Available online at www.sciencedirect.com

ScienceDirect

Journal of Economic Theory 187 (2020) 105011

JOURNAL OF Economic Theory

www.elsevier.com/locate/jet

Estimating information cost functions in models of rational inattention ☆

Ambuj Dewan [a,*], Nathaniel Neligh [b]

[a] *Neuroscience Institute, NYU School of Medicine, New York University, United States of America*
[b] *Economic Science Institute, Chapman University, United States of America*

Received 24 July 2018; final version received 2 February 2020; accepted 6 February 2020
Available online 11 February 2020

Abstract

Models of costly information acquisition have grown in popularity in economics. However, little is known about what form information costs take in reality. We show that under mild assumptions on costs, including continuity and convexity, gross payoffs to decision makers are non-decreasing and continuous in potential rewards. We conduct experiments involving simple perceptual tasks with fine-grained variation in the level of potential rewards. Most subjects exhibit monotonicity in performance with respect to potential rewards, and evidence on continuity and convexity of costs is mixed. Moreover, subjects' behavior is consistent with a subset of cost functions commonly assumed in the literature.

JEL classification: D81; D83

☆ We are grateful to Mark Dean and Navin Kartik for their guidance on this project. We thank Marina Agranov, David Rojo Arjona, Jushan Bai, Ethan Bromberg-Martin, Jisu Cao, Alessandra Casella, Yeon-Koo Che, Giorgio Coricelli, Chad Kendall, Judd Kessler, José Luis Montiel Olea, Pietro Ortoleva, Hashem Pesaran, Andrea Prat, Miikka Rokkanen, Christoph Rothe, Simon Smith, Michael Woodford, two anonymous referees, and seminar and colloquium participants at Columbia University and the University of Southern California for their helpful comments and suggestions. We would also like to thank the numerous colleagues who helped us test our experiments. We acknowledge the financial support of grants from the Columbia Experimental Laboratory for the Social Sciences and the Microeconomic Theory Colloquium at Columbia University. Portions of this research were completed while Dewan was a postdoctoral scholar research associate at USC Dornsife INET, University of Southern California. The authors affirm no financial interest in the outcomes of the experiments detailed herein. All data were collected with the approval of the Columbia University Institutional Review Board.

* Corresponding author.
E-mail address: Ambuj.Dewan@nyulangone.org (A. Dewan).

https://doi.org/10.1016/j.jet.2020.105011

Keywords: Experimental economics; Behavioral economics; Rational inattention; Decision theory

1. Introduction

It has been observed in many settings that people have a limited capacity for attention, and this has a strong impact on their decision-making. For example, Chetty et al. (2009) demonstrate that consumers underreact to non-salient sales taxes; De los Santos et al. (2012) show that people only visit a small number of websites before making online purchases; and Allcott and Taubinsky (2015) provide evidence that people do not fully account for energy efficiency when making purchasing decisions about light bulbs.[1] Several laboratory experiments have demonstrated evidence of limited attention, including Gabaix et al. (2006), Caplin and Martin (2017), and Dean and Neligh (2019).

An increasingly common explanation for this phenomenon is the theory of *rational inattention* (Sims, 2003, 2006; Caplin and Dean, 2015; Matějka and McKay, 2015). This theory posits that people rationally choose the information to which they attend, trading off the costs of paying more attention with the ensuing benefits of better decisions. This decision-making process occurs in two stages. In the first stage, the decision-maker chooses what information to acquire and pays costs accordingly. In the second stage, the decision-maker uses the information she acquired to make decisions. Some authors make minimal assumptions on first-stage costs and derive the resulting behavioral implications (e.g. Caplin and Dean, 2015; Chambers et al., 2019). Other authors assume a specific functional form for these costs, such as mutual information (Matějka and McKay, 2015; Steiner et al., 2017; Caplin et al., 2019), or channel capacity (Woodford, 2012a). However, little is known about what form these costs take in reality, and different assumptions on these costs can lead to starkly different predictions. For instance, the properties of the information cost function can determine the multiplicity of equilibria in a global game (Hellwig et al., 2012; Morris and Yang, 2019), or whether financial investments are diversified or concentrated (Van Nieuwerburgh and Veldkamp, 2010).

In this paper, we use a laboratory experiment to investigate these first-stage information costs,[2] a crucial input for a rational inattention framework. Subjects complete a series of tasks where they must identify the numerosity of an arrangement of randomly-placed dots, as in Saltzman and Garner (1948) and Kaufman et al. (1949), but without time pressure. Our tasks are also similar to the ball-counting tasks of Caplin and Dean (2014) and Dean and Neligh (2019).

This series of tasks has fine-grained variation in the level of potential rewards for a correct answer. For each reward level, we observe both the correct answer and the subject's response. We interrogate these data in two ways: (1) testing various properties of cost functions; (2) determining which functional forms for information costs are most consistent with observed behavior.[3]

[1] Other empirical studies that find evidence of limited attention include: Hossain and Morgan (2006), Pope (2009), Lacetera et al. (2012), and McDevitt (2014) (consumer choice); Bernard and Thomas (1989), Huberman (2001), DellaVigna and Pollet (2007), Malmendier and Shanthikumar (2007), Hirshleifer et al. (2009), and Ehrmann and Jansen (2017) (financial markets); Bartoš et al. (2016) (housing and labor); and Ho and Imai (2008) and Shue and Luttmer (2009) (voting behavior). For a survey that discusses additional field studies, see DellaVigna (2009).

[2] We also conducted an online experiment, the results of which we report in Supplementary Appendix S5.

[3] Using fine-grained variation in incentives provides several advantages for the second set of analyses. For details, please refer to Appendix Subsection A4.2.

For the first set of analyses, the data allow us to recover several properties of each subject's underlying information cost function. The analyses proceeds hierarchically. For each subject, we are concerned with: (1) whether an information cost function exists, i.e. whether the subject's behavior is consistent with a rational inattention framework; (2) if it exists, whether it produces behavior that is responsive to incentives; and (3) if it induces responsiveness, whether it is "well-behaved," i.e. continuous and convex. In the paper, we derive conditions on subjects' data for testing each of these properties.

The reason we are interested in well-behavedness is because continuity and convexity are important characteristics of many cost functions and are often assumed in economic analysis. The convexity[4] of an information cost function can greatly affect model predictions. For example, Van Nieuwerburgh and Veldkamp (2010) study a portfolio choice problem in which investors choose which assets to learn about and how much to learn about each of them. Depending on the convexity of the investor's utility and cost functions, it can be optimal for the investor to learn about all available assets or to simply concentrate their attention on a single asset; utility and cost functions that imply concave objective functions result in generalized learning, whereas those that imply convex objective functions result in specialized learning. Convexity also has implications for comparative statics in a model of rational inattention. As we prove in this paper, continuity and convexity together imply that gross payoffs (excluding information costs) change continuously in incentives. This has crucial implications for economic analysis: an elasticity-based approach to welfare analysis, which is based around local properties of agents' behavior, can be deeply misleading if there are discontinuities in that behavior. As an illustration of the importance of these properties, we show in Supplementary Appendix S6 that the violation of them can have profound implications in a simple contracting model.

In our sample, over 85% of subjects exhibit behavior consistent with having an information cost function, but just over half exhibit responsiveness. We also find that roughly one-third of responsive subjects (those whose performance on the tasks improves with increasing potential rewards) have behavior that is consistent with well-behaved cost functions.

The second important set of analyses in our paper fits various classes of cost functions to our subjects' data and selects the best fit for each subject. From the accuracy of subjects' responses for each reward level, we infer how their performance in the experimental tasks changes with potential rewards; put differently, we estimate a *performance function* that traces out the relationship between the potential reward and the probability of success. From this relationship, we can recover estimates of the parameters of a subject's information cost function. Comparing the subjects' performance functions to those predicted by a range of information cost functions allows us to find the best fit for each individual.

In this paper, we provide a general result for recovering well-behaved, differentiable information cost functions from performance functions, and we derive functional forms for the performance functions associated with a range of information cost functions, some well-behaved and some not. Of particular interest to us are cost functions that have previously been used in the economic literature: mutual information (cf. Matějka and McKay, 2015), which is a scaling of the expected reduction in entropy from a decision-maker's prior beliefs to their posterior beliefs; Tsallis entropy costs (cf. Caplin et al., 2019), which generalize mutual information; fixed costs for information acquisition (e.g. Grossman and Stiglitz, 1980; Barlevy and Veronesi, 2000;

[4] Note that a finite, convex function is continuous on the interior of the space on which it is defined; in most cases of interest, continuity will be a necessary condition for convexity.

Hellwig et al., 2012); and costs for increasing the precision of normally distributed signals (e.g. Verrecchia, 1982; Van Nieuwerburgh and Veldkamp, 2010). As we show in the paper: the first implies a logistic performance function; the second implies a sigmoid, inverse-S, or concave performance function; the third implies a binary performance function with two levels of performance; and the fourth implies a concave performance function. Of the set of models we estimate, we find that the data of the subjects who are responsive to incentives are best fit by one of these four models, with roughly two-thirds of subjects best fit by the first two models, a quarter of subjects best fit by the third model, and one-seventh of subjects best fit by the fourth model. Thus, while there is some heterogeneity in the population with respect to which cost functions best reflect human behavior, the set of potential cost functions that we need to consider can reasonably be reduced to four of the cost functions commonly found in the literature.

The main advantage of using an experiment to characterize information costs is that it allows us to observe many decisions from the same individual, over a small time frame, in an environment where we can control the information available to subjects, thereby giving us a rich data set from which to recover the properties and parameters of each individual's cost functions. This is simply not possible with an administrative data set that contains a small number of decisions made by each individual. The experimental methodology we use is also highly adaptable and can accommodate a wide range of information acquisition tasks that may be of interests to researchers. Thus, our approach can be seen as a "testing bed" for theories of limited attention. Specifically using perceptual tasks with clear correct and incorrect answers, rather than choices between goods or gambles, allows us to estimate information costs separately from gross utility.

Furthermore, to our knowledge, our paper is the first to use an experiment with fine-grained variation in incentives to infer properties of information cost functions. This fine-grained variation is essential for estimating subjects' performance functions, which is crucial for our model-fitting exercise. Although several papers have examined competing hypotheses of dynamic evidence accumulation using perceptual data (e.g. Ratcliff and Smith, 2004; Smith and Krajbich, 2019), and some have used such data to fit a single model of static information acquisition (e.g. Shaw and Shaw, 1977; Pinkovskiy, 2009; Cheremukhin et al., 2015; Dean and Neligh, 2019), ours is the first to run a model selection exercise between a large number of types of cost functions in a static rational inattention framework.[5] Moreover, in contrast to previous experimental work, the tasks in our experiment involve more than two options, which allows us to differentiate between information cost functions that are not readily distinguished from each other under simple binary choice.

Our experiment also provides significant advantages relative to experiments that use choice over gambles to study information costs (e.g. Pinkovskiy, 2009; Cheremukhin et al., 2015). In a perceptual task, the correct answer is objective and known to the experimenter, whereas in a choice between gambles, the "correct answer" is determined by the subject's preferences, which the experimenter does not directly observe and cannot easily be disentangled from the information costs entailed by the subject in reaching a decision. Moreover, because of our incentivization scheme, any subset of perceptual tasks in our experiment can be unambiguously ranked in terms

[5] Cheremukhin et al. (2015) assume a specific but flexible functional form for information costs and perform two types of model selection exercises. One is based on model fits and is used to select between expected utility and rank-dependent utility. The other is based on parameter estimates and is used to distinguish between additively separable Shannon entropy-based costs (Matějka and McKay, 2015) and a capacity constraint on mutual information (Sims, 2003), since the functional form they consider nests the former and approximately nests the latter. This analysis omits most of the information cost functions considered in our paper.

of potential rewards, which cannot in general be done with choices over gambles. Therefore, using perceptual tasks allows us to not only more cleanly estimate information costs but also to generate objective measures of performance and see how they vary with potential rewards, which as we demonstrate in Section 2 of the paper, is crucial for testing whether subjects even are rationally inattentive (i.e. have information costs) in the first place.

The remainder of the paper proceeds as follows. Section 2 presents the theoretical framework that we use in this paper. Section 3 introduces various models of cost functions and applies them to the tasks of our experiment. Section 4 presents our experimental design. Section 5 presents and discusses basic experimental results and categorizes subjects according to the behaviors they exhibit. Section 6 fits various models of cost functions to the subjects' data and runs a model selection exercise to determine which is the best fit for each subject. Section 7 concludes. A more general version of this paper's theoretical framework, most proofs, experimental instructions, and robustness checks are included in appendices that can be found in the online supplement. Additional experimental results and an application to the delegation of investment are presented in unpublished supplementary appendices.[6]

2. Theoretical framework

2.1. Uniform guess tasks

In this section, we present a simplified rational inattention framework that is adapted to the tasks we use in our experiment. A fuller treatment of the theory for a general discrete rational inattention framework can be found in Appendix A1.

Consider a task where there is some unknown true state of the world $\theta \in \Theta$ that a decision-maker (DM) has to identify, and learning about the true state is costly. There are n possible states, each of which is *a priori* equally likely, i.e. $\Pr(\theta) = \frac{1}{n}\ \forall \theta \in \Theta$. In other words, the DM's *prior belief* on the state of the world is uniform. The DM receives a reward r for correctly identifying the state and no reward for incorrectly identifying the state. Therefore, the DM's goal is to maximize her probability of correctly identifying the state, which we call her *performance*, net of whatever costs she incurs in gathering information about the true state. We refer to tasks with this setup as *uniform guess tasks*.

The DM's performance is determined by her choice of *information structure*, which lists how likely guessing each state is, given the true state. Denote by $a \in \Theta$ the DM's guess of the state. Formally, an information structure is a collection of conditional probabilities $(q_{i,j})$, $i, j = 1, \ldots, n$, where $q_{i,j} = \Pr(a = \theta_i | \theta = \theta_j)$.

When the DM makes her guess, she has a belief about the likelihoods of each of the possible states, given by $\Pr(\theta = \theta_i | a = \theta_j)$. Applying Bayes' rule, it can be shown that this is equal to $\frac{q_{i,j}}{\sum_{k=1}^{n} q_{k,j}}$. We refer to this probability distribution of the state of the world conditional on the DM's guess as her *posterior belief*.

The DM's guess is correct when $a = \theta$ and is incorrect when $a \neq \theta$. Therefore, her performance is:

$$P = \frac{1}{n} \sum_{i=1}^{n} q_{i,i} \tag{1}$$

[6] Supplementary appendices can be found at https://sites.google.com/view/ambuj-dewan/research.

The DM's objective is to choose P maximize:

$$rP - C(P) \tag{2}$$

where $r > 0$ is the reward, $P \in [0, 1]$ is the DM's chosen performance, and the function $C(\cdot)$ is her associated cost. Denote by $P(r)$ the DM's choice of P for a given r, and call the resulting mapping from reward to performance the *performance function*.

An example of a uniform guess task is the type of task we implement in our experiment. In this type of task, which we refer to as the "dots" task, the DM is shown a screen with a random arrangement of dots. Her goal is to determine the number of dots on the screen, which is between 38 and 42, inclusive, with each possible number equally likely. She receives a reward r for correctly guessing the number of dots and no reward otherwise. In our example, information costs could include the cost of effort exerted in counting dots, cognitive costs incurred in employing an estimation heuristic, or the opportunity cost of time spent trying to determine the number of dots.

2.2. *Testing for rational inattention*

In order to be able to say anything about the properties of the DM's information cost function, one must first determine whether such a function even exists. As Caplin and Dean (2015) demonstrate in their Theorem 1, observed behavior is consistent with a rational inattention framework with additively separable costs if and only if it satisfies their "no improving attention cycles" (NIAC) and "no improving action switches" (NIAS) conditions. Roughly speaking, NIAC ensures that attention is allocated efficiently, and NIAS ensures that guesses of the state are made optimally, given the information that the DM has obtained. We refer the reader to Caplin and Dean (2015) for formal definitions of these conditions, though the equivalent conditions we present in the propositions of this subsection will suffice for understanding the present paper.

In uniform guess tasks, the efficient allocation of attention can be thought of as paying more attention when it is more valuable to do so, i.e. when the rewards are higher. This is formalized in the following proposition.

Proposition 1. *The DM's behavior is consistent with NIAC iff $P(r)$ is non-decreasing in r.*

What NIAC rules out is negative responses to increased incentives, e.g. by being stressed out by higher stakes.

In a uniform guess task, making optimal guesses means that a correct guess is more likely than any individual incorrect guess. In other words, the DM cannot perform better by switching her guesses. This is formalized in the following proposition.

Proposition 2. *The DM's behavior is consistent with NIAS iff $\forall\, r$, $\forall\, x \in \Theta$, and $\forall\, y \in \Theta$,* $\Pr(\theta = x|a = x) \geq \Pr(\theta = y|a = x)$.

Put differently, NIAS is satisfied in uniform guess tasks if and only if the DM's empirical posterior beliefs are maximized at the guessed state. What this rules out is the systematic misuse of information, e.g. by mentally exchanging two states of the world.

Taking these results together, a DM completing a set of uniform guess tasks is rationally inattentive iff the conditions of Propositions 1 and 2 are satisfied.

2.3. Responsiveness

A set of behaviors that is trivially consistent with rational inattention is one where the DM selects the same performance level for each reward. This is consistent with frameworks such as signal detection theory, where the DM's information structure is exogenously given. More interesting are cases where the DM does modify her behavior in response to changes in the level of incentives.

Definition 1. A DM is *responsive (to incentives)* in a uniform guess task if for some $r_2 > r_1$, $P(r_2) > P(r_1)$.

Put differently, a DM is responsive to incentives if $P(r)$ exhibits an observable region of strict increase.

2.4. Continuity and convexity

Continuity and convexity are assumptions made on costs in much of economic analysis. In a rational inattention framework, continuity of the cost function implies that gathering a small amount of additional information increases the total cost of information by only a small amount, and convexity implies that the marginal cost of information is increasing; the more information is acquired, the harder it is to acquire additional information. These properties have testable implications for the DM's behavior. Denote by $P^*(r)$ the DM's optimal choice of performance for each r.

Definition 2. $C(\cdot)$ is *well-behaved* if it is continuous and convex on $[0, 1]$, is strictly increasing and strictly convex on $\left(\frac{1}{n}, 1\right)$, and has a global minimum at $\frac{1}{n}$.

Proposition 3. *If $C(\cdot)$ is well-behaved, then $P^*(r)$ is continuous.*

Therefore, assuming the DM is utility-maximizing, one can reject the well-behavedness of $C(\cdot)$ if it is observed that her performance function $P(r)$ is discontinuous.

3. Cost functions

The space of admissible cost functions is vast. Indeed, any cost function $C : [0, 1] \longrightarrow \bar{\mathbb{R}}$ leads to behavior consistent with NIAS and NIAC. In this subsection, we introduce the classes of cost functions that are most relevant for our analysis and derive their behavioral implications. We also outline how to recover these cost functions from data that fit the corresponding performance functions.

3.1. A general recovery result for differentiable cost functions

If C is differentiable, then the DM's problem can be solved by appealing to the calculus. Not only does this allow one to solve for the DM's optimal performance function, but it also allows an analyst who does not observe C to recover it from observed behavior.

Proposition 4. *Suppose that C is well-behaved and differentiable. Then $P^*(r) = (C')^{-1}(r)$ for all r such that $C'\left(\frac{1}{n}\right) < r < \lim_{x\uparrow 1} C'(x)$. Moreover, $P^*(r) = 1$ for $r \geq \lim_{x\uparrow 1} C'(x)$.*

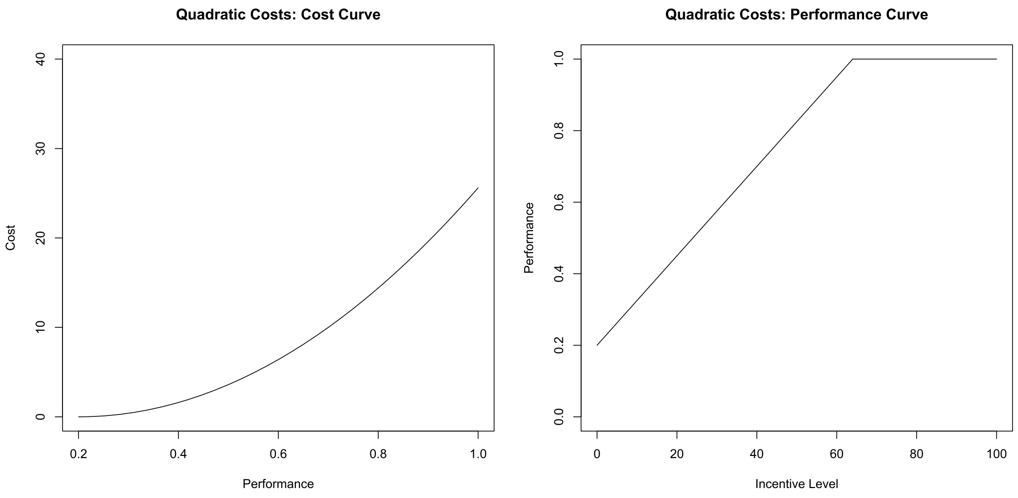

Fig. 1. Quadratic costs. The left panel shows the cost function for $c = 40$ and $d = 0.2$, and the right panel shows the resulting performance curves.

This follows from taking the first-order condition in the maximization of (2) and the fact that performance cannot exceed 1. Therefore, assuming the DM is utility-maximizing, her cost function can be recovered by inverting and integrating her observed performance function, provided that her performance is strictly increasing and continuous in incentives.

For example, suppose that C is quadratic:

$$C(P) = \begin{cases} 0, & P \leq d \\ c(P-d)^2, & P > d \end{cases} \tag{3}$$

where $\frac{1}{n} \leq d < 1$. d represents the amount of information that is freely available to the DM, and c regulates the marginal cost of information. Applying Proposition 4, we have:

$$P^*(r) = \begin{cases} \frac{r}{2c} + d, & r \leq 2c(1-d) \\ 1, & r > 2c(1-d) \end{cases} \tag{4}$$

This performance function is affine (where performance would not exceed 1), and it is depicted in Fig. 1 along with the corresponding cost function. Note that (3) can be recovered from (4) by inverting and integrating the non-constant segment. This procedure is general, but it does not always yield a closed form for the recovered C. For instance, if $P^*(r)$ is a polynomial of degree 5 or higher (where it is non-constant), then a general algebraic closed form does not exist for C.

3.2. Entropy-based cost functions

One way of modeling the cost of information is to measure how much uncertainty or "randomness" the DM reduces when she acquires information. Learning is effortful, and greater reductions of uncertainty require greater effort. The reduction of uncertainty is usually measured as a difference between her *prior* uncertainty and her *posterior* uncertainty. Formally, let $H : \Delta^{n-1} \longrightarrow \mathbb{R}_{\geq 0}$ be concave. Denoting her belief by p and her information structure by q, this difference is:

$$H(p) - \mathbb{E}[H(p|q)] \tag{5}$$

In general rational inattention problems, this form of cost for the information structure q is called *posterior-separable* (Gentzkow and Kamenica, 2014; Caplin et al., 2019). In uniform guess tasks, we can write (5) as:

$$H\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) - \frac{1}{n}\sum_{j=1}^{n}\left[\sum_{i=1}^{n} q_{i,j}\right] H\left(\frac{q_{1,j}}{\sum_{k=1}^{n} q_{k,j}}, \ldots, \frac{q_{n,j}}{\sum_{k=1}^{n} q_{k,j}}\right) \tag{6}$$

Of course, the choice of H is important; different H functions measure "randomness" in different ways. One popular choice is *Shannon entropy* (Shannon, 1948). It is defined as follows:

$$H^S(p) := -\alpha \sum_{i=1}^{n} p_i \ln(p_i) \tag{7}$$

where α is a strictly positive constant. The posterior-separable cost function that uses Shannon entropy is known as *mutual information*, and it has been widely studied in the literature (Matějka and McKay, 2015; Caplin et al., 2019). In uniform guess tasks, by substituting (7) into (6), mutual information can be written as:

$$\alpha\left[\ln(n) + \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n} q_{i,j} \ln\left(\frac{q_{i,j}}{\sum_{k=1}^{n} q_{k,j}}\right)\right] \tag{8}$$

There exist several generalizations of Shannon entropy. The one we study here is known as *Tsallis entropy* (Tsallis, 1988), and it is defined as follows:

$$H^T(p) := \frac{\alpha}{\sigma - 1}\left(1 - \sum_{i=1}^{n} p_i^{\sigma}\right) \tag{9}$$

for $\sigma \neq 1$, where α and σ are strictly positive constants. It can be shown that $H^T(p)$ converges pointwise to $H^S(p)$ as $\sigma \to 1$. Thus, Tsallis entropy generalizes Shannon entropy. In uniform guess tasks, by substituting (9) into (6), the posterior-separable cost function that uses it can be written as:

$$\frac{\alpha}{\sigma - 1}\left[1 - n^{1-\sigma} - \frac{1}{n}\sum_{j=1}^{n}\left(\sum_{i=1}^{n} q_{i,j}\right)\left(1 - \sum_{i=1}^{n}\left(\frac{q_{i,j}}{\sum_{k=1}^{n} q_{k,j}}\right)^{\sigma}\right)\right] \tag{10}$$

for $\sigma \neq 1$ and as (8) for $\sigma = 1$.

It can be shown that in the case of uniform guess tasks and Tsallis entropy costs, it is optimal for $q_{i,i}$ to be the same for all i and $q_{i,j}$ to be the same for all $i \neq j$.[7] Therefore, we can rewrite (8) and (10) in terms of performance as:

$$C(P) = \begin{cases} \dfrac{\alpha}{\sigma - 1}\left[P^{\sigma} + (n-1)^{1-\sigma}(1-P)^{\sigma} - n^{1-\sigma}\right], & \sigma \neq 1 \\ \alpha\left[\ln(n) + P\ln(P) + (1-P)\ln\left(\dfrac{1-P}{n-1}\right)\right], & \sigma = 1 \end{cases} \tag{11}$$

[7] For a formal statement and proof of this result, see Lemma A2 in Appendix Subsection A2.5. This implication does not in general hold empirically; for details, see Supplementary Appendix S1.

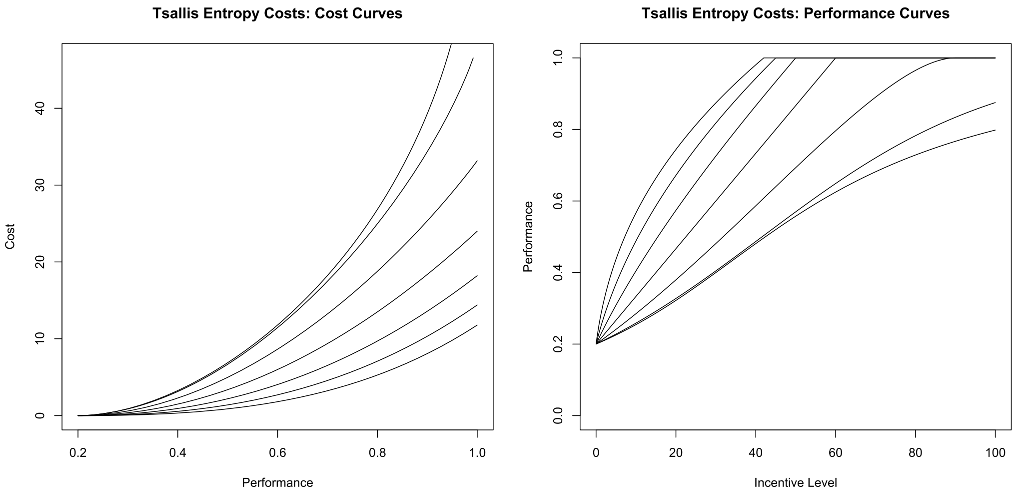

Fig. 2. Tsallis entropy costs. The left panel shows the cost function for $\sigma \in \{0.5, 1, 1.5, 2, 2.5, 3, 3.5\}$, going clockwise, and the right panel shows the resulting performance curves for those values of σ, going counterclockwise. α is set at 30.

Table 1
Properties of Tsallis performance functions for different values of σ.

σ	Shape	$= 1$ for $r \geq \frac{\alpha\sigma}{\sigma-1}$?
$(0, 1)$	Sigmoidal	No
1	Logistic	No
$(1, 2)$	Sigmoidal	Yes
2	Affine	Yes
$(2, 3)$	Inverse-S	Yes
3	Square root	Yes
> 3	Concave	Yes

where $\sigma = 1$ is the special case of Shannon entropy costs. With this cost function, we can obtain the optimal performance function that solves (2).

Proposition 5. *When $n \geq 3$, the performance function associated with Tsallis costs can be characterized as follows:*

- *For $\sigma \neq 1$, the performance function is $P^*(r) = \min\{\tilde{P}, 1\}$, where $\tilde{P}$ is a non-negative solution to $r + \frac{\alpha\sigma}{\sigma-1}\left[\left(\frac{1-\tilde{P}}{n-1}\right)^{\sigma-1} - \tilde{P}^{\sigma-1}\right] = 0$.*
- *For $\sigma = 1$, $P^*(r) = \frac{\exp(\frac{r}{\alpha})}{n-1+\exp(\frac{r}{\alpha})}$.*

The shapes of these cost functions and the corresponding performance functions are displayed in Fig. 2, and properties of the performance functions are listed in Table 1. The σ parameter allows for much flexibility in the performance function, with sigmoidal curves for low σ (between 0 and 2) and concave curves for high σ (greater than 3). For sufficiently high σ (greater than 1), perfect performance is attained for a high enough reward, $\frac{\alpha\sigma}{\sigma-1}$.

3.3. Normal signals

Some authors, such as Verrecchia (1982) and Van Nieuwerburgh and Veldkamp (2010), have assumed that the DM receives normally distributed signals about the underlying state of the world, which she uses to update her prior beliefs, and she pays a higher cost for a more precise signal. In this subsection, we present such a setup.

Let $\Theta \subset \mathbb{R}$, so that we can order its elements from smallest to largest as $\theta_1 < \theta_2 < \ldots < \theta_n$, and suppose that the DM receives signals $\hat{m} \sim N(\theta, s^2)$ about the state of the world θ. The DM can choose the precision $\zeta^2 := s^{-2}$ of these signals, and she pays a cost $K(\zeta)$ accordingly, where K is increasing, convex, and differentiable.[8]

Suppose that the distance between consecutive states is constant so that $\exists\, \eta$ such that $\theta_i - \theta_{i-1} = 2\eta$ for $i \geq 2$. Then it can be shown that the DM's problem is:

$$\max_{\zeta \in [0,\infty)} \frac{r}{n} \left[2\Phi(\zeta\eta) + (n-2)\left(2\Phi(\zeta\eta) - 1\right)\right] - K(\zeta) \tag{12}$$

Each choice of ζ induces a performance $P = \breve{P}(\zeta) := \frac{1}{n}\left[2\Phi(\zeta\eta) + (n-2)\left(2\Phi(\zeta\eta) - 1\right)\right]$. This allows us to rewrite (12) in the form of (2) by rewriting the cost of information to be a function of P rather than ζ. Because it can be shown that $\breve{P}(\cdot)$ is one-to-one, this is accomplished by setting $C(P) = K(\breve{P}^{-1}(P))$. It can also be shown that the resulting $C(\cdot)$ is strictly convex. We then have the following proposition:

Proposition 6. *A DM with normal signals and cost of information $C(\cdot)$ such that $K(\cdot)$ is increasing, is convex, and has non-negative third derivative*[9] *has a strictly concave performance function.*

This type of performance function is depicted in the right-hand panel of Fig. 3, for the case of linear K.

3.4. Fixed costs

Another common model of information costs in the literature is "all-or-nothing" costs, where the DM begins with no information but can become completely informed about the state of the world if she pays a cost (e.g. Grossman and Stiglitz, 1980; Hellwig et al., 2012). Here, we generalize this form of costs by allowing for the DM to receive some information for free and pay a fixed cost to receive more information; we do not stipulate that she must become fully informed.[10]

We can represent this situation as follows. Let there exist $\underline{q}, \bar{q}$ such that $\frac{1}{n} \leq \underline{q} < \bar{q} \leq 1$ and:

$$C(P) = \begin{cases} 0, & P \leq \underline{q} \\ \kappa, & P \in (\underline{q}, \bar{q}] \\ \infty, & P > \bar{q} \end{cases} \tag{13}$$

[8] Note that K is defined as a function of the positive square root of the precision. However, for the sake of parsimony, we will refer to it as the "cost of precision."

[9] This assumption on the third derivative is a technical assumption. It holds if, for instance, K is linear in precision (i.e. quadratic in the square root of precision), as we assume later in the paper.

[10] A similar modeling assumption is made by Admati and Pfleiderer (1988) in an asset market model, where a trader can choose either to remain uninformed about asset returns or to receive a noisy signal about returns at a fixed cost.

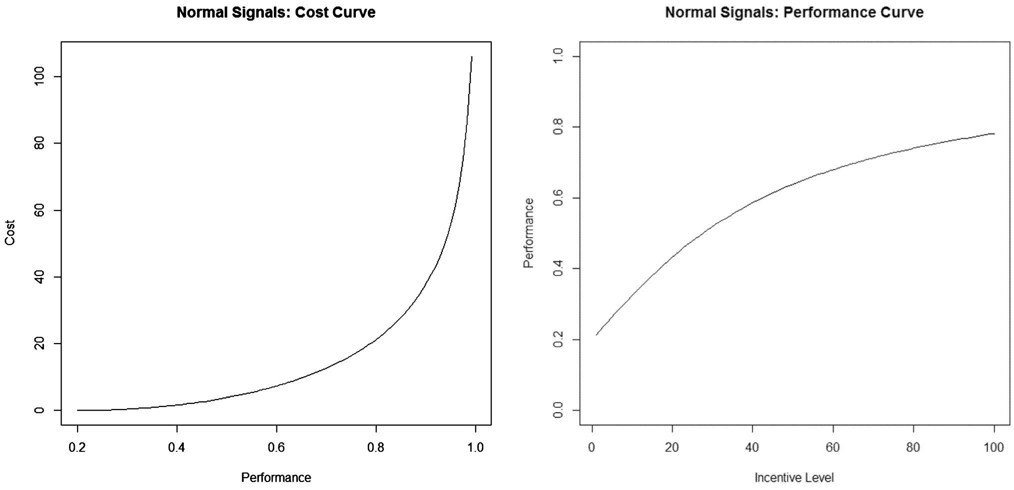

Fig. 3. Normal signals with cost of precision given by $K(\zeta) = 4\zeta^2$. The left panel shows the cost function, and the right panel shows the resulting performance curve.

According to this cost function, the DM can receive information up to an accuracy of $\underline{q}$ for free, but she must pay a fixed cost κ to acquire information up to an accuracy of $\bar{q}$.

Cost functions with fixed costs such as these can be seen as representing dual-system cognitive processes (cf. Stanovich and West, 2000; Kahneman, 2003). In such processes, a small amount of information may be acquired at a very low cost, but there is a fixed cost to acquiring more information. This implies a discontinuity in the cost function between information structures with "low" informativeness and those with "high" informativeness.

In uniform guess tasks, the DM is willing to pay the cost κ of acquiring information only when the rewards are sufficiently high, i.e. when $r\bar{q} - \kappa \geq r\underline{q}$. This implies a binary performance function: for $r \leq \frac{\kappa}{\bar{q} - \underline{q}}$, the DM acquires no information and achieves $\underline{q}$, and for $r > \frac{\kappa}{\bar{q} - \underline{q}}$, the DM acquires enough information to achieve $\bar{q}$.

Cost functions of this subclass are easily recoverable from data by estimating the relationship depicted in the right panel of Fig. 4 and finding the incentive level threshold at which the DM's performance level jumps.

3.5. Other non-convexities

Other non-convex cost functions can also produce discontinuous performance functions. In fact, the cost function need not even be discontinuous for this occur. To illustrate this, consider a DM who has a cost function C that is concave in performance, as depicted in Fig. 5.

Net payoffs are maximized when the positive distance between gross payoffs and costs is largest. For low reward levels (such as r_1), this happens at the no-information performance level, 0.2. For high reward levels (such as r_2), this happens at the full-information performance level, 1. In this manner, just as in the fixed-cost case, a binary performance function obtains, with the DM acquiring no information if the incentive is low and acquiring full information if the incentive level is high.

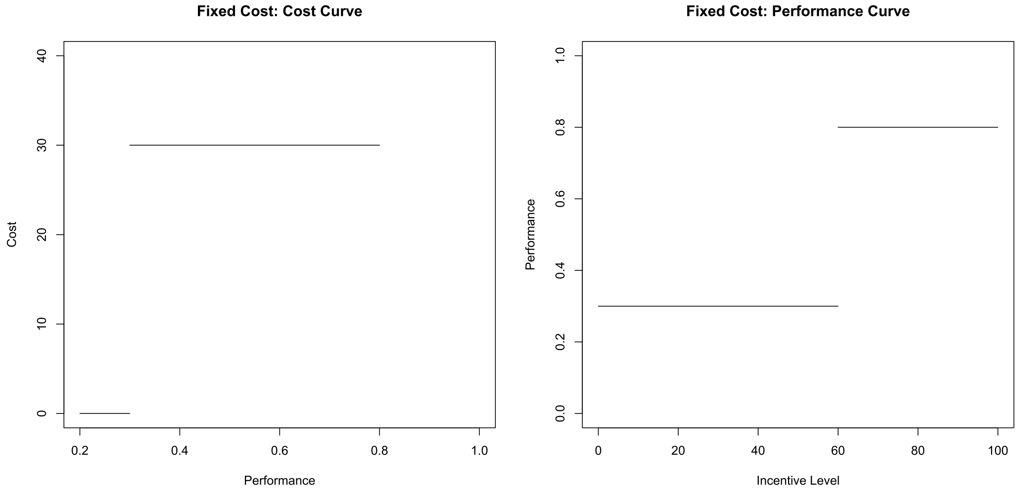

Fig. 4. Fixed cost for information acquisition. The left panel shows the cost function, and the right panel shows the resulting performance curve. Parameters are $\kappa = 30$, $\underline{q} = 0.3$, and $\bar{q} = 0.8$.

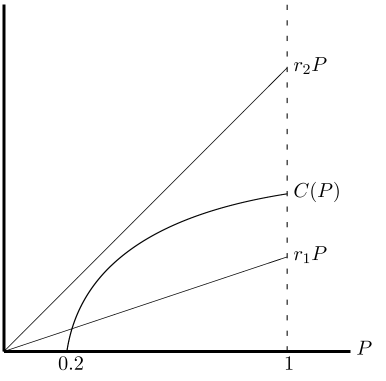

Fig. 5. Concave costs.

More complicated performance functions are also possible. Consider a richer representation of a dual-system cognitive process (cf. Kahneman, 2003) than was presented in the preceding subsection:

$$C(P) = \begin{cases} 0, & P \leq d_1 \\ c_1(P - d_1)^2, & d_1 < P \leq d_2 \\ c_1(d_2 - d_1)^2, & d_2 < P \leq d_3 \\ c_2(P - d_3)^2 + c_1(d_2 - d_1)^2, & P > d_3 \end{cases} \tag{14}$$

where $c_1 > 0$, $c_2 > 0$, and $\frac{1}{n} \leq d_1 < d_2 \leq d_3 < 1$. According to this cost function, information is available for free up to a performance level of d_1. In the interval $(d_1, d_2]$, performance can be adjusted up to a level of d_2. This represents the "automatic" system of the dual-system process, and can be thought of as a subconscious process to which the brain can variably allocate mental resources. Exerting effort beyond d_2 can be seen as actively thinking about the problem at

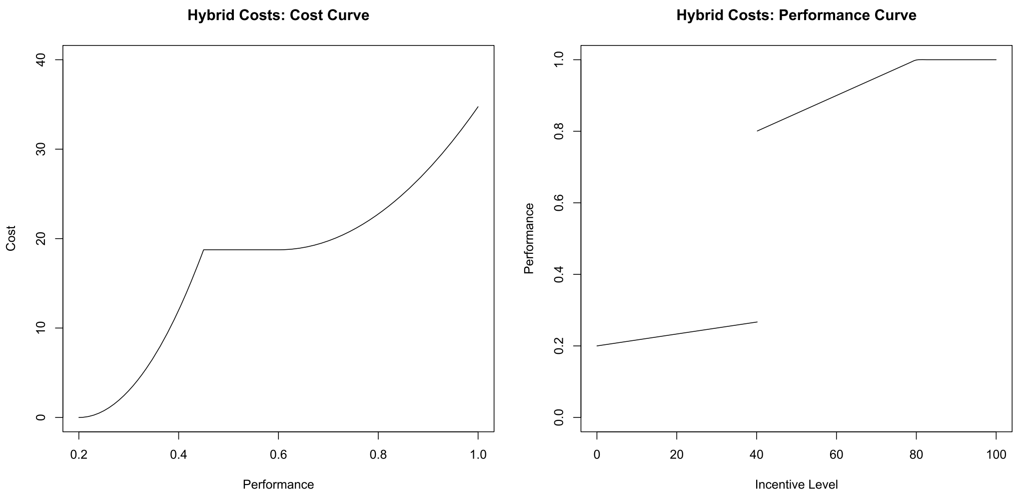

Fig. 6. Hybrid costs. The left panel shows the cost function, and the right panel shows the resulting performance curve. Parameters are $c_1 = 300$, $c_2 = 100$, $d_1 = 0.2$, $d_2 = 0.45$, and $d_3 = 0.6$.

hand, or engaging the "controlled" system of the dual-system process. Thinking allows a performance of at least d_3 to be achieved, with higher performance levels attainable with more effort. This "hybrid" cost function that concatenates two quadratic cost curves induces a discontinuous performance function.

Proposition 7. *Suppose that*[11]*:*

$$d_3 \in \left(\frac{c_1 d_1 + c_2 - \sqrt{(c_1 d_1 + c_2)^2 + (c_1 + c_2)(c_1(1 - 2d_1 - (d_2 - d_1)^2) - c_2)}}{c_1 + c_2}, \frac{c_1 d_1 + c_2 + \sqrt{(c_1 d_1 + c_2)^2 + (c_1 + c_2)(c_1(1 - 2d_1 - (d_2 - d_1)^2) - c_2)}}{c_1 + c_2} \right)$$

Then the cost function (14) *yields the following performance function:*

$$P^*(r) = \begin{cases} \frac{r}{2c_1} + d_1 & r < \delta \\ \min\left\{ \frac{r}{2c_2} + d_3, 1 \right\}, & r \geq \delta \end{cases} \tag{15}$$

where $\delta = \frac{c_1(d_2 - d_1)^2}{d_3 - d_1}$ *if* $c_1 = c_2$ *and* $\delta = \frac{2c_1 c_2}{c_1 - c_2}\left[\sqrt{(d_3 - d_1)^2 + \frac{(c_1 - c_2)(d_2 - d_1)^2}{c_2}} - (d_3 - d_1)\right]$ *if* $c_1 \neq c_2$.

This performance function consists of two affine segments separated by a jump discontinuity, and it flattens out once the upper bound of perfect performance is reached. It is depicted in Fig. 6 along with its corresponding cost function.

11 This condition on d_3 ensures that the performance function has two separate regions of strict increase rather than just one.

Table 2
Properties of cost functions.

Cost Function	Continuous	Convex	Performance Function
Differentiable and well-behaved	Yes	Yes	Inverse of derivative
Quadratic	*Yes*	*Yes*	*Affine*
Tsallis entropy	Yes	Yes	Sigmoid/inverse-S/concave (SIC)
Mutual information	*Yes*	*Yes*	*Logistic*
Normal signals	Yes	Yes	Concave
Dual-process	Can be	No	Discontinuous
Fixed costs	*No*	*No*	*Binary*
Hybrid	*Yes*	*No*	*Piecewise affine*

Note: Performance-function properties of normal-signal costs are for state spaces with equidistant spacing.

3.6. Summary

Table 2 summarizes the properties of the classes of cost functions discussed in this section and lists the corresponding performance functions. While each cost function generates a unique performance function, recovery of a cost function from a performance function is not in general unique, as illustrated in Subsection 3.5. However, a best-fitting performance function can be selected based on a DM's observed behavior, and this datum can be used to determine which of a set of plausible classes of cost functions could have generated the observed behavior. This is the exercise we perform in Section 6.

4. Experimental design

4.1. Description

The experiment we implemented involved a series of perceptual tasks, each for a potential reward. In each of these tasks, subjects were shown a screen with a random arrangement of dots and were asked to determine the number of dots on the screen. The number of dots was between 38 and 42, inclusive, and each number was equally likely.[12] Subjects were informed of these facts; there was no deception or withholding of information about the structure of the tasks. Subjects also completed a second set of tasks involving the identification of angles. We refer to the first type of task as the "dots" task and to the second as the "angle" task. For the sake of brevity, we relegate the description and results of the "angle" task to Supplementary Appendix S4.

Each task had a potential reward in an experimental currency called "points." At the start of each task, subjects were shown this reward, which we refer to as the *incentive level*, in large characters for three seconds (e.g. Fig. 7), before it was replaced with the random dot arrangement (e.g. Fig. 8). Displaying the incentive level before the dot arrangement ensured that subject looking at the screen would see the incentive level before being able to start the task. While the dot arrangement was on screen, the incentive level continued to be displayed to the right of the screen, ensuring that subjects would not have to memorize this number. Subjects then had as

[12] These numerosities were selected to be in line with previous experiments with similar tasks (e.g. Caplin and Dean, 2014).

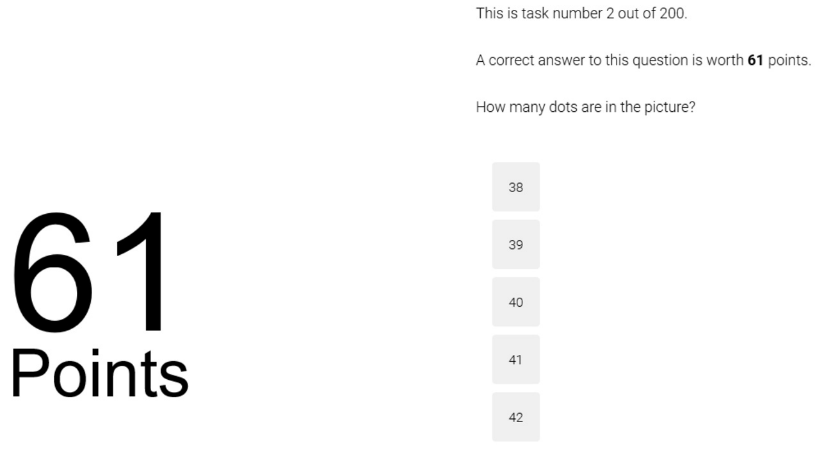

Fig. 7. Incentive display for a task.

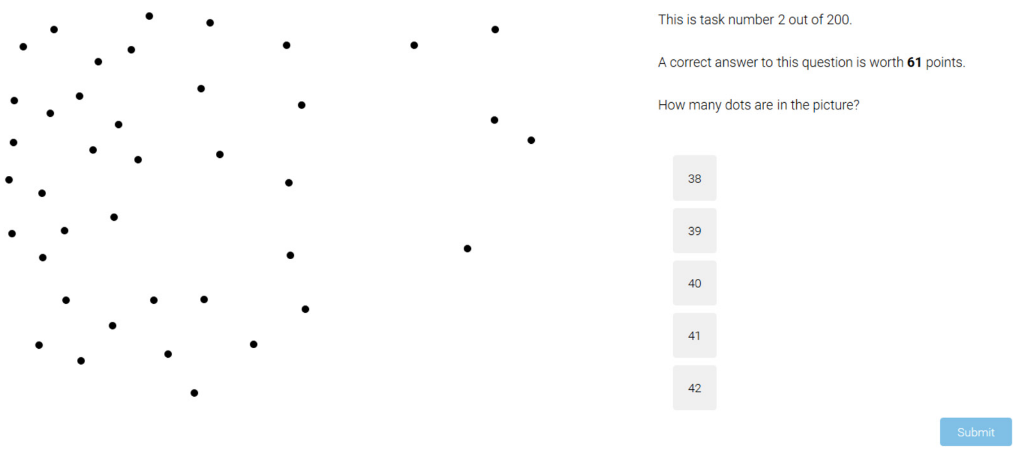

Fig. 8. Arrangement of dots for a task.

much time as they desired to determine the number of dots on the screen before proceeding to the next task. If they answered correctly, then they earned the potential reward; if not, then they earned no points for that task. Feedback was not given until the end of the experiment.

Subjects completed 200 tasks, each at an integer incentive level between 1 and 100, inclusive. They were randomly shown either all 100 "dots" tasks or all 100 "angle" tasks first. Blocks of tasks were balanced by incentive level to ensure roughly the same level of variation in incentives throughout the experiment. Subjects were first shown each of the 50 odd incentive levels between 1 and 100 in a random order, and were then shown each of the 50 even incentive levels between 1 and 100 in a random order. This was repeated (in a different random order) for the next 100 tasks.[13]

[13] A detailed explanation of the advantages and disadvantages of this kind of fine-grained incentive structure can be found in Appendix Subsection A4.2. To summarize: while fine-grained incentives present some drawbacks in terms of

Experimental earnings were determined as follows. One task from the first half the experiment and one task from the second half of the experiment were randomly selected for payment. The incentive level of each selected task determined the probability of winning one of two monetary prizes. For example, if the first selected task had an incentive level of 84 and was answered correctly, and the second selected task had an incentive level of 33 and was answered incorrectly, then this would give the subject an 84% probability of winning the first prize and a 0% probability of winning the second prize. Determining earnings in this manner ensured that expected earnings were linear in the incentive level, which obviated the need to elicit risk preferences.[14] In other words, this ensured that under the assumption of expected utility theory, the subjects' utilities (excluding information costs) were known to us (up to a multiplicative constant).[15] Thus, the estimated relationship between performance and incentive level for each subject could be considered a valid estimate of their performance function, without the need to apply any additional transformation.

As mentioned above, subjects completed 200 tasks in total: 100 "dots" tasks and 100 "angle" tasks. They either completed all the "dots" tasks or all the "angle" tasks first, and this order was randomly determined.[16] For 41 subjects, the prizes were $10 US, and for 40 subjects, the prizes were $20 US. In addition, subjects were paid a $10 participation fee.

All sessions were conducted at the Columbia Experimental Laboratory in the Social Sciences (CELSS) at Columbia University, using the Qualtrics platform. We ran 8 sessions with a total of 81 subjects, who were recruited via the Online Recruitment System for Economics Experiments (ORSEE) (Greiner, 2015).

4.2. Discussion

Our experimental design has several beneficial features as compared to previous experimental work in limited attention. Firstly, the departure from experiments involving binary choice with two states of the world (e.g. Ratcliff and Smith, 2004; Cheremukhin et al., 2015; Dean and Neligh, 2019) allows for distinguishing between certain types of cost functions in a manner that would otherwise not be possible. For instance, if there were only two states (i.e. two different possible numbers of dots), then entropy-based cost functions would not yield performance curves with inflection points, and it would therefore be difficult to distinguish between Tsallis entropy costs with $\sigma \in (0, 2)$ and normal signal precision costs. Furthermore, having more than two states, some of which were closer to each other than others (e.g. 39 is closer to 38 than to 42) allows us to study perceptual distance, which we discuss further in Supplementary Appendix S1.

Secondly, using perceptual tasks instead of value-based choices, such as choices over gambles (e.g. Pinkovskiy, 2009; Cheremukhin et al., 2015), allows for cleaner identification of information costs. On any given trial, the true state of the world is known to the experimenters; we need

testing cost function properties, they provide significant benefits when it comes to recovering a subject's cost function from observed behavior.

[14] This binary lottery incentivization technique was pioneered by Roth and Malouf (1979).

[15] We relax the assumption of expected utility theory and allow for incentives to be probability-weighted in Appendix Subsection A4.3. Our qualitative results remain largely unchanged. However, it should be noted that a limitation of our design is that probability weighting of incentives and the relationship between performance and incentives cannot be separately nonparametrically identified. Thus, our robustness results rely on specific (albeit flexible) functional forms of probability weighting. Further discussion of this issue can be found in Appendix Subsection A4.3.

[16] In the online version of this experiment, subjects completed 200 "dots" tasks and no "angle" tasks. Results and further details can be found in Supplementary Appendix S5.

not infer it from choice data. Thus, any choice on a subject's part can be cleanly classified as either a correct choice or a mistake, and this classification is then used to estimate information costs. This stands in contrast to experiments with choice over gambles, where it is necessary for the analyst to simultaneously use choice data to estimate utility parameters and use the estimated utility to construct the classification of correct responses and mistakes that is used to estimate information costs.

Moreover, choices over gambles present a couple of conceptual issues in a rational inattention framework. In such a framework, the amount of attention paid to a particular decision problem depends on the marginal benefit of selecting one option over another. However, in choices over gambles, these benefits are not known to the DM *ex ante*, and practically speaking, they cannot be calculated independently of knowing which gamble is preferable to another. By contrast, in perceptual tasks with known rewards, the marginal benefit of answering correctly relative to answering incorrectly is known *ex ante*, regardless of what the correct answer actually is. Furthermore, if utility over gambles (excluding additively separable information costs) deviates from expected utility theory — for instance, with probability weighting — then this can itself be seen as the result of a rationally inattentive process (Woodford, 2012b; Steiner and Stewart, 2016), theoretically but not necessarily practically distinct from the additively separable information costs that cause choice mistakes, thereby calling into question the descriptive validity of the model.

Finally, using fine-grained variation in incentives allows us to see how subjects' behavior changes in response to small changes in potential rewards. To illustrate, suppose that a subject's utility of money were linear, and they were participating in the $10 prize treatment. In that case, since an incentive level of 63 gives a 1% higher chance of receiving the prize for answering correctly than an incentive level of 62, the former incentive level is worth 10 cents more to the subject than the latter. Observing how a subject's behavior changes in response to these small differences in incentives allows us to reliably trace out their performance curve and classify that curve according to the information cost function that generated it.

5. Categorization results

In this section, we present the first set of results of our laboratory experiments. We perform an individual-level analysis to classify subjects according to whether their behavior is consistent with rational inattention, responsiveness to incentives, and well-behavedness of their cost functions. Additional experimental results related to demographics and aggregate behavior are provided in Supplementary Appendix S2.

5.1. Choice data

The data we are most interested in for each task t are the incentive level r_t, the true state of nature θ_t, and the subject's response a_t. For each task, define the subject's *correctness* as $y_t := \mathbb{1}_{\{\theta_t\}}(a_t)$. That is, y_t takes the value 1 if the subject correctly determined the state of the world in task t and 0 otherwise.

We are primarily interested in the relationship between correctness and incentive level. We can think of the pattern of successes and failures that we observe as being generated by some underlying data-generating process that for every possible reward level tells us the probability of answering correctly. We denote this probability by $P_t := \Pr(y_t = 1|r_t) = \Pr(a_t = \theta_t|r_t)$ for each task t; in other words, the underlying data-generating process is the performance function. (See

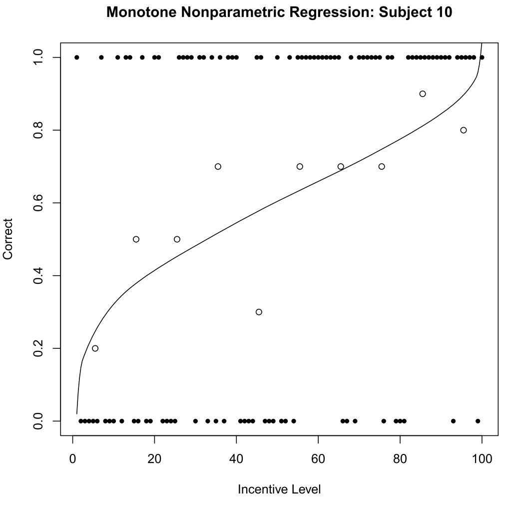

Fig. 9. Isotone nonparametric regression of correctness on incentive level for Laboratory Subject 10 (Dette et al., 2006). Circles indicate average success rate within each bin of 10 incentives.

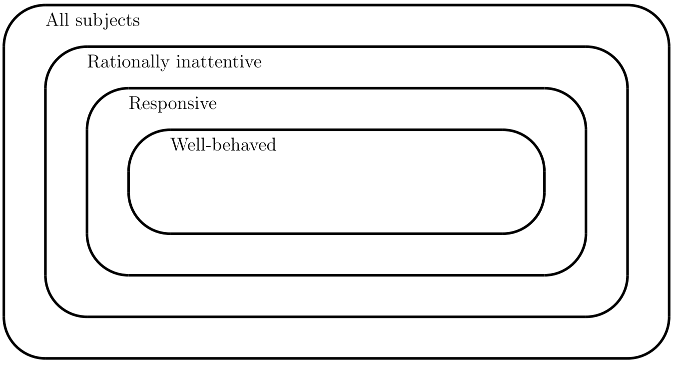

Fig. 10. Categorization of subjects.

Fig. 9.) Using the correctness data allows us to infer whether subjects have behavior consistent with having an information cost function, and if so, what its properties are.

We are able to categorize subjects according to whether their behavior adheres to these properties. First, we classify them by whether or not they are rationally inattentive. Then, we classify rationally inattentive subjects by whether or not they are responsive to incentives. This subset of subjects is the subset of greatest interest to us; these are the subjects for whom we can estimate performance functions and back out corresponding information cost functions. Finally, we classify responsive subjects according to whether or not their behavior is consistent with well-behaved cost functions. This categorization scheme is illustrated in Fig. 10.

5.2. Rational inattentiveness

We now proceed with the individual-level categorization exercise.

Before testing the properties of the subjects' cost functions, it is necessary to determine whether there exists a cost function that rationalizes their data in the first place. To that end, we test the necessary and sufficient "no improving attention cycles" and "no improving action switches" conditions by testing the equivalent conditions established in Subsection 2.2.

5.2.1. No improving attention cycles

As demonstrated in Proposition 1, a subject satisfies NIAC in our experiment if and only if their probability of correctly guessing the state is non-decreasing in the reward. This implies that rationally inattentive subjects have non-decreasing performance functions.

At this point, a clarification is in order. As we showed in Proposition 1, NIAC holds in a set of uniform guess tasks iff for any pair of incentive levels (r_1, r_2) with $r_1 > r_2$, we have that $P^*(r_1) \geq P^*(r_2)$. Observationally, this means that the subject had more correct answers under incentive level r_1 than incentive level r_2. However, in our experiment each subject is given each incentive level only once. Therefore, the empirically-observed probabilities of answering each decision problem correctly are either 0 or 1. If were to apply the NIAC condition directly to our data, this would mean that the only subjects whose behavior is consistent with NIAC would be those who always answer incorrectly up to some incentive threshold after which they always answer correctly. Given the stochasticity of choice under limited attention, this scenario is implausible.

Therefore, rather than strictly interpreting our data as stochastic choice data and making direct pairwise comparisons of decision problems to test NIAC, we adopt an estimation-based approach. We flexibly estimate the performance function given correctness data and see if this estimate is significantly different from a non-decreasing function, in which case we reject NIAC. In theory, unless there is some reward threshold below which the subject is never correct and above which the subject is always correct, the fit of a monotone performance function can be improved by adding peaks and troughs. The question, then, that we wish to pose is not whether a non-monotone or decreasing function can fit the data, but whether we can reject the hypothesis that a non-decreasing function explains the data.

To test for weak positive monotonicity, we employ a method developed by Doveh et al. (2002) and compare the estimation of an unrestricted cubic polynomial regression of correctness on incentive level for each subject to one with a positive derivative restriction.[17] The null hypothesis for this test is that the response function is monotonic. At the 5% level, we fail to reject positive monotonicity for 77 out of 81 lab subjects (95.1%).[18] Examples of polynomial regressions and correctness data for two subjects, one who rejects NIAC and one who fails to reject NIAC, are depicted in Fig. 11.

[17] Several other methods in the statistical and econometric literatures have been devised to test for the monotonicity of regression, including but not limited to Bowman et al. (1998), Ghosal et al. (2000), Hall and Heckman (2000), Birke and Dette (2007), and Chetverikov (2019), most of which are nonparametric. We use Doveh et al.'s (2002) parametric test because it is less prone to rejecting monotonicity when there are outliers, e.g. a lone failure in a region of success, or vice versa.

[18] The optimization in the computation of the restricted regression for lab subject 35 failed to converge, and so we did not perform the test for them. For that subject, a one-tailed t-test of the coefficient on incentive level in a linear regression of correctness on incentive level failed to reject the null of the coefficient being non-negative at the 5% level, and so we classify them as having a non-decreasing performance function.

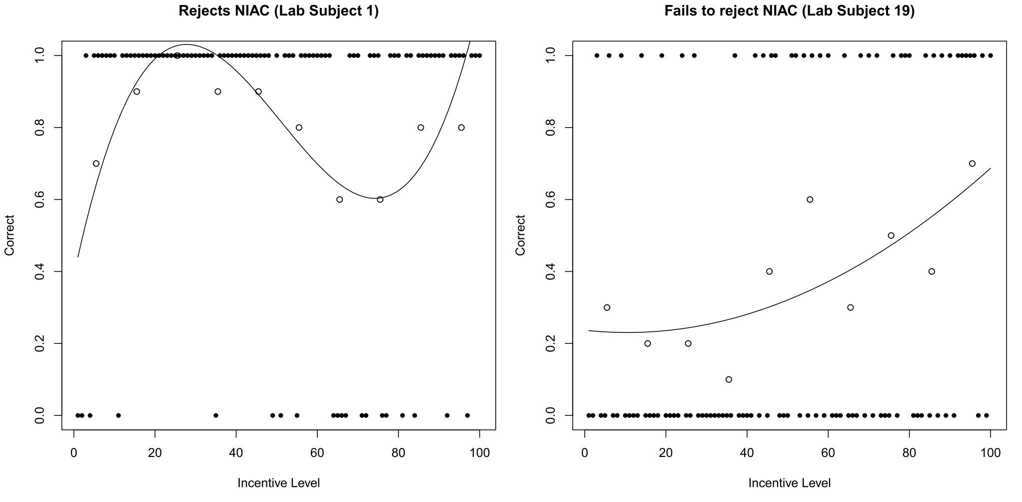

Fig. 11. Unrestricted cubic polynomial regression of correctness on incentive level for Subjects 1 and 19. The former rejects NIAC (and therefore rejects rational inattentiveness), and the latter fails to reject NIAC. Circles indicate average success rate within each bin of 10 incentives.

5.2.2. *No improving action switches*

To test for the second necessary and sufficient condition for rational inattentiveness, NIAS, we cannot simply examine the estimated performance function; following Proposition 2, we must look at the posterior probabilities of each state given each response. We employ a bootstrap procedure. For each subject and action, we calculate the empirically observed distribution of true states, i.e. we calculate $\Pr(\theta = y|a = x)$ for each x and y.[19] We then simulate 499 bootstrap samples for each distribution. If the most common true state is the one corresponding to the action in at least 5% of samples for each action for a given subject, then that subject fails to reject NIAS. In other words, we check that $\Pr(\theta = x|a = x)$ is maximized at $\theta = x$, for each $a = x$, in at least 5% of samples. Overall, we find that 74 out of 81 (91.3%) laboratory subjects fail to reject NIAS.

Overall, 70 out of 81 (86.4%) laboratory subjects fail to reject both NIAC and NIAS. We refer to these subjects as "rationally inattentive," or simply "rational," subjects.

5.3. *Responsiveness*

Of the subjects who fail to reject rational inattentiveness, some of them may have flat response functions, i.e. while they could be rationally inattentive, they do not actually respond to incentives (within the range of incentives presented to them).

To determine which subjects are responsive to incentives, for each subject who failed to reject rational inattentiveness, we run a linear weighted least squares regression of correctness on incentive level and run a one-sided t-test of the coefficient on incentive level with the null of

[19] It should be noted that strictly speaking, the NIAS condition applies separately to each decision problem that the DM faces. Since each subject faces each incentive level only once, they actually face 100 different decision problems. For that reason, we test a slightly weaker condition: whether an individual exhibits overall systematic misuse of information. Overall systematic misuse of information implies systematic misuse of information in at least one decision problem.

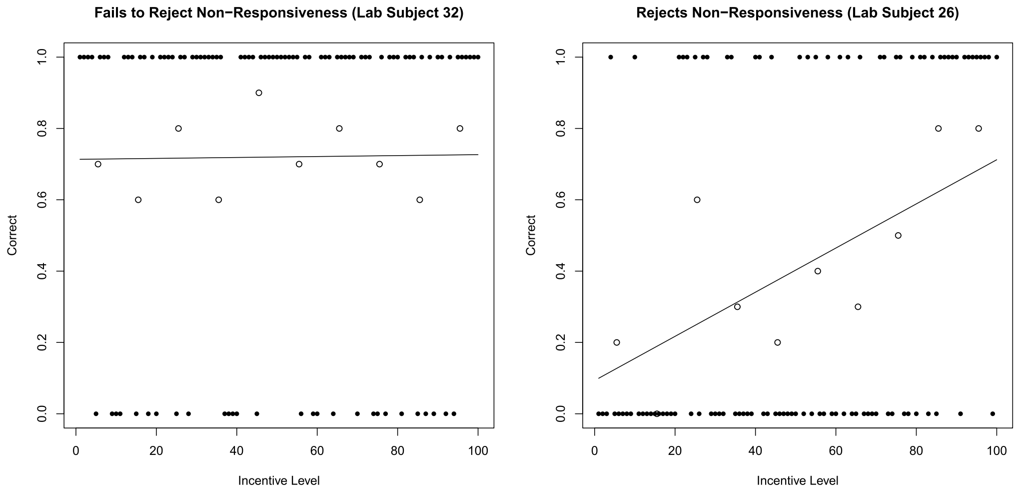

Fig. 12. Linear regressions of correctness on incentive level for two subjects. The left panel shows an unresponsive subject, and the right panel shows a responsive one. Circles indicate average success rate within each bin of 10 incentives.

non-positivity, i.e. non-responsiveness to incentives. However, this is insufficient to capture all responsive subjects; a subject may be responsive only within a small range of incentives. To address this issue, for each subject, we repeat this procedure on incentive levels 1 through 50 and on incentive levels 51 through 100.[20] If a subject has a significantly positive coefficient on incentive level in any of these three regressions, then we classify them as responsive.[21]

At the 5% significance level, 42 out of 70 lab subjects (60.0%) who fail to reject rational inattentiveness are responsive to incentives. Examples of full-sample linear regressions and correctness data for two subjects, one who fails to reject non-responsiveness and one who rejects non-responsiveness, are depicted in Fig. 12.

5.4. *Well-behavedness*

If the assumptions of Proposition 3 are satisfied, i.e. the cost function is well-behaved, then the performance function should be continuous in r. Therefore, observing a discontinuity in the performance function indicates a violation of convexity.

Strictly speaking, one cannot definitively observe a discontinuity without an infinite data set; a continuous function with a sufficiently steep slope at points of potential discontinuity can always be used to fit finite data. Therefore, for each subject, the question we wish to answer is whether it is more plausible that a discontinuous performance function or a continuous performance function generated their correctness data. This implies a statistical test where the null hypothesis is that the performance function belongs to some class of discontinuous functions, and the alternative is that the performance function belongs to some class of continuous functions.

20 Further sample splitting leads to the spurious detection of responsiveness; it leads to some subjects with >95% success being classified as responsive.

21 We must consider the full-sample regressions in tandem with the split-sample regressions. If we considered only the split-sample regressions, then we would classify subjects who have binary-response performance functions with thresholds around 50 as non-responsive.

Table 3
Categorization of subjects.

Category	Of All Subjects	Of R.I. Subjects	Of Resp. Subjects
All subjects	81 (100%)	—	—
R.I. subjects	70 (86.4%)	70 (100%)	—
Resp. subjects	///	42 (60.0%)	42 (100%)
W.B. subjects	///	///	13 (31.0%)

Note: "R.I." = rationally inattentive; "Resp." = responsive; "W.B." = well-behaved, i.e. subjects whose behavior is consistent with continuous, convex cost functions. — denotes that the column category is a subset of the row category, and /// denotes that the row category is defined only on a subset of the column category.

We test for the presence of a discontinuity by applying a likelihood ratio test. We estimate a step function of the form[22]:

$$P_t = \beta_0 + \beta_1 \mathbb{1}_{[\delta,\infty)}(r_t) \tag{16}$$

where β_0, β_1, and δ are the parameters to be estimated and compare its likelihood to an estimation of the following logistic relationship:

$$P_t = \frac{\beta_1}{1 + \exp(-\lambda(r_t - \delta))} + \beta_0 \tag{17}$$

In (16), δ is the location of the discontinuity, whereas in (17), it is a location parameter that determines the midpoint of the curve's upward sloping portion. It can be shown that (16) is the pointwise limit of (17) as λ goes to infinity. Therefore, (16) can be seen as the restricted null model, and a likelihood ratio test comparing these models is effectively a test of the null hypothesis that $\lambda = \infty$, i.e. it is a test against the null hypothesis that there is a jump discontinuity. Since we are performing this test only on responsive subjects, our estimates of β_1 for each subject should be positive, and therefore this procedure should not detect spurious downward jump discontinuities for those subjects.[23]

Using this test, at the 5% level we cannot reject that 29 out of 42 responsive lab subjects (69.0%) have discontinuities in their response functions.[24]

5.5. Summary of categorization

Table 3 summarizes the results of preceding subsections. Each cell indicates the number and percentage of row category subjects in the column category. It should be noted that the vast majority (86.4%) of subjects are rationally inattentive, and moreover, most rationally inattentive subjects are responsive (60.0%).

22 We use a variant of the procedure of Bai and Perron (1998) for this estimation, *imposing* a discontinuity and determining its location instead of using their algorithm to determine whether such a discontinuity is present.

23 Several procedures for detecting discontinuities have been proposed in the econometric literature. See, for example, Andrews (1993), Andrews and Ploberger (1994), Bai and Perron (1998), and Porter and Yu (2015). All of these procedures are designed to detect both positive and negative jump discontinuities, and so they are vulnerable to the detection of spurious negative jumps in our setting. A clarification is in order here. Bai and Perron (1998) propose both an estimation procedure and a testing procedure for models with structural breaks with unknown discontinuity points. We use their estimation procedure to estimate (16), but we do not use their testing procedure.

24 As a robustness check, we also ran this test at the 10% level to gain additional statistical power. In this case, we cannot reject that 27 out of 42 responsive lab subjects (64.2%) are not well-behaved.

Table 4
Performance functions estimated and their corresponding cost functions.

	Cost Function	Performance Function	Ref.
1	Very high or infinite marginal costs	Constant	N/A
2	Simple dual-process or concave	Binary	3.4/3.5
3	Hybrid dual-process	Affine with break	3.5
4	Quadratic in performance	Affine (without break)	3.1/3.2
5	Convex on the order of $P^{\frac{3}{2}}$	2nd degree polynomial	3.1
6	Convex on the order of $P^{\frac{4}{3}}$	3rd degree polynomial	3.1
7	Shannon mutual information	Logistic	3.2
8	Posterior-separable with Tsallis entropy	Sigmoid/inverse-S/concave (SIC)	3.2
9	Normal signals with linear cost of precision	Concave	3.3

Notes: The “Ref.” column indicates in which subsections of Section 3 the relevant theoretical treatment can be found. Model 1 is included as a robustness check.

6. Model selection

In this section, for each responsive subject we fit several possible parametric functional forms for performance functions, each of which can be generated by some cost function. These models are listed in Table 4.[25] Estimating equations, estimation methods, and mappings from estimated parameters to cost function parameters are detailed in Table 5. In contrast to existing experimental papers that estimate information costs on laboratory data (e.g. Pinkovskiy, 2009; Cheremukhin et al., 2015; Dean and Neligh, 2019), we estimate not just models that nest Shannon mutual information, but also models that do not, such as fixed costs for information (Model 2) and normal signal costs (Model 9).

Since the models are non-nested and are estimated using different methods, we cannot use a traditional auxiliary regression method for model selection. To determine which model is the best fit for each responsive subject, we estimate each model for each such subject and then compare their Akaike Information Criteria (AIC) (Akaike, 1974), selecting the model that yields the lowest AIC. The results of this selection are given in Table 6.[26]

All responsive subjects are best fit by binary (fixed costs), logistic (mutual information), SIC (Tsallis entropy costs) or concave (normal signals with linear precision cost) performance. The first implies some sort of non-convexity or discontinuity in the cost function, whereas the latter three are consistent with convex cost functions. Figs. 13, 14, 15, and 16 show what these performance functions look like for four subjects, each best fit by a different model.

Table 7 shows the average estimated AIC and rank of each model in the selection exercise.[27] Models 2 (binary), 7 (logistic), and 8 (SIC) have the lowest ranks on average. Flexible polynomial

[25] The reason that we do not consider the channel capacity cost function is because since the prior distribution in our task is uniform, channel capacity would be consistent with the same behavior as mutual information (cf. Section 1.2.3 of Woodford, 2012a).

[26] As a robustness check, we also perform the analysis with the small sample-corrected AIC (AICc), where AICc $= AIC + \frac{2k(k+1)}{T-k-1}$, T is the number of tasks, and k is the number of parameters in the model (Technically, the small sample correction should depend on the underlying model, but this particular correction formula is said to be appropriate for a wide variety of settings. For more information, refer to Subsection 7.4.1 of Burnham and Anderson, 2002). Our qualitative findings are completely unaffected. In particular, none of the subjects have a different best-fitting model under the AICc than under the AIC.

[27] A lower rank implies a lower AIC and therefore a better fit.

Table 5
Estimating equations for performance functions.

	Perf. F'n	Estimating Equation	Method	Estimated Cost Function
1	Constant	$P_t=\beta_0$	OLS	N/A
2	Binary	$P_t=\beta_0+\beta_1\mathbb{1}_{[\delta,\infty)}(r_t)$	BP98	$\hat{C}(P)=\begin{cases}0, & P\le\hat\beta_0\\ \hat\delta(\hat\beta_1-\hat\beta_0), & P\in(\hat\beta_0,\hat\beta_0+\hat\beta_1]\\ \infty, & P>\hat\beta_1\end{cases}$
3	Affine w/ break	$P_t=\beta_0+\beta_1\mathbb{1}_{[\delta,\infty)}(r_t)+\beta_2 r_t+\beta_3 r_t\cdot\mathbb{1}_{[\delta,\infty)}(r_t)$	BP98	$\hat{C}(P)=\begin{cases}0, & P\le\hat\beta_0\\ \frac{1}{2\hat\beta_1}(P-\hat\beta_0)^2, & \hat\beta_0<P\le\hat d_2\\ \frac{1}{2\hat\beta_1}(\hat d_2-\hat\beta_0)^2, & \hat d_2<P\le\hat\beta_0+\hat\beta_2\\ \frac{1}{2\left(\hat\beta_1+\hat\beta_3\right)}(P-(\hat\beta_0+\hat\beta_2))^2+\frac{1}{2\hat\beta_1}(\hat d_2-\hat\beta_0)^2, & P>\hat\beta_0+\hat\beta_2\end{cases}$ where $\hat d_2=\hat\beta_0+\sqrt{\hat\beta_1\hat\delta\left(\hat\beta_3\hat\delta+2\hat\beta_2\right)}$
4	Affine	$P_t=\beta_0+\beta_1 r_t$	WLS	$\hat{C}(P)=\frac{1}{2\hat\beta_1}(P^2-0.04)-\frac{\hat\beta_0}{\hat\beta_1}(P-0.2)$
5	2nd deg. poly.	$P_t=\beta_0+\beta_1 r_t+\beta_2 r_t^2$	WLS	$\hat{C}(P)=\frac{\hat\beta_1}{2\hat\beta_2}(0.2-P)\pm\frac{1}{24\hat\beta_2}\left(\left(4\hat\beta_2(P-\hat\beta_0)+\hat\beta_1^2\right)^{\frac{3}{2}}-\left(4\hat\beta_2(0.2-\hat\beta_0)+\hat\beta_1^2\right)^{\frac{3}{2}}\right)$ where the $+$ is taken if $\hat\beta_2>0$ and the $-$ is taken if $\hat\beta_2<0$
6	3rd deg. poly.	$P_t=\beta_0+\beta_1 r_t+\beta_2 r_t^2+\beta_3 r_t^3$	WLS	Omitted for legibility
7	Logistic	$P_t=\frac{1}{4\exp(-\frac{r_t}{\alpha})+1}$	MLE	$\hat{C}(P)=\hat\alpha\left[\ln(n)+P\ln(P)+(1-P)\ln\left(\frac{1-P}{n-1}\right)\right]$
8	SIC	$r_t+\frac{\alpha\sigma}{\sigma-1}\left[\left(\frac{1-P_t}{4}\right)^{\sigma-1}-P_t^{\sigma-1}\right]=0$	MLE	$\hat{C}(P)=\begin{cases}\frac{\hat\alpha}{\hat\sigma-1}\left[P^{\hat\sigma}+(n-1)^{1-\hat\sigma}(1-P)^{\hat\sigma}-n^{1-\hat\sigma}\right], & \hat\sigma\ne1\\ \hat\alpha\left[\ln(n)+P\ln(P)+(1-P)\ln\left(\frac{1-P}{n-1}\right)\right], & \hat\sigma=1\end{cases}$
9	Concave	$P_t=\frac{8}{5}\Phi\left(\frac{\zeta^*(r_t)}{2}\right)-\frac{3}{5}$, where $\alpha\zeta^*=\frac{2}{5}r_t\phi\left(\frac{\zeta^*}{2}\right)$	MLE	$\hat{C}(P)=\hat\alpha\breve{P}^{-1}(P)$, where $\breve{P}(\zeta):=\frac{1}{5}\left[2\Phi\left(\frac{1}{2}\zeta\right)+3\left(2\Phi\left(\frac{1}{2}\zeta\right)-1\right)\right]$

Notes on estimation: "BP98" = Bai and Perron (1998). Model 2 is estimated using a variant of their algorithm where a structural break is imposed rather than detected. Notes on nesting of models: Model 2 nests Model 1. Model 3 nests Models 2 and 4. Model 6 nests Model 5, which nests Model 4, which nests Model 1. Model 8 nests Model 7.

Table 6
Model Selection for Responsive Subjects.

Model	Binary (2)	Logistic (7)	SIC (8)	Concave (9)
Number of Subjects	10 (23.8%)	19 (45.2%)	7 (16.7%)	6 (14.3%)

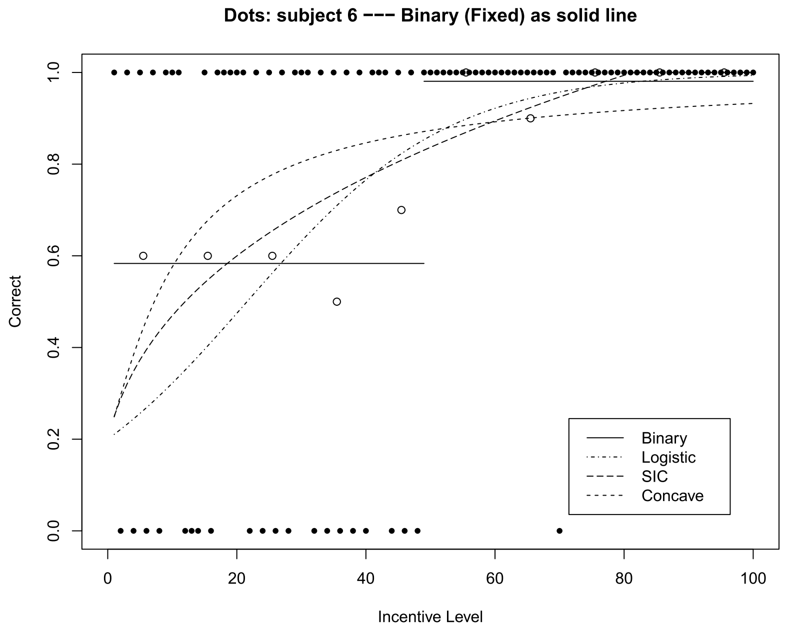

Fig. 13. Fits of binary, logistic, SIC, and concave performance for Subject 6, with the best-fitting binary (fixed cost) model as a solid line. Solid dots indicate correctness for each incentive. Circles indicate average success rate within each bin of ten incentives.

fits do quite poorly; the average rank of a cubic performance function (Model 6) is higher than that of the constant performance model (Model 1).

Note that the average AIC and rank for Model 7, the logistic performance function, is only slightly lower than that of Model 2, the binary performance function, despite the fact that substantially more subjects are best fit by Model 7 than by Model 2. This may be due to the binary model being a decent fit when the best-fitting model is logistic, but the logistic model being a poor fit when the best-fitting model is binary. When the logistic model is the best fit, the average rank of the binary model is 3.368; however, when the binary model is the best fit, the average rank of the logistic model is 4.600. Note also that the average rank of Model 9 (normal signals with linear precision cost) is fairly high at 5.119. This indicates that when Model 9 is not the best fit for a subject, it is a poor fit.

We remark that notably more subjects are classified as not well-behaved than are best fit by discontinuous performance functions. One possible explanation for this is that since our test for well-behavedness has discontinuity as its null hypothesis, strong evidence against discontinuity is required to classify a subject as well-behaved.[28] Another possible explanation is that some

[28] We expand upon this point and provide power calculations in Appendix Subsubsection A4.2.1.

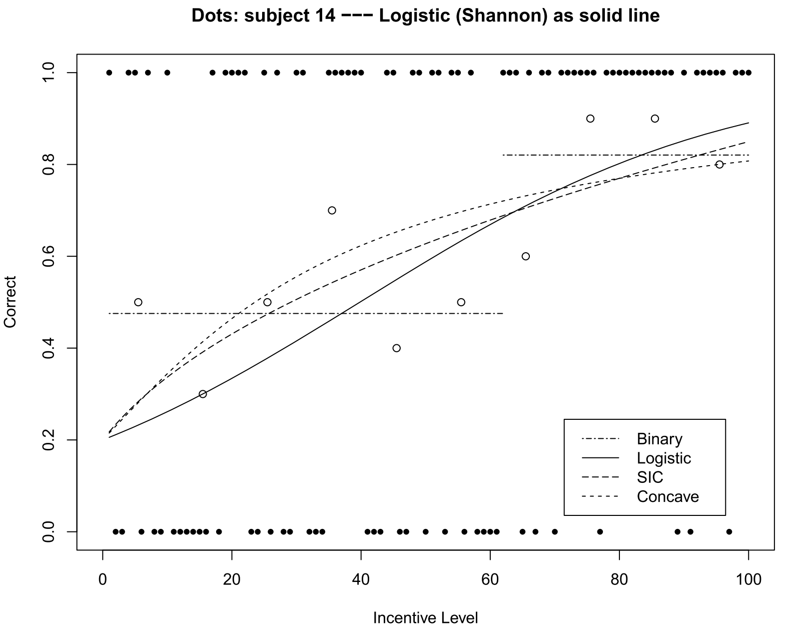

Fig. 14. Fits of binary, logistic, SIC, and concave performance for Subject 14, with the best-fitting logistic (Shannon) model as a solid line. Solid dots indicate correctness for each incentive. Circles indicate average success rate within each bin of ten incentives.

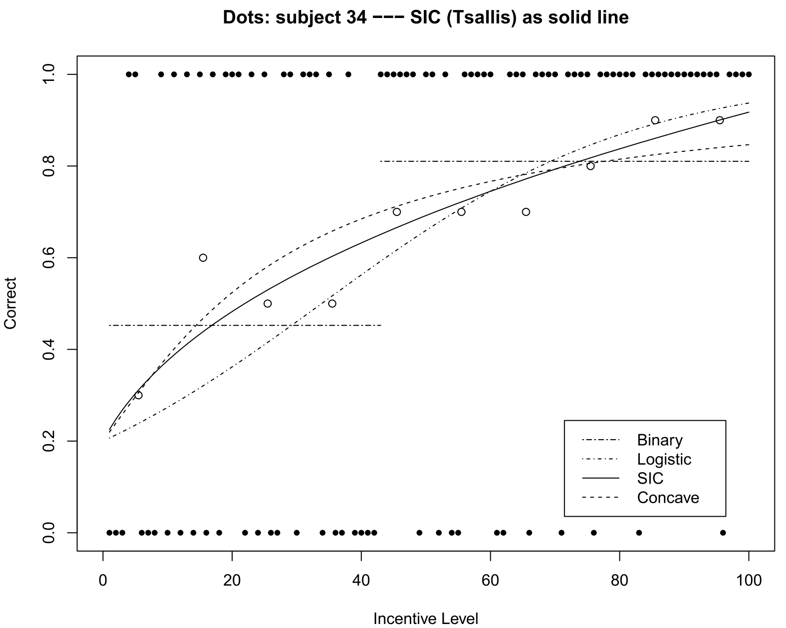

Fig. 15. Fits of binary, logistic, SIC, and concave performance for Subject 34, with the best-fitting SIC (Tsallis) model as a solid line. Solid dots indicate correctness for each incentive. Circles indicate average success rate within each bin of ten incentives.

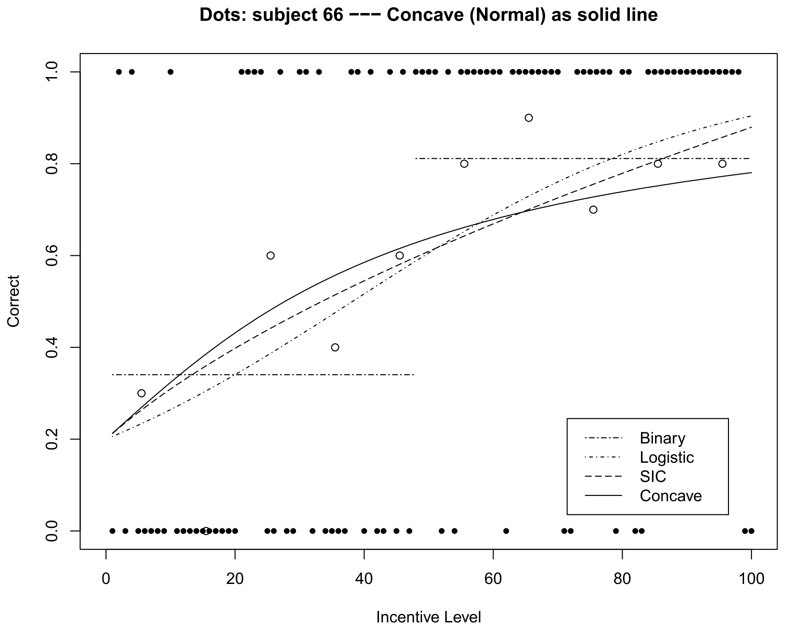

Fig. 16. Fits of binary, logistic, SIC, and concave performance for Subject 66, with the best-fitting concave (normal) model as a solid line. Solid dots indicate correctness for each incentive. Circles indicate average success rate within each bin of ten incentives.

Table 7
Average AIC and Rank for Estimated Models.

	Model	AIC	Rank
1	Constant	131.165	7.476
2	Binary	114.060	2.881
3	Affine with break	119.982	4.429
4	Affine	117.832	5.833
5	Quadratic	131.123	6.595
6	Cubic	132.929	7.643
7	Logistic	116.008	2.714
8	SIC	113.672	2.310
9	Concave	121.967	5.119

subjects' actual cost functions generate a discontinuous performance function with non-affine components that roughly approximates logistic or SIC performance. Distinguishing between these explanations based on our current data set would be difficult. Future experiments that use a different incentive structure could potentially accomplish this.[29]

7. Conclusion

This paper has provided a schema for testing properties of and estimating information cost functions in a rational inattention framework. To the extent that the presence or absence of char-

[29] We outline what such an incentive structure could look like in Appendix Subsubsection A4.2.3.

acteristics such as continuity and convexity can have an impact on people's decisions, it is worth knowing whether their cost functions satisfy such conditions. Decision-makers' cost functions are not directly observable, so instead we must infer their characteristics from observed behavior. We conducted a set of experiments that allowed us to implement tests of the properties of interest and perform a model selection exercise.

These experiments reveal substantial heterogeneity in behavior. Most subjects are rationally inattentive, but only about half are actually responsive to incentives. Many subjects have behavior that is consistent with continuous, convex cost functions, but a substantial fraction do not. Moreover, there is considerable heterogeneity in how subjects adjust their attention in response to incentives, though this heterogeneity is limited to four classes of cost functions: fixed costs, mutual information, Tsallis entropy costs (which nest mutual information), and normal signals are the only best-fitting cost functions for responsive subjects in terms of performance.

The fact that there is a significant presence of both binary performance and continuous performance functions in the population has important implications for economic modeling. In Supplementary Appendix S6, we present an application of rational inattention to a principal-agent framework of investment delegation and show that the principal's optimal payment schedule crucially depends on the shape of the agent's information cost function, and moreover, equilibrium robustness in this model relies on continuity; if an agent's information cost function is discontinuous, infinitesimal deviations from the optimal contract can lead to large welfare losses for the principal. Our experimental results also indicate that if a modeler wishes to use a single cost function for all agents for the sake of simplicity, then Tsallis costs, which have the lowest average rank and AIC, may be a good compromise, due to their flexibility.

Three possible avenues for future experimental research present themselves. The first is to obtain more detailed data on what subjects are actually paying attention to. Eyetracking has already been used in several economics experiments (e.g. Wang et al., 2010; Krajbich et al., 2010; Arieli et al., 2011) to track subjects' gaze, which allows researchers to find out what visual information the subjects are acquiring. Tracking subjects' mouse movements in computer-based tasks (e.g. Gabaix et al., 2006; Reeck et al., 2017) is another potential approach, since those movements indicate to which areas of their computer monitors they are paying attention. The second is to use choice data in tandem with reaction time data to fit models of dynamic information acquisition (e.g. Ratcliff and Smith, 2004; Clithero, 2018; Webb, 2018). This would also allow researchers to determine to what extent subjects trade off speed and accuracy in their decision-making. The third is to exploit the structure of the state space and data on subjects' mistakes to study how subjects perceive distance and dissimilarity (e.g. Natenzon, 2019; Pomatto et al., 2019; Chong et al., 2019).

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jet.2020.105011.

References

Admati, Anat R., Pfleiderer, Paul, 1988. A theory of intraday patterns: volume and price variability. Rev. Financ. Stud. 1 (1), 3–40.

Akaike, Hirotugu, 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19 (6), 716–723.

Allcott, Hunt, Taubinsky, Dmitry, 2015. Evaluating behaviorally motivated policy: experimental evidence from the lightbulb market. Am. Econ. Rev. 105 (8), 2501–2538.

Andrews, Donald W.K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61 (4), 821–856.
Andrews, Donald W.K., Ploberger, Werner, 1994. Optimal tests when a nuisance parameter is present only under the alternative. Econometrica 62 (6), 1383–1414.
Arieli, Amos, Ben-Ami, Yaniv, Rubinstein, Ariel, 2011. Tracking decision makers under uncertainty. Am. Econ. J. Microecon. 3 (4), 68–76.
Bai, Jushan, Perron, Pierre, 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66 (1), 47–78.
Barlevy, Gadi, Veronesi, Pietro, 2000. Information acquisition in financial markets. Rev. Econ. Stud. 67 (1), 79–90.
Bartoš, Vojtěch, Bauer, Michal, Chytilová, Julie, Matějka, Filip, 2016. Attention discrimination: theory and field experiments with monitoring information acquisition. Am. Econ. Rev. 106 (6), 1437–1475.
Bernard, Victor L., Thomas, Jacob K., 1989. Post-earnings-announcement drift: delayed price response or risk premium? J. Account. Res. 27, 1–36.
Birke, Melanie, Dette, Holger, 2007. Testing strict monotonicity in nonparametric regression. Math. Methods Stat. 16 (2), 110–123.
Bowman, A.W., Jones, M.C., Gijbels, I., 1998. Testing monotonicity of regression. J. Comput. Graph. Stat. 7 (4), 489–500.
Burnham, Kenneth P., Anderson, David R., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer.
Caplin, Andrew, Dean, Mark, 2014. Revealed preference, rational inattention, and costly information acquisition. Working paper.
Caplin, Andrew, Dean, Mark, 2015. Revealed preference, rational inattention, and costly information acquisition. Am. Econ. Rev. 105 (7), 2183–2203.
Caplin, Andrew, Martin, Daniel, 2017. Defaults and attention: the drop out effect. Rev. Écon. 68 (5), 747–755.
Caplin, Andrew, Dean, Mark, Leahy, John, 2019. Rationally inattentive behavior: characterizing and generalizing Shannon entropy. Working paper.
Chambers, Christopher P., Liu, Ce, Rehbeck, John, 2019. Costly information acquisition. Working paper.
Cheremukhin, Anton, Popova, Anna, Tutino, Antonella, 2015. A theory of discrete choice with information costs. J. Econ. Behav. Organ. 113, 34–50.
Chetty, Raj, Looney, Adam, Kroft, Kory, 2009. Salience and taxation: theory and evidence. Am. Econ. Rev. 99 (4), 1145–1177.
Chetverikov, Denis, 2019. Testing regression monotonicity in econometric models. Econom. Theory 35 (4), 729–776.
Chong, Edmund, Moroni, Monica, Wilson, Christopher, Shoham, Shy, Panzeri, Stefano, Rinberg, Dmitry, 2019. Manipulating synthetic optogenetic odors reveals the coding logic of olfactory perception. bioRxiv, https://www.biorxiv.org/content/early/2019/11/16/841916.
Clithero, John A., 2018. Improving out-of-sample predictions using response times and a model of the decision process. J. Econ. Behav. Organ. 148, 344–375.
De los Santos, Babur, Hortaçsu, Ali, Wildenbeest, Matthijs R., 2012. Testing models of consumer search using data on web browsing and purchasing behavior. Am. Econ. Rev. 102 (6), 2955–2980.
Dean, Mark, Neligh, Nathaniel, 2019. Experimental tests of rational inattention. Working paper.
DellaVigna, Stefano, 2009. Psychology and economics: evidence from the field. J. Econ. Lit. 47 (2), 315–372.
DellaVigna, Stefano, Pollet, Joshua M., 2007. Demographics and industry returns. Am. Econ. Rev. 97 (5), 1667–1702.
Dette, Holger, Neumeyer, Natalie, Pilz, Kay F., 2006. A simple nonparametric estimator of a strictly monotone regression function. Bernoulli 12 (3), 469–490.
Doveh, E., Shapiro, A., Feigin, P.D., 2002. Testing of monotonicity in parametric regression models. J. Stat. Plan. Inference 107 (1–2), 289–306.
Ehrmann, Michael, Jansen, David-Jan, 2017. The pitch rather than the pit: investor inattention, trading activity, and FIFA World Cup matches. J. Money Credit Bank. 49 (4), 807–821.
Gabaix, Xavier, Laibson, David, Moloche, Guillermo, Weinberg, Stephen, 2006. Costly information acquisition: experimental analysis of a boundedly rational model. Am. Econ. Rev. 96 (4), 1043–1068.
Gentzkow, Matthew, Kamenica, Emir, 2014. Costly persuasion. Am. Econ. Rev. 104 (5), 457–462.
Ghosal, Subhashis, Sen, Arusharka, van der Vaart, Aad W., 2000. Testing monotonicity of regression. Ann. Stat. 28 (4), 1054–1082.
Greiner, Ben, 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. J. Econ. Sci. Assoc. 1 (1), 114–125.
Grossman, Sanford J., Stiglitz, Joseph E., 1980. On the impossibility of informationally efficient markets. Am. Econ. Rev. 70 (3), 393–408.

Hall, Peter, Heckman, Nancy E., 2000. Testing for monotonicity of a regression mean by calibrating for linear functions. Ann. Stat. 28 (1), 20–39.

Hellwig, Christian, Kohls, Sebastian, Veldkamp, Laura, 2012. Information choice technologies. Am. Econ. Rev. 102 (3), 35–40.

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

Ho, Daniel, Imai, Kosuke, 2008. Estimating causal effects of ballot order from a randomized natural experiment: the California alphabet lottery, 1978–2002. Public Opin. Q. 72 (2), 216–240.

Hossain, Tanjim, Morgan, John, 2006. ...Plus shipping and handling: revenue (non) equivalence in field experiments on eBay. B.E. J. Econ. Anal. Policy 6 (2), 1–30.

Huberman, Gur, 2001. Familiarity breeds investment. Rev. Financ. Stud. 14 (3), 659–680.

Kahneman, Daniel, 2003. Maps of bounded rationality: psychology for behavioral economics. Am. Econ. Rev. 93 (5), 1449–1475.

Kaufman, E.L., Lord, M.W., Reese, T.W., Volkmann, J., 1949. The discrimination of visual number. Am. J. Psychol. 62 (4), 498–525.

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

Lacetera, Nicola, Pope, Devin G., Sydnor, Justin R., 2012. Heuristic thinking and limited attention in the car market. Am. Econ. Rev. 102 (5), 2206–2236.

Malmendier, Ulrike, Shanthikumar, Devin, 2007. Are small investors naive about incentives? J. Financ. Econ. 85 (2), 457–489.

Matějka, Filip, McKay, Alisdair, 2015. Rational inattention to discrete choices: a new foundation for the multinomial logit model. Am. Econ. Rev. 105 (1), 272–298.

McDevitt, Ryan C., 2014. "A" business by any other name: firm name choice as a signal of firm quality. J. Polit. Econ. 122 (4), 909–944.

Morris, Stephen, Yang, Ming, 2019. Coordination and continuous stochastic choice. Working paper.

Natenzon, Paulo, 2019. Random choice and learning. J. Polit. Econ. 127 (1), 419–457.

Pinkovskiy, Maxim L., 2009. Rational inattention and choice under risk: explaining violations of expected utility through a Shannon entropy formulation of the costs of rationality. Atl. Econ. J. 37 (1), 99–112.

Pomatto, Luciano, Strack, Philipp, Tamuz, Omer, 2019. The cost of information. Working paper.

Pope, Devin G., 2009. Reacting to rankings: evidence from "America's Best Hospitals". J. Health Econ. 28 (6), 1154–1165.

Porter, Jack, Yu, Ping, 2015. Regression discontinuity designs with unknown discontinuity points: testing and estimation. J. Econom. 189 (1), 132–147.

Ratcliff, Roger, Smith, Philip L., 2004. A comparison of sequential sampling models for two-choice reaction time. Psychol. Rev. 111 (2), 333–367.

Reeck, Crystal, Wall, Daniel, Johnson, Eric J., 2017. Search predicts and changes patience in intertemporal choice. Proc. Natl. Acad. Sci. 114 (45), 11890–11895.

Roth, Alvin E., Malouf, Michael W., 1979. Game-theoretic models and the role of information in bargaining. Psychol. Rev. 86 (6), 574–594.

Saltzman, I.J., Garner, W.R., 1948. Reaction time as a measure of span of attention. J. Psychol. 25 (2), 227–241.

Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27 (3), 379–423.

Shaw, Marilyn L., Shaw, Peter, 1977. Optimal allocation of cognitive resources to spatial locations. J. Exp. Psychol. Hum. Percept. Perform. 3 (2), 201–211.

Shue, Kelly, Luttmer, Erzo, 2009. Who misvotes? The effect of differential cognition costs on election outcomes. Am. Econ. J. Econ. Pol. 1 (1), 229–257.

Sims, Christopher A., 2003. Implications of rational inattention. J. Monet. Econ. 50 (3), 665–690.

Sims, Christopher A., 2006. Rational inattention: beyond the linear-quadratic case. Am. Econ. Rev. 96 (2), 158–163.

Smith, Stephanie M., Krajbich, Ian, 2019. Gaze amplifies value in decision making. Psychol. Sci. 30 (1), 116–128.

Stanovich, Keith E., West, Richard F., 2000. Individual differences in reasoning: implications for the rationality debate? Behav. Brain Sci. 23 (5), 645–665.

Steiner, Jakub, Stewart, Colin, 2016. Perceiving prospects properly. Am. Econ. Rev. 106 (7), 1601–1631.

Steiner, Jakub, Stewart, Colin, Matějka, Filip, 2017. Rational inattention dynamics: inertia and delay in decision-making. Econometrica 85 (2), 521–553.

Tsallis, Constantino, 1988. Possible generalization of Boltzmann-Gibbs statistics. J. Stat. Phys. 52 (1–2), 479–487.

Van Nieuwerburgh, Stijn, Veldkamp, Laura, 2010. Information acquisition and under-diversification. Rev. Econ. Stud. 77 (2), 779–805.

Verrecchia, Robert E., 1982. Information acquisition in a noisy rational expectations economy. Econometrica, 1415–1430.

Wang, Joseph Tao-yi , Spezio, Michael, Camerer, Colin F., 2010. Pinocchio's pupil: using eyetracking and pupil dilation to understand truth telling and deception in sender-receiver games. Am. Econ. Rev. 100 (3), 984–1007.

Webb, Ryan, 2018. The (neural) dynamics of stochastic choice. Manag. Sci. 65 (1), 230–255.

Woodford, Michael, 2012a. Inattentive valuation and reference-dependent choice. Working paper.

Woodford, Michael, 2012b. Prospect theory as efficient perceptual distortion. Am. Econ. Rev. 102 (3), 41–46.