

Experimental Tests of Rational Inattention*

Mark Dean[†] and Nathaniel Neligh[‡]

September 2018

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 based on 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

Attention is a scarce resource. The impact of attentional limits has been identified in many important economic settings.¹ This has lead to a wide range of modeling approaches aimed

*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 This paper reports on experiments similar to those in the unpublished working papers Caplin and Dean [2013] and Caplin and Dean [2014] and subsumes those parts of that paper that are common.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

[‡]Economic Science Institute, Chapman University, neligh@chapman.edu

¹For example, shoppers may buy unnecessarily expensive products due to their failure to notice whether or not sales tax is included in stated prices (Chetty *et al.* [2009]). Buyers of second-hand cars focus their

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

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

In this paper we use a sequence of four laboratory experiments to provide answers to these questions. The basic set up is a simple information acquisition task in which subjects are presented with a number of balls on the screen which can either be red or blue. They must then choose between different actions, the payoff of which 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 setup the four experiments vary different features of the decision problem, including the range of available actions, the value of correct choice and the prior probability of possible states. By repeatedly exposing the subject to each decision problem we can collect ‘state dependent stochastic choice’ (SDSC) data, or the probability that each action is chosen in each state. Such data is particularly useful for testing models of rational inattention, and learning about attention cost (see Caplin and Dean [2015]).

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

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]).

²We 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 this term 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’.

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

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

Complementary to this agnostic approach to attention costs, we also ask whether our data is in line with specific cost functions. Here we initially focus on costs which are linear in Shannon mutual information (henceforth the Shannon model). Mutual information was introduced to the economics literature as a measure of attention costs by Sims [2003], and has been justified on information theoretic grounds, related as it is to the average number of signals needed to generate a given set of posterior beliefs (see Sims [2003]). The resulting model has proved popular in theoretical and applied research.⁴ Costs which are linear in mutual information have received particular attention due to the link between resulting behavior and the Logit model of discrete choice (see Matejka and McKay [2015]). Compared to the general model, the Shannon model is extremely restrictive, with only a single parameter related to the marginal cost of information. Recent works by Matejka and McKay [2015] and Caplin *et al.* [2017]⁵ have highlighted a number of behavioral regularities implied by the Shannon model: The Locally Invariant Posteriors (LIP) and Invariant Likelihood Ratio (ILR) conditions restrict how optimal information acquisition responds to changes in prior beliefs and incentives respectively. The Shannon model also implies that behavior is invariant to the addition or subtraction of states which are identical in terms of the payoffs of all available actions - a property labeled Invariance Under Compression (IUC) by Caplin *et al.* [2017]. This property means the model is incommensurate with any notion of ‘perceptual distance’, by which some states are easier to differentiate than others.

Our sequence of experiments is designed to first determine whether subjects do actively adjust their attention, (experiment 1), then to test key features of the general and Shannon models, as outlined above. Each of the experiments 2-4 varies a basic property the decision problem (respectively incentives, prior beliefs and number of states), and so together provide a rich data set for studying attentional responses as well as providing simple tests of each of the key features of the Shannon model (respectively ILR, LIP and IUC). This structure allows us to take a ‘general to specific’ approach to inattention, by first identifying active adjustment of attention, then asking whether these changes are commensurate with any model in the rational inattention class, before finally asking in what way the behavior is

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

⁵See also Caplin and Dean [2013].

consistent with the Shannon model.

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

Experiment 2 examines the response of information acquisition to incentives using a simple two state/two action design in which we vary the benefit from choosing the correct action. This provides a simple test of the NIAC and NIAS conditions which characterize the general model, in an admittedly undemanding setting. The aggregate data strongly supports both conditions, and at the individual level 81% of subjects exhibit no significant violation of either. This implies that there is a notion of information costs which rationalizes the behavior of the majority of subjects. We find a high degree of heterogeneity in these costs across subjects. Both in the aggregate and at the individual level we find that subjects typically respond less strongly to incentives than predicted by Shannon costs.

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

Our final experiment tests the IUC condition of the Shannon model using a decision problem with multiple states but only two actions to choose from, the payoffs of which are the same in many different states. The Shannon model implies that behavior should be the same in all states which are payoff equivalent. This is incommensurate with the idea of a perceptual distance, by which some states are easier to differentiate between than others. We show that, in our experimental set up, IUC performs poorly: subjects are better at differentiating between states which are further apart (e.g. 60 vs 50 red balls) than those that are closer together (e.g. 51 vs 50 red balls).

Our experiments uncover behavior that is rationally inattentive but not qualitatively consistent with Shannon costs for at least two reasons: subjects respond too slowly to incentives and are affected by perceptual distance. In section 6 we fit to our data a variety of models that have the potential to solve these problems, in order to determine if they provide a quantitatively better fit of our subject’s choices. In order to better model responses to incentives we take two approaches. First, we consider a cost function from the class of ‘posterior separable’ models, introduced by Caplin and Dean [2013], which retain from the Shannon model the feature that costs are based on the expected change in some convex function of beliefs, but relax the assumption that this function must be Shannon entropy.⁶ Second we consider models in which Shannon costs are raised to a power. In order to tackle the issue of perceptual distance we make use of the ‘neighborhood’ structure of Hébert and Woodford [2017], by which each state is assigned to one or more neighborhoods, and the cost of differentiating between states depends on the number of neighborhoods they share. Our results support a model which makes use of a neighborhood structure and either generalized entropy (Shorrocks [1980]), or a nonlinear function of mutual information as a basis of costs.

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.* [2016] 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. Contemporaneous to this paper, Ambuehl *et al.* [2018] 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. Pinkovskiy [2009] and Cheremukhin *et al.* [2015] fit the Shannon model using data on stochastic choice between lotteries, but do not test the sharp behavioral predictions from that model as we do here. Bartoš *et al.* [2016] report the results of a field experiment which supports rationally inattentive behavior in labor and housing markets, 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

⁶Models in this class have recently been used for several economic applications - see for example Clark [2016], Morris and Strack [2017] and Hébert and Woodford [2017].

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 6.

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

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

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. The rational inattention model assumes that the DM solves this trade off optimally.

In each decision problem, the DM chooses an *information structure*: a stochastic mapping from objective states of the world to a set of subjective signals. While this formalization sounds somewhat abstract, it 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 $\pi : \Omega \rightarrow \Delta(\Gamma)$ that have finite support $\Gamma(\pi) \subset \Gamma$ and that satisfy Bayes' rule, meaning that for all $\omega \in \Omega$ and $\gamma \in \Gamma(\pi)$,

$$\gamma(\omega) = \Pr(\omega|\gamma) = \frac{\Pr(\omega \cap \gamma)}{\Pr(\gamma)} = \frac{\mu(\omega)\pi(\gamma|\omega)}{\sum_{v \in \Omega} \mu(v)\pi(\gamma|v)},$$

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,

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

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

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

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

Our assumptions on the data mean that G is observable but K is not. We use the 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*, which can be recovered from the data by assuming that each chosen action is associated with exactly one signal.⁷ Caplin and Dean [2015] provide a formal definition of the revealed information structure, but essentially it assumes that the DM used an information structure which consists of the posteriors described in equation 1 for each chosen act, with the probability of receiving that posterior given by the (unconditional) probability of choosing the associated act. 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.

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

We 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. These include those in which agents can either pay to receive all information or learn nothing,⁸ those that apply additive normal noise to the agent's information and then allows them to pay a cost to decrease the variance of that noise,⁹ in which the decision maker chooses a partition structure on the state space,¹⁰ or in which information is free up to a hard capacity constraint.¹¹ See Caplin and Dean [2015] for a discussion. The only substantive assumption is that the objective function is additively separable between gross utility and costs (see Chambers *et al.* [2018] 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. Stochasticity therefore derives from changes in preferences, rather than noise in the perception of the state of the world.

Typically the RUM has not been applied to situations in which there is an objective, observable state of the world, and there are many possible ways that the model could be adapted to such a situation.¹² 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

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

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

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

¹¹For example Sims [2003]

¹²For 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.

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 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 The Shannon Model

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

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

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

Note that we focus on the case in which costs are linear in Shannon mutual information. In section 5 we discuss costs which are nonlinear transforms of mutual information. An alternative mode 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.¹⁴

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

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

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

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

The Locally Invariant Posterior (LIP) property states that local changes in prior beliefs do not lead to changes in optimal posterior beliefs.¹⁵ Specifically, if, for some decision problem

¹³Recall that we identify a signal with its resulting posterior distribution.

¹⁴We 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 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.

¹⁵Again, see Caplin and Dean [2013] and Caplin *et al.* [2017] for further details.

(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) . This property holds for the broader class of posterior separable cost functions we discuss in section 5. We will test this proposition in experiment 3.

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 for all acts. Caplin *et al.* [2017] show that this ‘Invariance Under Compression’ property fully characterizes the Shannon model within a the broader class of posterior separable models described in section 5. 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.

3 Experimental Design

3.1 Set Up

We now introduce our experimental design which 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 of the world is determined by the number of red balls on the screen. Prior to seeing the screen, subjects are informed of the probability distribution over such states. Having seen the screen, they choose from a number of different actions whose payoffs are state dependent. As in the theory, a decision problem (DP) is defined by the prior distribution and the set of available actions. Figure A0.1 in the appendix shows a typical screenshot from the experiment.

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 exter-

nally 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 6 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.¹⁶

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

¹⁶Subjects 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.

revealed posteriors γ^a are sufficient.

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

3.2 Experiment 1: Testing for Responsive Attention

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

Table 1: Experiment 1						
DP	Payoffs					
	$U(a, 1)$	$U(a, 2)$	$U(b, 1)$	$U(b, 2)$	$U(c, 1)$	$U(c, 2)$
1	50	50	b_1	b_2	n/a	n/a
2	50	50	b_1	b_2	100	0

The structure of the two DPs is as follows. There are two equally likely states of the world - 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 2.¹⁷

Table 2: Treatments for Experiment 1		
Treatment	Payoffs	
	b_1	b_2
1	40	55
2	40	52
3	30	55
4	30	52

The incentive for gathering information in DP 1 is low. The subject can simply choose a

¹⁷We use multiple values in order to explore the parameter space somewhat - a priori we did not know the values of b_1 and b_2 that would generate violations of monotonicity.

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

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

Experiment 1 also provides a first test for the NIAS and NIAC conditions which characterize the general model. In the interests of brevity, we relegate a discussion of these tests to appendix A1.

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 each DP. Again, states 1 and 2 were represented by 49 and 51 red balls respectively.

Table 3: Experiment 2				
DP	Payoffs			
	$U(a, 1)$	$U(a, 2)$	$U(b, 1)$	$U(b, 2)$
3	5	0	0	5
4	40	0	0	40
5	70	0	0	70
6	95	0	0	95

The primary aim of this experiment is to provide estimates of the cost function associated

with information acquisition. However, in order for this to be meaningful it must be the case that behavior is rationalizable with some underlying 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 - which guarantees that subjects are using the information they have efficiently - requires that

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

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). If and only if this condition holds then a (resp. b) is the optimal choice of action given the posterior probabilities over states when a (b) was chosen. See Caplin and Dean [2015] section E for the derivation of the NIAS and NIAC conditions for experiments 2 and 3.

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

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

This condition guarantees that gross payoff cannot be increased by reallocating information structures across decision problems, and so there exists some cost function which rationalizes behavior.

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

$$\begin{aligned} \kappa &= \frac{\ln(\gamma_3^a(1)) - \ln(\gamma_3^b(1))}{5} = \frac{\ln(\gamma_4^a(1)) - \ln(\gamma_4^b(1))}{40} \\ &= \frac{\ln(\gamma_5^a(1)) - \ln(\gamma_5^b(1))}{70} = \frac{\ln(\gamma_6^a(1)) - \ln(\gamma_6^b(1))}{95} \end{aligned} \tag{2}$$

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)¹⁸ and two acts. Again there are 4 decision problems, each of which is repeated 50 times. Because this experiment made use of only two payoff levels, payment was made in cash, rather than probability points. Table 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

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

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

A natural alternative model is one of base rate neglect (see for example Tversky and Kahneman [1974]), in which subjects ignore changes in prior probabilities when assessing

¹⁸We 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.

alternatives. A DM who ignored the impact of changing priors on their posterior would be in danger of violating NIAS as $\mu(1)$ increases.

In contrast, the Shannon model 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.¹⁹ 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 trials.

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.²⁰

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.²¹ 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 $\{1, \dots, \frac{N}{2}\}$ and zero otherwise, while action b pays off \$10 in states $\{\frac{N}{2} + 1, N\}$ 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.

This illustrates an important behavioral feature of the Shannon model, which is that it lacks any sense of ‘perceptual distance’, by which some states are harder to distinguish than others. According to the Shannon model, subjects are no more likely to make mistakes in

¹⁹Of course, for Bayes’ rule to hold, it must also be the case that the unconditional probability of choosing option a increases.

²⁰We thank Sandro Ambuehl for this suggestion.

²¹This experimental design was developed a part of a distinct project on information acquisition in global games. See Dean *et al.* [2016].

states that are close to the threshold of $\frac{N}{2}$ than those that are far away.

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).²²

4 Implementation and Results

Subjects were recruited from the New York University and Columbia University student populations.²³ 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

²²A previous version of the paper reported the results of another treatment in which the state of the world was determined by the number of letters on the screen. We omit these results for brevity.

²³Using the Center for Experimental Social Science subject pool at NYU and the Columbia Experimental Laboratory in the Social Science subject pool at Columbia.

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 observing such data should fall as more data is collected, and so power increases.²⁴ 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 hypotheses above are defined in terms of inequalities. This is typically the case for the tests we employ. When testing against a null hypothesis which encompasses an entire region of the parameter space, there are two possible approaches. The Bayesian approach is to assign some prior to the parameter space and then update it using the data. The null is rejected if 95% of the posterior weight falls outside the null region. The frequentist approach simply treats the null hypothesis as a single point hypothesis placed at the location in the null region which is the most favorable to the null hypothesis. In this paper we will use this approach - so, in the case of Monotonicity, we will 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. 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,

²⁴ Assuming that the values are not equal in the population.

standard errors are corrected for clustering at the subject level.

4.2 Experiment 1: Testing for Responsive Attention

We now describe the results of our first experiment. Table 5 summarizes the results of the Monotonicity tests from experiment 1. The first panel reports the probability of choosing action b in state 1 with and without c available - aggregated across all subjects - and the p-values from a statistical test to determine whether the latter is significantly higher than the former. The second panel repeats the exercise for $P(b|2)$. The final column reports the fraction of subjects who show a significant violation of Monotonicity at the 5% level. 28 subjects took part in this experiment, evenly divided across the 4 treatments.²⁵

Table 5: Results of Experiment 1 ²⁶								
		$P(b 1)$			$P(b 2)$			
Treat	N	$\{a, b\}$	$\{a, b, c\}$	Prob	$\{a, b\}$	$\{a, b, c\}$	Prob	% Subjects
1	7	2.9	6.8	0.52	50.6	59.8	0.54	29
2	7	5.7	14.7	0.29	21.1	63.1	0.05	43
3	7	9.5	5.0	0.35	21.4	46.6	0.06	29
4	7	1.1	0.8	0.76	19.9	51.7	0.02	57
Total	28	4.8	6.6	0.52	28.3	55.6	<0.01	39

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

Results for NIAS and NIAC tests are reported in appendix A1. Broadly speaking, both conditions are supported by the data from experiment 1.

²⁵One 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.

²⁶P values from a logit regression of the choice of option b on dummies representing whether or not c was present and whether the state was 1 or 2. Standard errors clustered at the individual level.

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.²⁷

We begin by testing the NIAS and NIAC conditions which are necessary and sufficient for the general model. 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. The aggregate data firmly supports the NIAS condition.

Table 6: NIAS Test from Experiment 2 ²⁸			
DP	$P_j(a 1)$	$P_j(a 2)$	Prob
3	0.74	0.40	0.00
4	0.76	0.34	0.00
5	0.78	0.33	0.00
6	0.78	0.28	0.00

Figure 1 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. Most of the differences between DPs are significant at the 10% level.²⁹

²⁷Three 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.

²⁸Results of a logistic regression of choice of action a on a dummy for state 1. P value reported is that associated with the state 1 dummy. Standard errors clustered at the subject level.

²⁹Standard errors produced using a logit regression of correct choice on treatment, with standard errors clustered at the individual level. The success rate at 5 probability points is significantly different from that at 95 points at <0.1%, and different between 40 and 70 points at 10%. Behavior at 40 probability points is significantly different from 95 points at 10%. 70 probability points is not significantly different from either 40 or the 95 points.

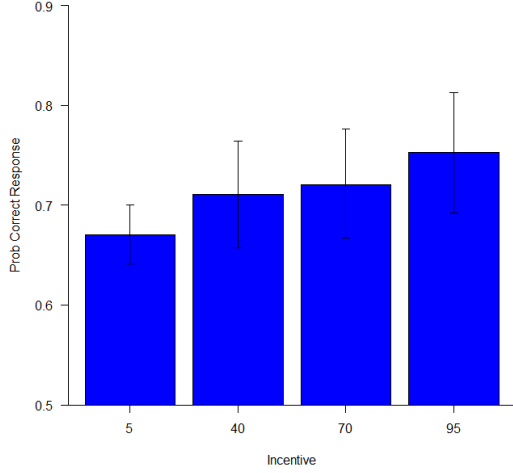


Figure 1: Probability of Correct Response by Decision Problem

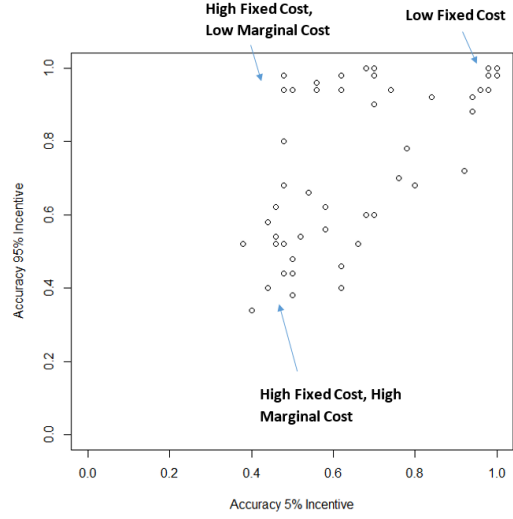


Figure 2: Individual Accuracy 5 Point vs 95 Point Reward

Table 7 reports the individual level data, and in particular the fraction of subjects who exhibit significant violations of the NIAS condition, the NIAC condition, both or neither. 81% of subjects show no significant violations of either condition.³⁰

Table 7: Individual Level Data from Experiment 2	
Violate	%
NIAS Only	2
NIAC Only	17
Both	0
Neither	81

³⁰We checked the NIAC condition and the NIAS conditions separately for each individual. The NIAS condition was tested by simply estimating a logit model regressing probability of choice of action a on state. If the coefficient was significantly negative that is considered a significant violation of NIAS.

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

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 2 gives some idea of the heterogeneity of those costs across subjects. It graphs the probability of choosing the correct action at the 5 point level vs the 95 point level for each subject. The fact that most points fall above the 45 degree line is the defining feature of rational inattention. However, within this constraint there is still a great deal of variation. Our data set includes ‘high fixed cost, high marginal cost’ individuals who gather little information regardless of reward: their accuracy is near 50% for the low and high reward levels. It also includes ‘low fixed cost’ subjects who have accuracy close to 100%, even in the low reward decision problem. Finally there are ‘high fixed cost, low marginal cost’ subjects, who actively adjust their accuracy as a function of reward.

We next examine the extent to which subjects behave as if their costs are in line with the Shannon model. Figure 3 shows the estimated cost parameter κ from each decision problem and in each state using aggregate data, based on the identity from equation (2). The Shannon model predicts that these should be equal. As we can see this is not the case: estimated costs are increasing in reward level: 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.³¹

³¹Cost estimates and standard errors based on an OLS regression of response on treatment and state dummies, which were then converted into cost estimates using the method discussed in section 3.3. Standard errors are clustered at the individual level.

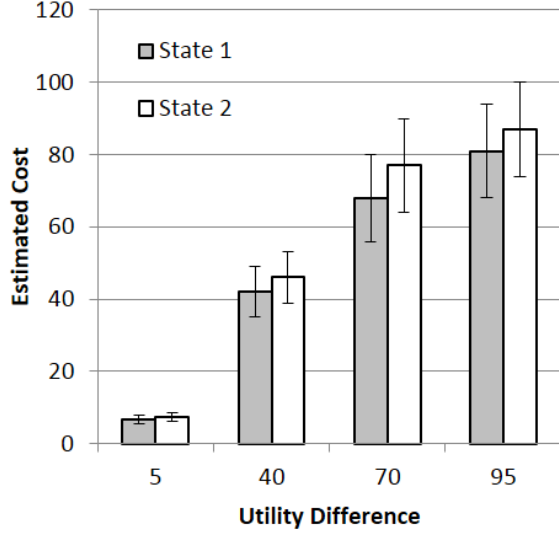


Figure 3: Estimated Costs for the Shannon Model

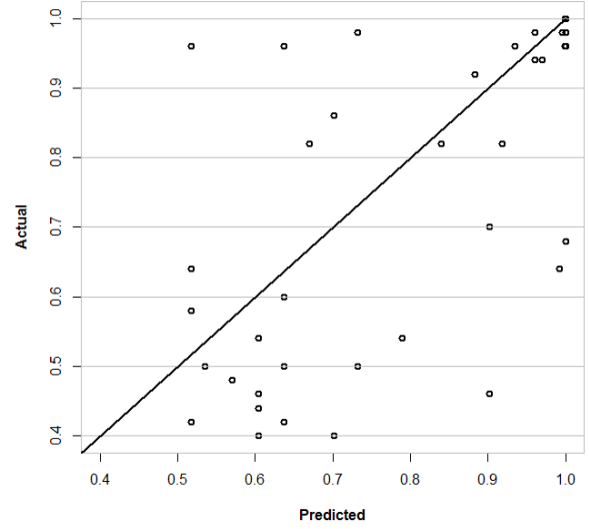


Figure 4: Predicted vs actual accuracy in the 70% payoff treatment

At the individual level we also see significant violations of the Shannon model. Figure 4 shows a scatter plot of the predicted vs actual accuracy for each subject in the 70 point DP, where the predictions are made using the Shannon model and the accuracy displayed at the 40 point level.³² The scatter plot shows more subjects below the 45 degree line (i.e. are less accurate than predicted) than above (more accurate than predicted).³³

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).³⁴ Of the 221 possible comparisons, we find 66 violations of the ‘too slow’ variety

³²We 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.

³³For 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.

³⁴For each person and each incentive level pair we regress correctness on incentive level and a dummy for the higher incentive level with no constant using a logit regression. Note that a logit regression of correctness on incentive level with no constant is equivalent to fitting a Shannon model in this case. Significant coefficients on the high incentive dummy mean significant violations of Shannon. Positive coefficients mean that accuracy

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 experiment 3. We first examine the extent to which the 54 subjects³⁵ 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. Table 8 indicates that subjects do on average change their behavior in response to changing priors, and that NIAS broadly 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.³⁶ 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.

Table 8: NIAS Test ³⁷				
DP	$P_j(a 2)$	Constraint on $P_j(a 1)$	$P_j(a 1)$	Prob
7	0.29	0.29	0.77	0.000
8	0.38	0.39	0.88	0.000
9	0.40	0.80	0.90	0.045
10	0.51	0.91	0.91	0.538

is responding too fast while negative coefficients mean it is responding too slow.

³⁵Data 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).

³⁶0% of individuals exhibit significant violations of NIAS at the 50% prior, 2% at the 60% and 75% priors and 11% at the 85% prior.

³⁷Tests based on an OLS regression of choice of action on state for each treatment to obtain estimates of $P_j(a|2)$ and $P_j(a|1)$. Standard errors clustered at the individual level. These estimates are then used in a test of the linear restriction implied by the NIAS model.

We next study the degree to which our data supports the predictions of the Shannon model in the form of the LIP condition. In order to do so, 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 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

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 also report results based on subjects whose posteriors are significantly above or below the relevant thresholds at a 5% level.³⁸

The first prediction of the posterior separable model is that, in DP i with prior $\mu_i(1)$, subjects who use a posterior $\gamma_7^a(1) < \mu_i(1)$ should exclusively choose action a , while those with $\gamma_7^a(1) > \mu_i(1)$ should choose both a and b , where γ_7^a refers to the posteriors revealed in DP 7 given the choice of a . Table 10 tests this ‘no learning’ prediction. The top panel divides subjects into those who have a threshold (i.e. point estimate of posterior belief from DP 7) above $\mu_i(1)$, and those for whose threshold is below $\mu_i(1)$ for $\mu_8(1) = 0.6$, $\mu_9 = 0.75$ and $\mu_{10} = 0.85$. For each of these decision problems, and each of these groups, it then reports the fraction of subjects who exclusively choose a . The second panel repeats the exercise but

³⁸For 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.

for subjects whose posteriors were significantly above or below the relevant threshold.

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
Point estimates	$\gamma_7^a(1) < \mu_i(1)$	35%	27%	29%
	$\gamma_7^a(1) \geq \mu_i(1)$	0%	7%	13%
Significant differences	$\gamma_7^a(1) < \mu_i(1)$	33%	46%	41%
	$\gamma_7^a(1) \geq \mu_i(1)$	3%	10%	14%

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, around 30% of the subject 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 . Looking at subjects whose point estimates are significantly different from the relevant threshold the condition does somewhat better.

The second part of the LIP condition states that, in each DP, subjects who are still actively gathering information³⁹ should use the same posteriors as they did in DP 7. Figure 5 tests this hypothesis. Panel a focusses on DP 8. It reports data exclusively on subjects who should be choosing both a and b in this DP according to the posterior separable model (i.e. those for whom $\gamma_7^a > 0.6$, 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. The LIP prediction is that these posteriors should be the same. Panels b and c repeat this analysis for DPs 9 and 10. Figure 5 shows some, but imperfect support for the LIP prediction: of the 6 comparisons, only one (the posterior following choice of b with the 0.75 prior) shows a significant difference at the 10% level.⁴⁰

³⁹i.e. those for whom their posterior beliefs from DP 7 are significantly above the prior in that DP.

⁴⁰Tests based on an OLS regression of choice of state on action for each treatment to obtain estimates of $P_j(1|a)$ and $P_j(2|b)$ as functions of the coefficients. Standard errors clustered at the individual level. Standard errors for the conditional probabilities were derived using the delta method.

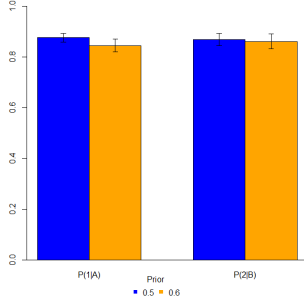


Figure 5a: Subjects with threshold above 0.6

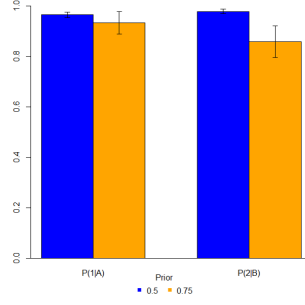


Figure 5b: Subjects with thresholds above 0.75

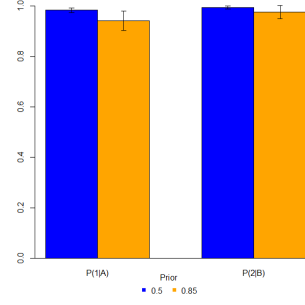


Figure 5c: Subjects with threshold above 0.85

4.5 Experiment 4: Symmetry

23 subjects took part in experiment 4.⁴¹ The results are summarized in figure 6, which shows the probability of choosing the correct action as a function of the state for each DP and each treatment.

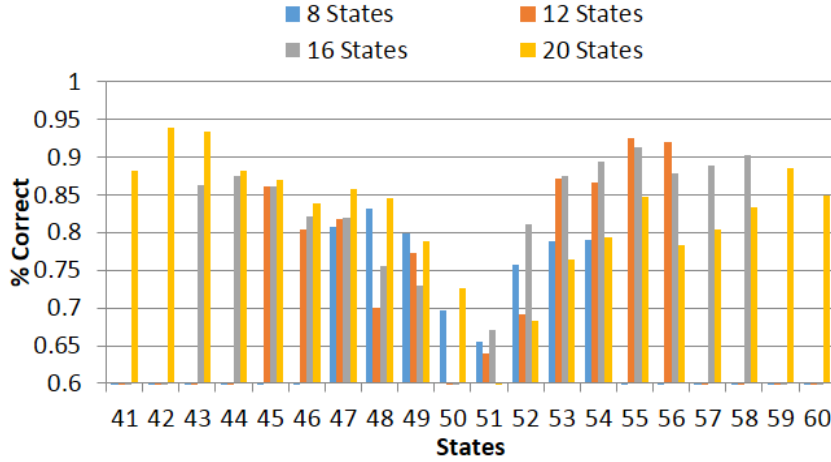


Figure 6: Experiment 4

This figure show clear and systematic violations of symmetry: subjects were more likely to

⁴¹Data from a single session which took place on 27th June 2013 at the CESS laboratory at NYU.

make mistakes in states near the threshold of 50. This observation is confirmed by regression analysis, which finds a significant and positive correlation between distance from threshold and probability of correct responses for each DP.⁴²

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) that the Shannon model has some significant difficulties in explaining our data, both in terms of the relationship it predicts between changing rewards and information gathering, and its unrealistic symmetry properties.

In this section we discuss some of the issues which could effect these conclusions. In particular, could aggregation and order effects be responsible for some of the results we find, and so be the reason we have rejected 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 (in extremis, all the red balls could appear on the left of the screen while all the blue ones appear on the right). Aggregating across individuals may also cause problems, because different individuals may have different costs of attention. 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

⁴²Results from an OLS regression. A distance measure was constructed as equal to the absolute difference between the number of balls of the more frequent color on the screen and 50. Choice was then regressed on distance, which action is correct, and DP, aggregating across decision problem. Standard errors clustered at the subject level. The estimated coefficient on distance is 0.032 ($P < 0.001$).

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: Data generated by aggregating across different cost or difficulty levels would respond more slowly to incentives than the Shannon model would predict under the assumption of no variation. 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. The first is to relax the assumption that costs are linear in mutual information. The cost function

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

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 nor, as it is not posterior separable, the LIP condition. It does, however, maintain the implication that payoff equivalent states will be treated equivalently in experiment 4.

An alternative approach is to maintain the ‘posterior separable’ structure of the Shannon cost function, but replace entropy with some other function of the prior and posterior (Caplin and Dean [2013]). In general these cost functions take the form

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

for some strictly convex function T .⁴³ Posterior separable cost functions satisfy LIP, but not necessarily ILR or IUC. They can therefore allow for different elasticities of attention with respect to incentives.

One possible choice for T is to replace Shannon entropy with some more general form. In the estimation below we use the functional form of Generalized entropy (Shorrocks [1980]):

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

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 Constant Relative Risk Aversion utility functions generalize log utility: the function changes continuously in the parameter ρ , with $\rho = 1$ (an affine transform of) Shannon entropy.⁴⁴ Generalized entropy does not imply ILR but does imply LIP: Caplin *et al.* [2017] show that this property characterizes the class of (uniformly) posterior separable cost functions. Again, costs based on Generalized entropy will lead payoff equivalent states to be treated equivalently in experiment 4.

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 [2017] 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’ $X_1 \dots X_I$. 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

$$K_N(\mu, \pi) = \sum_{i=1}^I \mu(X_i) \kappa_i \sum_{\gamma} \pi(\gamma|X_i) (-H(\gamma|X_i) - [-H(\mu|X_i)]) \quad (3)$$

where $\mu(X_i)$ is the prior probability of a state in neighborhood X_i , κ_i is the marginal cost of information in neighborhood i , $\pi(\gamma|X_i)$ is the probability of signal γ conditional on

⁴³In Caplin *et al.* [2016] this class of models is referred to as ‘uniformly posterior separable’ to differentiate them from a broader class of models in which the function T is allowed to vary with the prior.

⁴⁴Specifically, when $\rho = 1$, generalized entropy is equal to the maximal possible entropy minus the entropy of the observed distribution.

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.

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 [2017] show that such cost functions can be microfounded as the result of a process of sequential information gathering. Finally, this cost function is also posterior separable.

One important free parameter for the class of neighborhood based cost functions is the definition of the neighborhoods. Here we follow Hébert and Woodford [2017] 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 [2017] 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 maximum likelihood. 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. Details of the estimation procedure appear in appendix A4.

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

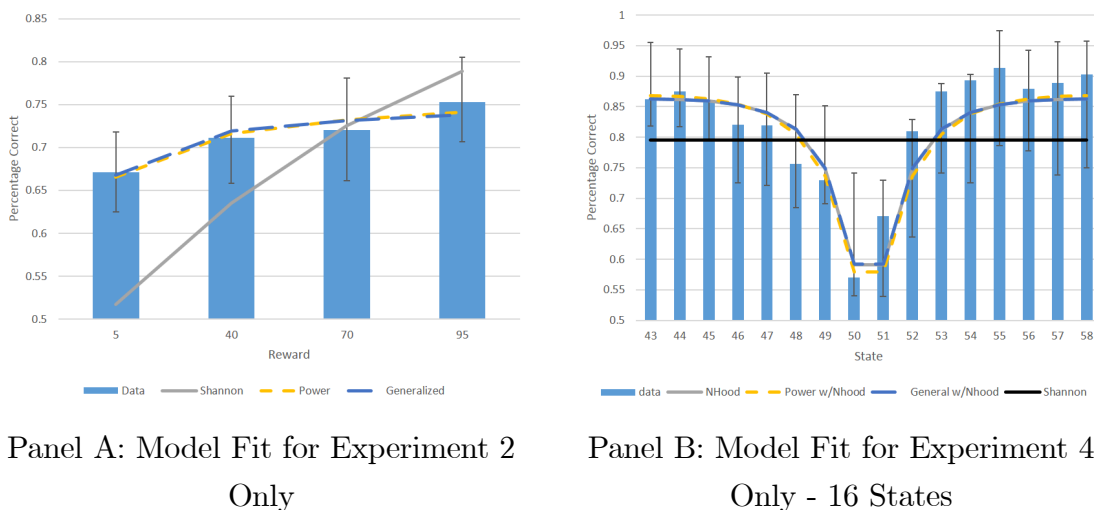
5.1.1 Aggregate Results

We begin by reporting the results from aggregate data, estimated from experiments 2 and 4 both separately and jointly. Figure 7 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.

Looking first at experiment 2, we see in figure 7 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.

Figure 7



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 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	25.08	0.38	-	-	1690	1675
Power w/NHood	299.50	99.40	2.98	-	670	647
Generalized w/NHood	0.05	2.92	-	13.01	647	624

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.⁴⁵ As can be seen from the model fits in appendix A4, the Neighborhood model is 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. Despite the fact that the Power with Neighborhood and Generalized with Neighborhood models provide rather different parameter results when estimated separately on experiments 2 and 4, both have intermediate parameter values that do a reasonable job in both cases. The Generalized with Neighborhood model does somewhat better at capturing the fall in accuracy in states close to the cutoff in experiment 4.

⁴⁵Note that in this experiment we assume linearity of utility across the two relevant reward levels - \$10 and \$40.

5.1.2 Individual Result

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 by 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 is 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.

6 Related Literature

Many papers have established the importance of attention limits in economically interesting contexts, including consumer choice,⁴⁶ financial markets,⁴⁷ and voting behavior.⁴⁸ 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

⁴⁶Chetty *et al.* [2009], Hossain and Morgan [2006], Allcott and Taubinsky. [2015], Lacetera *et al.* [2012], Pope [2009], Santos *et al.* [2012].

⁴⁷DellaVigna and Pollet [2007], Huberman [2015], Malmendier and Shanthikumar [2007], Bernard and Thomas [1989], Hirshleifer *et al.* [2009].

⁴⁸Shue 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.

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,⁴⁹ 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.⁵⁰ More recent work (Taubinsky [2013], Goecke *et al.* [2013], Khaw *et al.* [2016]) 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.* [2018] test two implications of the Shannon 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

⁴⁹An 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.

⁵⁰Though see Sanjurjo [2017].

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. Another concurrent unpublished study, Dewan and Neligh [2017] also finds violations of IUC in a different perceptual task.

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.⁵¹ 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 tasks⁵² 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.

⁵¹See for example Ratcliff and McKoon [2008] for an introduction to this class of models.

⁵²Probably 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.

7 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 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.

One obvious feature of the paper is that it studies information gathering in one particular setting. Understanding how the form of an information acquisition task affects behavior is an important avenue for future research. However, it is also important to note that the models that we test in make at least some claim to universality. Much like utility maximization or Nash equilibrium, one of the strengths of rational inattention is its portability - it can potentially explain behavior 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. For example, as reported in earlier versions of this paper, an alternative version of experiment 4 using stimuli with no natural perceptual distance does not lead to systematic violations of IUC, meaning the Shannon model may work well in that domain. We draw the analogy with experimental game theory, in which an understanding of which simple, stylized settings generate behavior consistent with Nash equilibrium has been important for developing richer models of strategic interaction.

References

- Hunt Allcott and Dmitry Taubinsky. Evaluating behaviorally motivated policy: Experimental evidence 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. Attention and selection effects. Technical report, 2018.
- Sandro Ambuehl. An offer you can't refuse? incentives change what we believe. 2017.
- Gadi Barlevy and Pietro Veronesi. Information acquisition in financial markets. *Review of Economic Studies*, 67(1):79–90, 2000.
- Vojtěch Bartoš, Michal Bauer, Julie Chytilová, and Filip Matějka. Attention discrimination: Theory and field experiments with monitoring information acquisition. *American Economic Review*, 106(6):1437–1475, 2016.
- Victor Bernard and Jacob Thomas. Post-earnings-announcement drift delayed price response or risk premium. *Journal of Accounting Research*, 27:1–36, 1989.
- Henry David Block and Jacob Marschak. *Contributions to Probability and Statistics*, volume 2, chapter Random orderings and stochastic theories of responses, pages 97–132. Stanford University Press, 1960.
- Kenneth H Britten, Michael N Shadlen, William T Newsome, and J Anthony Movshon. The analysis of visual motion: a comparison of neuronal and psychophysical performance. *Journal of Neuroscience*, 12(12):4745–4765, 1992.
- Isabelle Brocas, Juan D Carrillo, Stephanie W Wang, and Colin F Camerer. Imperfect choice or imperfect attention? understanding strategic thinking in private information games. *The Review of Economic Studies*, page rdu001, 2014.
- Andrew Caplin and Mark Dean. Behavioral implications of rational inattention with shannon entropy. NBER Working Papers 19318, National Bureau of Economic Research, Inc, August 2013.
- Andrew Caplin and Mark Dean. Revealed preference, rational inattention, and costly information acquisition. *NBER Working Papers*, (19876), Jan 2014.
- Andrew Caplin and Mark Dean. Revealed preference, rational inattention, and costly information acquisition. *The American Economic Review*, 105(7):2183–2203, 2015.
- Andrew Caplin and Daniel Martin. The dual-process drift diffusion model: Evidence from response times. *Economic Inquiry*, 2015.
- Andrew Caplin and Daniel Martin. A testable theory of imperfect perception. *The Economic Journal*, 125(582):184–202, Jul 2015.

- Andrew Caplin, Mark Dean, and Daniel Martin. Search and satisficing. *The American Economic Review*, 101(7):2899–2922, 2011.
- Andrew Caplin, Mark Dean, and John Leahy. Rational inattention, optimal consideration sets and stochastic choice. 2016.
- Andrew Caplin, Mark Dean, and John Leahy. Rationally inattentive behavior: Characterizing and generalizing shannon entropy. 2017. Unpublished.
- Christopher Chambers, Ce Liu, and John Rehbeck. Costly information acquisition. 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.
- Aubrey Clark. Contracts for information acquisition. 2016.
- Mark Dean, Stephen Morris, and Isabel Trevino. Endogenous information structures and play in global games. Mimeo, 2016.
- Stefano DellaVigna and Joshua Pollet. Demographics and industry returns. *American Economic Review*, 97(5):1667–1702, 2007.
- Ambuj Dewan and Nathaniel Neligh. Estimating information cost functions in models of rational inattention. 2017.
- Andrew Ellis. Foundations for optimal inattention. 2013.
- Xavier Gabaix, David Laibson, Guille Moloche, and Stephen Weinberg. Costly information acquisition: Experimental analysis of a boundedly rational model. *American Economic Review*, 96(4):1043–1068, 2006.
- Sen Geng. Decision time, consideration time, and status quo bias. *Economic Inquiry*, 54(1):433–449, 2016.
- Henry Goecke, Wolfgang J Luhan, and Michael WM Roos. Rational inattentiveness in a forecasting experiment. *Journal of Economic Behavior & Organization*, 94:80–89, 2013.
- David Green and John Swets. *Signal Detection Theory and Psychophysics*. Wiley, New York, 1966.

- Sanford Grossman and Joseph Stiglitz. On the impossibility of informationally efficient markets. *American Economic Review*, 70(3):393–408, 1980.
- Faruk Gul and Wolfgang Pesendorfer. Random expected utility. *Econometrica*, 74(1):121–146, 2006.
- Benjamin Hébert and Michael Woodford. Rational inattention with sequential information sampling. 2017.
- Christian Hellwig, Sebastian Kohls, and Laura Veldkamp. Information choice technologies. *The American Economic Review*, 102(3):35–40, 2012.
- David Hirshleifer, Sonya Seongyeon Lim, and Siew Hong Teoh. Driven to distraction: Extraneous events and underreaction to earnings news. *Journal of Finance*, 64(5):2289–2325, 2009.
- Daniel Ho and Kosuke Imai. Estimating causal effects of ballot order from a randomized natural experiment: The california alphabet lottery, 1978-2002s. *Public Opinion Quarterly*, 72(2):216–240, 2008.
- Tanjim Hossain and John Morgan. ...plus shipping and handling: Revenue (non) equivalence in field experiments on ebay. *The B.E. Journal of Economic Analysis & Policy*, 6(2):1–30, 2006.
- Gur Huberman. Familiarity breeds investment. *Review of Financial Studies*, 14(3):659–680, Jun 2015.
- Mel Win Khaw, Luminita Stevens, and Michael Woodford. Discrete adjustment to a changing environment: Experimental evidence. Technical report, National Bureau of Economic Research, 2016.
- Ian Krajbich, Carrie Armel, and Antonio Rangel. Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, 13(10):1292–1298, 2010.
- I Krajbich, D Lu, C Camerer, and A Rangel. The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in Psychology*, 3:193–193, 2011.
- Nicola Lacetera, Devin Pope, and Justin Sydnor. Heuristic thinking and limited attention in the car market. *American Economic Review*, 102(5):2206–2236, 2012.
- 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.
- 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. Rationally inattentive seller: Sales and discrete pricing. *The Review of Economic Studies*, 83(3):1125–1155, 2016.
- Daniel McFadden. *Frontiers in Econometrics*, chapter Conditional logit analysis of qualitative choice behavior. Academic Press, New York, 1974.
- Jordi Mondria. Portfolio choice, attention allocation, and price comovement. *Journal of Economic Theory*, 145(5):1837–1864, 2010.
- Stephen Morris and Philipp Strack. The wald problem and the equivalence of sequential sampling and static information costs. 2017.
- John W Payne, James R Bettman, and Eric J Johnson. *The Adaptive Decision Maker*. Cambridge University Press, 1993.
- Maxim Pinkovskiy. Rational inattention and choice under risk: Explaining violations of expected utility through a shannon entropy formulation of the costs of rationality. *Atlantic Economic Journal*, 37(1):99–112, March 2009.
- Luca Polonio, Sibilla Di Guida, and Giorgio Coricelli. Strategic sophistication and attention in games: an eye-tracking study. *Games and Economic Behavior*, 94:80–96, 2015.
- Devin Pope. Reacting to rankings: Evidence from "america's best hospitals". *Journal of Health Economics*, 28(6):1154–1165, 2009.
- Roger Ratcliff and Gail McKoon. The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, 20(4):873–922, April 2008.
- Roger Ratcliff and Philip Smith. A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, 111(2):333–367, 2004.
- Roger Ratcliff, Philip Smith, Scott Brown, and Gail McKoon. Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, 20(4):260–281, April 2016.

- L Felice Reddy, William P. Horan, Deanna M. Barch, Robert W. Buchanan, Eduardo Dunayevich, James M. Gold, Naomi Lyons, Stephen R. Marder, Michael T. Treadway, Jonathan K. Wynn, Jared W. Young, and Michael F. Green. Effort-based decision-making paradigms for clinical trials in schizophrenia: Part 1-psychometric characteristics of 5 paradigms. *Schizophrenia Bulletin*, 41(5):1045–1054, September 2015.
- Ricardo Reis. Inattentive producers. *Review of Economic Studies*, 73(3):793–821, 2006.
- Arthur J Robson. The biological basis of economic behavior. *Journal of Economic Literature*, 39(1):11–33, 2001.
- Adam Sanjurjo. Search with multiple attributes: theory and empirics. *Games and Economic Behavior*, 2017.
- Babur De Los Santos, Ali Hortaçsu, and Matthijs Wildenbeest. Testing models of consumer search using data on web browsing and purchasing behavior. *American Economic Review*, 102(6):2955–2980, Oct 2012.
- 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. From intentions to actions: A model and experimental evidence of inattentive choice. *Working Paper*, 2013.
- Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131, 1974.
- Stijn van Nieuwerburgh and Laura Veldkamp. Information immobility and the home bias puzzle. *Journal of Finance*, 64(3):1187–1215, 06 2009.
- Don van Ravenzwaaij, Martin Mulder, Francis Tuerlickx, and Eric-Jan Wagenmakers. Do the dynamics of prior information depend on task context? an analysis of optimal performance and an empirical test. *Frontiers in Psychology*, 3(132), May 2012.

- Robert Verrecchia. Information acquisition in a noisy rational expectations economy. *Econometrica*, 50(6):1415–1430, 1982.
- Ming Yang. Coordination with flexible information acquisition. *Journal of Economic Theory*, 158:721–738, 2015.
- Angela Yu. . *Encyclopedia of Computational Neuroscience*, chapter Decision Making Tasks. Springer, 2014.
- Trisha Van Zandt, Hans Colonius, and Robert Proctor. A comparison of two response time models applies to perceptual matching. *Psychometric Bulletin & Review*, 7(2):208–256, 2000.

Online Appendix - For Online Publication

Appendix A0: Experimental ScreenShot and Instructions

Experimental Screenshot

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
<input type="radio"/>	A	10	0
<input type="radio"/>	B	0	10
<input type="radio"/>	C	5	5

[← Previous](#)[Next →](#)

Figure A0.1: Typical Screenshot

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.**

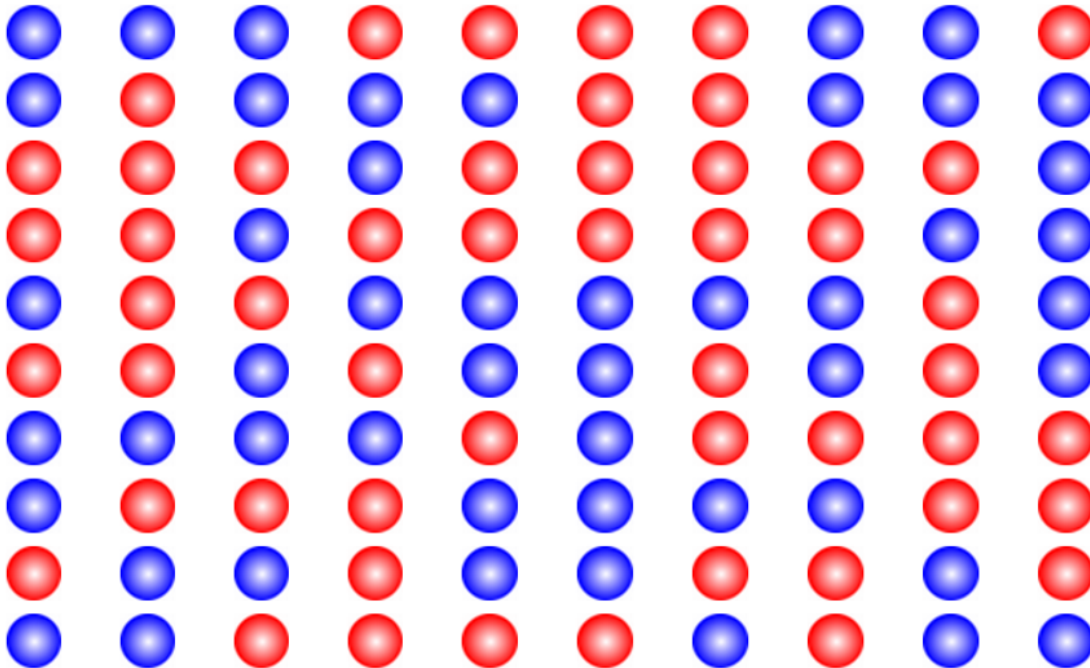
← [Home](#)

[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 $\frac{3}{4}$ chance that there will be 49 red dots on the screen, and a $\frac{1}{4}$ 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
<input type="radio"/>	A	10	0
<input type="radio"/>	B	0	10
<input type="radio"/>	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 →](#)

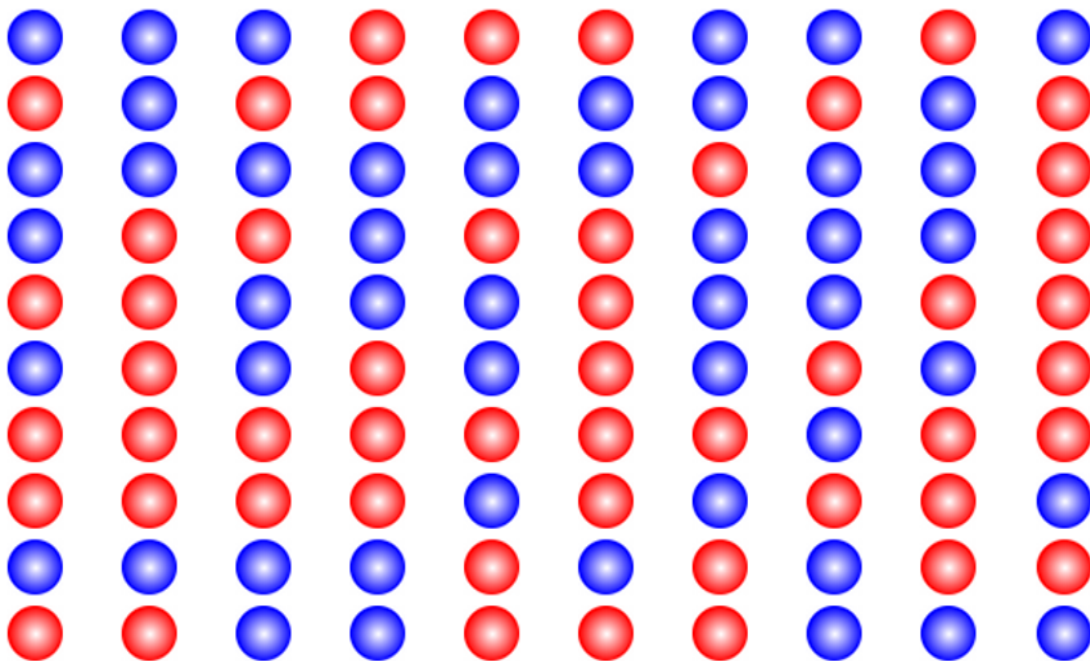
Individual Decision-Making Experiment

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
<input type="radio"/>	A	10	0
<input checked="" type="radio"/>	B	0	10
<input type="radio"/>	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
<input checked="" type="radio"/>	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
<input checked="" type="radio"/>	A	70	0
<input type="radio"/>	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
<input checked="" type="radio"/>	A	5	0
<input type="radio"/>	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
<input checked="" type="radio"/>	A	40	0
<input type="radio"/>	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
<input type="radio"/>	A	95	0
<input type="radio"/>	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

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

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

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

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

$$\begin{aligned} &G(\mu, \{a, b\}, \pi) + G(\mu, \{a, b, c\}, \hat{\pi}) \\ &\geq G(\mu, \{a, b\}, \hat{\pi}) + G(\mu, \{a, b, c\}, \pi) \end{aligned} \tag{4}$$

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

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

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

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

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 4 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.⁵³ The first column reports the mean value for the LHS of the tests described in table A1.1. Recall

⁵³Estimate for the first row generated by constructing, for each choice and each individual

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

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.⁵⁴

At the individual level we observe only a small number of significant violations of NIAS or NIAC. Of the 28 tests of NIAS in DP 1 we find 3 violations. In DP 2 of the 168 tests we find 6 violations. For NIAC, we find 2 significant violations in 28 tests (note each individual provides a single opportunity to test NIAC, as they face only 2 decision problems).

where $1(\text{choose } a)$ 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 frequency 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.

⁵⁴Point estimates and standard errors calculated as in the NIAS tests above.

Appendix A2: Shannon without Subjects who Violate NIAS or NIAC

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

Figure A2.1 shows estimated costs 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

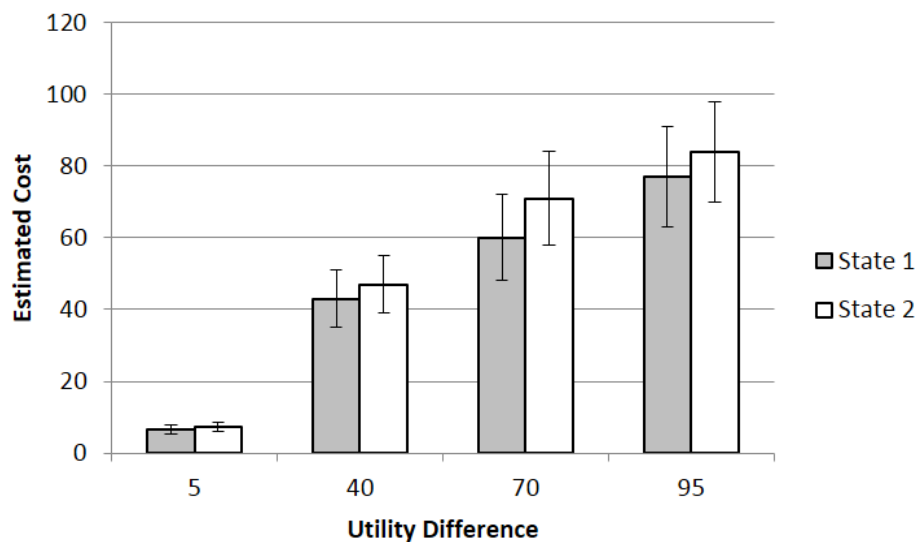


Figure A2.1: Estimated Costs - Consistent Subjects Only

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.

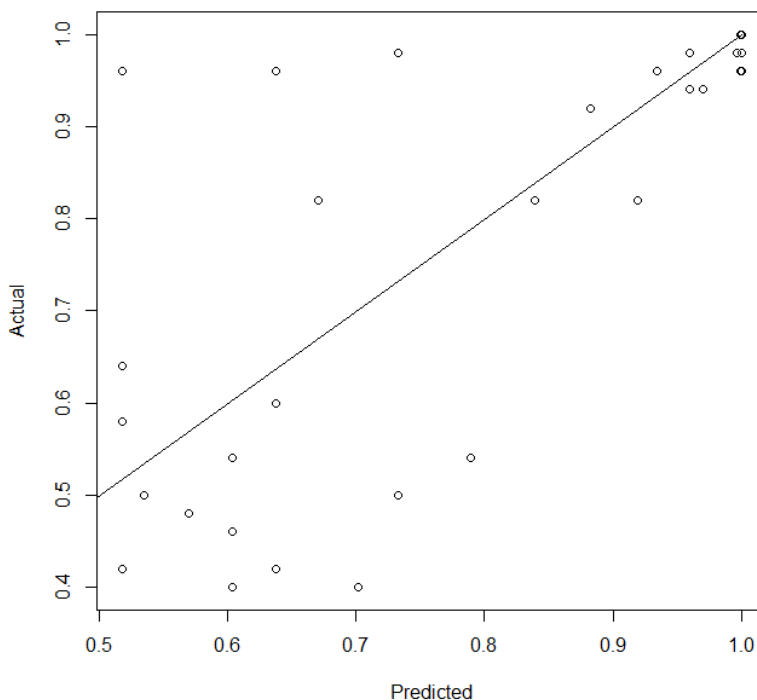


Figure A2.3: Predicted vs actual accuracy in the 70% payoff treatment

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.

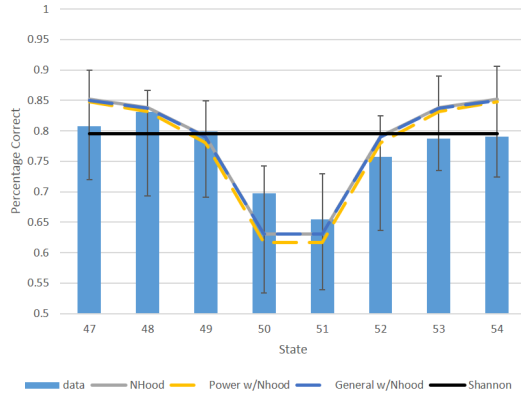
For the individual level model fits for experiments 1 and 3, the likelihood function also included an ‘error term’: an additional free parameter which represented a player’s probability of uniformly randomizing over all available moves. This was done because some of the models generate very stark predictions 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

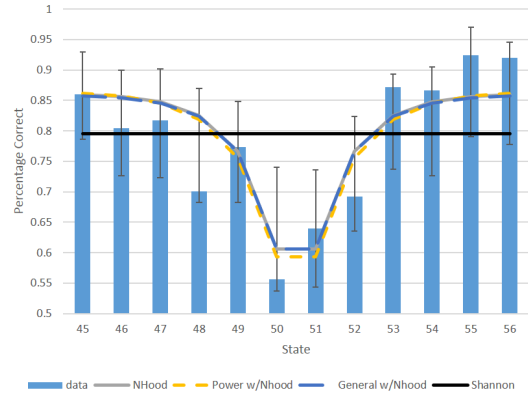
in experiment 4, because we know that the accuracy in the true prediction of the model will always be monotonically decreasing as the number of red and blue balls in a state get closer together.

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.

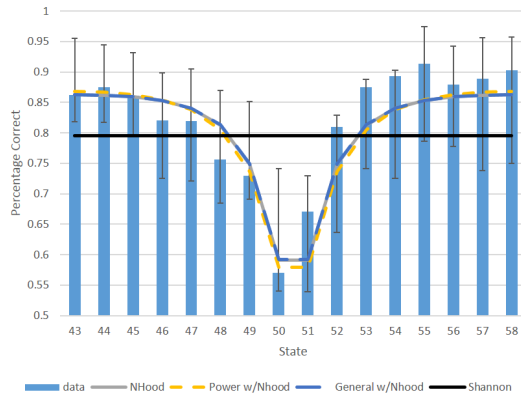
Table A4.1: Estimation Results for Experiment 4: All Treatments



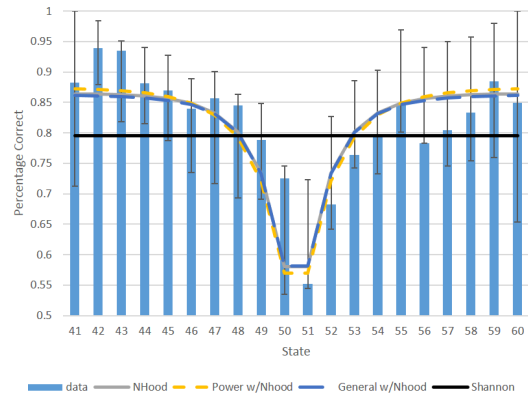
8 States



12 States



16 States



20 States

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

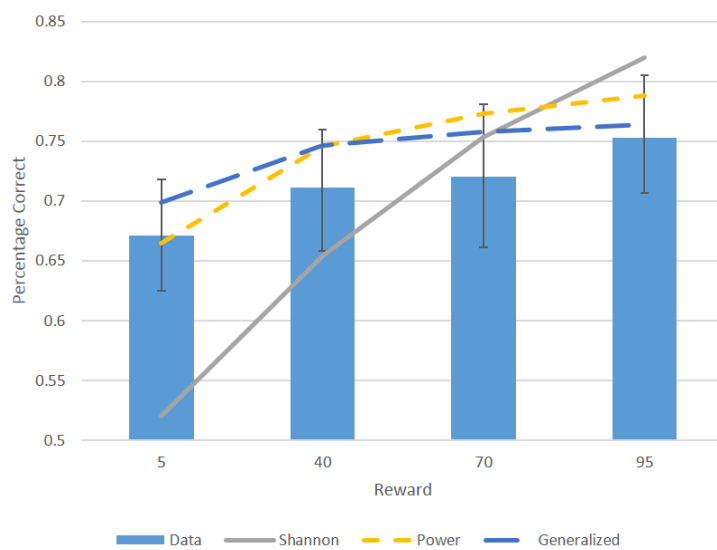
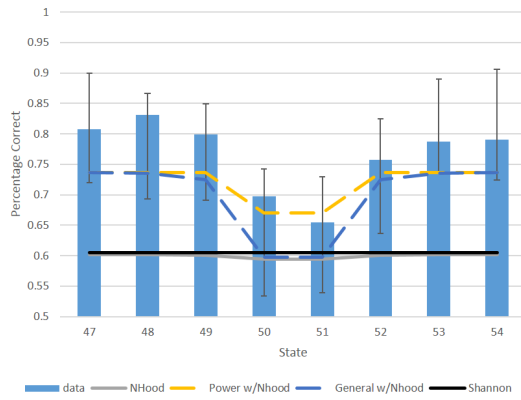
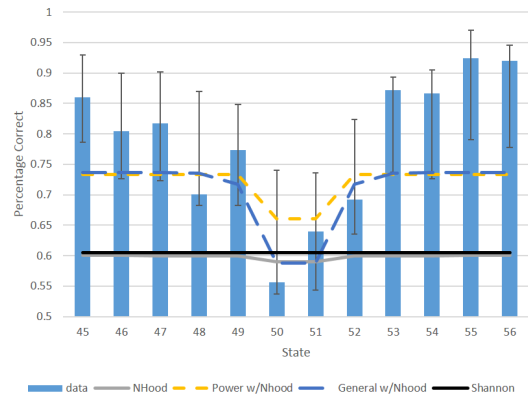


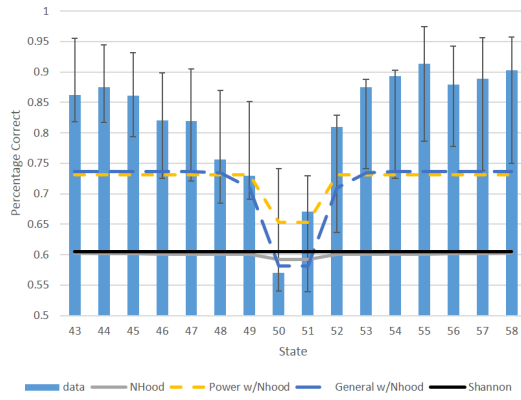
Table A4.3: Results from Joint Estimation: Experiment 4



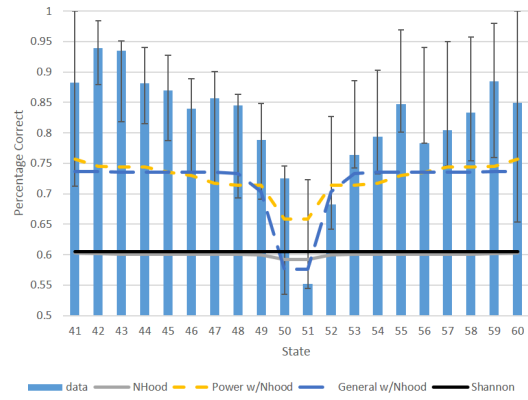
8 States



12 States



16 States



20 States