
EXPERIMENTAL EVIDENCE ON RATIONAL INATTENTION

ANTON CHEREMUKHIN, ANNA POPOVA,
AND ANTONELLA TUTINO

**RESEARCH DEPARTMENT
WORKING PAPER 1112**



Federal Reserve Bank of Dallas

Experimental Evidence on Rational Inattention

Anton Cheremukhin Anna Popova Antonella Tutino*

December 13, 2011

Abstract

We show that rational inattention theory of Sims (2003) provides a rationalization of choice models à la Luce and gives a structural interpretation to probability curvature parameters as reflecting costs of processing information. We use data from a behavioral experiment to show that people behave according to predictions of the theory. We estimate attitudes to risk and costs of information for individual participants and document overwhelming heterogeneity in these parameters among a relatively homogeneous sample of people. We characterize, both theoretically and empirically, the aggregation biases this heterogeneity implies and find these biases to be substantial.

JEL: D81, D03, C91, C44.

Keywords: Discrete Choice, Behavioral Experiments, Shannon's Information Theory, Rational Inattention.

*Anton Cheremukhin: Federal Reserve Bank of Dallas, 2200 N Pearl St, Dallas TX 75201, chertosha@gmail.com, 214-922-6785. Anna Popova: University of Illinois at Urbana-Champaign, 603 East Daniel St, Champaign, IL 61820, apopova2@illinois.edu. Antonella Tutino: Federal Reserve Bank of Dallas, 2200 N Pearl St, Dallas TX 75201, tutino.antonella@gmail.com, 214-922-6804. We are grateful to Michel Regenwetter for access to the data, encouragement and helpful comments. We also thank Tony Marley and Duncan Luce for questions and suggestions which helped improve the draft of the paper. All remaining errors are our own. Anna's work and data collection were supported by National Science Foundation grant SES # 08-20009 (PI: M. Regenwetter, University of Illinois at Urbana-Champaign), entitled *A Quantitative Behavioral Framework for Individual and Social Choice*, awarded by the Decision, Risk and Management Science Program. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of their colleagues, the National Science Foundation, the University of Illinois, the Federal Reserve Bank of Dallas or the Federal Reserve System. First draft: July 15, 2011.

1 Introduction

When a problem is so involved that no rational formulation is available, then some quantification is still possible by the coefficients of correlation of contingency and the like. But such statistical procedures constitute an acknowledgement of failure to rationalize the problem and to establish functions that underlie the data.

– Louis Leon Thurstone (1929), p. 224

There has been a long standing divide between rational choice theory and behavioral decision theory. Rational choice theory postulates that people are rational optimizing agents capable of making decisions quickly and precisely even in an uncertain environment. Behavioral decision theory describes people's behavior by documenting empirically observed choices. Oftentimes, behavioral decision literature has challenged the predictions of rational choice literature. In particular, behavioral experiments seem to contradict the view that every time and everywhere people optimally choose their actions. In a classical study Mosteller and Nogee (1951) showed that people are not as consistent in their choice as rational choice theory would predict.¹ To model this phenomenon several theories have been proposed.

Louis Thurstone's (1927) discriminial process theory explains the apparent behavioral variations by associating a response strength to each choice option and by assuming that response strength oscillates over time leading to variations in observed behavior. Similarly, a generalization of this approach to more than two alternatives, random utility theory, developed among others by Block and Marschak (1960), Yellott (1977) and Falmagne (1978), states that people change their attitude towards alternatives in a random fashion. We can infer from the epigraph above that Thurstone thought of it as somewhat unsatisfactory since it characterizes statistical properties of observed behavior without really explaining why such behavior took place.

Another prominent theory proposed by Duncan Luce (1959) assumes that response strengths are constant while there is variability in the choice process. Luce's choice theory postulates a relationship between response strength and the corresponding probability of choice based on the property of independence from irrelevant alternatives. While leading to parsimonious functions describing choice probabilities, this approach has been criticized on several grounds. The two main objections are that the predicted probabilities are very sensitive to the order in which possible irrelevant alternatives are eliminated (see Saari (2005)), and that the implied curvature parameters are hard to interpret. Nonetheless, this second theory has been extensively used for modeling various microeconomic phenomena, from choice of geographical location in McFadden (1978) to choice under risk in Holt and Laury (2002) and Birnbaum (2008) among others. There are numerous examples in the psychology and economics literatures

¹For discussions of these issues see among others Block and Marschak (1960), Barbera and Pattanaik (1986), Simon (1986), Rieskamp, Busemeyer and Mellers (2006), and Regenwetter et. al. (2010).

that highlight the need for a theoretical framework capable of rationalizing the empirical finding that outcomes of discrete choice are probabilistic.²

The aim of this paper is to fill the void left by the theory in explaining these experimental results. In particular, this paper proposes a way to reconcile the predictions of rational choice theory with the observed choices described by the behavioral decision literature. The theory we propose is based on Rational Inattention Theory of Sims (2003). At the core of rational inattention is the idea that people have limited cognitive capacity to process information. Knowing that they have limited ability to process information useful for decision-making, they choose to pay attention to those pieces of information that are most relevant to their utility and disregard the rest. The selection of useful information is dictated by their utility.

To measure the amount of information that people can process in a given unit of time, we use the analogy that people's brains and their cognitive ability of processing information can be thought of as akin to an information processing device, such as a computer or a telephone line. These devices can bring a limited amount of information from a sender to a receiver in a finite period of time. For instance, we rely on the transmission rate of the computer to load the content of an email. Similarly, we rely on our brain to understand the content of the email and, possibly, produce a reply. As most information processing devices, the human brain cannot absorb nor react to information infinitely quickly and precisely. Hence, people's choices are prone to error. By error we mean that given two alternatives, A and B, the person might choose alternative A over alternative B even though B yields higher value than A.

We model a person's limited ability to process information using Shannon's (1948) information theory and, in particular, Shannon's channel capacity.³ In our model, people choose their preferred amount of information and allocation of attention (jointly called information structure) under the constraint that they have limited capacity to process information. We capture the idea that choosing an information structure is costly by associating a utility cost to the capacity of the channel, measured by the amount of information intrinsic to the chosen information structure.

This cost represents the effort that people put in processing the chosen information structure. For instance, in the example with an email, a person may want to select how carefully and precisely he wants to respond to an email depending on how important the email is. The more he cares about producing a well crafted response, the more attention he needs to pay to the content of the email. Thus, the person chooses how much information in the email he wants to process to map it into his email response.

²A by no means comprehensive list of applications of Luce's choice model and related random utility models includes choices of geographical location, migration, housing, college, occupation, labor force participation, number of children, number and brand of automobile, shopping location, recreation location, means of travel and travel destination. For a lengthy discussion of these models and their use see Cramer (2003).

³For a discussion of uses of information and knowledge and other related issues see Section 7.

Given that processing information is costly, a rational decision-maker faces a trade-off between the costs and the benefits of processing pieces of information. He selects the bits of information that are the most relevant for his utility and ignores the rest, accepting potential errors stemming from this disregard of information. So long as these errors are taken into account when choosing the optimal information structure, a rationally inattentive agent lives comfortably in the optimizing framework postulated by rational choice theory. Furthermore, he may behave in a way that is consistent with the choices documented by the behavioral decision literature.

From a theoretical standpoint, we find that rational inattention theory (1) provides a rationalization of choice models à la Luce;⁴ (2) provides a structural interpretation to probability curvature parameters as reflecting costs of processing information; (3) augmented by a theory of value, allows for joint structural estimation of attitude to risk and costs of processing information using experimental data.

These theoretical results are important for at least two reasons. First, when applied to experimental data they make it possible to establish a direct link between microeconomic estimates of costs of information processing and macroeconomic predictions of models with rationally inattentive agents. These estimates could be useful for calibrating macroeconomic models based on rational inattention theory. Since to date rational inattention theory has been lacking direct empirical evidence, our estimates themselves constitute an empirical contribution.

Second, it may seem that choice models à la Luce have a vast body of empirical support and thus parameter estimates from existing microeconomic studies could be routinely used in macroeconomic models. However, a caveat is in order. In the overwhelming majority of empirical studies, theories of choice have been studied by analyzing responses pooled across subjects.⁵ In principle, this approach cannot distinguish between two cases. On the one hand, all people might face costs of processing information and individually behave according to the predictions of rational inattention theory. On the other hand, the same data could be generated by people with different preferences all facing no cost of processing information whatsoever. When pooling responses across subjects one could equally well get an aggregate response reminiscent of that predicted by Luce's choice model. Thus, it is nearly impossible to distinguish between these two extreme cases when only pooled responses are observed.

This possibility calls for a test of predictions of rational inattention theory at the individual level. To experimentally test the predictions of rational inattention theory and measure costs of processing information, experiments should

⁴The relationship between rational inattention theory and the logit model of discrete choice has been recently independently discovered by Matějka and McKay (2011). They study a special case of our setup with non-stochastic choice options. Our results apply to a more general set of probabilistic choice options such as gambles.

⁵Hey and Orme (1994), Harless and Camerer (1994), Holt and Laury (2002) and Birnbaum (2008) are prominent examples among this majority, while Regenwetter et. al. (2011) belongs to rare exceptions.

be deliberately designed to set people in identical situations repeatedly, allowing the researcher to check whether the predicted choice frequencies can explain behavior at the individual level.

The empirical contribution of this paper is to use such a behavioral experiment to measure costs of processing information at the individual level. We use data collected in Michel Regenwetter's laboratory at the University of Illinois at Urbana Champaign to provide empirical support to our theoretical predictions that individuals: a) behave probabilistically; b) pay less attention and are thus more prone to error when differences between choice options are smaller.

The experimental design allows us to jointly estimate people's attitude to risk and costs of information processing at the individual level using likelihood methods. It allows us to characterize the role played by information-processing constraints and across-person heterogeneity in explaining variations in behavior at the aggregate level. We document a striking amount of heterogeneity in both dimensions. Both the cost of information processing and attitude to risk vary by orders of magnitude within a relatively homogeneous sample of participants.

Using our estimates, we document how heterogeneity leads to a substantial aggregation bias, i.e. a difference in the behavioral response of an average agent from our sample and the properties of a fictional "representative agent", characterized by the behavioral response aggregated across all subjects in our sample. We use Jensen's inequality to prove theoretically that we should expect this "representative agent" to be characterized by more inattentive behavior than the average individual. This suggests that the direction of aggregation bias in the cost of information observed in the laboratory is a consequence of heterogeneity over costs of processing information.

Recently, Luce (2010) has emphasized heterogeneity in risk aversion as a pivotal link between theoretical and empirical findings. Drawing from a vast literature in psychology and economics,⁶ Luce (2010) makes the case for partitioning individuals into different risk types. He argues that such a partition might be useful to empirically evaluate and compare the theoretical predictions of models of utility representation, in particular cumulative prospect theory of Tversky and Kahnemann (1992). Our model based on rational inattention theory delivers structural forms for probability distributions that we use to assess and compare theories of value such as expected utility (EU), cumulative prospect theory (CPT) and the transfer of attention exchange (TAX) model.

The paper is organized as follows. Section 2 provides a discussion of rational inattention as a theory of choices. A simple numerical example illustrates how the theory works. Section 3 introduces our model and characterizes the optimal choice probability. Section 4 describes the experiment and the methodology applied to experimental data. Section 5 shows the results and compares different theories of values. Section 6 characterizes, both empirically and theoretically, the aggregation bias resulting from observed heterogeneity. Section 7 discusses alternative approaches to modeling information in behavioral decision theory

⁶See, for instance, Tversky and Kahneman (1992), Marley and Luce (2005) and Birnbaum (2008), Regenwetter et al. (2011).

and in the context of our model. Section 8 concludes.

2 Rational Inattention as a Theory of Error

Consider a choice between two options, A and B . Figure 1 displays on the vertical axis the probability of choosing option A whereas on the horizontal axis it displays the value differential between the two options. The dashed line shows the predictions of the model according to existing rational choice theories and the solid line represents a typical observed mean response averaged across a sample population in an experimental setting approximated by Luce’s choice model.

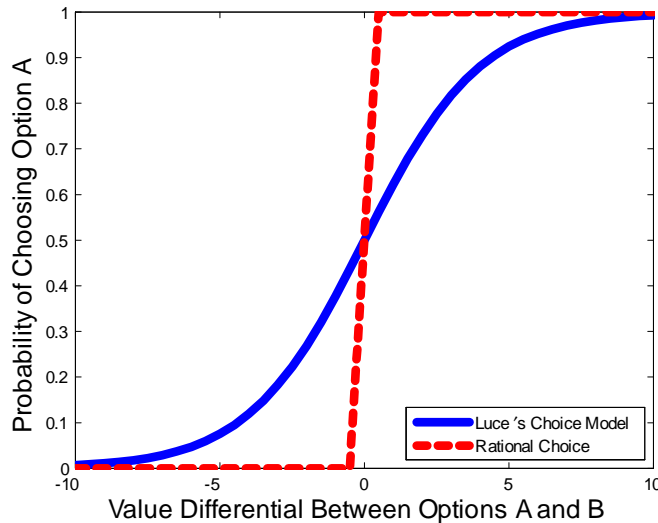


Figure 1. Data averages and rational theory predictions.

The observed frequency of choosing option A is at odds with predictions of existing rational theories: in reality, people do make errors. The frequency of making an error depends on the difference in values of the two options. The bigger this difference, the more consistent are people in their choices. The goal of this paper is to construct a model where the observed frequency of choosing option A can be fully taken into account as an outcome of rational choice.

To explain the observed choice frequencies we use insights from the rational inattention literature of Sims (2003), (2006), Tutino (2011).⁷ We think that this framework can help explain individual and aggregate deviations from the rational choice paradigm documented by the behavioral decision literature.

⁷An early description of the ideas behind rational inattention theory can be found in Sims (1998). An accessible exposition of rational inattention theory can be found in Wiederholt (2010).

Rational inattention theory is based on (1) a statistical measure of uncertainty intrinsic to choice between options and (2) people's value of those options as determined by their utility. The measure of uncertainty of the options is based on Shannon's (1948) entropy. Entropy is measured in bits of information. This measure is parsimonious, yet general, as it depends solely on the probabilistic nature of the options. Shannon's channel capacity (1948) is used to measure the amount of information processed when choosing between options. The amount of information processed through the channel, equal to the reduction in uncertainty achieved through the choice process, depends only on the joint distribution between the choice options presented to the decision maker and his behavior. This reduction in uncertainty is also measured in bits of information.

The idea that processing information is costly is captured in rational inattention theory by associating a utility cost to the capacity of the channel. This cost represents the effort that people put in processing information. Thus, rational inattention theory augments the standard theory of value, which attributes utilities to different choices, with a utility cost associated with processing information.

According to rational inattention theory, the decision maker faces a trade-off between the benefits of choosing the best option precisely and the costs of identifying the best option. The key difference from previous rational theories comes from the fact that rational inattention theory allows the decision maker to rationally choose how much information to process and maps this choice into his choice frequency. The decision maker is able to select the pieces of information that are the most relevant for his utility and ignore the rest. So long as potential errors stemming from this disregard of information are taken into account by the decision-maker, inattentive behavior is a natural outcome of the optimizing framework postulated by rational choice theory.

We use these ideas from rational inattention theory to model choices between discrete probabilistic options. The main prediction of our model is that decision makers with limited ability to process information choose information about options according to how much they care about each specific option. If they perceive that the difference between the options is so small that it is not worth to pay close attention, they will be indifferent between the two options and choose randomly. The more they care about one option over another -say, option A over option B-, the more attention they will pay. As a result, they will more frequently choose option A over option B. In this respect, one can think of rational inattention as a rationale for people's errors. Our model predicts that people pay less attention and, thus, make errors more frequently when they care less about the options.

To see the intuition behind our application of rational inattention theory, consider an illustrative example. Two cards are randomly drawn from a deck and placed on the table. A player's payoff is determined by the dollar values written on the cards and by his decision of which of the two cards to pick up. The player is allowed to pick up only one of the cards, the one on the left or the one on the right, which will determine his payoff. For simplicity we assume

that the player is risk-neutral.

We can interpret the idea that the player does not know the values of the cards ex ante by thinking of them in a probabilistic sense. For instance, consider a deck with an equal number of three types of cards with payoffs of \$18, \$9 and \$0 respectively written on them. We assume that the player knows the payoffs and the probability distribution of the cards. We can interpret the idea that the player needs to process information before making his decision, as if before processing information he perceives the cards to be face down.

Before picking up a card, he can choose to process some information about the outcome by flipping some of the cards. He can take a look at none of the cards (strategy $S=0$), or just at one card (strategy $S=1$), or at both cards (strategy $S=2$). The more cards he flips the larger the amount of information he will be processing.

We define the amount of information processed as the reduction in uncertainty associated with the probability distribution over the potential payoffs. As we shall see, each of the strategies $S \in \{0, 1, 2\}$ is associated with a probability distribution over payoffs $[18, 9, 0]$. We denote this probability distribution $[p_{18}, p_9, p_0]$, where all probabilities are non-negative and sum up to one. Note that each strategy S implies different probability distributions. The uncertainty associated with each strategy S is equal to the entropy, denoted $H(S)$, of the corresponding probability distribution. Entropy is computed as follows:

$$H(S) = p_{18} \log_2 p_{18} + p_9 \log_2 p_9 + p_0 \log_2 p_0.$$

Ex ante uncertainty, equivalent to strategy $S = 0$, is represented by a uniform distribution. When all cards have the same probability $\frac{1}{3}$ the corresponding value of entropy equals $H(0) = \log_2 \frac{1}{3}$. The amount of information processed for each strategy S equals the reduction in uncertainty associated with strategy S net of ex ante uncertainty, $H(0)$. We denote $I(S)$ the amount of information processed when playing strategy S . It is computed as follows:

$$I(S) = H(S) - H(0) = p_{18} \log_2 p_{18} + p_9 \log_2 p_9 + p_0 \log_2 p_0 - \log_2 \frac{1}{3}.$$

We assume that the player faces a cost of θ dollars per bit of information to process. The player ex-ante plans which card (left or right) to pick up conditional on his strategy S and the corresponding amount of information that he gets. Thus, the player faces a tradeoff between the expected payoff and the cost of processing information. His optimal strategy reflects the optimal amount of information to process and may result in not always picking the card with the highest payoff.

Following the rational inattention literature, we define the expected value associated with strategy S as the difference between the expected payoff and the cost of information:

$$V(S) = 18p_{18} + 9p_9 + 0p_0 - \theta I(S).$$

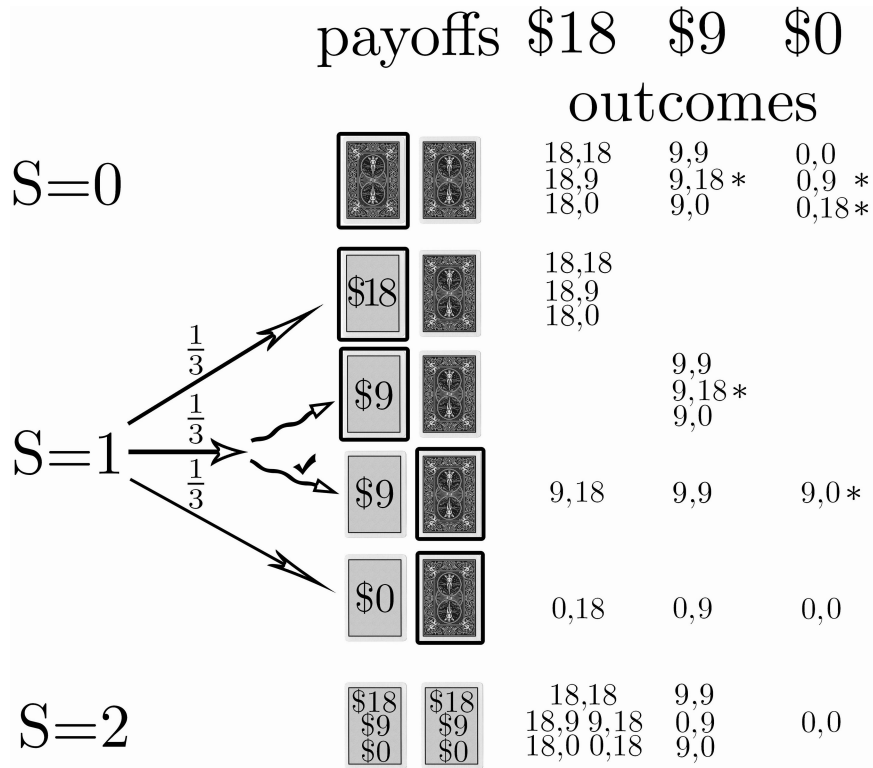


Figure 2. Decision Tree of the Game.

Let us consider each potential strategy in turn, derive the corresponding probability distribution and use it to compute the expected value of the strategy. Figure 2 describes the tree of consequences for each strategy illustrating the corresponding outcomes and potential payoffs. Strategies are denoted $S=0$, $S=1$ and $S=2$ on the left. In each case the card which is picked up is highlighted by a box. Columns on the right represent final payoffs (\$18, \$9 and \$0 at the top) and pairs of cards which would lead to those payoffs under each strategy. For instance, under strategy $S=0$ the card on the left⁸ is picked up (highlighted). For example, if the card on the left is 0, and the card on the right is 0, 9 or 18, the player gets \$0 (right-most column). Outcomes when the card with a strictly lower payoff is picked up are marked with stars in Figure 2.

We shall start with two extremes. First, consider the strategy of flipping none of the cards (strategy $S=0$). In this case the player can randomly pick one of the cards, which gives him a uniform distribution over payoffs $[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}]$. In this case the probability of making an error, i.e. picking the card which yields a strictly lower payoff is $\frac{1}{3}$. The associated amount of information is zero, and the expected value of this strategy is:

⁸We could assume either left or right here, without loss of generality.

$$V(0) = \frac{1}{3}18 + \frac{1}{3}9 + \frac{1}{3}0 - 0\theta = 9.$$

Second, consider the other extreme of flipping both cards (strategy S=2). In this case the player can always pick up the better one of the two. The probability of making an error is zero. This strategy gives the player \$18 in 5 cases out of 9, \$9 in 3 cases out of nine, and leaves him with \$0 only if both cards contain \$0. Therefore, the associated expected value is:

$$V(2) = \frac{5}{9}18 + \frac{3}{9}9 + \frac{1}{9}0 - \theta \left(\frac{5}{9} \log_2 \frac{5}{9} + \frac{3}{9} \log_2 \frac{3}{9} + \frac{1}{9} \log_2 \frac{1}{9} - \log_2 \frac{1}{3} \right) = 13 - 0.23\theta.$$

Finally, consider the strategy of flipping only one of the cards (strategy S=1) and the associated conditional choices. Without loss of generality, let the card be the one of the left. In case this card is \$18, it is always better to pick it up and get \$18 for sure. In case the card on the left is \$0, it is better to pick up the other card, as the prospect of getting something is better than the prospect of getting nothing for sure.⁹ In case the card on the left gives \$9, the expected payoffs on the left and on the right are equal, but picking up the card on the right is associated with higher uncertainty. This implies a lower amount of information processed, and hence a higher expected value. As a result, in case the card on the left gives \$9, it is better to pick up the card on the right.

By multiplying the probabilities of getting each of the cards on the left by the conditional probabilities associated with picking left or right in each case, we arrive at the probability distribution $\left[\frac{5}{9}, \frac{2}{9}, \frac{2}{9}\right]$ associated with the strategy of looking at just one card:

$$\frac{1}{3}[1, 0, 0] + \frac{1}{3}\left[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right] + \frac{1}{3}\left[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right] = \left[\frac{5}{9}, \frac{2}{9}, \frac{2}{9}\right].$$

The expected value of strategy $S = 1$ is then:

$$V(1) = \frac{5}{9}18 + \frac{2}{9}9 + \frac{2}{9}0 - \theta \left(\frac{5}{9} \log_2 \frac{5}{9} + \frac{2}{9} \log_2 \frac{2}{9} + \frac{2}{9} \log_2 \frac{2}{9} - \log_2 \frac{1}{3} \right) = 12 - 0.15\theta.$$

Figure 3 compares the expected values associated with each of the strategies. It makes clear that different values of information cost will lead to different optimal strategies. Table 1 summarizes for each of the three strategies, the corresponding probability distribution over payoffs, the expected value of the strategy, the range of costs of information when the strategy is optimal and the probability of picking the wrong card.

⁹Note that in case the cost of information is so high that the extra uncertainty associated with a gamble on the right is higher than the incremental expected payoff of \$9 on the left, strategy S=0 also strictly dominates strategy S=1.

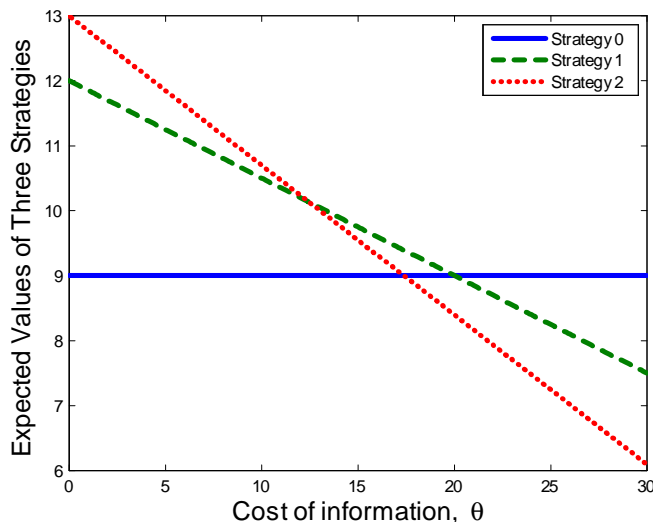


Figure 3. Expected Values of the Strategies

Strategy	Prob Dist	Info	Value	Cost of Info	P(error)
0	$\frac{3}{9}, \frac{3}{9}, \frac{3}{9}$	0	9	$\theta > 20.1$	$\frac{1}{3}$
1	$\frac{2}{9}, \frac{2}{9}, \frac{2}{9}$	0.15	$12 - 0.15\theta$	$11.9 < \theta \leq 20.1$	$\frac{1}{9}$
2	$\frac{3}{9}, \frac{3}{9}, \frac{1}{9}$	0.23	$13 - 0.23\theta$	$0 \leq \theta \leq 11.9$	0

Table 1. Properties of the Strategies

Our example illustrates three points. First, a lower cost of processing information leads to more information being processed. Second, a lower cost of information is associated with a lower probability of making an error. Third, even when it is feasible to eliminate errors completely and make a fully informed choice, it may be optimal to leave room for error if information processing is costly.

The behavior of the player in our example illustrates a way a person could process information about any two options. He could be playing a mind game of asking himself questions about the two options, which will give him pieces of information necessary to make the choice. Oftentimes, the person would be able to economize on time and effort instead of determining precisely which option is better. He would economize more in situations when the outcomes are closer to each other.

Note that in our example by asking himself other types of questions like "is the left card 18\$?" or, more generally, by optimally choosing the perceived distribution of payoffs written on the cards, the player is in principle able to break available information into even finer pieces. A more general cognitive process which could represent decision making in this case could be described as follows.

Before processing any information about choice options, the decision maker represents uncertainty about the options in the form of some probability distribution over payoffs. He evaluates the options starting with this prior distribution. One can think of the process of evaluation as a randomly drawn sequence of attributes that the person pays attention to. Evaluation using each attribute gives the decision maker new information about the options. He uses this information to update his prior and obtain a tighter posterior at each step.

Paying attention to every next attribute involves as much effort. Meanwhile each attribute makes the posterior tighter lowering the information gain from the next attribute. At some step the decision maker will feel that thinking more about the options is counterproductive. This will happen when the benefit from drawing another attribute and tightening the posterior even further is not worth the cognitive effort of thinking of another attribute. At this point the person will stop and pick the option with the highest weight in his current posterior.

Of course, the cognitive representation we just described is by no means unique. However, our example illustrates that the brain works as an information processing device. Independent of the particular cognitive process, it will have the features of Shannon's channel and will be subject to constraints placed by Shannon's information theory.

We use rational inattention theory, which relies on the analogy with Shannon's channel to describe the observed outcome without going into the details of particular representations of gambles and attributes and their optimal choice in the mind of a person. We follow the rational inattention literature in adopting the assumption that coding of information is done efficiently, while processing information through the channel is costly. By coding we mean the process through which people 'read' the information presented to them. We expand on coding issues in Section 7.

Rational inattention theory provides a general framework for describing choices when processing information is costly. It does so abstracting from the details of coding information through specific questions and payoff distributions. In the next section we use rational inattention theory to describe a general model of discrete choice among gambles and derive the optimal choice probabilities.

3 Theoretical Framework

This section formally establishes the theoretical environment of the paper. Consider a decision-maker faced with a choice among K gambles. Each gamble $k \in \{1, \dots, K\}$ has J_k possible outcomes with X_{kj} representing payoffs and p_{kj} respective probabilities:

$$k : \quad \{X_{kj}, p_{kj}\}_{j=1}^{J_k}, \quad \sum_{j=1}^{J_k} p_{kj} = 1, \quad j \in \{1, \dots, J_k\}.$$

Following rational inattention theory, we allow the decision-maker to choose a probability distribution $\{s_k\}$ over options, where s_k denotes his probability of choosing gamble k :

$$\{s_k\}_{k=1}^K, \quad \sum_{k=1}^K s_k = 1, \quad k \in \{1, \dots, K\}.$$

We also allow the decision-maker to have a prior bias towards one option against any other, which may stem from the way these gambles are presented. We denote the decision-maker's prior distribution over options, $\{s_{k0}\}$, defined in the same way as $\{s_k\}$.

The amount of information processed by the decision-maker reflected in his choice of s_k is captured by Shannon's relative entropy, which captures the distance between the prior distribution over payoffs, $\{s_{k0}p_{kj}\}$, and the posterior distribution over payoffs, $\{s_k p_{kj}\}$. We denote the amount of information, I , which is an implicit function of the chosen probability distribution. The amount of information measured by the relative entropy is computed as follows:

$$I(s_k) = \sum_{k=1}^K \sum_{j=1}^{J_k} s_k p_{kj} \log_2 \left(\frac{s_k p_{kj}}{s_{k0} p_{kj}} \right) \quad (1)$$

Shannon's relative entropy can be interpreted as a reduction of uncertainty about the gambles that the decision-maker can achieve by choosing s_k . The more information he is able to process, the larger the reduction of uncertainty.

Lemma 1 *The amount of information, I , is independent of gamble probabilities p_{kj} .*

Proof.

$$I(s_k) = \sum_{k=1}^K \sum_{j=1}^{J_k} s_k p_{kj} \log_2 \left(\frac{s_k}{s_{k0}} \right) = \sum_{k=1}^K s_k \log_2 \left(\frac{s_k}{s_{k0}} \right).$$

Derivation of this result relies on the assumption that the decision-maker is unable to alter the objective probabilities of events, p_{kj} , and that these probabilities are well-defined for each gamble in a Kolmogorov sense. ■

We assume that the decision maker has a cost associated with processing information. We model it using a cost function, $C(I)$, which is an increasing

function of the amount of information, I . We assume that the cost enters additively into the objective function. Then, the objective is to maximize the expected value of the gambles, $V(k)$, $k \in \{1, \dots, K\}$, net of the cost of information, $C(I)$. The decision-maker's problem amounts to:

$$\begin{aligned} & \max_{s_k} \sum_{k=1}^K s_k V(k) - C(I(s_k)), \\ \text{s.t.} \quad & I(s_k) = \sum_{k=1}^K s_k \log_2 \left(\frac{s_k}{s_{k0}} \right), \\ & \sum_{k=1}^K s_k \leq 1, \quad s_k \geq 0. \end{aligned} \tag{2}$$

Theorem 2 *If the cost of information, $C(I)$, is a differentiable increasing convex function of the amount of information, then the optimal choice probabilities are given by:*

$$s_k = \frac{s_{k0} \exp \left(\frac{V(k)}{\theta / \ln 2} \right)}{\sum_{m=1}^K s_{m0} \exp \left(\frac{V(m)}{\theta / \ln 2} \right)}. \tag{3}$$

where $\theta = \frac{\partial C(I)}{\partial I}$ is the derivative of the cost function with respect to the amount of information, evaluated at the chosen amount of information.

Proof. First, note that information, $I(s_k)$ is a strictly convex function of the probability distribution $\{s_k\}$. This follows from the fact that this function is twice differentiable, and its Hessian is a diagonal matrix which contains only non-negative elements.

Second, since $C(I)$ is increasing and convex in I , convexity of information with respect to probabilities $\{s_k\}$ guarantees that the composite function $C(I(s_k))$ is also a convex function of the probability distribution $\{s_k\}$. This in turn implies that the objective function of the decision-maker is concave in the choice variable $\{s_k\}$.

Maximization of a concave function with respect to a linear constraint with a non-zero gradient and a set of non-negativity constraints leads to a unique solution satisfying the first-order condition:

$$V(k) - \frac{\theta}{\ln 2} \left(\ln \frac{s_k}{s_{k0}} + 1 \right) - \lambda = 0.$$

where $\theta = \frac{\partial C(I)}{\partial I}$ is the derivative of the cost function and λ is the Lagrange multiplier associated with the constraint that probabilities sum up to one. Note that this equation holds for all $k \in \{1, \dots, K\}$. We can combine first-order conditions for any pair of k and m to obtain:

$$\frac{s_k}{s_m} = \frac{s_{k0}}{s_{m0}} \exp \left(\frac{V(k) - V(m)}{\theta / \ln 2} \right).$$

By further rearranging and summing up over s_m we obtain the optimal probability (3). ■

This result is a generalization of the choice model of Luce (1994). To see this, assume no prior bias regarding the gambles,¹⁰ suppose that $K = 2$ and the two gambles considered are labeled A and B . Then the formula reduces to:

$$s_A = \frac{1}{1 + \exp\left(\frac{v(B)-v(A)}{\theta/\ln 2}\right)}. \quad (4)$$

However, our result is more general since it can account for the fact that people can have a prior bias towards one option against the other which stems from the way gambles are presented to the decision-maker. In this paper we do not specify a particular theory of inefficient coding which could lead to a prior bias. Instead, we assume that coding is efficient and take the prior as given. The design of the experiment we use in our empirical analysis allows us to eliminate any prior bias and thus abstract from concerns associated with possibly inefficient coding of information by participants of the experiment.

Rational inattention theory predicts that the decision maker should flip a biased coin when making his choice. The bias of the coin is endogenous. It depends on the trade-off between the marginal benefit of being more attentive and the marginal cost of processing more information, captured by the expression $\theta = \frac{\partial C(I)}{\partial I}$. In the empirical part of the paper we make the assumption that the cost function, $C(I)$, is linear in the amount of information, I . This transforms θ into a preference parameter which fully characterizes the costs of processing information of the decision-maker.

Note that the structural forms (3) and (4) are derived from first principles. Thus, rational inattention theory provides a micro-foundation for the logit specification of Luce's choice model, widely used in the literature. It also predicts that individuals should behave as described by these curves when faced with repeated choices. Our model also gives an intuitive interpretation to the curvature parameters of Luce's choice model as reflecting costs of processing information.

Our theory complements the theory of value, which determines valuations of gambles $V(k)$ depending on the payoffs X_{kj} and their objective probabilities, p_{kj} . In this paper we estimate individual information costs considering several theories of value for two-branch gambles. We follow the literature in assuming that individual valuations of sure outcomes are given by the utility function with constant relative risk aversion:

$$U(X) = \alpha \frac{X^{1-\gamma}}{1-\gamma}, \quad (5)$$

where α is a positive constant, $\gamma \in R$ represents risk-aversion of the decision-maker. Expected utility theory assumes that values of gambles are computed by weighting these utilities with objective probabilities of payoffs:

¹⁰That is, assume that $\{s_{k0}\}$ is uniform and equals $\{\frac{1}{K}\}$.

$$V(k) = \sum_{j=1}^{J_k} p_{kj} U(X_{kj}). \quad (6)$$

Another theory of value, cumulative prospect theory (CPT), developed in Kahneman and Tversky (1979) and Tversky and Kahneman (1992), assumes that subjective probabilities are not equal to objective probabilities, but linked to them through a probability weighting function. We use the following specification of the weighting function in our empirical analysis:

$$w(p) = \frac{(1 - \delta) p^\phi}{\left((1 - \delta) p^\phi + (1 - p)^\phi \right)^{\frac{1}{\phi}}}, \quad (7)$$

where $\phi > 0$ and $0 \leq \delta \leq 1$ are behavioral parameters. In the case of two gambles, the weight $w(p)$ corresponds to the outcome with a higher payoff, while the weight $1 - w(p)$ is attached to the lower payoff. Then, values of gambles are determined by:

$$V(k) = \sum_{j=1}^{J_k} w(p_{kj}) U(X_{kj}). \quad (8)$$

A prominent alternative weighting theory is the transfer of attention exchange (TAX) theory developed by Birnbaum (1974) and Birnbaum and Chavez (1997). The special version of this theory assumes that the weight of the higher valued outcome is computed as follows:

$$w(p) = \frac{\left(1 - \frac{\delta}{3}\right) p^\phi}{p^\phi + (1 - p)^\phi}, \quad (9)$$

where $\phi > 0$ and $0 \leq \delta \leq 3$ are also behavioral parameters. Both of these theories boil down to expected utility when $\phi = 1$, and $\delta = 0$. Also, when $\gamma = 0$ agents are risk-neutral. Thus, we have 3 parameters $\{\gamma, \phi, \delta\}$ which characterize various theories of value, and a single cost parameter θ that summarizes agent's limited ability to process information.

In order to convert the cost of information θ from utils per bit to dollars per bit we use the conversion factor: $\sum_{k=1}^K s_{k0} \frac{\partial V(k)}{\partial X_k}$ for each gamble. This conversion factor makes the parameter α irrelevant, as it scales both the utility function and the cost of information. This is consistent with intuition, as the individual scale of utility affects the absolute cost of information in utils per bit, but does not affect the relative cost measured in dollars per bit.

4 Methodology

4.1 Experimental Setup

In order to assess the relevance of the cost of information and compare the models described in the previous section, we use experimental data collected in Michel Regenwetter’s laboratory at the University of Illinois at Urbana Champaign in the summer of 2009. The main property of the experimental setup useful to us is that each participant was asked the same question repeatedly a large number of times. The collected experimental data contains observed frequencies with which each subject chose from each pair of gambles. This data allows us to test rational inattention theory at the individual level.

Experimental data available to us contains answers of \mathcal{N} individuals who were repeatedly asked to compare \mathcal{M} pairs of gambles. Each individual faced each gamble \mathcal{L} times. The $\mathcal{M} \times \mathcal{L}$ overall gambles per individual were shuffled to make sure that the experiment is not plagued by memory effects.

N	$X_1, \$$	p_X	$X_2, \$$	$Y_1, \$$	p_Y	$Y_2, \$$
1	29.38	0.65	1.19	18.00	0.68	3.21
2	27.98	0.42	18.89	25.44	0.47	3.90
3	26.44	0.52	1.92	26.03	0.34	5.77
4	25.05	0.24	24.01	25.32	0.66	10.56
5	23.64	0.71	10.78	25.03	0.98	6.86
6	20.76	0.80	11.61	12.42	0.93	8.14
7	19.38	0.23	2.46	12.57	0.96	0.73
8	18.02	0.39	4.97	15.01	0.49	14.26
9	16.66	0.60	9.03	16.32	0.19	10.87
10	19.58	0.48	15.17	26.39	0.45	10.07
11	13.88	0.41	5.05	8.91	0.70	8.67
12	29.83	0.38	12.47	25.10	0.85	22.74
13	21.78	0.72	11.16	21.30	0.66	20.91
14	9.61	0.17	6.49	9.87	0.31	4.17
15	16.11	0.20	8.10	22.75	0.13	6.18
16	6.88	0.53	6.69	13.86	0.90	0.96
17	24.08	0.90	5.02	23.74	0.07	14.41
18	18.56	0.99	1.70	27.68	0.97	2.16
19	22.51	0.88	0.00	19.30	0.71	0.73
20	22.57	0.70	0.12	11.53	0.79	2.81

Table 2. Gamble payoffs and probabilities.

Each question contained two gambles, A and B, with parameters $\{X_1, p_X, X_2, 1 - p_X\}$ and $\{Y_1, p_Y, Y_2, 1 - p_Y\}$.¹¹ Gambles were randomly uniformly drawn from the whole domain of potential gambles following the procedure proposed by Rieskamp (2008). One advantage of this procedure is that it does not favor any particular theory by construction. Thus, the gamble space generates no a

¹¹In the dataset $\mathcal{N} = 40$, $\mathcal{M} = 20$, $\mathcal{L} = 60$.

priori bias to the estimates of the parameters.¹² Second, this gamble selection procedure guarantees that costs associated with coding information about the gambles to be processes are roughly the same for all gambles. This puts all choices on equal grounds and eliminates any bias associated with the possibility of inefficient coding.

Outcomes of the gambles were selected from a uniform distribution over $[0,30]$ in 0.01 increments. Probabilities were selected from a uniform distribution over $[0,1]$ in 0.01 increments. About 59% of these gambles were screened because either one gamble showed first-order stochastic dominance over the other or one gamble had at least double the expected value of the other. 20 pairs were randomly selected from the remaining gambles. Table 2 presents the gambles used in the experiment.

The experiment was conducted on laptop computers in a laboratory space at the University of Illinois at Urbana-Champaign. Forty individuals participated in the study, split roughly half-half by gender and all approximately of college-age. The experiment was conducted over two sessions, separated by at least four days for each participant. Each session was not restricted in time and took roughly one hour to complete. At the beginning of each session a clear description was given of what to expect from the experiment and several practice gambles were played. The participants were also warned that each pair of gambles could be selected at the end of the session to be played for real, making clear that their choice in each pair of gambles could affect their final payoff.

Participants were presented with a sequence of gamble pairs, one pair at a time. Probabilities were displayed in the form of pie charts. Participants could choose only one gamble from each pair. Gamble pairs were ordered by the computer quasi-randomly, with the condition that the same pair was never presented twice in succession. Over the course of a session, each gamble pair was presented 30 times, so participants made 600 choices in each of the two sessions. At the end of each session, one gamble out of 600 chosen by the participant was randomly selected by the computer and played for real. The outcome of the chosen gamble was paid to the participant together with a \$5 flat payment. The average payment was \$20.97 per session.

4.2 Estimation of Parameters

This experimental data allows us to estimate jointly the values of all four parameters $\{\theta, \gamma, \phi, \delta\}$ for each of \mathcal{N} individuals, or any subset of parameters for any restricted version of the theory. To correctly jointly estimate the parameters we need to write down a likelihood function. The likelihood function of the data is the density of a binomial distribution with s_A being the binomial probability. The likelihood of option A being chosen x times and option B being chosen y

¹²This setup is important. For instance, consider the classical Experiment I from Tversky (1969). The gambles for that experiment were selected in a subspace of all gambles which have almost the same expected payoff. Using such a set of gambles in our experimental setup would give us no ability to identify parameters characterizing costs of information.

times given the deep parameters $\omega = \{\theta, \gamma, \phi, \delta\}$ and parameters of the question $\zeta = \{X_1, X_2, p_X, Y_1, Y_2, p_Y\}$ is given by:

$$L(x, y|\omega, \zeta) = \binom{x}{x+y} (s_A)^x (1 - s_A)^y.$$

Hence, the log likelihood of the data is given by:

$$\log L(x, y|\omega, \zeta) = \log \binom{y}{x+y} + x \log s_A(\omega, \zeta) + y \log (1 - s_A(\omega, \zeta)), \quad (10)$$

where

$$s_A(\omega, \zeta) = \frac{1}{1 + \exp\left(-\frac{V(A)-V(B)}{\theta/\ln 2}\right)}, \quad (11)$$

We use this specification to estimate ω by maximizing the sum of log likelihoods defined as:

$$\Lambda_a = \sum_{i=1}^{\mathcal{M}} \log L(x_i, y_i|\omega_a, \zeta_i). \quad (12)$$

To evaluate model fit it is also useful to compute the likelihood of an "unrestricted" model which allows for individual parameters $s_{a,i}$ for each question i for each participant a . This model has as many parameters as data points. Estimates of parameters of this model will be equal to the observed probabilities, i.e. $s_{a,i} = \frac{x}{x+y}$. Therefore, log likelihood of the unrestricted model is computed as follows:

$$\log L^{UR}(x, y) = \log \binom{y}{x+y} + x \log \frac{x}{x+y} + y \log \frac{y}{x+y}, \quad (13)$$

where $0 \log 0$ is always replaced by 0.

To compare models we sum up individual likelihoods, and then penalize the joint likelihood for over-parameterization using the Bayes Information Criterion (BIC):

$$BIC = -2 \sum_{a=1}^{\mathcal{N}} \Lambda_a + n \ln O. \quad (14)$$

where n is the total number of estimated parameters, and O is the total number of observations.

5 Results

We apply our estimation procedure to 8 statistical models all based on three theories of value described earlier. The estimation results are presented in Table 3. The first row (EU) represents the Expected Utility model, where parameters $\{\theta, \gamma\}$ were estimated for each participant individually, and then the joint likelihood was computed for this model by summing individual likelihoods. Similarly, CPT stands for Cumulative Prospect Theory with three parameters $\{\theta, \gamma, \phi\}$ when $\delta = 0$ while CPT+ stands for an extended version of this theory with four parameters $\{\theta, \gamma, \phi, \delta\}$. TAX stands for the Transfer of Attention Exchange model which also has four parameters. Each row describes the support of each estimated parameter and the restriction on each parameter which was not estimated.

We also estimate versions of EU and CPT+ with restricted values of the cost of processing information. Rows 5 (EU n) and 6 (CPT+ n) of Table 3 represent cases when costs are made very low. Rows 7 (EU r) and 8 (CPT+ r) represent cases when costs were fixed at 3 bits per cent.¹³ Finally, row 9 (UR) of Table 3 shows the likelihood of the unrestricted model. We restrict the support for parameters $\{\phi, \delta\}$ to ranges commonly assumed in the literature. The cost of processing information is restricted to be non-negative, while the parameter of risk aversion can take any real value. Using likelihoods for each of the specifications we compute the Bayes Information Criterion statistic for each specification.

Estimates of these nine statistical models allow us to answer several important empirical questions. First, we can test whether our model is a better approximation of the data compared to other rational choice models. Comparison of lines 1-4 to lines 5-6 illustrates the overwhelming support by the data of rational inattention theory. Models with low information costs (for instance, when the inverse cost is fixed to be above 20 bits per cent) are strongly rejected by the data compared to models with higher costs. Models of deterministic choice, a particular case of our model when the cost of processing information is zero, have log likelihood of minus infinity. This means that models based on rational inattention theory are a much better description of individual choices in this experimental setup compared to models of deterministic rational choice. The data favors the intuition that agents process only part of the available information and, thus, rationally choose to make errors from time to time.

Second, we can assess the role played by heterogeneity in information costs. Comparing lines 1-4 to 7-8 reveals that the assumption of identical shadow cost across participants is strongly dominated by models which allow for heterogeneity in costs of processing information. Thus, heterogeneity is a statistically significant outcome of our exercise.

Third, comparing lines 1-4 with each other allows us to compare different theories of value. Table 3 reveals that more parsimonious models are subsequently dominated by less parsimonious models in spite of the fact that less

¹³This value characterizes the average response aggregated across subjects.

parsimonious specifications are penalized for over-parametrization. Among the least parsimonious models, the TAX model of Birnbaum and Chavez (1997) seems to fit the data better than the extended version of cumulative prospect theory. Column 9 in Table 3 shows the numbers of participants of the experiment, best described by each of 8 theories. It shows that the TAX model best describes 23 out of 40 participants of the experiment, CPT+ best describes 7 participants, CPT best describes 6, and EU best describes 4.

Fourth, comparison of lines 1-4 to line 9 of Table 3 demonstrates that the TAX model comes relatively close to the unrestricted model in terms of fit. Column 10 of Table 3 shows that for 10 participants out of 40 the TAX model fits the data better than the unrestricted model. In fact, for 2 participants out of 40 the EU model seems to be the best description of the data. We conclude from these observations that the rational inattention model we develop in this paper overall fits the data relatively well.

Note, that we do not focus on comparing models within individuals as it appears not important for our main results. Because careful comparison of value theories within subjects is not the goal of this paper, we leave this as an avenue for future research. In what follows, we restrict our analysis to estimates from three most likely specifications each with a different number of parameters: EU, CPT and TAX.

Cross-sectional estimates of the parameters of these three models are presented in Figures 4, 5 and 6 in the form of histograms and non-parametric density curves. Means, medians and standard deviations of the cross-sectional distributions of parameters are also presented in each graph.

These estimates enable us to obtain robust measures of costs of processing information. Independently of the theory of value, the inverse cost of processing information of most participants of the experiment falls in the region between 1 and 15 bits per cent, with the mean in the vicinity of 5-8 bits per cent and a standard deviation of approximately 4 bits per cent. Our estimates also show that people are extremely heterogeneous not only in costs of processing information but also in attitudes to risk. Risk aversion consistently varies across individuals from as high as +2, which is a relatively high value of aversion to risk compared to other experimental studies,¹⁴ to as low as -4, which is a very high value of risk loving attitude.

There are similarly large variations in all other parameter estimates. Estimates of the curvatures of likelihood functions as well as estimates of restricted versions of each model indicate that only a tiny fraction of variation in these parameters across subjects can be attributed to measurement error. Most point estimates are relatively precise with relatively small standard errors.

From the same figures we can see that this heterogeneity results in a significant aggregation bias when aggregating across subjects. We define a representative agent to be the fictional agent representing responses pooled across the sample population as if represented by a single agent. Parameter estimates

¹⁴The standard estimates of Holt and Laury (2002), who aggregate across subjects, are in the range of positive 0.3-0.5.

of this representative agent (RA) are also depicted in Figures 4, 5 and 6. The inverse cost of information characterizing the representative agent is around 3 bits per cent in each case, which is roughly half that of either an average or a median agent in the sample. Similarly, although the average agent is risk loving in both CPT and TAX models, the representative agent is characterized by substantial aversion to risk.

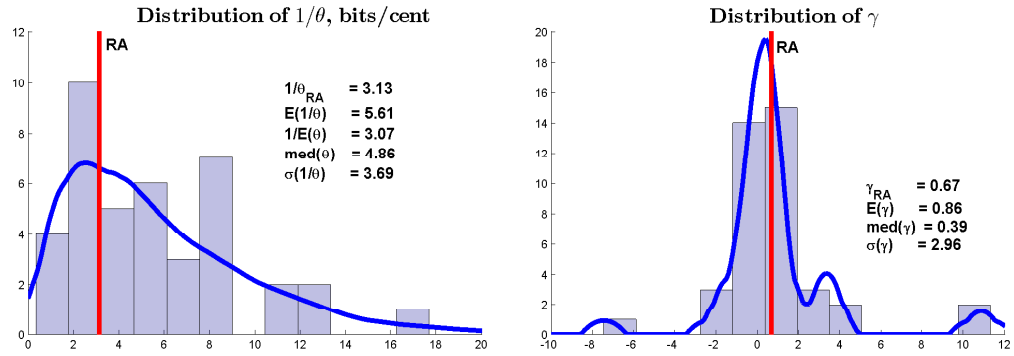


Figure 4. Estimates of parameters under EU.

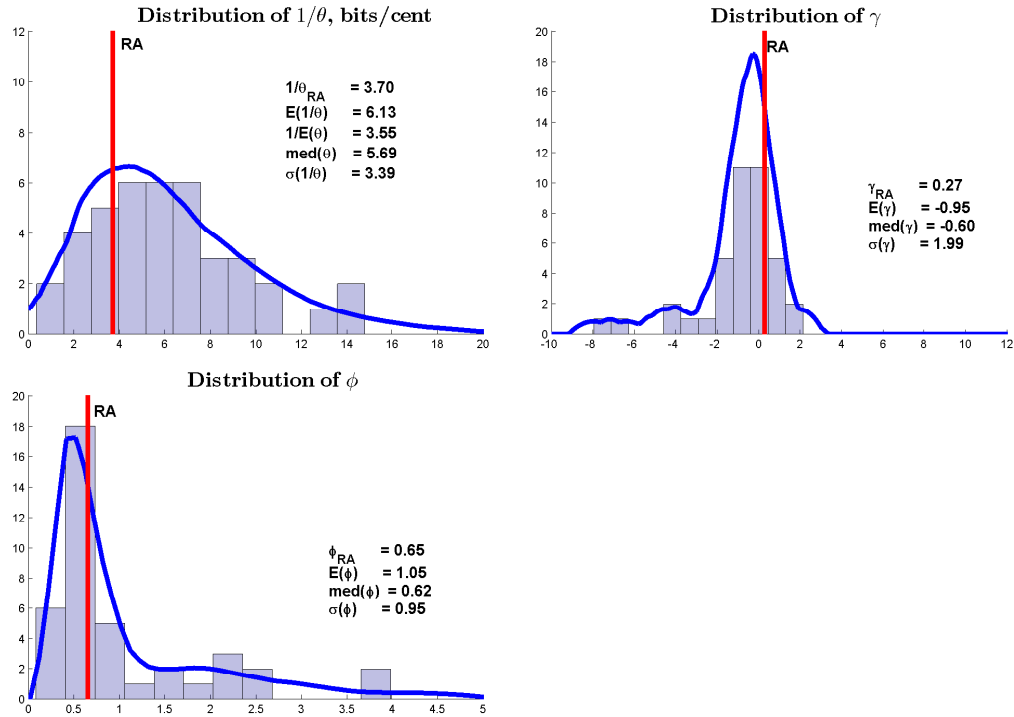


Figure 5. Estimates of parameters under CPT

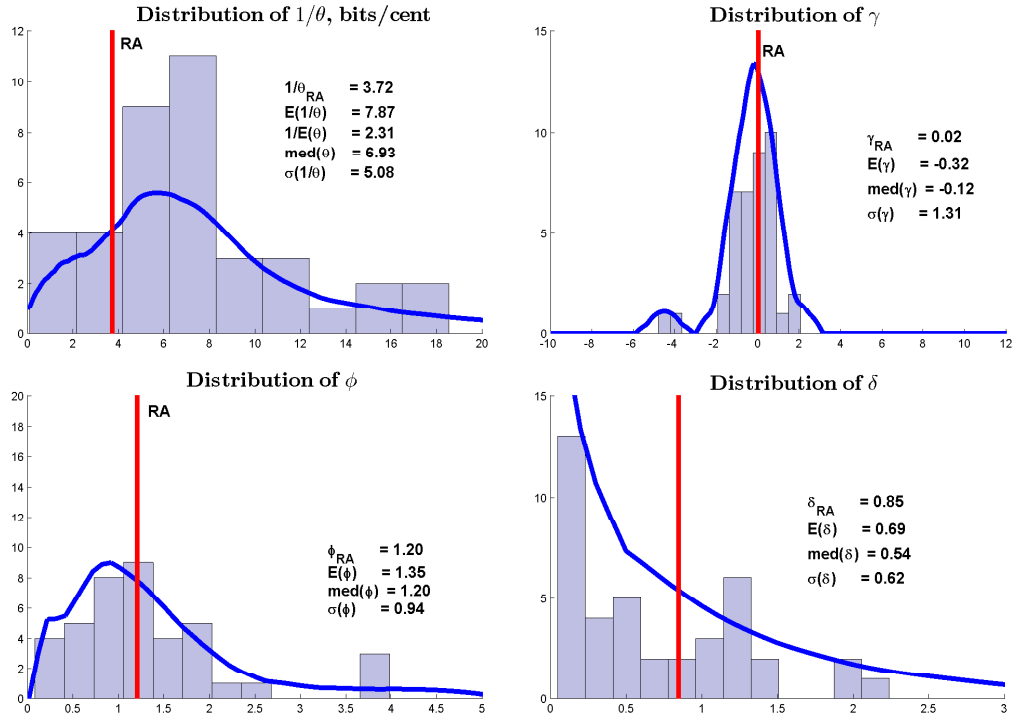


Figure 6. Estimates of parameters under TAX

Model	θ	γ	ϕ	δ	LL	n	BIC	# of participants	
								$\succ [1-8]$	$\succ UR$
1. EU	R^+	R	1	0	-9916	80	20693*	3	2
2. CPT	R^+	R	R^+	0	-7565	120	16424*	6	2
3. CPT+	R^+	R	R^+	$[0, 1]$	-6382	160	14490	7	3
4. TAX	R^+	R	R^+	$[0, 3]$	-5631	160	12986*	23	10
5. EU _n	1/25	R	1	0	-45380	40	91191	0	0
6. CPT+ _n	1/25	R	R^+	$[0, 1]$	-36920	120	75133	0	0
7. EU _r	1/3	R	1	0	-12729	40	25890	1	2
8. CPT+ _r	1/3	R	R^+	$[0, 1]$	-9127	120	19978	0	2
9. UR	—	—	—	—	-1368	800	11360	30	40

Table 3. Likelihood analysis of alternative models.

6 Aggregation

We start with a novel theoretical result about aggregation of rationally inattentive agents.

Theorem 3 *When agents differ (only) in their cost of information θ_i , the inverse cost of information of the representative agent is always biased downwards compared to the average across inverse costs of information of individual agents:*

$$\frac{1}{\theta_{RA}} < \frac{1}{N} \sum_{i=1}^N \frac{1}{\theta_i}.$$

Proof. Consider a pair of gambles which gives each agent a value differential $x \in R$. Denote parameter of the gamble pair $a = e^{-x \ln 2} > 0$ and inverse cost $\psi_i = \frac{1}{\theta_i} > 0$. According to predictions of rational inattention theory, the choice probability of agent i over gamble pair x is given by:

$$y_i = \frac{1}{1 + e^{-\frac{x \ln 2}{\theta_i}}} = \frac{1}{1 + a^{\psi_i}}.$$

The representative agent's choice probability is computed by averaging across agents:

$$y_{RA} = \frac{1}{N} \sum_{i=1}^N y_i = \frac{1}{N} \sum_{i=1}^N \frac{1}{1 + a^{\psi_i}}.$$

His inverse cost of information is then computed inverting the function:

$$y_{RA} = \frac{1}{1 + a^{\psi_{RA}}}.$$

Consider the function $f(z) = \frac{1}{1 + a^z}$ on $z > 0$. This function is strictly increasing and concave for $a < 1$, strictly decreasing and convex for $a > 1$, equals $\frac{1}{2}$ when $a = 1$. This last case happens only when when $x = 0$, when both sides are $\frac{1}{2}$ so θ_{RA} is undefined. Consider the case $a < 1$ first. By Jensen's inequality for any (unequal) values z_j in the domain and for any strictly positive weights a_j a concave function $f(z)$ satisfies:

$$f\left(\frac{\sum_j a_j z_j}{\sum_j a_j}\right) > \frac{\sum_j a_j f(z_j)}{\sum_j a_j}.$$

Hence,

$$\frac{1}{1 + a^{\psi_{RA}}} = \frac{1}{N} \sum_{i=1}^N \frac{1}{1 + a^{\psi_i}} < \frac{1}{1 + a^{\frac{1}{N} \sum_{i=1}^N \psi_i}}.$$

Since the function $f(z)$ is strictly increasing in z it follows that

$$\frac{1}{\theta_{RA}} < \frac{1}{N} \sum_{i=1}^N \frac{1}{\theta_i}.$$

Similarly, when $a > 1$ the function $-f(z)$ is strictly increasing and concave.

Hence,

$$\frac{1}{1 + a^{\psi_{RA}}} = \frac{1}{N} \sum_{i=1}^N \frac{1}{1 + a^{\psi_i}} > \frac{1}{1 + a^{\frac{1}{N} \sum_{i=1}^N \psi_i}}.$$

Since the function $f(z)$ is now strictly decreasing in z it again follows that

$$\frac{1}{\theta_{RA}} < \frac{1}{N} \sum_{i=1}^N \frac{1}{\theta_i}.$$

Note that the bias disappears only if all agents have identical costs of information ($\theta_i = \theta_j$) or when the two options being compared are identical ($a \rightarrow 1$).

■

The key implication of Theorem 3 is that we should expect aggregate behavior to appear as if produced by a much more inattentive representative agent compared to the average individual. Thus, we should expect to encounter seemingly even more inattentive behavior in the aggregate than in the laboratory.

Parameter	Rep Agent	Mean Agent	Cluster I	Cluster II
inv cost of info, $\frac{1}{\theta}$	3.1	5.64	9.3	3.2
Risk aversion, γ	0.67	0.86	1.80	0.23
N agents	1	40	16	24

Table 4. Parameters by cluster under EU.

Parameter	Rep Agent	Mean Agent	Main	Gambler
inv cost of info, $\frac{1}{\theta}$	3.7	6.1	6.6	2.3
Risk aversion, γ	0.27	-0.95	-0.23	-3.8
CPT parameter, ϕ	0.65	1.05	0.59	2.37
N agents	1	40	32	8

Table 5. Parameters by cluster under CPT.

Parameter	Rep Agent	Mean Agent	Main	Gambler
inv cost of info, $\frac{1}{\theta}$	3.7	7.9	7.2	1.7
Risk aversion, γ	0.02	-0.32	-0.04	-1.87
CPT param, ϕ	1.20	1.35	1.17	4.0
TAX param, δ	0.85	0.69	0.54	0.01
N agents	1	40	37	3

Table 6. Parameters by cluster under TAX.

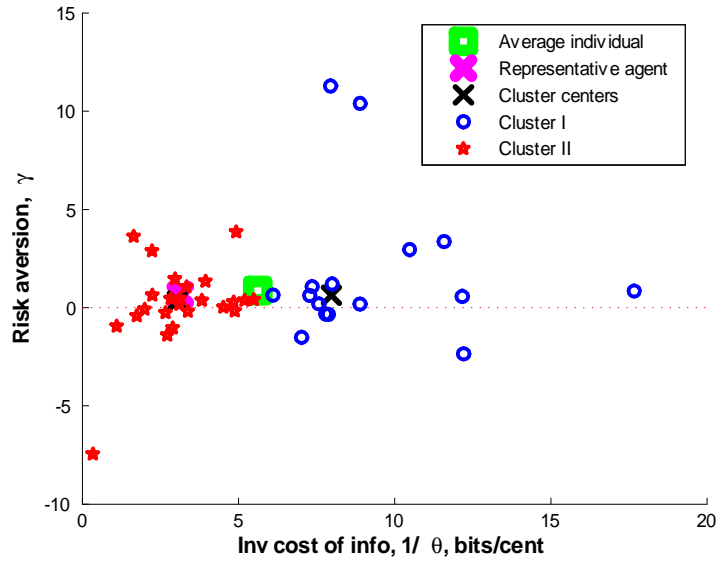


Figure 7. Clustering of parameters under EU.

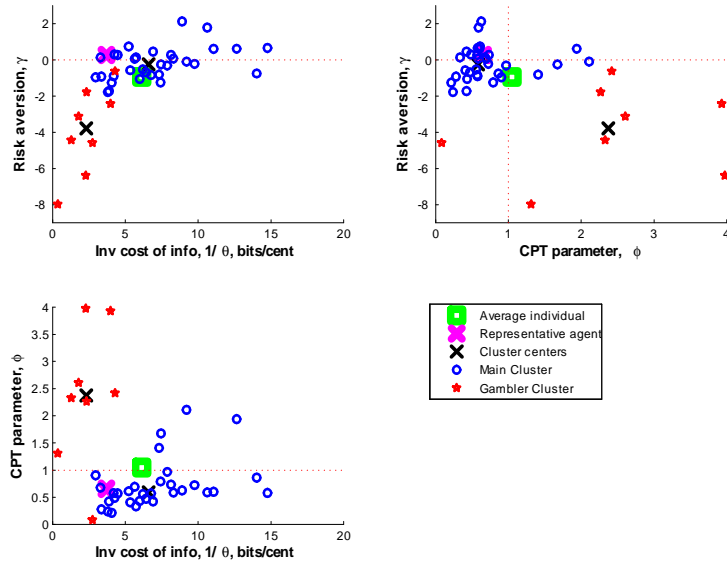


Figure 8. Clustering of parameters under CPT.

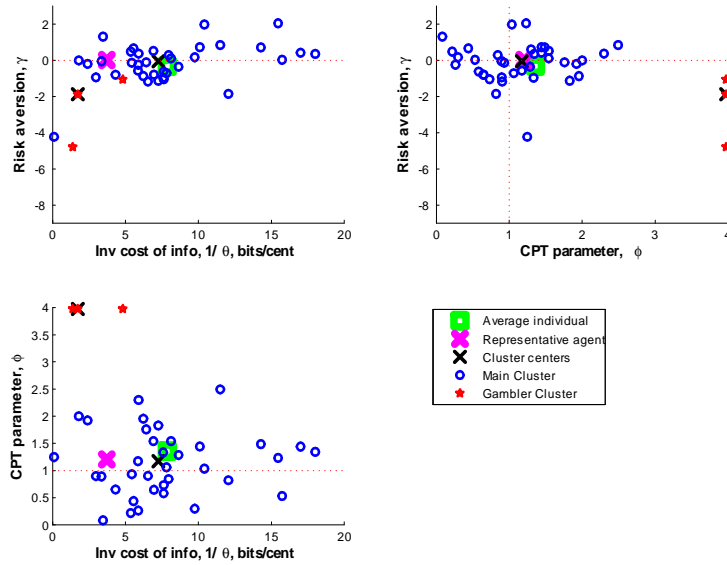


Figure 9. Clustering of parameters under TAX.

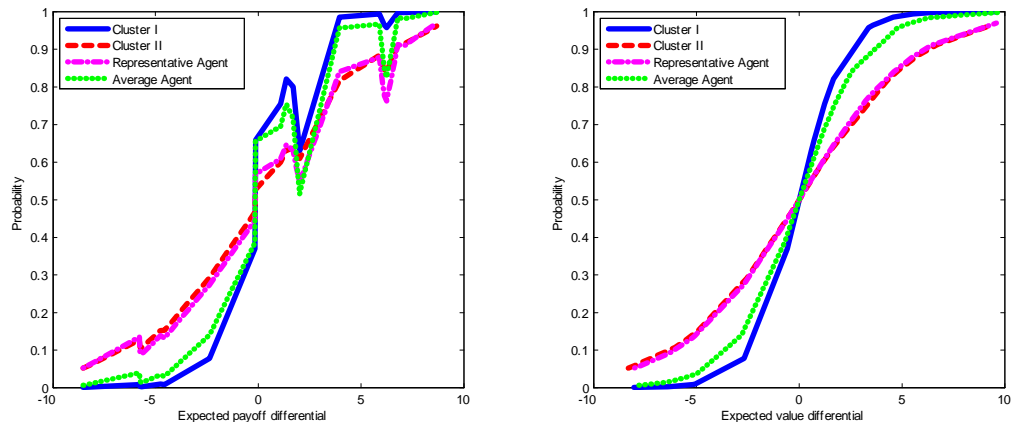


Figure 10. Predicted probabilities overall and by cluster under EU.

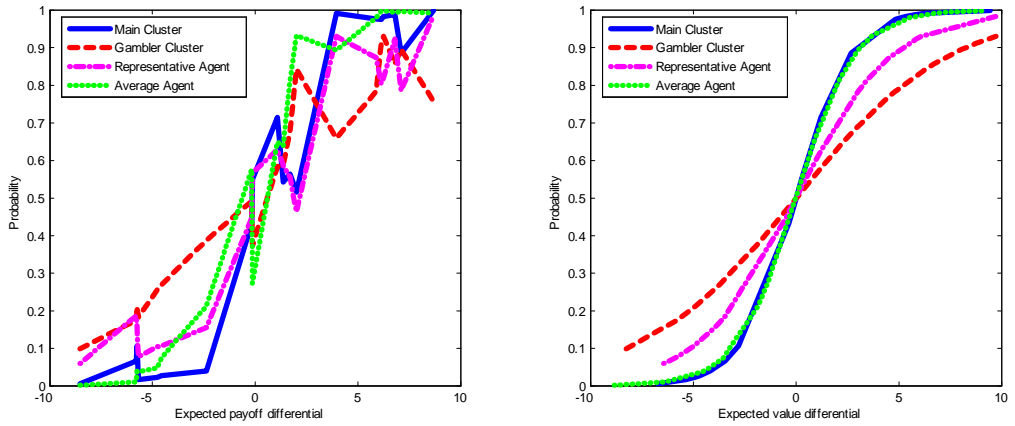


Figure 11. Predicted probabilities overall and by cluster under CPT.

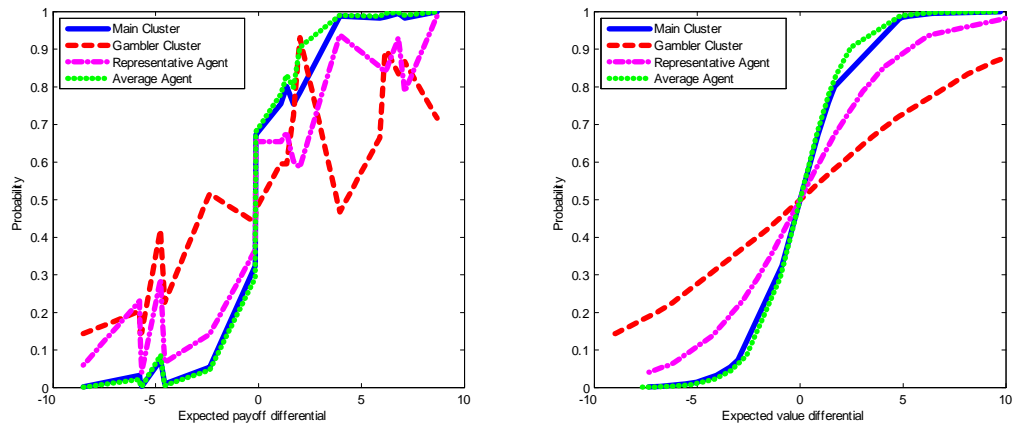


Figure 12. Predicted probabilities overall and by cluster under TAX.

Our empirical analysis allows us to assess the amount of heterogeneity and the associated aggregation bias. For illustrative purposes we sub-divide 40 agents in our dataset into 2 clusters by k-means using estimates of parameters $\{\theta, \gamma, \phi\}$ for each of the three models considered earlier. These clusters are presented in circles and stars in Figures 7, 8 and 9. Tables 4, 5 and 6 present the parameters of groups of agents for all three models considered.

Such classification is particularly stark for the CPT model and the TAX model. In each of these models the first cluster (circles) represents the main body of agents, roughly consistent with standard theories of value. The second group (stars), consisting of just a few agents, could be called gamblers. They are characterized by a lot of randomization (high cost of information) and a passion for risk taking. They also tend to put more weight on lower outcomes compared to objective probabilities ($\phi > 1$). In the same figures, we plot the average agent in our sample and the representative agent, meaning the pooled response across the sample population.

Results in this section demonstrate that, when responses are pooled across subjects, heterogeneity introduces a significant aggregation bias into the estimates of preferences. To illustrate how this bias works we plot in Figures 10, 11 and 12 the predicted choice probabilities as functions of gamble-specific expected payoff differentials and participant-specific expected value differentials. It is easy to see from these figures that in each model the probabilistic curve describing the representative agent (dash-dotted line) is shifted out substantially from the response of the average agent in the population (dotted line) towards the response of a small group of the most inattentive agents (dashed). These figures confirm the intuition obtained from our theoretical result and illustrate the size of the bias.

7 Coding, knowledge and models of heuristics

Our model based on rational inattention theory focuses on a decision maker's ability to act in an uncertain environment with limited processing capacity. Our model postulates that the decision maker, aware of his limited processing capacity, selects the information structure that conveys the highest utility. As a result, our model predicts that a rationally inattentive decision maker optimally chooses the amount of uncertainty that he is willing to tolerate by evaluating costs and benefits of processing information. This section makes three observations on the rational inattention model we proposed.

First, one important assumption of our model is that the only source of uncertainty faced by the decision maker is in the distribution of options ("states of the world"). We do not address any form of cognitive bias that might emerge from presenting the options as made of different numbers of branches (e.g. three instead of two) or of different probability representations (e.g. pie charts as opposite to percentages). These cognitive biases are treated in Shannon's information theory as coding and decoding problems. That is, when evaluating the capacity of a channel -human brain, in our case-, a prominent branch of

information theory is concerned about the optimal design and compression of inputs and outputs of the channel. Albeit we recognize that such a cognitive bias may be sizeable in experimental studies, we choose not to model this bias explicitly and assume that the coding is always efficient.

Second, we want to highlight the difference between information and knowledge in our model. Some studies have interpreted information as equivalent to knowledge. For instance, Gigerenzer and Goldstein (2011) describe recognition and evaluation as the two processes that constitute information use for decision making. They describe recognition as the process of accessing memory, -i.e., previous knowledge-, and evaluation as the process of comparing choice options to objects in the knowledge base. The decision maker does not acquire new information or produce new knowledge when using this heuristic process. In our model, recognition corresponds to the prior of the participant about the gamble he faces. Before processing any information, this prior knowledge is measured by the uncertainty (or entropy) of the gambles. Then, evaluation corresponds to processing information about the gambles in order to reduce uncertainty. Thus, in our model, evaluation is the process of acquiring information and forming new knowledge.

Third, we want to emphasize the difference between rational inattention models and models of heuristics as advocated by, *inter alia*, Cokely, Schooler and Gigerenzer (2010), as well as models based on Decision Field Theory, as advocated by Busemeyer and Johnson (2004). The reason why we use rational inattention theory to describe people’s behavior is due to the fact that its statistical foundations make the model general and universally applicable. So long as we can characterize the distribution of the state variables, we can measure ex-ante uncertainty. So long as we can postulate a theory of value, we can predict and measure the optimal reduction of uncertainty of the decision maker. We are concerned about the ability of the model to produce predictions consistent with observed behavior. We do not take a stand on whether this modeling strategy replicates the cognitive process that occurs in people’s brains when they make decisions. This area of research goes beyond the scope of our paper.

8 Conclusion

In this paper we propose rational inattention theory of Sims (2003) as an explanation to probabilistic outcomes of discrete choices. We show that rational inattention theory provides a rationalization of choice models à la Luce and gives a structural interpretation to probability curvature parameters as reflecting costs of processing information. We use data from a behavioral experiment to test the predictions of our theory. Experimental evidence strongly supports the predictions of our model that: (a) individual people behave probabilistically when repeatedly faced with the same choices; (b) the larger the difference in valuations between options, the smaller the frequency of erroneous choice.

Our theoretical model allows us to estimate jointly parameters of attitude to risk and cost of information for individual participants of the experiment. We

provide microeconomic evidence for rational inattention theory and obtain the first set of microeconomic estimates of its parameters. Our estimation procedure allows us to obtain direct estimates of costs of processing information from the laboratory, which could be used as guidance to calibrate macroeconomic models with rationally inattentive agents.

In addition, we document an overwhelming degree of heterogeneity among a relatively homogeneous sample of participants of the experiment, both in their attitude to risk and in their costs of processing information. We characterize, both theoretically and empirically, the aggregation biases resulting from this heterogeneity and find them to be substantial. We formally show that in an economy populated by rationally inattentive agents the fictional 'representative agent' would be much more inattentive than the average agent.

Our results and methodology are directly applicable to any setup involving discrete choices. The corresponding model, with or without a rational explanation, has been widely and successfully used in many fields of social science including economics, psychology, finance, marketing, political science. In light of these findings, it seems worth to augment theories of value with rational inattention theory and broaden its applications in different fields of social sciences.

References

- Barbera, Salvador, and Prasanta K. Pattanaik.** 1986. "Falmagne and the Rationalizability of Stochastic Choices in Terms of Random Orderings." *Econometrica*, 54(3): pp. 707–715.
- Birnbaum, M. H.** 1974. "The nonadditivity of personality impressions." *Journal of Experimental Psychology*, 102: pp. 543–561.
- Birnbaum, M. H., and A. Chavez.** 1997. "Tests of theories of decision making: Violations of branch independence and distribution independence." *Organizational Behavior and Human Decision Processes*, 71: pp. 161–194.
- Birnbaum, Michael H.** 2008. "New paradoxes of risky decision making." *Psychological Review*, 115(2): pp. 463–501.
- Block, H. D., and J. Marschak.** 1960. "Random orderings and stochastic theories of responses." In . , ed. W. Hoeffding W. Madow & H. Mann I. Olkin, S. Ghuyre, Chapter in Contributions to Probability and Statistics, pp. 97–132. Stanford University Press.
- Busemeyer, J. R., and J. G. Johnson.** 2004. "Computational models of decision making." In . , ed. D. Koehler and N. Harvey, Chapter in Handbook of Judgment and Decision Making, pp. 133–154. Blackwell Publishing Co.
- Cokely, Edward T., Schooler Lael J. Gerd Gigerenzer.** 2010. "Information use for decision making." In . , ed. M.N. Maack and M.J. Bates, Chapter in Encyclopedia of Library and Information Sciences, Third Edition, pp. 2727–2734. Taylor and Francis Group.

- Cramer, J. S.** 2003. *Logit models from economics and other fields*. Cambridge University Press.
- Falmagne, J.-Cl.** 1978. “A Representation Theorem for Finite Random Scale Systems.” *Journal of Mathematical Psychology*, 18: pp. 52–72.
- Goldstein, Daniel G., and Gerd Gigerenzer.** 2011. “The beauty of simple models: Themes in recognition heuristic research.” *Judgment and Decision Making*, 6(5): pp. 392–395.
- Harless, David W., and Colin F. Camerer.** 1994. “The Predictive Utility of Generalized Expected Utility Theories.” *Econometrica*, 62(6): pp. 1251–1289.
- Hey, John D., and Chris Orme.** 1994. “Investigating Generalizations of Expected Utility Theory Using Experimental Data.” *Econometrica*, 62(6): pp. 1291–1326.
- Holt, Charles A., and Susan K. Laury.** 2002. “Risk Aversion and Incentive Effects.” *American Economic Review*, 92(5): pp. 1644–1655.
- Kahneman, Daniel, and Amos Tversky.** 1979. “Prospect Theory: An Analysis of Decision under Risk.” *Econometrica*, 47(2): pp. 263–292.
- Luce, R. Duncan.** 1959. *Individual Choice Behaviours: A Theoretical Analysis*. New York: J. Wiley.
- Luce, R. Duncan.** 1994. “Thurstone and sensory scaling: then and now.” *Psychological Review*, 107: pp. 271–277.
- Luce, R. Duncan.** 2010. “Behavioral Assumptions for a Class of Utility Theories: A Program of Experiments.” *Journal of Risk and Uncertainty*, 41: pp. 19–27.
- Marley, A.A. J., and R. Duncan Luce.** 2005. “Independence properties vis-à-vis several utility representations.” *Theory and Decision*, 58: pp. 77–143.
- Matějka, Filip, and Alisdair McKay.** 2011. “Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model.” *unpublished manuscript*.
- McFadden, D.** 1978. “Spatial Interaction Theory and Planning Models.” In . . . , ed. Lundqvist L. Snickars F. Karlqvist, A. and J. Weibull, Chapter Modelling the choice of residential location. Amsterdam: North-Holland.
- Mosteler, Frederick, and Philip Nogee.** 1951. “An Experimental Measurement of Utility.” *Journal of Political Economy*, LIX(5): pp. 371–404.
- Regenwetter, Michel, Jason Dana, and Clinton P Davis-Stober.** 2010. “Testing Transitivity of Preferences on Two-Alternative Forced Choice Data.” *Frontiers in Psychology*, 1.

- Regenwetter, Michel, Jason Dana, and Clinton P Davis-Stober.** 2011. "Transitivity of Preferences." *Psychological Review*, 118: pp. 42–56.
- Rieskamp, Jörg.** 2008. "The probabilistic nature of preferential choice." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(6): pp. 1446–1465.
- Rieskamp, Jörg, Jerome R. Busemeyer, and Barbara A. Mellers.** 2006. "Extending the Bounds of Rationality: Evidence and Theories of Preferential Choice." *Journal of Economic Literature*, XLIV: pp. 631–661.
- Saari, Donald G.** 2005. "The profile structure for Luce's choice axiom." *Journal of Mathematical Psychology*, 49: pp. 226–253.
- Shannon, Claude E.** 1948. "A Mathematical Theory of Communication." *Bell System Technical Journal*, 27: pp. 379–423, 623–656.
- Simon, Herbert A.** 1986. "Rationality in Psychology and Economics." *The Journal of Business*, 59(4): pp. S209–S224.
- Sims, Christopher A.** 1998. "Stickiness." *Carnegie-Rochester Conference Series on Public Policy*, 49(1): pp. 317–356.
- Sims, Christopher A.** 2003. "Implications of Rational Inattention." *Journal of Monetary Economics*, 50(3): pp. 665–690.
- Sims, Christopher A.** 2006. "Rational Inattention: Beyond the Linear-Quadratic Case." *American Economic Review Papers and Proceedings*, 96(2): pp. 158–163.
- Thurstone, Louis Leon.** 1927. "A Law of Comparative Judgement." *Psychological Review*, 34: pp. 273–286.
- Thurstone, Louis Leon.** 1929. "Theory of Attitude Measurement." *Psychological Review*, 36: pp. 222–241.
- Tutino, Antonella.** 2011. "Rationally Inattentive Macroeconomic Wedges." *The Journal of Economic Dynamics and Control*, 35(3): pp. 344–362.
- Tversky, Amos.** 1969. "Intransitivity of Preferences." *Psychological Review*, 76(1): pp. 31–48.
- Tversky, Amos, and Daniel Kahneman.** 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty*, 5(4): pp. 297–323.
- Wiederholt, Mirko.** 2010. "Rational Inattention." *prepared for the The New Palgrave Dictionary of Economics*.
- Yellott, J. I., Jr.** 1977. "The relationship of Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution." *Journal of Mathematical Psychology*, 15: pp. 109–144.