violating NIAC if these restrictions were jointly rejected and the point estimates indicated a violation.

Finally there are subjects, who actively adjust their accuracy as a function of reward.

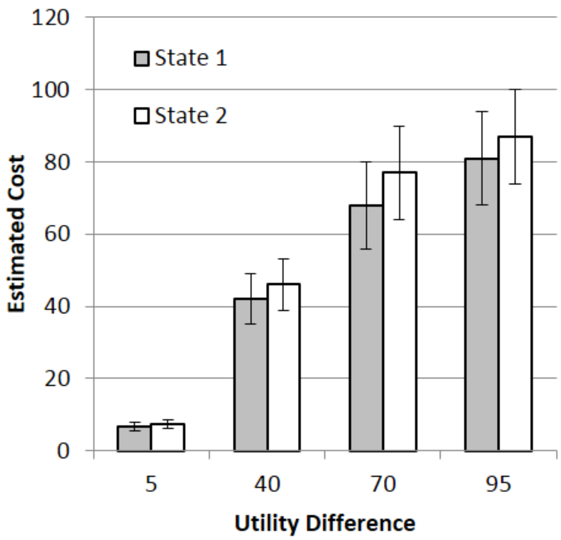


Figure 4: Estimated cost parameter from the Shannon model in each state and at each reward level using aggregate data from experiment 2. Bars show standard errors clustered at the individual level.

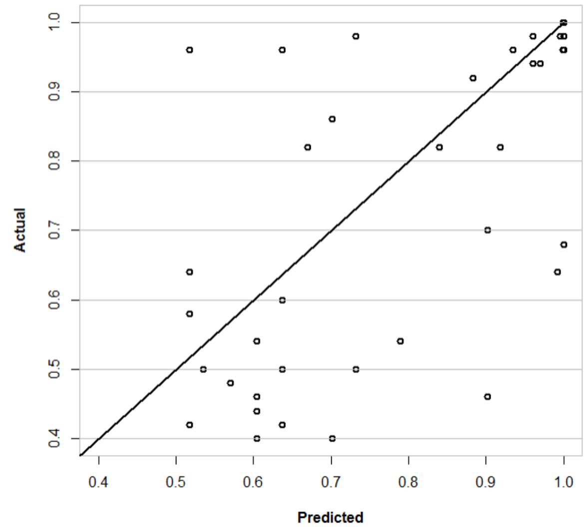


Figure 5: Scatter plot of subject level accuracy in the 70% payoff treatment vs predicted accuracy based on the Shannon model and behavior in the 40% accuracy treatment.

We next examine the extent to which subjects behave as if their costs are in line with the Shannon model. Figure 4 shows the estimated cost parameter κ from each decision problem and in each state using aggregate data, based on the identity from equation (3) and the parameters from the OLS version of regression (6).[25] The Shannon model predicts that these should be equal. As we can see this is not the case: estimated costs are increasing in reward: they 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 increasing their accuracy too slowly in response to changing incentives relative to the predictions of the Shannon model.

At the individual level we also see significant violations of the Shannon model. Figure 5 shows a scatter plot of the predicted vs actual accuracy for each subject in the 70 point DP,

[25]Standard errors for the costs are calculated using a nonllinear transform of those associated with regression (6), using the delta method.

where the predictions are made using the Shannon model and the accuracy displayed at the 40 point level.[26] The scatter plot shows more subjects (27) below the 45 degree line (i.e. are less accurate than predicted) than above (11 - more accurate than predicted).[27]

For each subject and pair of reward levels we can test for significant violations of the Shannon model which indicate 'too slow' adjustment (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). For each person and each incentive level pair we perform a logit regression of a dummy variable indicating correct choice on incentive level and a dummy for the higher incentive level with no constant, which is equivalent to fitting the Shannon model (see Dewan and Neligh [2020] section 3.2). 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. Of the 221 possible comparisons,[28] 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.

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

Next we report the results from the 54 subjects who took part in experiment 3.[29] Table A2.3 in appendix 2 reports the results of OLS and logit regressions of the choice of a on state and

[26] 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.

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

[28] Again, we discard observations in which the subject had a point estmate worse than random choice at the low incentive level.

[29] Data from 3 sessions: 1st October 2012 at the CESS laboratory at NYU (24 subjects), and 25th July and 12th August 2016 at the CELSS laboratory at Columbia (7 and 23 subjects respectively).

decision problem dummies, i.e.

$$1(\text{choice}_{i,j} = a) = \beta_0 + \beta_{2,7} 1(\text{state=2} \cap \text{DP=7}) + \sum_{k=1}^{2} \sum_{m=8}^{10} \beta_{k,m} 1(\text{state=k} \cap \text{DP=m}) + \varepsilon_{i,j} \quad (7)$$

with errors clustered at the subject level.

We first examine the extent to which subjects in experiment 3 obeyed NIAS. Table 8 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 p-value for the null hypothesis that NIAS is violated in the aggregate data, based on an F-test of the linear restriction specified in equation (4), using the coefficients from the OLS regression reported in table A2.3. Table 8 indicates that subjects do on average change their behavior in response to changing priors, and that 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. This pattern is repeated at the individual level, where we see only a small number of subjects exhibiting significant violations of NIAS: 0% at the 50% prior, 2% at the 60% and 75% priors and 11% at the 85% prior.

Table 8: NIAS Test

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	0.001
10	0.51	0.91	0.91	0.606

Each row reports the probability that a is chosen in state 2 in that DP (column 2), the constraint implied by NIAS on the probability that a is chosen in state 1 (column 3), and the actual probability that a is chosen in state 1 (column 4) in the aggregate data. Column 5 reports the probability associated with the null hypothesis that the columns 3 and 4 are equal

These results show that subjects are not completely ignoring the changing priors in the experiment, nor is any base rate neglect strong enough to lead to frequent NIAS violations. Indeed if subjects ignored the change in prior then the resulting data would have violated NIAS. To show this we take the estimated conditional choice probabilities $P(a|1)$ and $P(a|2)$ from DP 7 and use them to simulate behavior in DP 10. Of 100 simulations, 82% exhibited significant violations of NIAS at the aggregate level.

These results also show that any overconfidence the subjects have in the strength of their signals is not enough to cause violations of NIAS. As one benchmark for how demanding these tests are, we can conduct the following thought exercise. Consider a subject who believes that they are getting signals equal to the accuracy we observe in the aggregate data, but in fact are getting signals whose accuracy is given by $\hat{P}_j(a|i) = \lambda P_j(a|i) + (1-\lambda)0.5$ for $j \in \{7, ..10\}$ and $i \in \{1, 2\}$. One can calculate the minimum λ for the resulting choice data to be consistent with NIAS.[30] Based on the aggregate data, these bounds are 0 for DP7, 0.36 for DP8, 0.77 for DP9 and 0.98 for DP10.

The support we find for NIAS is perhaps surprising, given previous work demonstrating base rate neglect. Indeed, the early work of Tversky and Kahneman [1974] showed almost complete base rate neglect which, as we have demonstrated, would be enough to lead to a violation of NIAS. One possibility is that subjects do better because the information they receive is presented naturalistically, rather than numerically - i.e. by counting balls on a screen. There is evidence that this can help statistical reasoning (see Gigerenzer and Hoffrage [1995], Brase [2009] and Garcia-Retamero and Hoffrage [2013]), though it is worth noting that information on the prior is represented numerically in our experiment. Another possible reason is the number of repetitions the subjects do, giving them a lot of experience.

We next study the degree to which our data supports the predictions of uniform posterior separability in the form of the LIP condition. In order to do so, we need to estimate the posteriors associated with each choice of action in each decision problem for each individual. We do so by running equivalents to the OLS version of regression (7) for each subject. Estimates for the posteriors are then calculated using the associated coefficients and Bayes rule, with standard errors for the posteriors calculated using the delta method.

We first divide subjects 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 9 shows this categorization based

[30] Using equation (4), the relevant bound is given by

$$\lambda \geq \frac{1}{2} \frac{2\mu(1) - 1}{\mu(1)(P_j(a|1) - 0.5) + (1 - \mu(1)\,(0.5 - P_j(a|2))}$$

on the point estimates:

Table 9: 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
Number and fraction of subjects whose estimated posteriors associated with the choice of a in DP 7 fall into each range		

Of course, it is possible that these point estimates incorrectly categorize subjects as they are only noisy estimates of the true conditional probabilities. We therefore report results based on subjects whose posteriors are significantly above or below the relevant thresholds at a 5% level.[31]

The first prediction of the uniformly 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 γ_7^a refers to the posteriors revealed in DP 7 given the choice of a. Table 10 tests this 'no learning' prediction. It divides subjects into those who have a threshold (i.e. point estimate of posterior belief from DP 7) significantly above $\mu_i(1)$, and those for whose threshold is significantly 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

[31] For the 0.6 prior 9 subjects were significantly below and 28 significantly above. For the 0.75 prior 15 below and 10 above. For the 0.85 prior 24 were below and 8 above.

these groups, it then reports the fraction of subjects who exclusively choose a.

Table 10: Testing the 'No Learning' Prediction:				
Fraction of subjects who never choose b				
		$\mu(1)$		
		DP8	DP9	DP10
		0.6	0.75	0.85
Significant differences	$\gamma_7^a(1) < \mu_i(1)$	33%	46%	41%
	$\gamma_7^a(1) \geq \mu_i(1)$	3%	10%	14%
Each column reports the fraction of subjects who exclusively choose a in the associated decision problem. The top row is subjects whose estimated posterior from choosing a in DP7 is significantly below the prior in the relevant DP, while the bottom row is subjects whose posterior is significantly above				

Table 10 shows that, while it does not perfectly match our data, the 'no learning' prediction does produce the correct comparative statics. In each DP, 33%-46% of the subjects who should exclusively choose a based on their point estimates do so, higher than the equivalent fraction for those who should be choosing both a and b. As we discuss in section 4.6, changes in estimation costs between trials could lead to violations of the no learning prediction.

The second part of the LIP condition states that, in each DP, subjects who are predicted to be gathering information should use the same posteriors as they did in DP 7. Figure 6 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 uniformly posterior separable model (i.e. those for whom $\gamma_7^a > 0.6$, significant at the 5% level). 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. These estimates are based on the OLS version of equation (7) run on the relevant subjects, with the coefficients and standard errors transformed as discussed above. The LIP prediction is that these posteriors should be the same. Panels b and c repeat this analysis for DPs 9 and 10. Figure 6 shows our data is relatively well described by the LIP prediction: of the six comparisons, only one (the posterior following choice of b with the 0.75 prior) shows a significant difference at the 10% level based on a nonlinear Wald test of the hypothesis that the relevant posteriors are equal. However, a test of the joint hypothesis that all six conditions hold simultaneously is rejected at the 5% level.

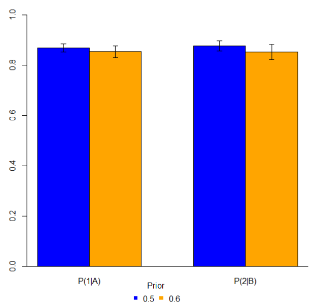


Figure 6a: Subjects with posterior above 0.6

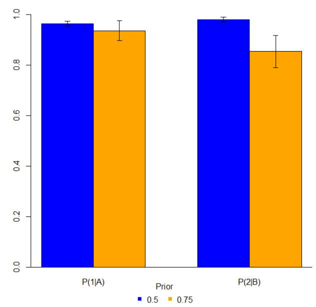


Figure 6b: Subjects with posterior above 0.75

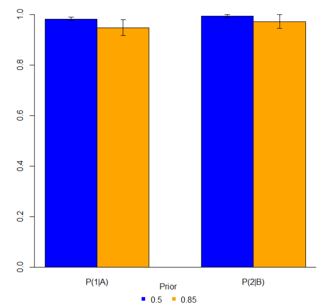


Figure 6c: Subjects with posterior above 0.85

Figure 6 - Estimates of posterior beliefs following choice of each action in each decision problem. “Subjects with posterior above x” refers to subjects whose estimated posterior following the choice of a in DP 7 are significantly above x. Bars show standard errors clustered at the individual level.

4.5 Experiment 4: Symmetry

23 subjects took part in experiment 4.[32] The results are summarized in figure 7, which shows the probability of choosing the correct action as a function of the state for each DP and each

[32] Data from a single session which took place on 27th June 2013 at the CESS laboratory at NYU.

treatment.

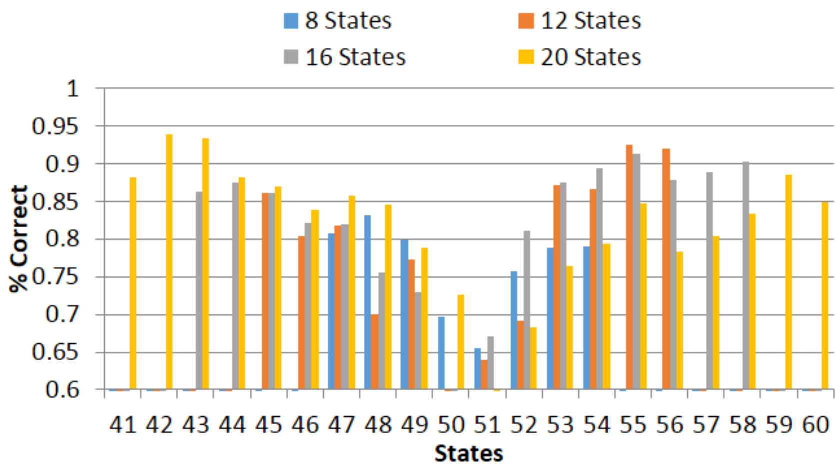


Figure 7: Experiment 4 - probability of choosing the correct action in each state.

This figure show clear and systematic violations of symmetry: subjects were more likely to make mistakes in states near the threshold of 50. This observation is confirmed by a logit regression of the form

$$1(\text{choice}_{i,j} \text{ is correct}) = \beta_0 + \beta_1 d_j + \varepsilon_{i,j} \tag{8}$$

where d_j is the absolute value of the number of red balls on the screen minus 50, and with standard errors clustered at the subject level. The estimate for the coefficient β_1 is 0.032 (P<0.001).

4.6 Discussion

Our overall conclusions from this set of experiments are: (1) that 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 simple environments we consider; (3) there are qualitative similarities between our data and the LIP condition that characterizes uniformly posterior separable models; (4) 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 symmetry properties, which are unrealistic in this information gathering environment.

In this section we discuss some of 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 the Shannon model? As noted above, we make use of two types of aggregation: within subject across decision problems and between subjects. 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 varies from trial to trial in order to prevent learning. It could be that some configurations are easier to understand than others. Aggregating across individuals may also cause problems, because different individuals may have different costs of attention. Of the two, we expect the latter 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 are clustered by color, the variance of which will be low when the number of balls is large. While plausibly more susceptible to variation, aggregate level data is useful because it provides us with much more power to detect differences in behavior across decision problems.

For most of the tests that we perform neither type of aggregation presents 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.

The exceptions are the test of the ILR condition in experiment 2 and the test of LIP in experiment 3. In the former case it is true that variability in information costs or difficulty could lead to violations of the predictions of the Shannon model in the direction we observe: In the presence of such variation we can show using equation (3) that estimated information costs are:

$$\hat{\kappa} = \frac{u}{\ln\left(E\left(\frac{\exp(u/\kappa)}{1-\exp(u/\kappa)}\right)\right) - \ln\left(E\left(\frac{1}{1-\exp(u/\kappa)}\right)\right)}$$

where u is the reward for making the right selection and E is the expectation operator taken with respect to the distribution of true costs κ. If this distribution is degenerate, then estimated information costs equal the true costs. If the distribution is not degenerate, then $\hat{\kappa}$ can be above or below the average of κ, but is increasing in u.

As we believe cost variation is higher than difficulty variation it is encouraging that we find responses to be too slow at both the individual and aggregate level. While it may be that variations in difficulty are causing the Shannon model to fail at the individual level, the fact that this occurs in an experimental situation where we believe costs to be relatively stable means that the model is likely to have problems in other applications as well.

In the case of the LIP condition, variability in difficulty 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 for LIP 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. Repeating the analysis of section 4.3 while dropping the first block for each subject does not significantly change our conclusions: We still find few violations of NIAS and NIAC (88% of subjects exhibit no significant violations of either condition), and response to incentives is still slower than predicted by the Shannon model (for example, when comparing the 40% and 70% reward levels we find 25 subjects who respond significantly more slowly than Shannon predicts, versus 2 who are significantly too fast).

5 Alternative Cost Functions

Our evidence so far has offered support for the general rational inattention model (in the simple settings we study), but shows substantial violations of important features of the Shannon model. The IUC and ILR conditions are most clearly violated, while the evidence against LIP is arguably more mixed. This raises the question of whether there are cost functions other than Shannon which would do a better job of explaining our data. In this section we introduce some (non-exclusive) proposals from the literature, and discuss which of the problems with Shannon they solve. We then estimate a collection of these alternative models to determine which, if any, offer a significant improvement over Shannon.

We begin by discussing two natural approaches which allow more flexibility in fitting attentional responses to incentives, as measured in experiment 2. There may be good reasons for doing so, especially if subjects are gathering information by examining balls one at a time - a process akin to sequential sampling.[33] As pointed out by Morris and Strack [2017], costs that are linear in entropy can by microfounded through the sequential sampling approach, but requires the perhaps implausible assumption that the cost of sampling each ball decreases in the variance of the posterior. If sampling costs are instead constant then behavior would be better modelled by costs that are convex in mutual information. The presence of intrinsic incentives for choosing the right option in experiment 2, which are then partially crowded out by monetary incentives, would also lead to behavior consisted with convex information costs.

Our first approach is to simply relax the assumption that costs are linear in mutual information. The cost function

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

has two parameters: κ and σ. It allows for either decreasing or increasing marginal costs of mutual information. The resulting model does not imply the ILR condition will hold in experiment 2 nor, as it is not uniformly posterior separable, the LIP condition will hold in experiment 3. It does, however, maintain the implication that payoff equivalent states will be treated equivalently in experiment 4.

[33] As pointed out by one anonymous reviewer, a sequential sampling lens may make more sense than one in which the cost of information is truely based on entropy, and in which these costs are convex within each decision problem, but separable across decision problems.

An alternative approach is to consider other models in the uniformly posterior separable class. In the estimation below we replace Shannon entropy with the functional form of Generalized entropy (Shorrocks [1980]):

$$K^{Gen}_{\rho,\kappa}(\mu,\pi) = \kappa \sum_{\gamma\in\Gamma(\pi)} \pi(\gamma)\left[T^{Gen}_{\rho}(\gamma)\right]$$

$$T^{Gen}_{\rho}(\gamma) = \begin{cases} \left(\frac{1}{(\rho-2)(\rho-1)|\Gamma|}\sum_{\Gamma}\hat{\gamma}^{2-\rho} - 1\right) \text{ if } \rho \neq 1 \text{ and } \rho \neq 2; \\ \frac{1}{|\Gamma|}\left(\sum_{\Gamma}\hat{\gamma}\ln\hat{\gamma}\right) \text{ if } \rho = 1; \\ -\frac{1}{|\Gamma|}\left(\sum_{\Gamma}\ln\hat{\gamma}\right) \text{ if } \rho = 2. \end{cases}$$

where $\hat{\gamma} = \gamma|\Gamma|$.

This again adds a single parameter - ρ - to the Shannon model. Generalized entropy generalizes Shannon entropy in a manner similar to the way in which the Constant Relative Risk Aversion utility function generalizes log utility: the function changes continuously in the parameter ρ, with $\rho = 1$ being (an affine transform of) Shannon entropy.[34] Generalized entropy does not imply ILR, but does imply both LIP and that payoff equivalent states will be treated equivalently in experiment 4. For our purposes, the key feature is that it can allow for the marginal cost of information to increase more or less quickly than the Shannon model, depending on whether ρ is greater or less than 1.[35]

Less obvious is how to modify the Shannon model in order to accommodate perceptual distance effects of the type demonstrated in experiment 4. However, a recent paper by Hébert and Woodford [2021] offers one promising solution. They propose a class of 'neighborhood based' cost functions. In order to construct these costs, the state space is divided into I 'neighborhoods' $X_1...X_I$. An information structure is assigned a cost for each neighborhood based on the expected change in entropy between prior and posteriors conditional on being in that neighborhood. The total cost of the information structure is then the sum of costs across all neighborhoods. An example of a neighborhood based function with entropy costs is

$$K_N(\mu,\pi) = \sum_{i=1}^{I} \mu(X_i)\kappa_i \sum_{\gamma} \pi(\gamma|X_i)\left(-H(\gamma|X_i) - [-H(\mu|X_i)]\right) \tag{9}$$

where $\mu(X_i)$ is the prior probability of a state in neighborhood X_i, κ_i is the marginal cost

[34] Specifically, when $\rho = 1$, generalized entropy is equal to the maximal possible entropy minus the entropy of the observed distribution.

[35] Other generalizations of Shannon entropy, such as Tsalis entropy, also have this feature. Our choice of the generalized entropy function was essentially for computational convenience.

of information in neighborhood i, $\pi(\gamma|X_i)$ is the probability of signal γ conditional on a state in X_i occurring and $H(\gamma|X_i)$ is the entropy of the posterior generated by signal γ conditional on a state in X_i occurring. This cost function would imply that the ILR condition would hold in experiment 2, but symmetry would potentially be violated in experiment 4.

These cost functions have a number of nice features. First and foremost they allow for it to be more expensive to differentiate between some states than others: the cost of differentiating between two states depends on which neighborhoods they share. Second, Hébert and Woodford [2021] show that such cost functions can be microfounded as the result of a process of sequential information gathering. Finally, this cost function is also uniformly posterior separable.

One important free parameter for the class of cost functions is the definition of the neighborhoods. The theory itself does not give a guide as to how these should be defined, and whether or not a particular neighborhood is important in determining costs. Here we follow Hébert and Woodford [2021] and consider a model with two classes of neighborhood: first, a global neighborhood which contains all states, and second a collection of local neighborhoods which contain adjacent states (i.e. one neighborhood will contain 40 and 41 red balls, the next 41 and 42 red balls, etc). We further restrict the costs associated with all the local neighborhoods to be the same, meaning that this model has two parameters: the marginal cost of information in the global neighborhood κ_g and in the local neighborhoods κ_l.

5.1 Estimation

We now report the results of estimating models based on the classes above on our data. For the aggregate data we consider four model variants summarized in table 11 below.

First, as a baseline we will estimate the Shannon model. This allows us to measure the improvement in fit afforded by our alternative models, all of which nest Shannon as a special case. Next we estimate the Neighborhood model of Hébert and Woodford [2021] with entropy costs. This model should improve fit in experiment 4, as it allows for a notion of perceptual distance. However, as the model reduces to Shannon in the setting of experiment 2 it will do nothing to improve the fit of attentional responses to incentives. We therefore estimate two variants of the Neighborhood model. The 'Power with Neighborhood' model raises the mutual information cost in each neighborhood to a power, while the 'Generalized with Neighborhood' function keeps the neighborhood cost structure but replaces Shannon entropy with Generalized entropy as a measure of uncertainty.

We estimate the above models on both aggregate and individual level data using a two stage maximum likelihood procedure, as described in appendix A4. At the aggregate level we report results from experiment 2 and experiment 4, both jointly and separately. We do not report aggregate results from experiments 1 and 3 because here aggregation can easily generate behavior that has zero probability under models in the Shannon class even if the individual level data is consistent (for example choosing a, b and c with positive probability in experiment 1 or violating LIP in experiment 3). At the individual level we report results for all four experiments, and also add an 'Inattentive' model which assumes the subject gathers no information and selects the best option given prior beliefs. We do so because a number of subject seem to adopt such behavior - completing the experiment very rapidly and almost always making the same selection.

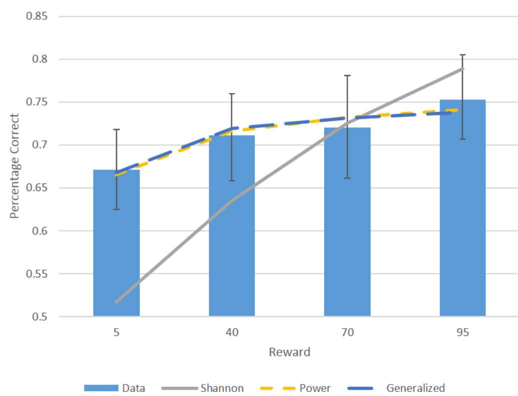
In comparing the models we rely primarily on measures which allow for non-nested model comparison while balancing parsimony and goodness of fit - namely the Bayesian Information Criterion (BIC) and the Akieke Information Criterion (AIC). However, the Shannon and Neighborhood models are nested in both the Power and Generalized versions, so we can also perform likelihood ratio tests when comparing these cost functions. Standard errors are clustered at the individual level.

5.1.1 Aggregate Results

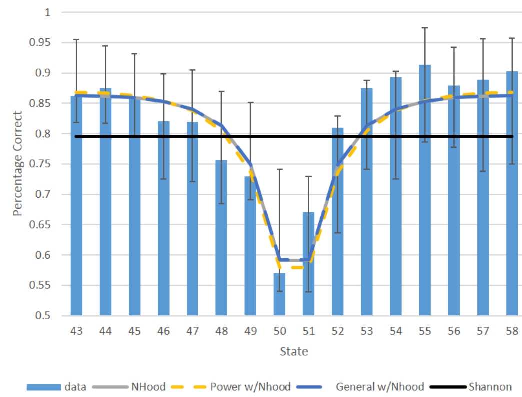
We begin by reporting the results from aggregate data, estimated from experiments 2 and 4 both separately and jointly. Figure 8 panel A shows the fitted values from the estimation on experiment 2 only. Note that, in the setting of experiment 2, all models collapse to their no-neighborhood versions, as the two states in this experiment are only jointly found in the global neighborhood. Panel B shows the best fit from the 16 state treatment from the estimation on experiment 4 only. Equivalent graphs for the other treatments in this experiment can be found in appendix A4. Table 12 reports parameter estimates and AIC and BIC scores from the models estimated separately and jointly on the two experiments.

Model fits from the joint estimation can also be found in appendix A4.

Figure 8



Panel A: Model Fit for Experiment 2 Only



Panel B: Model Fit for Experiment 4 Only - 16 States

Blue columns indicate aggregate data, with bars indicating standard errors clustered at the individual level. Lines represent the best fit of each model.

Looking first at experiment 2, we see in figure 8 that, as anticipated, the Shannon model predicts a faster response to incentives than is seen in the data. Both the Power and Generalized models work well as a solution to this problem, allowing for a much flatter expansion path. When fit to experiment 2 alone the two models give essentially identical estimates. In table 12 we see that both the AIC and BIC are much lower for the Power and Generalized models than for the Shannon model, indicating that the better fit these models provide is worth the addition of one parameter. Likelihood ratio tests confirm this story: for both the Power and Generalized models the restriction to the Shannon model is rejected at <0.01%. There is little to choose between the fit of the Power and Generalized models.

Looking next at experiment 4, we see that the important difference is between models that do not allow neighborhoods, and those that do, with the estimates from the latter class essentially the same, and well able to match the fact that subjects are better at the task for states further away from the cut off. The AIC and BIC show that all the neighborhood based models do much better than the no-neighborhood Shannon model, but that the difference between these models is small. Likelihood ratio tests favor each of the neighborhood based models over the Shannon model at <0.01%. The restriction of the Power or Generalized models to the Neighborhood model is not rejected at the 10% level

Table 11: Models for Estimation		
Model	Cost Function	Parameters
Shannon	$\kappa_g \left(\sum_{\gamma\in\Gamma(\pi)} \pi(\gamma)\left[-H(\gamma)\right] - \left[-H(\mu)\right]\right)$	κ_g
Neighborhood	$\sum_{i=1}^{I} \mu(X_i)\kappa_i \sum_{\gamma} \pi(\gamma\vert X_i)\left[-H(\gamma\vert X_i)\right] - \left[-H(\mu\vert X_i)\right]]$	κ_g, κ_l
Power w/Nhood	$\sum_{i=1}^{I} \mu(X_i)\kappa_i \left(\sum_{\gamma\in\Gamma(\pi)} \pi(\gamma\vert X_i)\left[-H(\gamma\vert X_i)\right] - \left[-H(\mu\vert X_i)\right]\right)^{\sigma}$	$\kappa_g, \kappa_l, \sigma$
Generalized w/Nhood	$\sum_{i=1}^{I} \mu(X_i)\kappa_i \sum_{\gamma} \pi(\gamma\vert X_i)\left[T_{\rho}^{Gen}(\gamma\vert X_i)\right] - T_{\rho}^{Gen}(\gamma\vert X_i)]$	κ_g, κ_l, ρ

Table 12: Parameter Estimates - Aggregate Data						
Model	κ_g	κ_l	σ	ρ	BIC	AIC
Experiment 2 Only						
NHood	28.82	-	-	-	379	372
Power	7728.00	-	4.23	-	55	41
Generalized	0.16	-	-	13.41	56	42
Experiment 4 Only						
Shannon	7.38	-	-	-	485	479
NHood	5.40	5.04	-	-	326	313
Power w/NHood	4.98	5.63	0.94	-	334	315
Generalized w/NHood	5.36	4.99	-	1.05	334	315
Experiment 2 and 4						
Shannon	23.49	-	-	-	1689	1681
NHood	3.24	23.93	-	-	738	722
Power w/NHood	14.31	88.90	1.99	-	447	423
Generalized w/NHood	0.01	1.80	-	8.41	421	397
Columns 2-5 report parameter estimates from the models described in table 11 Columns 6 and 7 report the Akaike information criterion and Bayes information criterion respectively						

When estimated jointly on the data from experiments 2 and 4, both the Generalized with Neighborhood and Power with Neighborhood models do much better than either the Neighborhood or Shannon models, with the former doing best according to both the AIC and BIC criteria.[36] As can be seen from the model fits in appendix A4, the Neighborhood model is

[36]Note that in this experiment we assume linearity of utility across the two relevant reward levels - $10 and $40.

unable to match the response to incentives in experiment 2. Moreover, because global costs have to be so high to even remotely match the data from that experiment, local costs have to be very small, meaning that it also does a poor job of matching the data from experiment 4.

Using the joint estimates on the data from experiments 2 and 4, we can test the hypothesis that the Power with Neighborhood and Generalized with Neighborhood models have the same parameters in the two experiments. A log likelihood test rejects the null hypothesis of equal parameters, suggesting some degree of misspecification. Intuitively, this must be because the accuracy when there are 51 or 49 red balls on screen in experiment 4 is different to that in experiment 2 when the reward levels are similar. In principle this could be because we assume linear utility to convert the prizes in the two experiments onto the same scale, but we suspect that this is not what is going on, as it would require a significant degree of risk *loving* to reconcile the two experiments. Instead we suspect it is due to the fact that subjects may be using different strategies in the two experiments that may lead to different costs in differentiating between close together states.

5.1.2 Individual Results

Table 13: Individual Results

Experiment	NHood	Power w/NHood	Gen w/NHood	Shannon	Inattentive
1	36	10	25	-	29
2	29	19	23	-	29
3	35	43	10	-	13
4	43	26	4	17	9

Percentage of subjects best described by each model according to the AIC

Table 13 shows the fraction of subjects who are best fit by each model according to the AIC criterion in each experiment. Broadly speaking these results replicate those from the aggregate analysis: in experiment 2 a significant fraction (42%) of subjects are better described by either the Power or Generalized models than by the Shannon model (a further 29% are best described as inattentive). In experiment 4, 73% of subjects are best described by a model that allows for a neighborhood structure. The results from experiment 1 and 3 are similar to those from experiment 2. The fact that the power function performs best in experiment 3 is notable, as this is the only model which does not imply the LIP condition.[37]

6 A Second Experimental Environment

The above results were derived from subjects who were gathering information from a particular stimulus - displays of colored balls on a screen. A natural question is the extent to which our results are specific to that environment. In order to provide a partial answer, we ran a set of follow up experiments with the same underlying structure as experiments 1-4 above, but with an information gathering task that was completely different - namely solving simple equations.

As we shall see below, the picture that emerges is broadly the same to that reported above: subjects actively manage their attention in a way that is consistent with rational inattention, but not with the more restrictive conditions of the Shannon model. Relative to the latter they again respond too slowly to changes in incentives, and exhibit failures of symmetry. The main difference is that there is less evidence in support of the LIP condition.

6.1 The Experimental Task

For a typical question in this new environment subjects were faced with a number of simple equations on the screen. Some of these equations were correct, and some were false. The state for that question was determined by the number of equations which were solved correctly. A similar design has previously been used by Ambuehl *et al.* [2020]. Appendix A5.1 includes a typical screenshot as well as the instructions presented to the subjects.

Other than the stimuli, our aim was to make the experiments match as closely to our 'main' experiment as possible. Two main differences are worth noting. First, due to Covid restrictions, the experiments were run using an online version of the laboratory environment: groups of around 20 subjects, recruited from the standard CELSS subject pool, took part remotely using video conferencing software. The experiment otherwise followed standard laboratory protocols, with instructions read out at the start and anonymous payment at the end.[38] Importantly, subjects were told that they were not allowed to leave the zoom session

[37]Similar results are obtained if the Bayesian Information Criterion is used intead of the AIC. The only notable difference is that the fraction of subjects best explained by the Shannon model in experiment 3 increases to 48%, while the percentage best explained by the power function reduces to 30%

[38]Using either Venmo or Amazon giftcards.

until all the participants had completed the experiment.

Second, we found in pre testing that subjects we taking significantly longer on each question in this new design relative to the 'balls' design from the main experiment - approximately twice as long per question. In order to keep the experiment at a manageable length, we therefore capped the number of treatments at 2, rather than the 4 used in main experiments 2 and 3. This means that, while we can perform all of the qualitative tests from section 4, but not the model fitting exercise from section 5 on the new data. As before, the order of decision problems was block randomized.

Data was collected in 12 experimental sessions between September 2021 and January 2022.

6.2 Experiment 2.1: Testing for Responsive Attention

Experiment 2.1 mimics experiment 1, which was designed to elicit violations of Monotonicity. Subjects were presented with nine equations. There were two equally likely states of the world, defined as 4 or 5 equations correct out of 9. As in experiment 1, in DP 2.1, the subject has the choice between the sure-thing option a, which pays 50 probability points, and an option b which pays less than a in state 1, but more in state 2 (i.e. $b_1 < 50 < b_2$ - with the precise values described in table 2). In DP 2.2, the option c is added, which paid 100 points in state 1 and 0 in state 2. Payment was in probability points with a prize of \$20. Each subject faced 75 repetitions of each DP. 35 subjects in from 2 sessions took part in this experiment.[39]

Table 14 summarizes the results of experiment 2.1, using the method of analysis described in section 4.2.

[39]Subjects were not evenly divided across the four treatments due to the online implementation of the experiment

Table 14: Results of Experiment 2.1								
		$P(b\|1)$			$P(b\|2)$			
Treat	N	$\{a,b\}$	$\{a,b,c\}$	Prob	$\{a,b\}$	$\{a,b,c\}$	Prob	% Subjects
1	13	12.0	10.2	0.41	76.7	76.6	0.99	38
2	5	6.1	9.0	0.40	81.0	88.3	0.55	20
3	8	3.4	3.1	0.49	68.2	78.5	0.41	50
4	9	8.6	11.7	0.52	66.0	81.0	0.29	56
Total	35	8.4	8.9	0.74	73.0	80.0	0.07	42

Panel 1 reports fraction of observations in which b was chosen when the true state was 1 and options a and b were available (column 1) or when a, b and c were available (column 2). Column 3 reports the probability associated with the null hypothesis that the probabilities in column 1 and 2 are equal. Panel 2 repeats the exercise for observations in which the true state is 2. The final column reports the fraction of subjects who choose b significantly more in state 2 when c is available.

Overall, we again see violations of Monotonicity, albeit weaker than those observed in the original design. Pooling all observations, we see that the choice of action b in state 2 increases from 73% to 80% when c is available, significant at the 10% level. Overall, 42% of subjects exhibit violations of Monotonicity. The weaker effect seems to be mainly due to the fact that subjects are collecting more information in the 'sums' design than the 'balls' design: the 73% of subjects choosing b in state 2 when c is not present compares to 28% of subjects using the previous stimuli.

6.3 Experiment 2.2: Changing Incentives

Experiment 2.2 is again designed to test subjects' response to changing in incentives. Subjects were presented with 7 equations, with two equally likely states represented by 3 and 4 equations correct. Subjects could choose between two actions, each of which paid off a positive amount in one state and zero in the other. In DP 2.3 the 'correct' answer gave 5 points, while in 2.4 it gave 95 points. Payment was in probability points with a prize of \$20.[40] 55 subjects from 4 sessions took part in this experiment.

We analyzed the data from experiment 2.2 using the procedures detailed in section 4.3.

[40] This compares to the \$40 prize used in experiment 2. The smaller amount was used partly because to keep the payment per question the same, and partly because, in pretesting, the higher prize lead to extremely high accuracy in all subjects.

Detailed results for the experiment are shown in Appendix 5.2. At the individual level, we find very few significant violations of either NIAS or NIAC (see table A5.3). We can also reject the hypothesis that NIAS is violated at the aggregate level (table A5.2). However, as we see in Figure 9, the response to incentives is smaller than it was in experiment 2: the probability of choosing the correct outcome increases from 82% in the 5 probability point treatment to 84% in the 95 point treatment. While the point estimates are in line with the NIAC hypothesis, they are not significantly different from each other. Unsurprisingly, this means that the aggregate data is also inconsistent with the hypothesis that costs are linear in Shannon mutual information. Figure 10 replicates the analysis of figure 3, and once again shows that estimated information costs are much higher in the 95 point treatment than in the 5 point treatment.

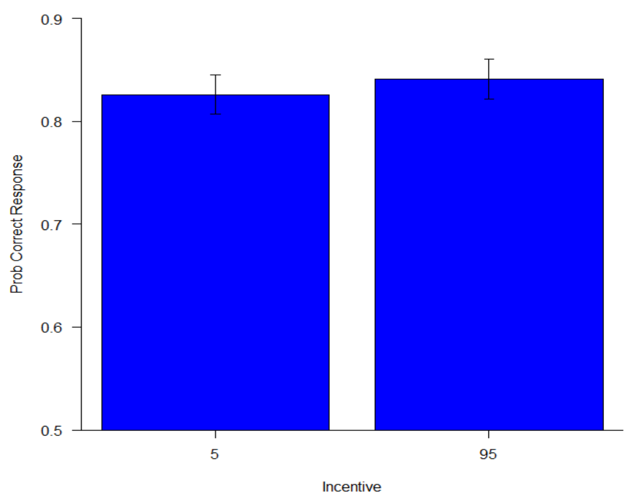


Figure 9: Probability of correct response by decision problem - aggregate data from experiment 2.2. Bars show standard errors, clustered at the individual level.

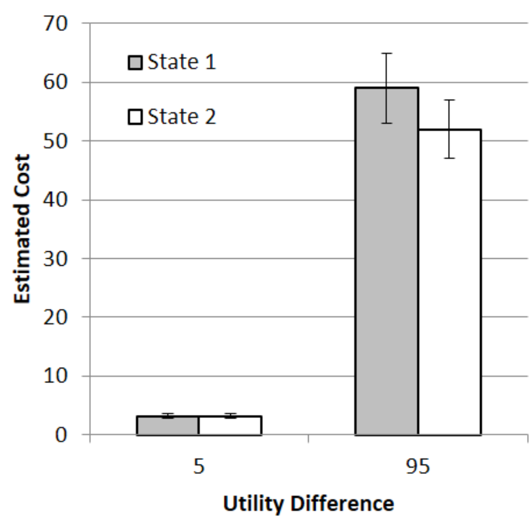


Figure 10: Estimated cost parameter from the Shannon model in each state and at each reward level using aggregate data from experiment 2.2. Bars show standard errors clustered at the individual level.

6.4 Experiment 2.3: Changing Priors

Experiment 2.3 replicates the design of experiment 3 to study the impact of changing priors. There are again 2 states, represented by 3 or 4 equations correct out of 7, and two actions, one (action a) which paid \$10 in state 1 and the other (action b) \$10 in state 2. There were two decision problems, each of which were repeated 50 times. In DP 2.5 each state was equally likely, while in DP 2.6 state 1 had a prior probability of 0.85. 53 subjects in 3 sessions took part in experiment 2.3.

We analyze the data using the same protocol described in section 4.4.

Table 15 reports the results of the NIAS tests for experiment 2.3, in the manner of table 8. Again, the condition is satisfied even in the case of the asymmetric prior.

Table 15: NIAS Test

DP	$P_j(a\|2)$	Constraint on $P_j(a\|1)$	$P_j(a\|1)$	Prob
2.5	0.19	0.19	0.88	0.000
2.6	0.30	0.88	0.96	0.009

Each row reports the probability that a is chosen in state 2 in that DP (column 2), the constraint implied by NIAS on the probability that a is chosen in state 1 (column 3), and the actual probability that a is chosen in state 1 (column 4) in the aggregate data. Column 5 reports the probability associated with the null hypothesis that the columns 3 and 4 are equal

In order to test the predictions of the LIP condition we divide subjects based on whether their posterior beliefs when choosing action 1 in DP 2.5 were greater or less than 0.85. Based on point estimates 60% of subjects fall into the first category (30% significantly so), and 40% in the second (16% significantly so)

Table 16 replicates the test of the 'no learning' prediction from table 10. Recall that those subjects whose posterior beliefs are below 0.85 in DP 2.5 should always choose a in DP 2.6, while those with posterior beliefs above should pick both a and b. Table 16 reports these results for experiment 2.3. More subjects always choose a in the former category than the latter, significant at the 1% level.

Table 16: Testing the 'No Learning' Prediction:		
Fraction of subjects who never choose b		
		$\mu(1)$
		DP10
		0.85
Significant differences	$\gamma^a_{2.5}(1) < 0.85$	56%
	$\gamma^a_{2.5}(1) \geq 0.85$	0%
Fraction of subjects who exclusively choose a in the associated decision problem. The top row is subjects whose estimated posterior from choosing a in DP7 is significantly below 0.85 while the bottom row is subjects whose posteriors significantly above		

Figure 11 replicates the test of LIP reported in figure 6 for the first experimental environment. Recall that the posterior beliefs of subjects who should be choosing both a and b based on the above classification should be the same in DP 2.5 and 2.6. Figure 11 shows estimated posteriors conditional on the choice of a and b for subjects whose beliefs in DP 2.5 were significantly above the 0.85 threshold. While the LIP condition holds following the choice of a, beliefs following the choice of b are significantly different at the 1% level.

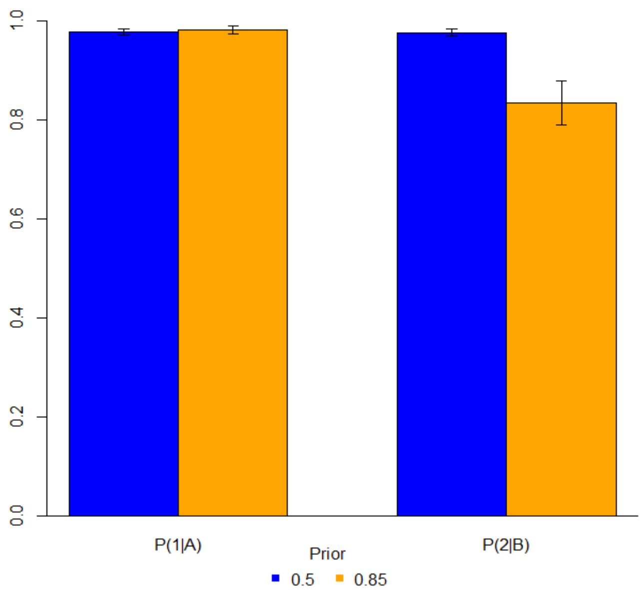


Figure 11: Posterior beliefs for subjects with threshold above 0.85. Bars inidcate standard errors clustered at the individual level.

6.5 Experiment 2.4: Invariance Under Compression

Our final experiments tests IUC in the new experimental environment. Subjects were presented with nine equations. The number of correct equations was between 2 and 7, all equally likely, giving rise to 6 possible states. Subjects could choose between action a, which paid \$10 if there were fewer than 5 equations correct (and zero otherwise) and action b, which paid \$10 if 5 or more equations were correct. 34 subjects in 2 sessions took part in the experiment.

Figure 12 shows the fraction of correct choices in each state (i.e. number of correct equations). While the pattern is messier than found in the first experimental environment, a regression of the form of equation (8) show that there is a significant, positive relationship between distance from the threshold and accuracy (coefficient 0.08, p<0.001).

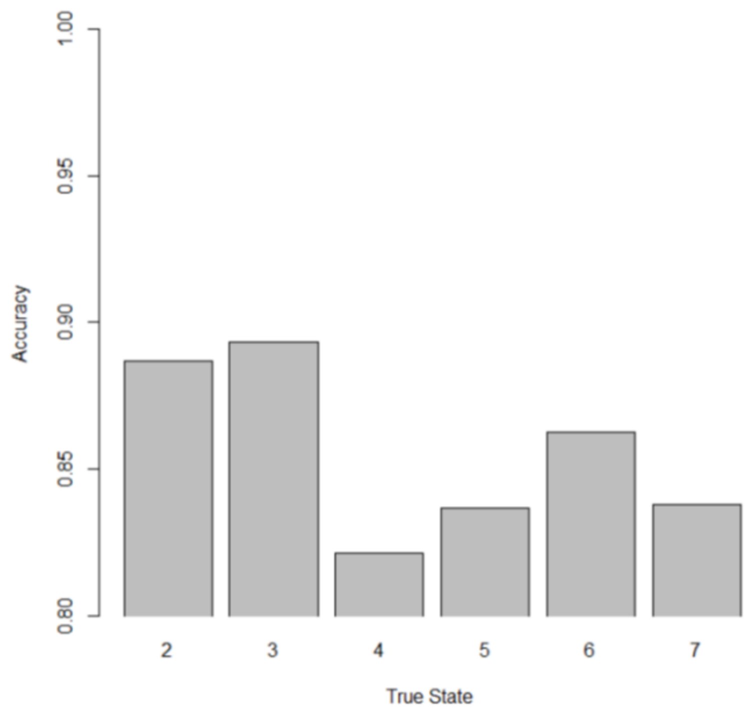


Figure 12: Experiment 2.4 - probability of choosing the correct action in each state.

7 Related Literature

Many papers have established the importance of attention limits in economically interesting contexts, including consumer choice,[41] financial markets,[42] and voting behavior.[43] 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

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

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

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

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.

Two recent papers have tested whether people adjust their attention to tax rates in a simulated shopping environment. Taubinsky and Rees-Jones [2018] show that underreaction to tax rates decreases on average as tax rates are saliently increased, suggesting that subjects are rationally adjusting their attention. Morrison and Taubinsky [2019] use a more detailed data to test a rich set of predictions regarding individual differences in attention and how these respond to incentives.

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 our 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,[44] 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.[45] More recent work (Taubinsky [2013], Goecke *et al.* [2013], Khaw *et al.* [2017]) has also focussed on the dynamics of information 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, while Cheremukhin *et al.* [2015] report 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.

Contemporaneous to our work, Ambuehl *et al.* [2020] test two implications of the Shan-

[44] 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 generally used the data to compare behavior to rational benchmarks.

[45] Though see Sanjurjo [2017].

non model: that decision makers for whom acquiring and processing information is more costly respond more strongly to changes in incentives for participating in a transaction with unknown but learnable consequences, and decide to participate based on worse information. They find strong support for both predictions. However, they also show that these predictions hold for a much broader class of information cost function, so these results are complementary to our findings in support of more general models of rational inattention. Ambuehl [2017] also tests a distinct implication of posterior separability, finding support for the prediction that higher incentives for participation in a transaction with unknown but learnable consequences cause people to skew their information acquisition towards participation, and thus lead to an increase in false positives and a decrease in false negatives, over and above what would be observed without flexible information acquisition.

Three other papers, concurrent to ours, test subsets of the properties that we consider in this paper. Dewan and Neligh [2020] and Caplin *et al.* [2020] use fine grained variations in rewards to recover the 'expansion path' of attention in a manner similar to our experiment 2.[46] Both paper report that subjects by and large follow NIAS and NIAC, but otherwise report significant heterogeneity in behavior. Denti [2020] makes use of the data from this paper to test for posterior separability in experiments 1 and 2. The latter is consistent while the former is not. None of these papers study the impact of changing priors.

Finally, there are a set of papers which test certain features of rational inattention in real world settings, including recruitment (Bartoš *et al.* [2016]), baseball (Bhattacharya and Howard [2021]), forecasting (Gaglianone *et al.* [2020]) and search for tax information (Hoopes *et al.* [2015]). While these papers are important in demonstrating and measuring attention adjustment in the field, they are unable to perform the type of fine grained tests that we can do using our experimental data.

There is also a large experimental literature on 'real effort tasks', in which subjects have the opportunity to complete tasks that require either physical or mental effort, and get paid for doing so according to a well specified schedule (see Charness *et al.* [2018] for a review). While sharing some similarities with our experiments, a key distinction is that, in rational inattention tasks, the reward to effort is not explicitly specified. Subjects make choices between alternatives, the payoffs of which depend on some underlying state. They are then free to engage in tasks that allow them to learn about that state, but exactly what tasks they can do, and the rewards for doing them, are things that the subjects must figure out for themselves. Arguably, testing whether people can do this is one of the central points of

[46] Both also have features that make certain tasks easier or harder in the manner of experiment 4.

experiments such as ours

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.[47] 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[48] 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.

8 Conclusion

In this paper we have provided experimental evidence that, when faced with informational constraints, people do adjust their attention in response to prior beliefs and incentives. Moreover, in simple settings, they do so broadly in line with a model of rational inattention, meaning that they act as if they are selecting information in order to maximize utility net of costs. These costs, however, do not seem to be well described as a linear function of Shannon mutual information. Our aggregate data is better matched by a model that allows

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

[48] 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.

for a 'neighborhood' structure, and which uses either a generalized form of entropy, or a nonlinear function of mutual information as a basis for costs.

Two immediate questions arise regarding our findings. First, which of our results are generalizable beyond our experimental tasks, and to what settings? Second, do these results change the way in which we think the rational inattention model should be applied to economic problems? While any answer to these questions clearly requires some speculation, we provide some conjectures below, in part as a potential spur to future research.

In broad terms, we can think of there as being three key features of an informational environments that could determine behavior: (1) The nature of the stimulus in which information is encoded, (2) the technology the decision maker has to extract information from that stimulus (including any constraints such as time limits) and (3) the task for which that information is to be used. We have presented results from two experimental environments. In both cases the state was determined by a number (number of red balls/number of equations correct), and we would expect it to be easier to differentiate states that were further apart according to this number. In both experiments the subjects only had their own senses and background knowledge available to determine the state. They had no external time limit, and were explicitly focusing on the task in hand. Importantly, in both experiments the value of information was extremely salient, as the subjects were provided with the value of different options in different states of the world.

How robust are our findings to changes in these factors? Our first finding is that subjects' actively adjust their attention, and that this can lead to violations of monotonicity. We conjecture here that there is nothing particularly special about our environment that lead to this conclusion, beyond the fact that subjects were able to gather more information at reasonable cost, and that they understood the rewards for doing so. Of course, if subjects were asked to extract information from a source in a language they did not speak, or faced a very strict time limit, then they might not be able make adjustments to the amount of information they acquire.[49]

Our second finding is that subjects behave broadly in line with the NIAS and NIAC conditions, and so appear to be optimizing relative to some information cost. Here we expect the key determinant to be the simplicity of the task at hand. In each experiment there were a small number of payoff relevant states and actions, and the payoff of each action in each state was clear. Moreover, the information inherent in learning activities (looking at a ball,

[49] See Dewan and Neligh [2020] for an example of a task in which higher incentives lead subjects to work harder, but did not increase their accuracy.

solving an equation) was relatively clear. In more complex settings we would expect to see violations of NIAS and NIAC, because solving the rational inattention problem would be harder. Understanding what these look like is, we believe, an important avenue for future research.

Third, our main experiment provided some mixed support for the LIP condition. As we saw in the follow up experiment, it may be that this finding is quite task dependent. One clue as to when such behavior might be apparent comes from the work of Morris and Strack [2017], and Hébert and Woodford [2019], who show that uniformly posterior separable information cost functions (which imply LIP) are consistent with optimal sequential information acquisition. Environments such as our first experiment, in which subjects could potentially sequentially sample balls one at a time, may therefore be more likely to give rise to LIP.

Fourth and fifth, we show that our data contradicts the response to incentives and the payoff invariance property of the Shannon model, and that richer parametric forms do a better job of explaining behavior. In the former case, while we do not believe our parameter estimates would be robust to different environments, we believe that generically, the simple linear version of the Shannon model is likely to be unable to capture behavioral responses. In the latter case, we imagine that whether or not IUC is satisfied will depend on the interplay between the perceptual environment and the available technology. In both our experiments, the perceptual environment clearly had the property that some states were easier to differentiate between than others. Moreover, the subjects did not have access to any technology that would allow them to collect information in a more efficient manner, by discarding this payoff irrelevant information. One could imagine that if, for example, the subjects were allowed to write computer programs which then advised them which option to pick, it might be that IUC would be recovered, as such a program could strip out all payoff irrelevant information.

The above discussion gives clues as to the economic environments to which our results would seem most relevant: those in which there is a natural concept of perceptual distance, that the decision makers are limited in the technology they have at their disposal, and where the benefits of information acquisition are clear.

To make the discussion more concrete, we can consider two cases in which violations of IUC have important economic implications: global games and welfare analysis. In the first case, Morris and Yang [2021] show that information cost functions that violate IUC - essentially by allowing for perceptual distance - admit a unique (inefficient) equilibrium, while Shannon costs allow for a multiplicity of outcomes. In the second case Angeletos and

Sastry [2019] show that cost functions which satisfy IUC will lead to economies in which the standard welfare theorems hold, while those that allow for perceptual distance will generically not.

Our findings, and the above discussion, allow us to start thinking about the situations in which each case would apply. For example, typical applications of global games include currency crises and pricing debt. In both these cases the underlying state has a natural concept of perceptual distance (the likelihood of an underlying project being successful; the fundamental state of the economy). However, in both cases it might be that sophisticated investors have access to technology that allows them to adapt their information acquisition in order to disregard payoff irrelevant information. In contrast, an application to bank runs also has a state which encodes perceptual distance (the return on deposits), but retail investors might not have access to such technology. In the economy of Angeletos and Sastry [2019], consumers are inattentive to prices. This seems to be a case in which there is a perceptual distance, and consumers are unlikely to have any technological solutions which allow them to ignore payoff irrelevant information and so we might expect IUC to fail.

The implications of the finding that the Shannon model does not well describe responses to incentives seems likely to be both more broadly applicable, and more vague. Generically, it means that the logit-like solution identified by Matejka and McKay [2015] will not apply. These have been heavily used in the literature (see for example Fosgerau *et al.* [2020], Matveenko and Mikhalishchev [2021]), but whether or not this formulation is crucial to the results, or merely analytically convenient, is likely to be question dependent.

Finally, there are a number of situations in which it has been shown that the ability of people to actively adjust their attention in response to incentives has implications for outcomes that are relatively robust to the precise form of the cost function - for example Bartoš *et al.* [2016] on discrimination and Matějka and Tabellini [2021] on electoral competition). Our result - that people can and do actively adjust their attention - holds promise for these models, particularly in cases where the value of information gathering is obvious. Again, in these examples the requirements in that regard seem relatively undemanding - for example recognizing that information on a job candidate who is ex ante expected to be poor is unlikely to lead them to be hired.

References

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

Sandro Ambuehl and Shengwu Li. Belief updating and the demand for information. *Games and Economic Behavior*, 109:21–39, 2018.

Sandro Ambuehl, Axel Ockenfels, and Colin Stewart. Who opts in? *Rotman School of Management Working Paper*, (3154197), 2020.

Sandro Ambuehl. An offer you can't refuse? incentives change what we believe. 2017.

George-Marios Angeletos and Karthik Sastry. Inattentive economies. Technical report, National Bureau of Economic Research, 2019.

Dan Ariely, Uri Gneezy, George Loewenstein, and Nina Mazar. Large stakes and big mistakes. *The Review of Economic Studies*, 76(2):451–469, 2009.

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.

Vivek Bhattacharya and Greg Howard. Rational inattention in the infield. *American Economic Journal: Microeconomics*, 2021.

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.

Gary L Brase. Pictorial representations in statistical reasoning. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 23(3):369–381, 2009.

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, Dániel Csaba, John Leahy, and Oded Nov. Rational inattention, competitive supply, and psychometrics. *The Quarterly Journal of Economics*, 135(3):1681–1724, 2020.

Andrew Caplin, Mark Dean, and John Leahy. Rationally inattentive behavior: Characterizing and generalizing shannon entropy. *Journal of Political Economy*, 130(6):000–000, 2022.

Andrew Caplin. Measuring and modeling attention. *Annual Review of Economics*, 8(1), 2016.

Christopher P Chambers, Ce Liu, and John Rehbeck. Costly information acquisition. *Journal of Economic Theory*, 186:104979, 2020.

Gary Charness, Uri Gneezy, and Austin Henderson. Experimental methods: Measuring effort in economics experiments. *Journal of Economic Behavior & Organization*, 149:74–87, 2018.

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.

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.

T Denti. Posterior-separable cost of information. Technical report, working paper, 2020.

Ambuj Dewan and Nathaniel Neligh. Estimating information cost functions in models of rational inattention. *Journal of Economic Theory*, 187:105011, 2020.

Mogens Fosgerau, Emerson Melo, Andre De Palma, and Matthew Shum. Discrete choice and rational inattention: A general equivalence result. *International economic review*, 61(4):1569–1589, 2020.

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.

Wagner Piazza Gaglianone, Raffaella Giacomini, João Victor Issler, and Vasiliki Skreta. Incentive-driven inattention. *Journal of Econometrics*, 2020.

Delphine Gandini, Patrick Lemaire, and Stéphane Dufau. Older and younger adultsâĂŹ strategies in approximate quantification. *Acta psychologica*, 129(1):175–189, 2008.

Rocio Garcia-Retamero and Ulrich Hoffrage. Visual representation of statistical information improves diagnostic inferences in doctors and their patients. *Social Science & Medicine*, 83:27–33, 2013.

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

Matthew Gentzkow and Emir Kamenica. Costly persuasion. *The American Economic Review*, 104(5):457–462, 2014.

Gerd Gigerenzer and Ulrich Hoffrage. How to improve bayesian reasoning without instruction: frequency formats. *Psychological review*, 102(4):684, 1995.

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.

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

John R. Hauser and Birger Wernerfelt. An evaluation cost model of consideration sets. *Journal of Consumer Research*, 16(4):pp. 393–408, 1990.

Benjamin M Hébert and Michael Woodford. Rational inattention when decisions take time. Technical report, National Bureau of Economic Research, 2019.

Benjamin Hébert and Michael Woodford. Neighborhood-based information costs. *American Economic Review*, 111(10):3225–55, 2021.

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.

Jeffrey L Hoopes, Daniel H Reck, and Joel Slemrod. Taxpayer search for information: Implications for rational attention. *American Economic Journal: Economic Policy*, 7(3):177–208, 2015.

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.

Daniel Kahneman and Amos Tversky. *Judgment under Uncertainty: Heuristics and Biases*, chapter Evidential impact of base rates, pages 153–160. Cambridge University Press, 1982.

Mel Win Khaw, Luminita Stevens, and Michael Woodford. Discrete adjustment to a changing environment: Experimental evidence. *Journal of Monetary Economics*, 91:88–103, 2017.

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.

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. *Games and Economic Behavior*, 104:131–145, 2017.

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 and Guido Tabellini. Electoral competition with rationally inattentive voters. *Journal of the European Economic Association*, 19(3):1899–1935, 2021.

Filip Matějka. Rationally inattentive seller: Sales and discrete pricing. *The Review of Economic Studies*, 83(3):1156–1188, 2015.

Andrei Matveenko and Sergei Mikhalishchev. Attentional role of quota implementation. *Journal of Economic Theory*, 198:105356, 2021.

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

Richard McKelvey and Thomas Palfrey. Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10:6–38, 1995.

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

Stephen Morris and Ming Yang. Coordination and continuous stochastic choice. *The Review of Economic Studies*, 2021.

William Morrison and Dmitry Taubinsky. Rules of thumb and attention elasticities: Evidence from under-and overreaction to taxes. *The Review of Economics and Statistics*, pages 1–45, 2019.

Rosemarie Nagel. Unraveling in guessing games: An experimental study. *The American economic review*, 85(5):1313–1326, 1995.

Luigi Paciello and Mirko Wiederholt. Exogenous information, endogenous information and optimal monetary policy. *The Review of Economic Studies*, 83:356–388, 2014.

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.

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.

Anthony F Shorrocks. The class of additively decomposable inequality measures. *Econometrica: Journal of the Econometric Society*, pages 613–625, 1980.

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 and Alex Rees-Jones. Attention variation and welfare: theory and evidence from a tax salience experiment. *The Review of Economic Studies*, 85(4):2462–2496, 2018.

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.

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.

Online Appendix - For Online Publication

Appendix A0: Instructions

Individual Decision-Making Experiment

Instructions

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

Please turn off cellular phones now.

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

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

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

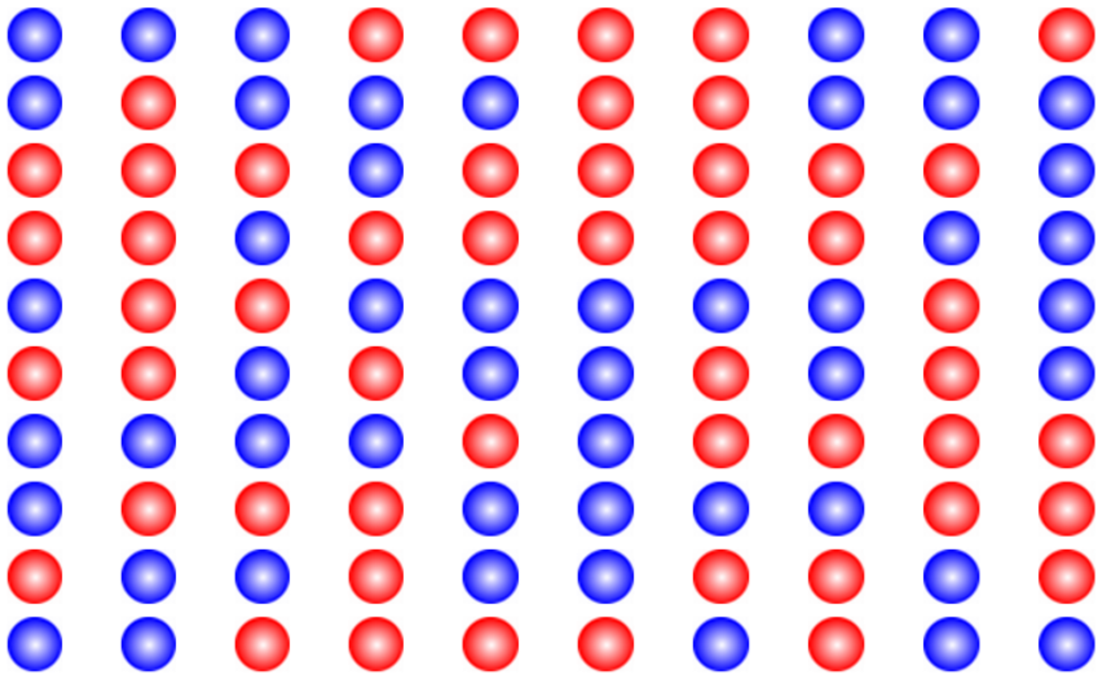
After you have completed the experiment, **please remain quietly seated until everyone has completed the experiment**.

← Home Next →

Individual Decision-Making Experiment

Instructions

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



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

← Previous Next →

Individual Decision-Making Experiment

Instructions

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

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

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

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

← Previous Next →

Individual Decision-Making Experiment

Instructions

Example Question

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

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

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

← Previous   Next →

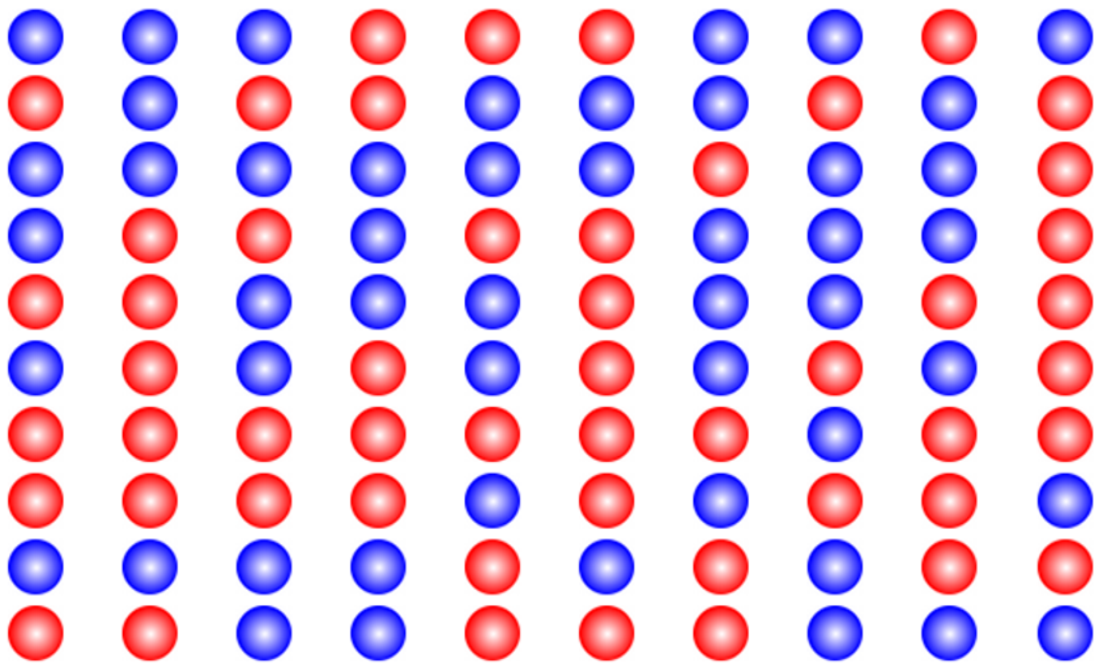
Individual Decision-Making Experiment

Instructions

Example Question

Remember:

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



Please select from the following options:

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

← Previous Next →

Individual Decision-Making Experiment

Instructions

Payment

For this question, you chose the following option:

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

There were 51 red dots on the screen.

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

← Previous Next →

Individual Decision-Making Experiment

Instructions

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

Block 1

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

You will be asked to choose between the following options:

	Option	Pay if there are 49 red dots	Pay if there are 51 red dots
○	A	70	0
○	B	0	70

Block 2

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

You will be asked to choose between the following options:

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

Block 3

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

You will be asked to choose between the following options:

	Option	Pay if there are 49 red dots	Pay if there are 51 red dots
○	A	40	0
○	B	0	40

Block 4

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

You will be asked to choose between the following options:

	Option	Pay if there are 49 red dots	Pay if there are 51 red dots
○	A	95	0
○	B	0	95

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

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

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

← Previous Home →

In experiments in which subjects were paid in probability points the following was added

At the end of the experiment we will randomly select one question you have answered. You will receive points from that question based on your response and the number of red/blue balls in the corresponding image.

Experimental points will give you a chance to earn the prize of $xxx. For every point you receive your chance of receiving the prize will increase by 1%.

For example, imagine that the computer randomly selected the fourth question to reward. You earned 72 experimental points for that question. You would then receive 72 points which would mean you would have a 72% chance of winning the $30.00 prize.

Appendix A1: NIAS and NIAC for Experiment 1

Deriving the NIAS and NIAC Conditions

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

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

Table A1.1 Summarizes these conditions, not all of which will hold simultaneously.

Table A1.1: 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 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

$$\begin{aligned} & G(\mu, \{a, b\}, \pi) + G(\mu, \{a, b, c\}, \hat{\pi}) \\ \geq \; & G(\mu, \{a, b\}, \hat{\pi}) + G(\mu, \{a, b, c\}, \pi) \end{aligned} \tag{10}$$

where π is the revealed information structure 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 generates 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 estimate $G(\mu, \{a, b\}, \pi)$ 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 - i.e. when only a and b are present. 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 γ^a and γ^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 10 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 $\{a,b\}$.

Empirical Results

Table A1.2 reports the results of the NIAS tests for experiment 1 using aggregate data.Estimate for the first row generated by constructing, for each choice and each individual

$$\frac{1(choose\_a).1(\omega_1)}{P(1)}(50 - b_1) + \frac{1(choose\_a).1(\omega_2)}{P(2)}(50 - b_2) - 100 + (b_1 + b_2)$$

where $1(choose\_a)$ is a dummy which takes the value 1 if a is chosen, $1(\omega_i)$ 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 A1.2. P values were found using bootstrapping with standard errors clustered at the individual level.

Data for other rows constructed using the same method.

The first column reports the mean value for the LHS of the tests described in table A1.1. 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 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 A1.2: 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 average value of this expression is 0.234, significantly different from 0 at the 5% level.[50]

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

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

find 6 violations. For NIAC, we find 2 significant violations in 28 tests (note each individual provides a single opportunity to test NIAC, as they face only 2 decision problems).

Appendix A2: Additional Results for Experiments 2 and 3

A2.1 Regression Results for Experiment 2

Table A2.1 Regression results for Experiment 2		
Dummy	OLS	Logit
$\beta_{3,2}$	-0.40	-1.43
	(0.06)	(0.25)
$\beta_{4,1}$	0.02	0.12
	(0.03)	(0.18)
$\beta_{4,2}$	-0.40	-1.71
	(0.06)	(0.29)
$\beta_{5,1}$	0.04	0.20
	(0.03)	(0.19)
$\beta_{5,2}$	-0.40	-1.73
	(0.06)	(0.29)
$\beta_{6,1}$	0.04	0.23
	(0.04)	(0.24)
$\beta_{6,2}$	-0.46	-2.00
	(0.05)	(0.28)
Constant	0.74	1.05
	(0.04)	(0.20)
N	10,400	10,400
Regression of the choice of option a on state and decision problem dummies. Standard errors (in parenthesis) clustered at the individual level		

Table A2.2: NIAC Tests - Experiment 2		
Test	OLS	Logit
$\beta_{3,2} = \beta_{4,1} - \beta_{4,2}$	0.07	0.07
$\beta_{3,2} = \beta_{5,1} - \beta_{5,2}$	0.05	0.06
$\beta_{3,2} = \beta_{6,1} - \beta_{6,2}$	0.00	0.00
$\beta_{4,1} - \beta_{4,2} = \beta_{5,1} - \beta_{5,2}$	0.67	0.63
$\beta_{4,1} - \beta_{4,2} = \beta_{6,1} - \beta_{6,2}$	0.10	0.12
$\beta_{5,1} - \beta_{5,2} = \beta_{6,1} - \beta_{6,2}$	0.21	0.25
F test of the linear hypothesis implied by the NIAC conditions for Experiment 2, performed using the regression coefficients described in table 2.1. Each cell reports the probability associated with the null.		

A2.2 Shannon Tests Excluding Subjects who Violate NIAS and 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 κ using aggregate data , replicating the analysis of figure 3. 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

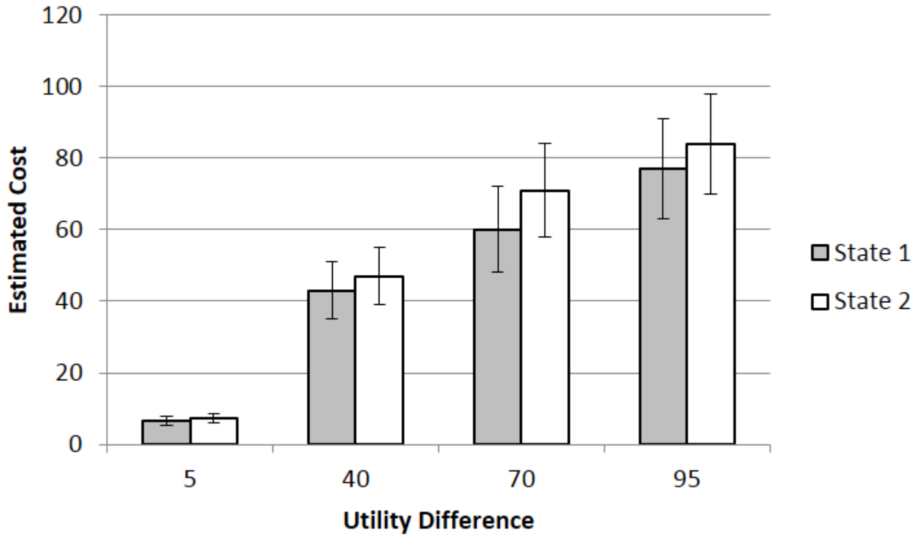


Figure A2.1: Estimated Costs - Consistent Subjects Only

Figure A2.2 replicates the individual level analysis of figure 4. 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 subjects exhibit 'too slow' violations only, 3 exhibit 'too fast violations' only and 21 have examples of neither.

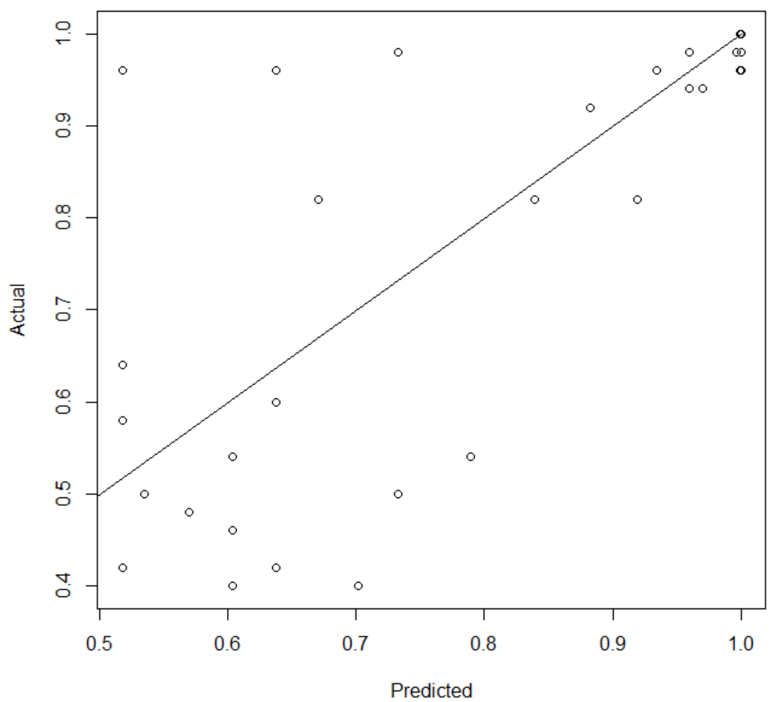


Figure A2.2: Predicted vs actual accuracy in the 70% payoff treatment - consistent subjects only

A2.3 Regression Results for Experiment 3

Table A2.3 Regression results for Experiment 3		
Dummy	OLS	Logit
$\beta_{7,2}$	-0.47 (0.05)	-2.09 (0.24)
$\beta_{8,1}$	0.11 (0.04)	0.76 (0.24)
$\beta_{8,2}$	–0.39 (0.07)	-1.67 (0.32)
$\beta_{9,1}$	0.13 (0.03)	0.95 (0.22)
$\beta_{9,2}$	-0.36 (0.06)	-1.58 (0.299)
$\beta_{10,1}$	0.14 (0.03)	1.09 (0.23)
$\beta_{10,2}$	-0.25 (0.07)	-1.14 (0.31)
Constant	0.77 (0.03)	1.19 (0.18)
N	10,800	10,800
Regression of the choice of option a on state and decision problem dummies. Standard errors (in parenthesis) clustered at the individual level		

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

Appendix A4: Estimation Strategy and Additional Results

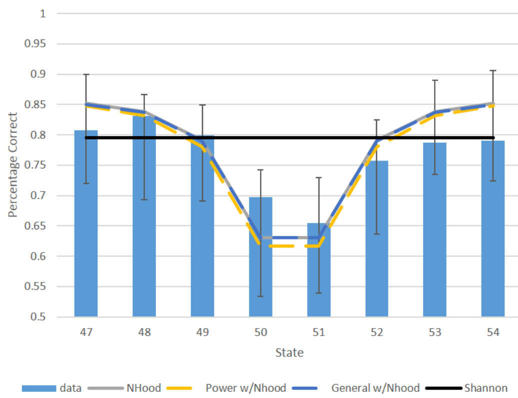
We estimated all maximum likelihood models using two stage numerical optimization. First, we have a function that, for a given set of parameters, finds the conditional choice probabilities by numerically maximizing the expected payoff function for a given choice problem net of the costs of information implied by the conditional choice probabilities and the model parameters. A second function takes the conditional choice probabilities from the first stage optimization and uses them to generate a likelihood for the observed data. This likelihood function is then numerically optimized using the mle function from the "stats4" package in R to find the parameters which best fit the data.

For the individual level model fits for experiments 1 and 3, the likelihood function also included an 'error term': an additional free parameter which represented a player's probability of uniformly randomizing over all available moves. This was done because some of the models generate very stark predictions in these settings with behaviors often being projected to occur with zero probability. In some cases, a model cannot be fit to a data set with finite likelihood regardless of parameters if we do not include this chance of random action.

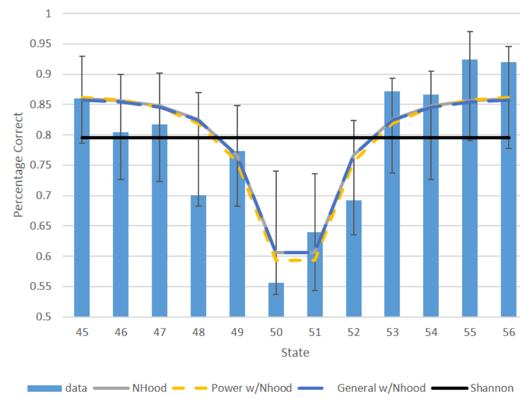
Where applicable, we took advantages of inherent features of the problem to improve the performance of the first stage estimation. For example, in experiment 2 we generated one accuracy parameter for each incentive level rather than separately numerically optimizing a probability of action A given state 1 and a probability of action b given state 2. We also imposed a monotonicity restriction on the predicted accuracy of the neighborhood models in experiment 4, because we know that the accuracy in the true prediction of the model will always be monotonically decreasing as the number of red and blue balls in a state get closer together.

Predicted behaviors were generated by plugging the parameters found in the maximum likelihood solutions back into the first stage optimizing functions. Confidence intervals in all cases were generated by running the same process on a bootstrap resampling of the data. In the aggregate data, errors were clustered at the individual level by resampling individuals rather than single observations. Likelihood ratio tests were done in the standard manner with p-values derived from the asymptotic chi-squared approximation of the distribution of the test statistic. More specifically, we employ the lr.test function from the "exTremes" package.

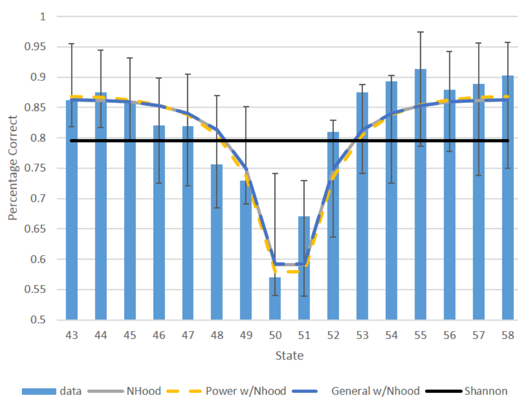
Table A4.1: Estimation Results for Experiment 4: All Treatments



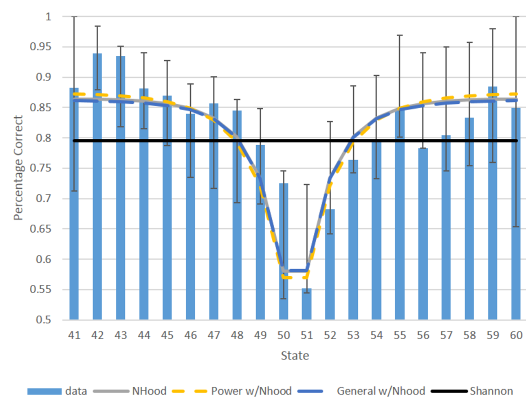
8 States



12 States



16 States



20 States

Table A4.2: Results from Joint Estimation: Experiment 2

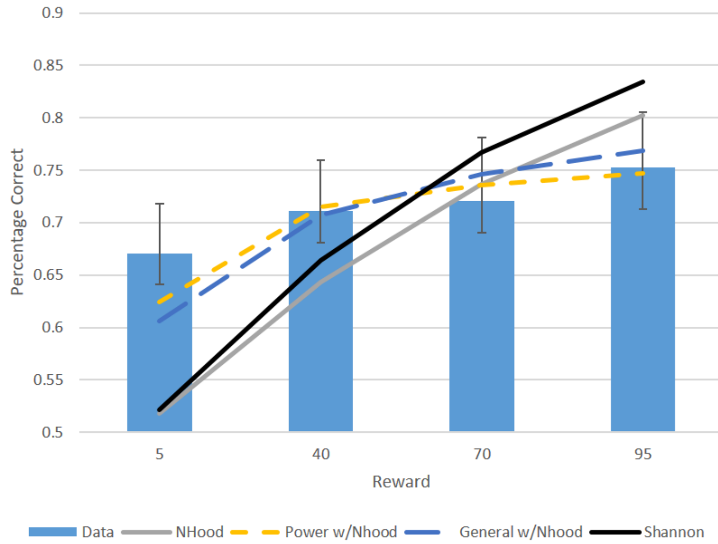
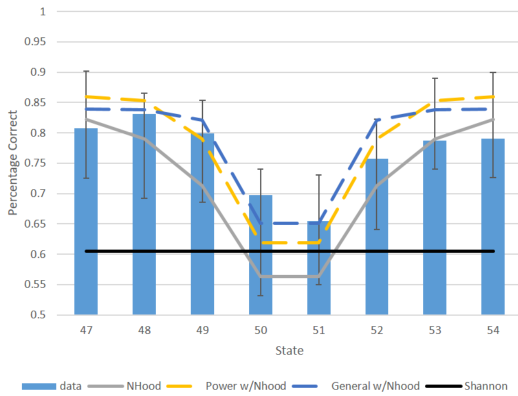
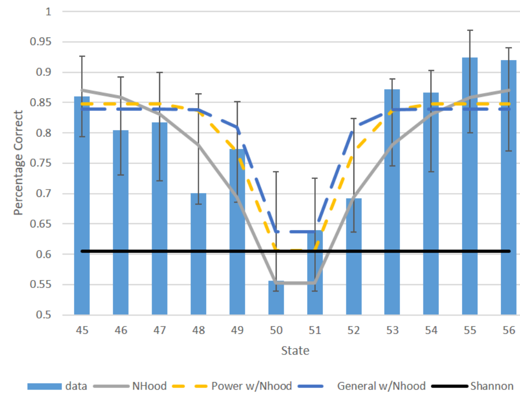


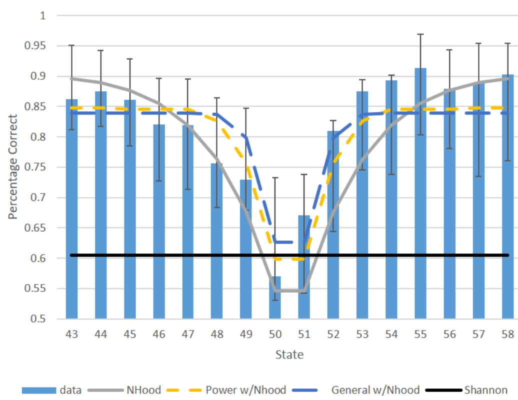
Table A4.3: Results from Joint Estimation: Experiment 4



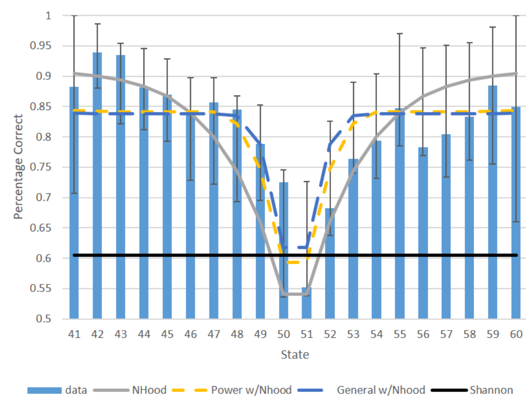
8 States



12 States



16 States

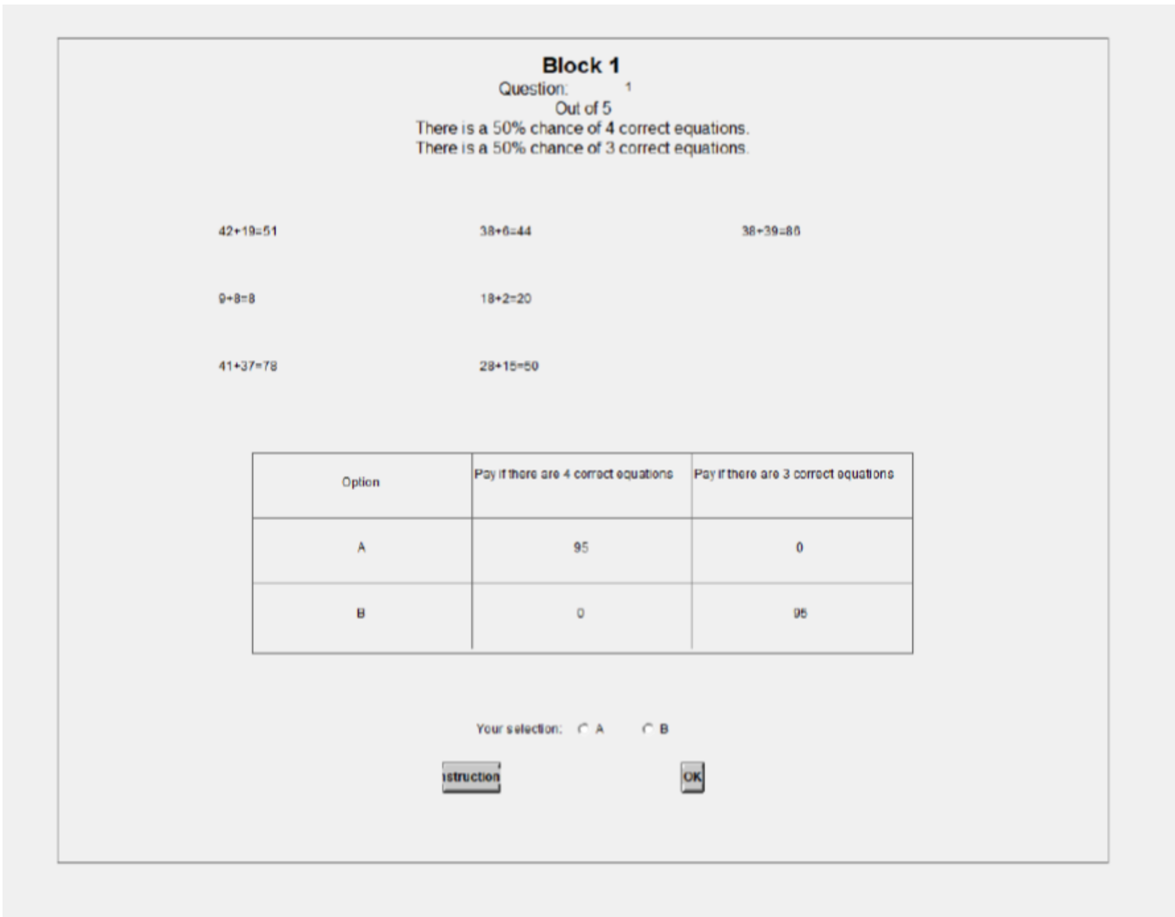


20 States

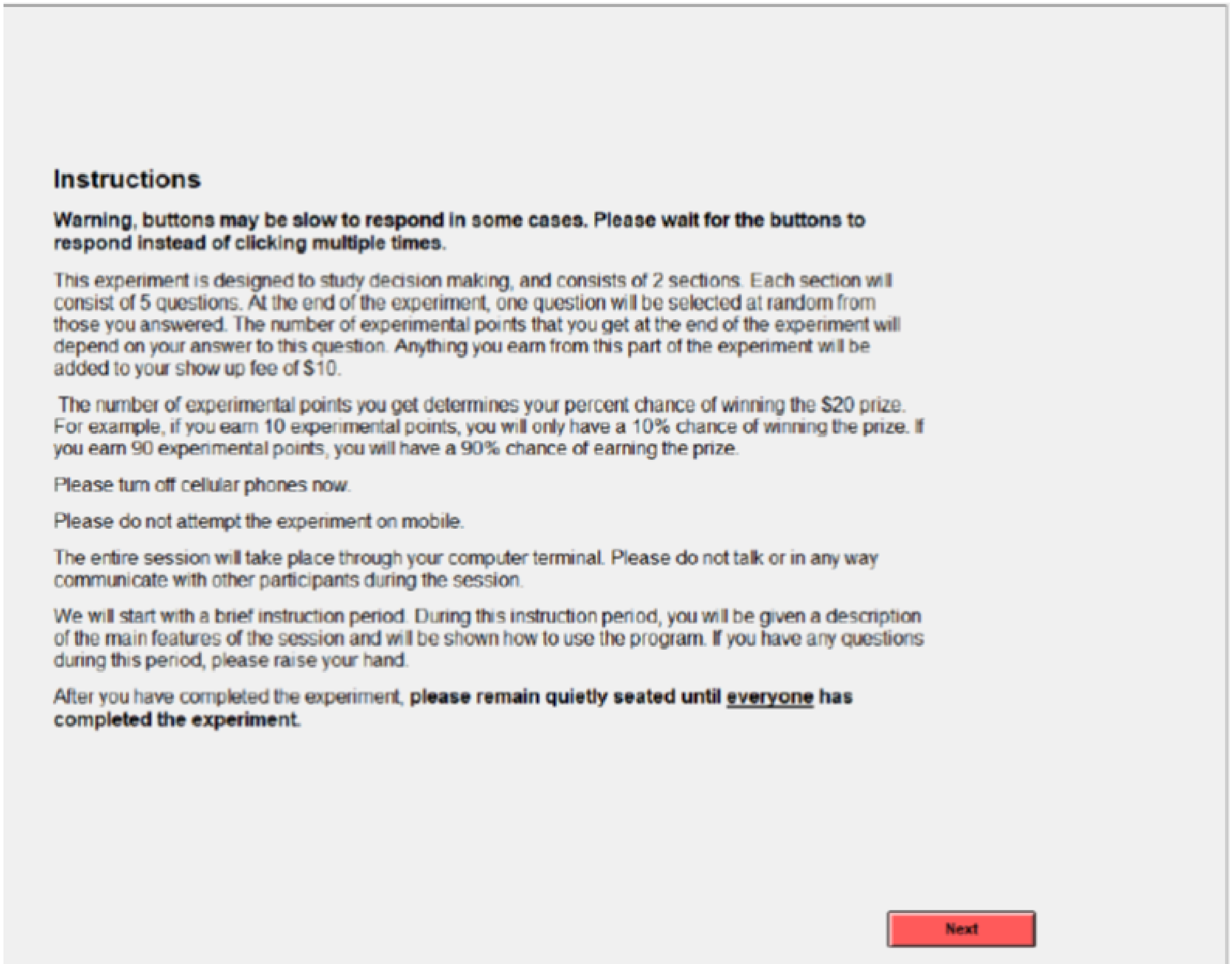
Appendix 5: A Second Experimental Environment

5.1 Experimental Screenshot and Instructions

5.1.1 Example Screenshot



5.1.2 Instructions



Instructions

Warning, buttons may be slow to respond in some cases. Please wait for the buttons to respond instead of clicking multiple times.

This experiment is designed to study decision making, and consists of 2 sections. Each section will consist of 5 questions. At the end of the experiment, one question will be selected at random from those you answered. The number of experimental points that you get at the end of the experiment will depend on your answer to this question. Anything you earn from this part of the experiment will be added to your show up fee of $10.

The number of experimental points you get determines your percent chance of winning the $20 prize. For example, if you earn 10 experimental points, you will only have a 10% chance of winning the prize. If you earn 90 experimental points, you will have a 90% chance of earning the prize.

Please turn off cellular phones now.

Please do not attempt the experiment on mobile.

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

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

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

Next

Instructions

For each question you will be shown 7 equations on a screen. Some of these equations will be correct while some will be incorrect. Here is an example of such a screen.

13+25=38	4+19=23	37+21=56
29+32=58	34+26=60	
17+24=41	6+19=34	

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

Previous | Next

Instructions

You will be asked to make a choice between two or more options. Each of these options will pay out a different number of experimental points, depending on how many correct equations are on the screen. The number of experimental points an option pays out in a state is presented in a grid like the one below.

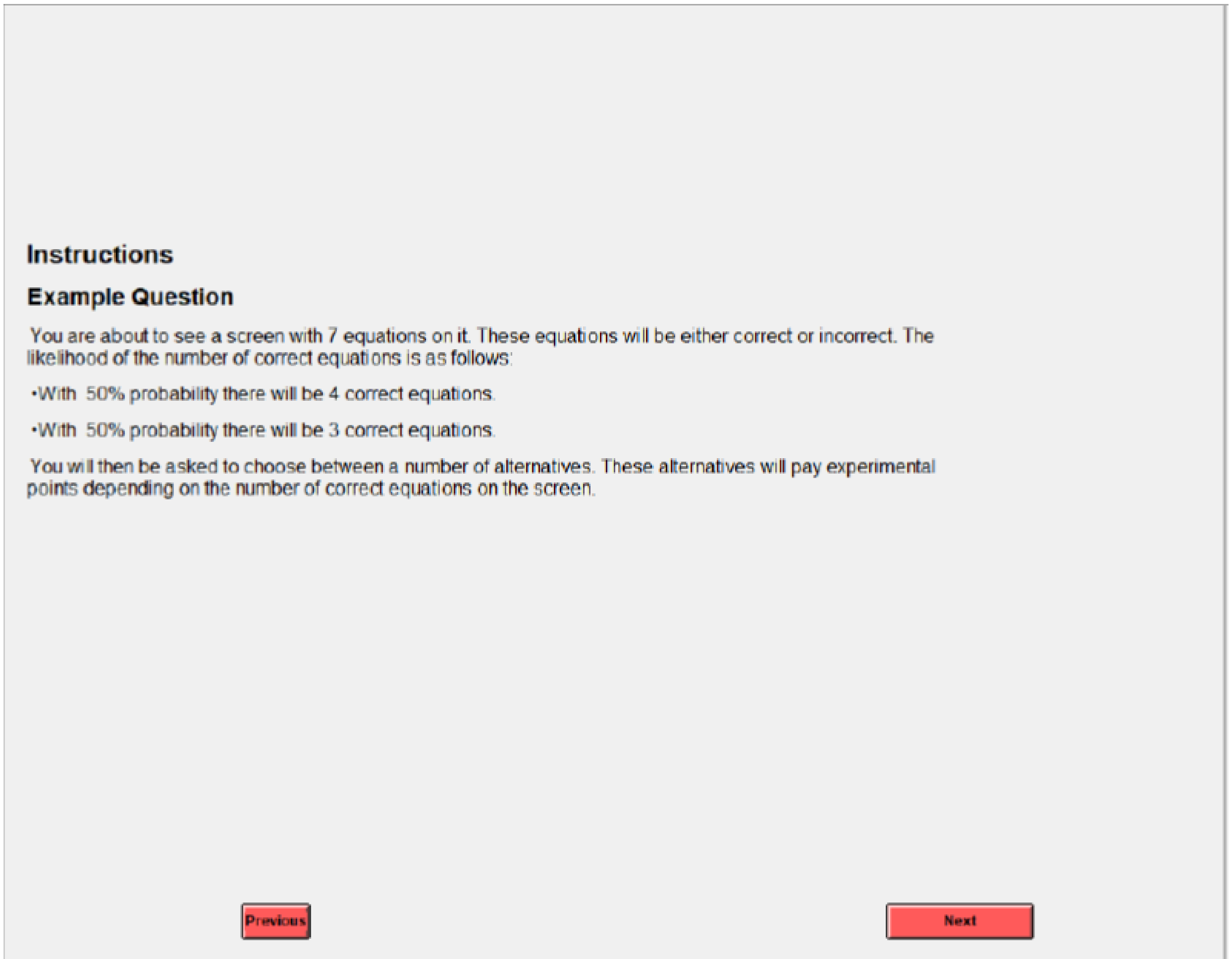
Option	Pay if there are 4 correct equations	Pay if there are 3 correct equations
A	10	0
B	0	10
C	5	5

Your selection: A B C

In this case, if you choose option A (and this question was the one selected for payment) then you would get 10 experimental points if there were 4 correct equations on the screen and 0 experimental points if there were 3 correct equations. If you chose B you would get 10 if there were 3 correct equations on the screen and 0 experimental points if there were 4 correct equations. If you chose option C you would receive 5 experimental points regardless of the number of correct equations on screen.

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

Previous Next


Instructions
Example Question
You are about to see a screen with 7 equations on it. These equations will be either correct or incorrect. The likelihood of the number of correct equations is as follows:
•With 50% probability there will be 4 correct equations.
•With 50% probability there will be 3 correct equations.
You will then be asked to choose between a number of alternatives. These alternatives will pay experimental points depending on the number of correct equations on the screen.
Previous
Next

Instructions

Example Question

Remember:

•With 50% probability there will be 4 correct equations.

•With 50% probability there will be 3 correct equations.

13+25=38	4+19=23	37+21=66
29+32=58	34+26=60	
17+24=41	6+19=34	

Options	Pay if there are 4 correct equations	Pay if there are 3 correct equations
A	10	0
B	0	10
C	5	5

Your selection: ○ A ○ B ○ C

Previous Next

Instructions

Payment

For this question, you chose the folloing option.

Option	Pay if there are 4 correct equations	Pay if there are 3 correct equations
C	5	5

There were 4 correct equations on the screen.

If this were the question that had been selected for payment, you would have received 5 experimental points in addition to your show up fee.

Previous

Next

Instructions

Here is a description of the questions that you will face in each of the 2 sections of the experiment.

Block 1

- With 50% probability there will be 4 correct equations
- With 50% probability there will be 3 correct equations
- You will be asked to choose between the following options.

Option	Pay if there are 4 correct equations	Pay if there are 3 correct equations
A	95	0
B	0	95

Block 2

- With 50% probability there will be 4 correct equations
- With 50% probability there will be 3 correct equations
- You will be asked to choose between the following options.

Option	Pay if there are 4 correct equations	Pay if there are 3 correct equations
A	5	0
B	0	5

Previous

Finish

5.2 Additional Results from Experiment 2.2

Table A5.1 Regression results for Experiment 2.2		
Parameter	OLS	Logit
$\beta_{2.5,2}$	-0.65	-3.11
	(0.04)	(0.31)
$\beta_{2.6,1}$	0.00	-0.01
	(0.02)	(0.17)
$\beta_{2.6,2}$	-0.69	-3.37
	(0.04)	(0.28)
Constant	0.82	1.54
	(0.03)	(0.19)
N	5,500	5,500
Regression of the choice of option a on state and decision problem dummies. Standard errors (in parenthesis) clustered at the individual level		

Table A5.2: NIAS Test from Experiment 2.2			
DP	$P_j(a\|1)$	$P_j(a\|2)$	Prob
2.3	0.82	0.17	0.00
2.4	0.82	0.14	0.00

Table A5.3 Individual Level Data from Experiment 2.2	
Violate	%
NIAS Only	0
NIAC Only	9
Both	0
Neither	91

5.3 Additional Results from Experiment 2.3

Table A5.4 Regression results for Experiment 2.3		
Parameter	OLS	Logit
$\beta_{2.5,2}$	-0.69	-3.45
	(0.04)	(0.34)
$\beta_{2.6,1}$	0.08	1.27
	(0.02)	(0.27)
$\beta_{2.6,2}$	-0.59	-2.88
	(0.06)	(0.39)
Constant	0.88	2.01
	(0.02)	(0.23)
N	5,300	5,300
Regression of the choice of option a on state and decision problem dummies. Standard errors (in parenthesis) clustered at the individual level		