Measuring and Modeling Attention

Andrew Caplin

Department of Economics, New York University, New York, NY 10003;
email: andrew.caplin@nyu.edu

Annu. Rev. Econ. 2016. 8:379–403

The *Annual Review of Economics* is online at
economics.annualreviews.org

This article's doi:
10.1146/annurev-economics-080315-015417

JEL codes: D8

Keywords

revealed preference, rational inattention, Bayesian updating, behavioral economics, imperfect information, costly information processing

Abstract

This article presents a selective review of economic research on attentional choice, taking an observation of Block & Marschak (1960) as its starting point. Because standard choice data conflate utilities and perception, they point out that it is inadequate for research in which attention is endogenous. The review focuses on their thesis that advances in our understanding of attention require modeling of novel choice-based data sets, and corresponding methods of measurement. By way of example, recent attentional research based on measuring and modeling state-dependent stochastic choice data is detailed. Next research steps in relation to strategic attention and the dynamics of learning are outlined. If the thesis of Block & Marschak is valid, engineering of new data sets will become an increasingly essential professional activity as attentional research advances.

1. INTRODUCTION

This article presents a selective review of research on measuring and modeling attentional choice and associated methods of model testing and model estimation. The key defining feature of the included material is a focus on the gap between information that is potentially available in the environment and information that is acted upon. This highly inclusive definition conveys the viewpoint that attentional research has a far longer history than is generally appreciated. It was Hayek (1937, 1945) who first opened economists' eyes to the importance of limits on knowledge and Stigler (1961) who first produced a model in which ignorance was optimal. Stigler's theory of search explicitly concerns paying to reduce the gap between potentially available information and private information. Moreover, although centrally focused on economic applications, material that is more traditionally seen as psychological or purely information theoretic in nature is covered where appropriate.

With its inclusive definition, this review must in principle incorporate the huge body of work that predates the recent takeoff in attentional research, which is based on the pioneering work of Sims (1998, 2003, 2010) on rational inattention (RI). The definition incorporates not only the single agent theory of imperfect information, but also the theory of asymmetric information, which imposes explicit barriers on the information internalized by at least one of the parties to a transaction. In addition, it incorporates vast tracts of decision theory, the theory of bounded rationality, behavioral economics, and even cognitive psychology (as stressed in Woodford 2012). Even information about the best production technology is hard to internalize, which has given rise to important theories of learning by doing and technology diffusion that can validly be viewed as reflecting attentional choice.

Given the breadth of the relevant research, it is clear that this review must be highly circumscribed.[1] The specific area of focus is the symbiotic relationship between advances in attentional modeling and advances in measurement. Economic models are typically tested using standard choice data. Unfortunately, such data provide very limited insight into the level of attention. Picking orange from the choice set {apple, orange} has an entirely different implication if the (objective) alternative of an apple was noticed and considered than if it was not. Yet standard economic data do not reveal what was noticed or considered, only what was chosen. This issue is fundamental and essentially universal. Observing final choices alone is inadequate once one makes allowance for incomplete information and its attentional grounding. As scientific researchers, this review argues that we have no choice but to face up to this limitation explicitly, and to work on data enrichments that liberate separate understandings of learning and choice.

The thesis that data enhancements are necessary when allowing for imperfect information is not original to this review. In fact it was stated explicitly by Block & Marshak (1960) more than 50 years ago, when introducing tests of random utility theory in standard data on stochastic choice. They defined "basic observations" as comprising the probability of choosing each available option in all possible choice sets, but were very much concerned with an implied identification problem and situations in which their data set would be insufficient:

> In particular, our operational approach seems to be unable to handle the following distinction that appears natural on grounds of common sense and may be important for predictions. If out of the pair

[1]It is fortunate in this regard that there are massive and wonderful surveys not only on search, on information asymmetries, but also on various aspects of attention on which I do not focus. I mention in particular the surveys by DellaVigna (2007) and Hellwig et al. (2012) and the work of Woodford (2012), which connects attention with experimental psychology and with models of salience and reference dependence.

$F = (a, b)$ of desirable objects a man chooses sometimes a and sometimes b, our introspection tells us that we may ascribe this to either or both of two different "causes":

1. He may have difficulty in perceiving all the relevant characteristics of the objects . . .
2. Even if he knew exactly the differences of the characteristics of the two objects, he might find them almost equally desirable . . . and he will vacillate as a result.

To disentangle the two "causes"—call them perceptibility and desirability (anticipated "satisfaction")—may be important if one wants to predict how people will act if perceptibility is kept constant while desirability varies, or vice versa. (Block & Marschak 1960, p. 99)

Given that the theory of attention is very much concerned with how changes in incentives change learning, Block & Marschak view it as necessary to include richer data when testing such theories. Indeed, they dedicate an entire section of their paper to consideration of alternative domains, albeit to little lasting effect:

> Our particular way of defining the class of basic observations and, correspondingly, of the general testable conditions is to some extent arbitrary . . . . by using a particular demarcation of the class of directly testable conditions (the one most closely corresponding to the nature of economic observations), we are able to carry out a reasonably complete analysis of the relevant logical relations. The study may thus serve as a start when similar attempts are made under another definition of basic observations. (Block & Marschak 1960, pp. 98–99)

Reading their remarkable words, it is hard to understand why there has been so little explicit focus on data enrichment appropriate to attentional modeling. Fortunately, there are notable exceptions to this broad pattern of neglect. Possibly the most well-developed form of data enhancement relevant to attention involves measurement of beliefs, as developed, for example, by Juster (1966) and Manski (1990, 2004) in surveys and Nyarko & Schotter (2002) in experiments. Going beyond this, efforts to test models of boundedly rational behavior are inducing economists to analyze nonstandard data reflective of attention, such as decision times and patterns of search. This review argues that these pioneering measurement efforts represent the tip of a very large iceberg.

Again following the lead of Block & Marschak (1960), this review stresses the value of specifying a particular ideal data set and deriving tests in this context. This is also the revealed preference approach of Samuelson (1938) and Afriat (1967). This approach seeks both to identify precisely the testable implications of broad classes of theories and to liberate parameter estimation when specification tests are passed. A defining feature of revealed preference methods is that they require one to specify upfront an "ideal" data set. Implicitly or explicitly, in the past this ideal data set has comprised stochastic choices alone, as indicated by Block & Marschak (1960). As they point out, the need to specify ideal data more richly is a defining feature of models of attentional choice that take their relationship to measurement seriously.

Conceptualizing ideal data sets in which the role of attention can be modeled and estimated is highly challenging. Ensuring that the corresponding data (or a suitable approximation to it) can be generated is equally challenging. So hard is it to identify new such data sets that it may be wise to reverse the standard process of connecting theory and data. Rather than first developing theories of attentional choice and then searching for corresponding data, the "dual" of this approach may be more fruitful. Appropriate models may be defined as those that have interesting implications in an expanded ideal data set that passes muster with the profession.

This review takes an explicit stand on a newly introduced data set that is particularly appropriate for attentional modeling. Specifically, it focuses on the value of measuring and modeling state-dependent stochastic choice (SDSC) data (Caplin & Dean 2015, Caplin & Martin 2015). Given that it was introduced into psychometric research by Weber 1996 (1834), the most striking thing is how long economists have taken to incorporate it into economic analysis. This review stresses the suitability of SDSC data for developing and estimating next generation models of attention. For example, it shows that appropriately defining the state of the world allows these data to liberate understanding of inattention in games. It shows also that appropriately enriching the interpretation of available actions enables it to shed light on the link between decision times and learning, as in the drift-diffusion model (DDM) of Ratcliff (1978). It argues also that SDSC data correspond perhaps more closely to "the nature of economic observations" than do standard stochastic choice data.

Whether or not one views SDSC data as the ideal data set for attentional research, this review argues that advances in measurement are absolutely required for the next stage of attentional research. As work on attention advances, it will therefore impact not only the substance but also the methods of economic analysis. Engineering of new data sets will become an increasingly essential professional activity. Proposing data enrichments that capture attention and identifying technologies that liberate them will call for research that bridges the gap between theoretical and applied economics, to the benefit of both.

Section 2 introduces key elements in models of attentional choice, focusing in particular on a canonical model of RI in the spirit of Sims (1998, 2003, 2010). Section 3 highlights how the demands of application led researchers to focus on attentional constraints and outlines corresponding enrichments of the model. Section 4 introduces research involving nonstandard data, including beliefs and process correlates of attention, and also introduces revealed preference methods of modeling and measuring attention. Section 5 introduces SDSC data for modeling of attentional choice, and Section 6 outlines possible applications to strategic settings and to learning. Section 7 concludes.

2. RATIONAL PRODUCTION AND CONSUMPTION OF INFORMATION

As currently envisaged, the attentional literature concerns situations in which it is hard to specify explicitly the constraint on becoming more informed and the precise steps that may be involved. In cases of this form, the constraints on knowledge typically reflect some private and hard to observe costs of accessing information that is in principle accessible. Who could argue with the idea that these are important constraints in the era of the Internet, search engines, information overload (Iyengar & Lepper 2000), and nudges (Thaler & Sunstein 2008). In this sense, what makes the recent attentional research new is the focus on internal cognitive constraints on information processing rather than external costs of information access.

This review takes a broader perspective. As conceptualized in this review, all models that take seriously the gap between subjective and potentially available information are attentional. The common feature of all such models is that each decision maker (DM) is envisaged as both the producer and consumer of information. Information is modeled as initially existing external to the DM in a form in which it has no direct value for conditioning choice. The first task for the DM is to summarize or categorize environmental information through some form of internalization. The resulting internal information is then used by the DM in arriving at a final choice. Note that this is true in models of imperfect and asymmetric information no less than in what is currently considered to be attentional research. These are attentional in the current view because there is no hard and fast line between information that must be paid for in a market or is essentially unavailable and information that can in principle be gathered privately through attentional effort.

It is the above definition of the subject matter that conveys its breadth of coverage, and its inclusion of many questions typically covered under the rubric of incomplete information. One might intuitively distinguish informational from attentional constraints based on whether there is some explicit payment or hard constraint on gathering information. In early models of search, gathering of information was conceived of as visible: visiting a store to get a price quote or a firm for a wage offer. Similarly, in buying a valuation signal for drilling rights in an auction, a costly geological investigation or test drill may have to be undertaken to aid in value estimation. Yet one cannot guarantee that information that is purchased is fully internalized. Neither can one guarantee that no additional learning was undertaken. Attentional choice is therefore of relevance even in cases in which it has been traditional to model information gathering as objective.

With respect to information asymmetries, it has been traditional to treat these as defined by the question at hand. It would be absurd to argue with the idea that the current owner of a car knows some facts about its quality that a random buyer does not. However, even here the actual nature of the difference in information may be very subtle, depending on the expertise of the owner in interpreting problems that have been experienced. Is there a $1,000 fix for the apparent failings of the car, or are the defects so overwhelming that it would be best to send the car to the scrap heap? Is the buyer more expert in this assessment than is the seller, or not? How much of this information can be inferred from a review of the service history of the car, how much in a quick drive, and how much in a longer inspection? If a long inspection is undertaken, is the buyer able to interpret the results? All of these unknowns imply that, even in cases in which there seems to be a natural asymmetry of information, it may be important to acknowledge the ability of the supposedly ignorant party to do additional private research. Not only does this suggest the need to model the costs of learning, but it also provides another hint of the difficulties in inference on which this review focuses. What data reveal the information held by two distinct parties who reach an agreement on a trade, when both may have invested to some extent in valuing the traded object?

Rather than try to distinguish them, this review treats them as all being subsumed in a broad general model in which costs and constraints of various kinds prevent agents from knowing all that they might in principle know about the underlying state of the world. Hence the modeling framework introduced in this section is rich enough to handle all forms of information cost, be they physical, mental, or involving monetary payment. If one envisions the DM's internalization of information as guided at least in part by the potential uses of this information, it is natural to connect them in a unified modeling framework. The most obvious analog to the classical theory of profit-maximizing firms producing goods in response to the willingness of consumers to pay is the theory of RI. In this theory, as in the classical theory of supply, production (of information) is motivated entirely by its likely end uses. There are many ways to formulate this general model. Technically, this section follows the approach of Caplin & Dean (2015), whose essentials are based on Sims (1998, 2003, 2010) and captures all models of costly information acquisition dating back at least to the theory of search (Stigler 1961). In the following section, several variations on the theme are outlined as appropriate for application.

The key questions for a microeconomic theory of attention involve first specifying how potentially available information in the environment is internalized, and then how it is used in decision making. Thaler & Sunstein (2008) note that businesses and policy makers may attempt to steer attention in particular ways. But in the end, there is no escaping the need for the DM to interpret the information that is thereby produced. Microeconomic theories of attention can be enriched but not replaced by modeling a sector that produces information or steers attention in particular directions.

2.1. The Value of Information

In the simplest general version of RI theory, there is a finite set of conceivable states of the world, $\omega \in \Omega$. The idea is that it is at least conceivable for the DM to know the true state.

When there is necessarily incomplete information, one can conceptualize the state as specifying all that is currently in principle knowable. It is assumed that the motivation for learning is indirect: Information is of value because of the role it plays in improving the quality of one or more decisions that ensue. In the simplest of cases, one considers a single decision problem, the choice in which hinges on learning. In a given decision problem, there is available a nonempty finite set of actions $A \subset \mathcal{A}$, with $\mathcal{A}$ standing for a larger class of potential actions, some of which are not available in the given context. It is assumed that this set of actions is known to the DM. In the standard RI model, the DM is an expected utility (EU) maximizer. The EU of each action a in each state ω is known and is specified by $u(a, \omega)$. For simplicity, EU is taken as bounded and normalized so that the utility of the best prize and the worst prize is 1 and 0, respectively.

A nonstandard aspect of the formulation above is the explicit separation of actions from their rewards (see Caplin & Martin 2015). This is of the essence in attentional modeling. As in so many other cases, Block & Marschak (1960) were fully aware of the need to separate choice of an alternative from any notion of the corresponding utility in the theory of attention. The fact that they did not do this explicitly reflected only the limits of their ambitions in the particular case at hand: "All of the various definitions of utility given in this paper will be related to empirical entities, called 'alternatives.' Each of these is identified precisely but combines the two aspects, information and desirability, in some unknown though presumably not-too-changeable fashion" (Block & Marschak 1960, p. 99).

Because the theory of attention is concerned largely with settings in which the link between information and desirability is highly changeable, they understood that it would require separation of the empirical entities, which we refer to as action choices, from desirabilities, which we refer to in the standard manner as utilities. Separating out actions from prizes allows in principle for any number of framing effects. To see this, note that one can use the model to capture the distinct choices associated with putting one prize in the top row of a display, and another in the second row.

How such manipulations in a display impact choice depends on beliefs. It would not be surprising if there were a prior belief that the top item in the list was more likely better, with the obvious implication that items higher in the list would be more likely to be considered and chosen. To capture this in a general manner, allow the DM to have prior beliefs $\mu \in \Delta(\Omega)$, which specify states $\Omega(\mu) \equiv \{\omega | \mu(\omega) > 0\}$ as possible, with $\Delta(\Omega)$ the set of such priors. As detailed below, information induces updating. Such updating can change the optimal choice. It is assumed that information is valued precisely and only because of the role it plays in improving the quality of the final decision. Overall, each decision problem is specified by (μ, A), the prior beliefs and the available actions, with $\mathcal{D}$ the set of such problems.

The importance of order effects in actual choice is merely one of the many different forms of framing that can be modeled when one treats actions as distinct from prizes and captures prior beliefs about the mapping. Additional examples of this are provided in Section 6 when discussing strategic applications, in which naming of actions is traditional, as when one formulates a model in terms of choosing the row or column of a matrix with a specific game defined by the prizes associated with any chosen such pair.

In addition to readily capturing many framing effects, the above formulation is strongly related to the vast literature on reference dependence, as Woodford (2012) stresses. In this respect, the key factor is that RI theory focuses on prior beliefs that can richly reflect the context in which the decision is interpreted by the DM.

2.2. The Internalization of Information

The goal in this section is to introduce a general purpose model of costly information acquisition. For our purposes, the most convenient approach is to treat the DM as choosing a set of possible posteriors as well as a mixed action strategy for each possible posterior. We limit ourselves to strategies that are Bayesian in that the expectation of the posteriors is equal to the prior.

Technically, a posterior-based strategy comprises a simple probability distribution over posteriors and mixed action strategies, with $\lambda = (Q_\lambda, q_\lambda)$ denoting these two separate aspects. Given prior $\mu \in \Delta(\Omega)$, the first component of the strategy is a set of posteriors $\gamma \in \Gamma(\mu)$ and their unconditional probabilities $Q_\lambda(\gamma)$. The second component of the strategy specifies the likelihood $q_\lambda(a|\gamma)$ of each action choice for each possible prior γ in $\Gamma(Q_\lambda)$. Given decision problem (μ, A) the feasible decision problems $\lambda \in \Lambda(\mu, A)$ are constrained by Bayes' rule, so that $\mu = \sum_{\gamma \in \Gamma(Q_\lambda)} \gamma Q_\lambda(\gamma)$, and by the fact that $q_\lambda(a|\gamma) > 0$ only if $a \in A$.

There are many equivalent formulations of the strategy space, starting with the information structures of Blackwell (1953). They are also the class of temporal lotteries introduced by Kreps & Porteus (1978) in studying preferences over the resolution of information, those employed by Benoît & Dubra (2011) in considering psychological evidence on overconfidence, and those employed by Kamenica & Gentzkow (2011) in the theory of "Bayesian persuasion."

It is envisaged that each information structure is subjectively costly to produce. As is standard, the cost of strategy $\lambda = (Q_\lambda, q_\lambda)$ depends only on the posterior distribution, not on the mixed action strategy, and is written as $K(\mu, Q_\lambda) \in \bar{\mathbb{R}}$. This function specifies the subjective cost in EU units of the corresponding distribution of posteriors. In the general formulation, one allows costs to be infinite to denote infeasibility, in particular imposing the Bayesian constraint in this manner. It is straightforward to translate standard models of costly information acquisition into this framework, albeit with an uncountable signal set, for example, when purchasing normally distributed signals at cost. In this case, higher costs result in a more precise set of posteriors.

2.3. Optimal Choice

The RI approach treats attention directly as an input-output device, outputting more informed posteriors from less informed priors. This keeps the model close to choice data, a point that is formalized in Sections 5 and 6. It liberates the selection of cost functions to match features of choice just as in production theory and classical choice theory. What is important to learn about a state is how it impacts rewards to available actions, so that the entire style of learning is shaped by its impact on the end result.

In basic RI theory, the DM faced with decision problem $(\mu, A) \in \mathcal{D}$ chooses $\lambda \in \Lambda(\mu, A)$ to balance additional rewards from improved resolution of uncertainty against costs of learning. One can directly compute the resulting choice-based EU as

$$U(\lambda) \equiv \sum_{\gamma \in \Gamma(Q_\lambda)} \sum_{a \in A} Q_\lambda(\gamma) q_\lambda(a|\gamma) \bar{u}(\gamma, a),$$

where $\bar{u}(\gamma, a) = \sum_{\omega \in \Omega(\mu)} \gamma(\omega) u(a, \omega)$. Given also $K \in \mathcal{K}$, the value of strategy $\lambda \in \Lambda(\mu, A)$ is therefore

$$V(\mu, \lambda|K) \equiv U(\lambda) - K(\mu, Q_\lambda).$$

This gives rise directly to the value function and to optimal strategies:

$$\hat{V}(\mu, A|K) \equiv \sup_{\{\lambda \in \Lambda(\mu, A)\}} V(\mu, \lambda|K);$$

$$\hat{\Lambda}(\mu, A|K) \equiv \{\lambda \in \Lambda(\mu, A) | V(\mu, \lambda|K) = \hat{V}(\mu, A|K)\}.$$

Note that this formulation involves the standard assumption that the disutility of an information system is separable from the utility of the prizes that result from the use of information. This is standard in all theories of costly information acquisition, starting with the theory of search. Different RI models involve differentially specifying the cost function. Of particular importance is the Shannon cost function, which specifies costs as an increasing linear function of the expected reduction in entropy between the prior and posterior,

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

Here $H(\gamma) = -\sum_{\omega \in \Omega(\mu)} \gamma(\omega) \ln \gamma(\omega) > 0$ is the Shannon entropy function extended to boundary points using the condition $\lim_{x \searrow 0} x \ln x = 0$.

There can be no doubt that those who are currently engaged in pursuing RI theory are heavily indebted to the remarkable conceptual and mathematical advances that information theory represents (Shannon 1948). In part, this debt relates to Shannon's reduced-form approach to capturing the difficulties of communicating information precisely, as a mapping from a prior to a distribution of posteriors. This appears highly prescient from the perspective of economic application and psychological reality. As Woodford (2012) stresses, neglecting the difficulties of encoding reality precisely has led economists to spend perhaps too long thinking about informational partitions when stochastic treatments are both more psychologically plausible and more closely related to the reality of choice. Glimcher (2011) also stresses the essential stochasticity of any theory of choice that respects fundamental neural constraints.

What is perhaps less easy to understand than the importance of stochasticity per se is the unique simplicity of the entropy function that Shannon introduced for apparently different reasons. A first key simplification that applies not only to the Shannon model but also to a far wider class of models is that one can, for many purposes, work with a very simple class of signals. Much as in the theory of implementation, in which where one can often limit attention to truthful mechanisms, so in RI theory, where one can often limit attention to cases in which the signals are identified with actions that are chosen. Given that in EU theory there is no direct advantage to mixing actions, and also that it is strictly suboptimal to choose the same action from two distinct posteriors given that Blackwell more informative structures are more costly, one can assume that there is a one-to-one mapping between an internal signal and the action that signal induces. This link between the optimal strategy and observed action choices is important from the operational viewpoint, as further discussed below.

What is more remarkable about the Shannon function is the separation it induces between unconditional and state-dependent action probabilities. Provided a generic condition on independence of payoffs is satisfied, the following complementary slackness conditions pin down the unique optimal action probabilities $P(a)$ with the Shannon cost function:

$$\sum_{\omega} \mu(\omega) \left[\frac{\exp(u(a, \omega)/\lambda)}{\sum_{b \in A} P(b) \exp(u(b, \omega)/\lambda)} \right] \leq 1 \text{ all } a \in A, \tag{1}$$

with equality if $P(a) > 0$. The full rationally attentive strategy is defined by state-dependent action probabilities $P(a|\omega)$ using a variant of the logit formula as specified in Section 6 (see Cover & Thomas 2006, Caplin et al. 2015a, Matějka & McKay 2015).

It is arguably reasonable in applications to begin with the Shannon model, not because of its inherent credibility, but rather because of its remarkable simplicity. To some extent, it is the "Cobb-Douglas" model of attention, and moves to richer models are best based on behavioral evidence, as in Woodford (2012). Clearly, better understanding of the Shannon model is required not only to judge where it does too much violence to reality, but also to identify methods of generalization that are analytically and computationally workable.

3. APPLICATIONS AND ADAPTATIONS

Just as all good families are rumored to have common elements, so do all good models of behavior designed for application. One of these is allowance for a gap between subjective and potentially available information. In this section, I discuss a number of different applied literatures and how they came face-to-face with questions of attentional choice. I outline also the modeling steps that were taken, somewhat on a case-by-case basis, to suitably capture attentional considerations. The universal importance of attention shows it to be in the trunk rather than on a branch of the economic tree. It rounds out the theory of incomplete information, which is no less a part of the durable foundations of the discipline than are the theories of production and consumption.

3.1. Costly Signal Acquisition and Sequentially Rational Inattention

Motivated by the observation that many buy goods that cost significantly more than clearly available alternatives, Stigler (1961) models price search as itself costly. To capture this in the RI model, one takes as the set of available actions A purchase of the given item at any of the locations at which it is available (assuming this to be known). The underlying state of the world is the mapping from such locations to prices, $\omega : A \rightarrow \mathbb{R}_+$. In the basic search model, prior beliefs about this price are independent and identically distributed across locations, so that the overall prior $\mu \in \Delta(\Omega)$ is the product of the univariate distributions over prices. The technology of learning is defined as uncovering a given price at a fixed cost $k > 0$. In the original batch version of the model, the decision is made ex ante on how many such goods to search. The utility of a given action is the negative of its price,

$$u(a, \omega) = -p(a).$$

The goal is to minimize the expected sum of learning costs and realized price, which is the minimum price identified in searching that given number of stores. Only if the DM is risk neutral can one interpret the learning costs as being measured in dollars: More generally, they must be measured in EU units as in the RI model above if the DM is not risk neutral.

What has made the search model so durable is precisely the fact that it was the first to openly embrace the gap between objective information and subjective information. Search theory has mushroomed since its introduction in large part because this gap is ubiquitous. There have been literally thousands of adaptations of the model that involve changing the precise definition of the underlying true state of the world, the technology and costs of learning, the nature of the prior, and the nature of the final utility function. Entirely analogous except for the probabilistic nature of the information revelation are the equally many models of costly signal acquisition, as when a potential bidder undertakes geological investigation before deciding how much to bid for the right to drill for oil in a given tract of land. Given the restrictions of space, two and only two of these adaptations are considered in this review: the move from batch to sequential search and allowance for goods with multiple unknown characteristics relevant to utility.

Arguably the most important and fundamental development of the batch model was the sequential search model of McCall (1970), which introduced the now standard reservation wage characterization of optimal strategies.[2] Although the RI model above is most straightforwardly seen as a generalization of the batch search approach of Stigler, adoption of the posterior-based approach allows easy connection to dynamic programming by defining Bellman equations based on recursive application of the cost function. Another key amendment to the standard search model arises when one can ex ante specify various different forms of uncertainty one faces when selecting among distinct goods that differ in more than one clear dimension. To capture this formally in the RI model, one redefines the state of the world as a mapping from action choice to some relevant n-dimensional vector of measurable characteristics, $\omega : A \longrightarrow \mathbb{R}^n$, whose EU is specified by the function $U : \mathbb{R}^n \longrightarrow \mathbb{R}$ over these characteristics:

$$u(a, \omega) = U(\omega(a)).$$

The prior is then a joint distribution over the characteristics associated with each available action choice. Commonly, one then changes the cost function for learning to involve one in which a fixed cost is paid to observe any given characteristic of any given good.

3.2. Attention and Bounds on Rationality

A key question in multidimensional search is when it is optimal to identify in depth the characteristics of one commodity after another, and when instead to compare many goods on a given attribute. These have been labeled alternative-based and attribute-based search, respectively, by Payne et al. (1993). It turns out that optimal strategies can involve a highly complex amalgam of these approaches (Gabaix et al. 2006). This has provoked interest in the development of boundedly rational models of attention and the decision process that typically take one or the other method of search as the starting point.

A pioneering model of boundedly rational search based implicitly on attentional costs is the theory of satisficing behavior, due to Simon (1955). This form of search is strictly alternative based, stopping at the first satisfactory alternative, as in the anchoring and adjustment model of Sauermann and Selten (outlined in Selten 1998) in which satisficing utility changes over time. The marketing literature has also generally adopted an alternative-based approach, focusing on making sure that a given good was considered. The consideration set literature matches in many ways the batch search model of Stigler (1961), albeit without the stress on optimization. In this literature, the key issue is which features of goods grab enough attention to ensure that a good will be among the considered options. With regard to how choices are in the end made among these options, there is less concern with matching any particular process. It is typical in this literature to use some form of random utility model, typically the logit model based on true utilities, to specify an additional layer of randomness in choice from the objects that are in the consideration set.

A major spur to applied work on limited attention was the finding of Madrian & Shea (2001) on how infrequently individuals adjusted default savings rates. This reflects in stark form the status quo bias to which Samuelson & Zeckhauser (1988) had alerted the profession. Although there are remaining controversies about how and why these effects occur, there is little doubt that attentional constraints play a key role (see Geng 2016).

Well before economists became interested in attention, perceptual psychologists focused on the distinction between externally available and subjectively perceived information. The Weber-Fechner laws of psychophysics highlight precisely the distinction between the relative weights of

[2] An argument can be made that the sequential approach to RI predates the batch approach. Wald (1973) introduced sequential analysis (and with it dynamic programming) during World War II precisely to take account of ongoing costs of learning of the quality of valuable products such as munitions.

objects that a subject is holding in either hand as known to the experimenter, and the subjectively perceived difference. Weber's observation that mistakes of judgment are very common when weights are proportionately close resulted in the concept of "just noticeable differences," whose importance to choice theory was highlighted by Tversky (1969) in generating examples of failures of transitivity.

Tversky (1972) also introduced an explicit theory of attention in multidimensional settings based on "elimination by aspects." He posits that the process of choice would inevitably impact the final choice, and argues that it is largely based on sequentially dropping goods clearly dominated in critical characteristics. Although there is some randomness that might play out in ordering attributes for inspection, Tversky hypothesizes that being noticeably worse in an attribute that draws attention would lead to elimination from consideration.

Many models of boundedly rational decision making that specify procedures in detail have issues of tractability in terms of deriving implications for behavior. This has inspired recent research that places tractability and availability of analytic solutions as a high priority. Leading the charge in this respect has been Gabaix (2014), who is pursuing a general purpose and tractable model of bounded rationality with which to rewrite the economic cannon. Spiritually, his sparsity-based model is a characteristics-based model in which it is not worthwhile to pay attention to the full list of characteristics, much as in the modern approach to machine learning.

An interesting question that theorists are starting to address is precisely where to place the boundary between rational and boundedly rational behavior. Models of boundedly rational behavior by definition involve choice mistakes. Patterns in these mistakes are generally used to motivate the corresponding models, as in the case of just noticeable differences. However, once one starts rationalizing these mistakes with a model of perception, one is very close to treating the apparent mistakes as themselves reasonable in light of architectural constraints associated with limited cognitive capacity (Woodford 2012). This raises the question of how far one can go in capturing various forms of seemingly anomalous behavior precisely using the RI framework, and using the evidence to define precisely which failures cannot be so captured. Of particular interest in this regard is the work of Natenzon (2016), who suggests that taking account of prior beliefs in the RI sense may provide an explanation for such otherwise paradoxical behaviors as the "compromise" effect. Caplin et al. (2015a) connect RI with the consideration set literature by showing that RI models can rationalize a decision to ignore many possible decisions on the grounds that they are unlikely to be worth choosing, as well as stochastic choice from the consideration set. Formulae are strikingly simple with the Shannon cost function.

An additional spur to models that hew closely to principles of optimality is that rigid models of boundedly rational behavior can produce absurdity. For example, consider a DM picking between two boxes, one of which contains $1 million and the other nothing. According to the consideration set literature, a DM who was only able to see the contents of the box containing nothing would choose it, as the other box is not in the consideration set. The RI model handles this very simply by specifying a prior over the underlying state that would in this case make it quite unnecessary to check the second box before choosing it. The general feature this points to is that prior beliefs matter a great deal in practice. As Woodford (2012) points out, this means that rationally inattentive behavior leads one to expect context effects, and in fact provides constraints on how they operate that are of possible interest.

3.3. Macroeconomic Applications

Adaptation to both anticipated and unanticipated changes in the environment is often slow. As a result, gradual adjustment to exogenous shocks is ubiquitous in areas ranging from adjustment of nominal prices to adjustment of financial portfolios (Ameriks & Zeldes 2000, Gabaix & Laibson 2001), to adoption of frontier technologies. In many models, such lags are motivated mechanically by the imposition of a timing convention (e.g., time-dependent adjustment of nominal prices). Another prevalent microeconomic foundation involves fixed costs of adjustment. These give rise to infrequent microeconomic adjustment that depends on the distance between current state variables and their optimal levels. There are many cases in which such state-dependent patterns of inertia speed up the macroeconomic response to shocks, particularly in the case of nominal pricing rigidity (Caplin & Spulber 1987, Golosov & Lucas 2007). Mankiw & Reis (2002) show that inattention can induce simple time-dependent rules, although this requires price setters to ignore the boom in sales that is the telltale sign that their price is out of line.

It was Sims (1998, 2003, 2010) who focused economists' minds once and for all on the question of how to incorporate attentional constraints into the analysis of microeconomic and macroeconomic adjustment. He introduced the Shannon model in its dual sense, treating processing time as the scarce resource, the input of which gives rise to clarity. Woodford (2009) introduced the current approach of modeling the cost as linear disutility. Colombo et al. (2014) consider a hybrid approach in which there is a convex cost to learning the precision of signals.

As in the literature on characteristics, many macroeconomic models involve uncertainty that can be modeled as multidimensional. In this context, Paciello & Wiederholt (2014) allow for cost functions that incentivize learning about two distinct sources of uncertainty jointly or separately. Van Nieuwerburgh & Veldkamp (2009, 2010) formulate models of multidimensional uncertainty that can only be resolved using different sources of information to explain home-country bias and underdiversification in investments.

In the field of technology adoption, many learning-based models have been invoked to explain failure of producers to adopt frontier technologies. Some involve standard search-type frictions or learning by doing. Others focus on informational barriers and spillovers across firms (e.g., Jovanovic & MacDonald 1994, Munshi 2003). Spillovers of this form are important to incorporate in models of RI. Often there are variables that are cheap or free to observe (e.g., prices, market aggregates, or decisions of neighbors) and that capture much information that one might otherwise have to incur cognitive costs to learn. How the presence of different forms of information resulting from the choices of others impacts attentional effort is a first-order question that is just starting to draw the interest of researchers (Caplin et al. 2015b). Again, the Shannon cost function turns out to have important simplifying properties.

Another important branch of the macroeconomic literature related to attentional modeling is the literature on nonrational expectations. It is well understood that private expectations are not always based on all information available to the model builder, and indeed that the model builder may be able to identify systematic errors in the expectational models that many hold (e.g., Marcet & Sargent 1989, Hommes 2013). In survey work, Manski (2004), in particular, has pioneered new measurements that provide insight into the nature of the expectations errors that may be made in practice. Although this literature is not always connected to current work on attention, there is nevertheless a strong relationship. Specifically, the model of RI indicates that the effort that individuals make to learn the truth is fundamentally related to its value in decision making. The same may hold for expectations. Those for whom expectations are most important may work harder to ensure their accuracy.

3.4. Attention and Utility

RI theory specifies beliefs as essentially supporting actors that provide no direct utility. They are important only for instrumental purposes in guiding choices among standard prizes. Yet there are many psychologically rich theories in which beliefs play a more critical role than this. This is certainly the case in models of anxiety, fear, and surprise, as captured in Caplin & Leahy (2001). In a standard application, this model can result in rejection of particularly worrying information about a threatening health state such as Huntington's disease. In such cases, attention may alter the balance between ex ante fear and ex post prizes, which in turn offers policy makers methods to change incentives related to disease prevention (Caplin 2003). This desire not to know may be considerably more general, as the survey of Golman et al. (2015) indicates.

In technical terms, it is relatively easy to adapt the RI model to allow for factors that induce nonstandard information preferences by changing the utility function to one that is itself defined on the domain of temporal lotteries. This domain was introduced into utility theory by Kreps & Porteus (1978), and is precisely the same domain on which the attention cost function is defined. In dynamic settings, another important change involves allowing for the ego protective and other roles of "beliefs as assets," in the sense of Bénabou & Tirole (2011): This might be necessary to understand apparently broad-based reluctance to seek out information that calls current beliefs into question.

Another fascinating and related set of phenomena involve reweightings of attributes induced by attentional considerations. Lichtenstein & Slovic (1971) show how different gambling environments induce an apparent shift in focus between prize amounts and associated probabilities in choice among lotteries. Rubinstein (1988) introduces an entirely different interaction between attention and utility. He considers models in which being dissimilar draws attention. If certainty is perceived as more qualitatively distinct from possibility than are correspondingly small differences in likelihood for two possible events, he shows that one can understand otherwise paradoxical patterns of choice among lotteries, such as those associated with the Allais paradox. Köszegi & Szeidl (2013) develop a model in which attention is drawn to attributes that differ the most. Bordalo et al. (2013) further develop this interaction between attention and utility in their general theory of salience-based reweightings of utility that derive from how sharply particular characteristics of products stand out.

The above only scratches the surface of the possible psychological enrichments on the RI model. Bénabou & Tirole (2002) model reasons to work harder to retain good rather than bad memories. Dynamic variants of the RI model appear feasible in which such biases are captured by allowing information that is not used to deteriorate in a stochastic manner. Material that one wished to forget would deliberately lie unused. Depending on the model details, one could also allow direct investment in forgetting that used costly attentional effort. Brocas & Carrillo (2008) and Alonso et al. (2014) develop particular models of mental function that have implications for information processing, whereas McFadden (1999) makes clear how very general is the operation of perceptual constraints in intermediating between reality and behavior.

3.5. Inattentive Play in Games

Issues of inattention are nowhere more apparent than in strategic settings. Yet game theorists have only recently started to model attentional constraints. As in the case of individual decision making, many approaches involve specifying bounds on rationality. Following Nagel (1995) and Ho et al. (1998), the literature on levels of sophistication has mushroomed. Jehiel (2005) also has developed a rich theory of play that involves grouping together objectively distinct strategies into subjective analogy classes.

In terms of applications, the consideration set literature is increasingly being developed in the context of classical models of industrial organization (Eliaz & Spiegler 2011, Spiegler 2011). Another important branch of this literature is due to Gabaix & Laibson (2006) and has to do with shrouded attributes of products that are deliberately made hard to notice and in which overcharging is therefore concentrated. Those designing Internet search engines, store layouts, and product brochures are intimately aware of their ability to manipulate choice by changing presentation. Eliaz et al. (2014) study cases in which the goal for an incumbent supplier is to stay below the DM's radar to avoid being subjected to a competitive price comparison.

There is every reason to expect rapid growth in the literature on attentional costs in strategic settings. Key model elements in a cost-based model of incomplete contracting were proposed by Dye (1985), and further formulation of this issue seems of great interest. In the context of search markets, there is also an interesting set of issues related to attentional constraints, and the possibility that a different microeconomic foundation of the costs of exploring options will give rise to more realistic models of problems in making and maintaining appropriate matches.

Policy makers are also increasingly players in the attention-grabbing (or avoiding) game. The work of Thaler & Sunstein (2008) on behavioral nudges was decisive in turning policy makers' attention to how to manipulate behavior using subtle features of the messaging either to encourage or discourage attention being given to a particular option. Allcott & Mullainathan (2010) summarize the importance of issues of presentation in breaking through the clutter to present clear messages to consumers. Caplin & Lowrance (2014) make the point that interactions among the public, the press, and the policy makers themselves may deflect public attention from potentially important policy changes. In many cases, this form of neglect is what policy makers most desire. It takes only the most casual observation to see how widespread this problem is.

Although there has been important early work on costly information processing in games, the general approach involving RI is just now arriving in strategic settings. Yang (2015) and Martin (2014) show that the fully flexible version of RI with Shannon costs is useful in characterizing equilibrium. Section 6 revisits the strategic modeling of RI.

4. MEASURING ATTENTION WITH NONSTANDARD DATA

4.1. Belief Data and Process Data as Reflections of Attention

The key point of this review is the need for data enrichment to study attention. As noted above, this point was forcefully made by Block & Marschak (1960). Moreover, as again indicated in Section 1, there is increasing recognition that data enrichments are needed to identify private beliefs and a fortiori levels of attention. Following Juster (1966) and Manski (1990), numerical probability assessments are now routinely gathered in such important surveys as the Health and Retirement Study and its international counterparts. As their use has grown, improvements in the technology of measurement have become more important, as in the "bins and balls" method of Delavande & Rohwedder (2008). This is starting to set up a virtuous circle. Improvements in measurement are liberating researchers to assess changes in probabilistic beliefs. Such assessments are of particular value in attentional research. For example, Wiswall & Zafar (2015) use sequential surveys to understand how the provision of objective information on returns to schooling alters understanding. This sheds an important if indirect light on which aspects of supplied information are internalized.

Just as survey measurement of beliefs is advancing, so too is corresponding measurement in experiments following the foundational work of Nyarko & Schotter (2002). As for surveys, increased confidence in the validity of such measurements has spurred research on how events change beliefs. Kuhnen (2015) experimentally investigated changes in beliefs about asset returns as a function of the realized history of such returns. She identified a particularly sharp and negative impact of realized losses on beliefs, which suggests the value of amendments to standard models of updating. It is as if losses not only grab attention, but may also hijack usual methods of updating.

Absent the data on beliefs, this swing to pessimism after a fall in stock prices might have been interpreted as loss aversion, for which little evidence was found in this case. This example illustrates the particular importance of data enrichments in exploring phenomena that are hard to explain with the standard EU model, but may themselves have many possible less standard rationalizations.

There is an important if currently underexplored link between the literature on measuring beliefs and that on costly attention. Consider in this regard the finding that those who are more optimistic about equity returns hold more stocks (Kézdi & Willis 2003). Caution is in order in interpreting this result, in particular if one is tempted to treat the beliefs as exogenous. It may instead be that they result at least in part from incentives. It is plausible that the public is not aware of the extent of the equity premium. In this particular case, prior beliefs may generally be more pessimistic than reality. Those with higher wealth and lower risk aversion and who are therefore more interested in investing in equities are also more likely to invest in learning. Having done so and received the good news, they are likely to invest relatively heavily in equities. The broad point is that the tolerable degree of error in expectations of the future may be endogenous in many settings.

In addition to being of interest in decision problems, belief data are increasingly being employed in analyzing strategic situations. In experiments on information disclosure, Jin et al. (2015) gather belief data to better understand why the unraveling result of Milgrom (1981) does not fully apply in practice. In their experiments, belief data indicate that information receivers are not skeptical enough of nondisclosure, and as a result, information senders sometimes hide the true state.

Another important set of data enrichments beyond measures of expectations relates to the process of search and choice. Gabaix et al. (2006) use the Mouselab interface to motivate a particular model of search. Reutskaja et al. (2011) use eye-tracking technology to support stochastic versions of consideration set models, as it was almost always the case that the chosen object was among those that had previously attracted gaze. Another form of process data is choice process data, which comprise not only the final choice, but also earlier provisional choices (Campbell 1978). The key advantage that choice process data provide is that they enable one to identify options that are rejected in addition to the ultimately accepted object. These data were gathered experimentally by Caplin et al. (2011) to identify a setting in which the satisficing model well-described behavior.

As with belief data, search data are increasingly being developed for strategic settings. Johnson et al. (2002) made pioneering use of the Mouselab interface in casting doubt on models with strategically sophisticated players. Their key empirical observation was that many failed even to uncover the payoffs of other players in the game. Modeling inattention is therefore a particularly high priority in strategic settings.

Choice process data may be of particular value in strategic settings, in part because it is so easy to implement. Agranov et al. (2015) incentivize players in Nagel's 2/3 guessing game to make provisional choices throughout a three-minute period of play. Dynamic patterns of individual play give clear insights into differences in internalization across different players in the game. Many players appear to increasingly internalize the structure of the game. By way of evidence, they change their decisions over time in a manner that both reflects ever higher levels of strategic sophistication and results in higher rewards. Yet a large minority display no such systematic pattern, making choices that continue to fluctuate widely through the entire period of play, even making dominated choices after considerable contemplation of the structure of the game. Looking at their final choices alone is entirely inadequate if one is interested in how choices are being made. More concretely, knowledge of the choice process would greatly help in predicting play after three minutes based on observed play for the first two minutes. Contemporaneous work of Recalde et al. (2013) on how dynamic patterns of play in public goods games are impacted by prior beliefs suggests that choice process data have the potential to reveal much about players' evolving understanding of the game they are playing.

4.2. Ideal Data and Revealed Preference Methods

Although belief and process data are extraordinarily important, they are harder to model than classical choice data, making them more difficult to incorporate into model estimation. It is therefore of interest to consider a different approach to data design that connects with the revealed preference method of Samuelson (1938). Samuelson places data directly as the center point of theoretical research. Following the operational approach of Bridgman, he defines the theory of utility maximization by its implications for a specified ideal data set (see Dixit 2012). The data set of interest was simply choice from budget sets, which all economists agreed was the essential observable in the classical theory of demand.

In the early period of utility theory, when meaning was still attached to marginal utility, many sought to uncover fundamental drivers of utility. Implicit in early thinking was a Benthamite process of reducing goods to their essential "experienced utility" components. The revealed preference approach cemented the move away from psychological origins firmly into the realm of behavior. There is much to be said in favor of the behavioral approach. The number of psychological factors that may enter the evaluation of a set of objects is limitless. Specification of choice data alone as the object of analysis greatly disciplines the modeling process. Indeed, it is also of interest to explore how attentional constraints impact this choice behavior. The pioneering work in this regard is by Manzini & Mariotti (2007), whose "rational short-list" work introduced revealed preference methods to models of boundedly rational decision making. Further progress in this respect has been made by Masatlioglu et al. (2011) in their work on "attention filters."

Although pioneering and important, revealed preference models of attention must place strong conditions on how attention is determined to derive sharp implications for standard choice data. This is clear from even the most trivial of cases, as beliefs about availability and utility are coequal in impacting observed choice. Perhaps the first clear statement of the importance of data enrichment is due to Block & Marschak (1960), who dedicate several sections of their pioneering paper on stochastic choice to a prescient discussion of the potential value of "varying the domain of testability."

This raises natural questions about how to pursue an operational approach to attention that may apply more broadly. Any such general operational approach to attention requires richer measurement to capture how the contemplation process impacts final decisions. One enriched data set that has been theoretically modeled to capture attention is choice process data. Caplin & Dean (2011) develop revealed preference characterizations of various alternative-based search models in these data. Yet even this data set has clear limits, as it is not well suited to model identification when there is any form of characteristics-based search. Another data set used to model inattention involves choice over choice sets. De Oliveira et al. (2013) characterize in these data a general model of RI, with related work by Ergin & Sarver (2010) and Ahn & Sarver (2013). This data set is valuable also for its ability to capture entirely different forces, such as possible self-control problems (Gul & Pesendorfer 2001) and preference for flexibility (Koopmans 1964, Kreps 1979). To date, little has been done to gather these data either in the laboratory or in the field, which is perhaps unfortunate.

5. STATE-DEPENDENT STOCHASTIC CHOICE DATA AND REVEALED PREFERENCE TESTS

5.1. The Data Set

This section introduces a data set that appears to be of particular value in the study of attention. This involves data on SDSC. Given decision problem $(\mu, A) \in \mathcal{D}$, SDSC data are a mapping from possible states to action probabilities,

$$P : \Omega(\mu) \to \Delta(A).$$

The set of such data sets is denoted $\mathcal{P}(\mu, A)$.

To see why this is such a natural data set from a theoretical point of view, it is instructive to round out the characterization of the solution to the Shannon model. As noted in Section 2, one first solves for optimal unconditional action probabilities using complementary slackness conditions. The full rationally attentive strategy is then defined using a variant of the logit formula to compute state-dependent action probabilities $P(a|\omega)$,

$$P(a|\omega) = \left[\frac{P(a)\exp(u(a, \omega)/\lambda)}{\sum_{b\in A} P(b)\exp(u(b, \omega)/\lambda)}\right].$$

Different cost functions have different implications for this data set. It seems possible then that one could characterize not only particular models of rationally inattentive choice but also entire model classes, as for standard choice data. As elaborated below, this is indeed the case: Caplin & Martin (2015) fully characterize Bayesian EU maximization in this data set for any given decision problem. Caplin & Dean (2015) characterize the general model of RI when looking at behavior in a set of distinct decision problems. In the latter case, the approach of Afriat is taken, so that the characterization applies to any finite set of decision problems, and does not aim to pin down the corresponding model objects uniquely.

An important question is how to generate SDSC data in practice. One part of this question concerns how a finite data set can be used to characterize probabilistic choice: In this respect, the situation is no different than in testing models of stochastic choice in which there is no state dependence (see the discussion in Block & Marschak 1960). The second part is how to measure the prior in practice. One option is to specify it experimentally. Experimental work with such data has been ongoing in psychometrics for more than 180 years, ever since Weber [1996 (1834)] randomly placed the heavier of two weights in one hand or another and asked for subjective perceptions. Explicit economic experiments generating the corresponding data are included in earlier working paper versions of Caplin & Martin (2015) and Caplin & Dean (2015). In field applications, the prior can be identified with the empirical distribution, provided successive draws are statistically independent. Martin (2015) tests a parsimonious general model of inattention to prices using scanner data on grocery purchases. In this example, the prior is the empirical distribution of prices and the SDSC is the stochastic choice of bundles at various prices, which is simply the stochastic demand function. In the case of Chetty et al. (2009), who analyze the impact on demand of inclusion or exclusion of sales tax in stated price, the prior would measure the proportion of the time sales tax was included in the stated price of goods, provided this was per se unpredictable in the corresponding field experiment.

Some might ask whether SDSC data are best regarded as standard choice data. If this is a question at all, it is not an important one. Certainly, it seems to involve nothing but measured choices in the standard sense. Yet it also has nonstandard aspects, in particular the assumption that the prior is included in the choice data set.

5.2. Bayesian Expected Utility Maximization

Caplin & Martin (2015) identify simple conditions on data $P \in \mathcal{P}(\mu, A)$ for a single decision problem $(\mu, A) \in \mathcal{D}$ to be consistent with selection of an optimal strategy by a Bayesian EU maximizing maximizer. The key data-based construct in stating this is the "revealed" posterior that is implied by the data in light of Bayes' rule,

$$\gamma_P^a(\omega) \equiv \frac{\mu(\omega)P(a,\omega)}{P(a)},$$

with $P(a) = \sum_{\omega\in\Omega}\mu(\omega)P(a,\omega)$ and $\Gamma(P)$ the set of such posteriors. Caplin & Martin show that Bayesian EU maximization is characterized by the impossibility of raising utility by switching wholesale from one action to another. This condition is formalized in the "no improving action switches" (NIAS) inequalities. These inequalities are not only necessary but sufficient for identifying a utility function and an optimizing choice strategy as the representation requires.

Axiom 1 (NIAS) Data set $P \in \mathcal{P}(\mu, A)$ satisfies NIAS if, given $a \in A$ with $P(a) > 0$,

$$\sum_{\omega\in\Omega}\gamma_P^a(\omega)\,u(a,\omega) = \max_{a\in A}\sum_{b\in A}\sum_{\omega\in\Omega}\gamma_P^a(\omega)\,u(b,\omega) \equiv \hat{u}(\gamma_P^a, A).$$

Note that one can treat both the attention strategy and the utility function as unknown. In many applications, it is simpler to treat the utility function as having been recovered already from standard choice data. In such cases, the NIAS conditions impose linear inequality constraints on the data. Caplin & Martin (2015) present three applications of the NIAS inequalities. The first involves making predictions for behavior that are robust to the exact form of signal processing. The second illustrates the bounds that the NIAS inequalities place on utility. The third application shows that simple models based on consideration sets all produce behaviors that violate the NIAS inequalities, and hence cannot be so rationalized in standard Bayesian models. The same applies to behaviors generated by procedural models of list order search proposed by Rubinstein & Salant (2006) and Salant & Rubinstein (2008) and the standard logit model of discrete choice.

5.3. Rational Inattention

In technical terms, the NIAS characterization is entirely straightforward. Its main value lies in confirming the potential value of SDSC data in modeling of attention. Further to this point, one can use this data set to characterize the theory of RI. The model of Bayesian EU maximization above does not specify how the information structure is chosen. RI theory goes one step further and specifies that it is optimally chosen. Testing this theory effectively requires data on a set of decision problems, as the theory restricts the response to attentional incentives. Caplin & Dean (2015) show that the only additional requirement beyond NIAS involves ruling out raising EU by moving chosen information structures between decision problems. Specifically, the "no improving attention cycles" (NIAC) condition implies that one cannot cycle attention strategies and increase maximal gross utility.

The key data-based construct needed to define this condition is the EU associated with the revealed attention strategy. Formally, the revealed posterior-based strategy $\lambda(P) = (Q_P, q_P)$ comprises the probability distribution over revealed posteriors and the mixed strategy as revealed in the data:

$$Q_P(\gamma) = \sum_{\{a\in A\}|\bar{\gamma}_P^a=\gamma\}} P(a);$$

$$q_P(a|\gamma) = \begin{cases} \frac{P(a)}{Q_P(\gamma)} & \text{if } \bar{\gamma}_P^a = \gamma; \\ 0 & \text{if } \bar{\gamma}_P^a \neq \gamma. \end{cases}$$

The optimal utility of this strategy for given choice data is

$$\hat{U}(\mu, A, P) \equiv \sum_{\gamma \in \Gamma(P)} Q_P(\gamma)\hat{u}(\gamma, A).$$

The NIAC condition applies to data for a set of distinct decision problems.

Axiom 2 (NIAC) Data set $P(k) \in \mathcal{P}(\mu, A(k))$ for $1 \leq k \leq K$ satisfies NIAC if, for any selection of decision problems $(\mu, A(m))$ for $1 \leq m \leq M$ with $A(1) = A(M)$,

$$\sum_{m=1}^{M-1} \hat{U}(\mu, A(m), P(m)) \geq \sum_{m=1}^{M-1} \hat{U}(\mu, A(m), P(m+1)).$$

5.4. Specification Tests and Parametric Models

As Varian (1982) and Choi et al. (2007) highlight, revealed preference methods liberate both specification tests for broad classes of models and parameter estimation in models that pass these tests. Caplin & Dean develop experiments that illustrate the former, with the general model of RI broadly passing the specification test in their particular experiment. Woodford (2012) illustrates the use of SDSC data to refine models of rational attention based on the psychometric experiments of Shaw & Shaw (1977). Noting the key role that the prior plays in determining how well decisions are made in different contingencies leads him to propose a cost function that penalizes learning in unlikely states by relating the cost of learning to the potential cost of information revelation. Undertaking further experiments of this kind and using properties of the data to further refine models are crucial next steps in attentional research.

As with standard choice data, SDSC data are of interest even when the theory that they are initially designed to test blatantly fails the specification test. In introducing the revealed preference approach, Samuelson (1938) made clear his view that the choice data were more fundamental than the model of utility maximization, which, if false, should be rejected in favor of models that more accurately characterize observed patterns of behavior. Almost 70 years later, Manzini & Mariotti (2007) introduced the "weak WARP" test precisely to characterize a class of boundedly rational models in these data. In just the same way, SDSC data may be of interest when the DM is not rationally inattentive, and even when Bayes' rule is not adhered to. This is taken up again in the next section.

6. TWO APPLICATIONS

This section shows that SDSC data provide a flexible framework for exploring the role of attention. It also shows that the precise data set gathered depends entirely on how one specifies the state of the world and the action set. It is the generality of these definitions that makes the data set of such high potential. This section briefly sketches two applications: that to inattentive play in games and that to the link between time input and learning.

6.1. Inattention in Games

SDSC data have particular potential in strategic settings. They may make it possible in principle to distinguish between the many theoretically reasonable forms of inattentive play. In this case, it is particularly important to specify what it is that the data set contains, noting that it may be entirely different from what any individual player gets to see. Treating the notion of ideal data seriously, the very best data set would involve an observer who sees the play of all players in all conceivable games, knowing precisely in each case the identities of the players, their objectively available strategies, and the outcome of the game. For the sake of concreteness, in what follows I discuss a far more limited data set, comprising a set of games that are played among a fixed group of N players, each of whom has a fixed set of available actions.

To begin, the overall action set is taken to be a product set of the action sets of all N players in the game,

$$\bar{A} \equiv \prod_{i=1}^{N} A^i.$$

Note that a particular action for a particular player is taken to be something that is globally defined regardless of the ensuing payoffs, such as picking the top row in a matrix game, or placing a monetary bid in a lottery. The state of the world $\omega \in \Omega$ corresponds to the game that is actually being played (as known to the data analyst). The prior $\mu \in \Delta(\Omega)$ specifies the relative frequencies with which distinct games are played.

SDSC data can in principle be used to test theories of game play. To simplify, assume that the observer knows the full set of utility functions $\{u^i\}_{i=1}^N$, hence the corresponding utility levels $u^i(\bar{a}, \omega)$ for all players in all games associated with all action profiles $\bar{a} = \{a^i\}_{i=1}^N \in \bar{A}$. Given this, one can specify first the theory of Nash play in each game in isolation for a fully attentive player. If this is rejected, a natural alternative involves only partial attention being paid on the part of players to the actual game that they are playing. For example, if the prior indicates that one game is particularly common, a player may reasonably choose corresponding optimal play always, even when the outside observer sees it to be mistaken. The bottom line is that the game that model builders specify inattentive participants to be playing may have little to do with the game as subjectively perceived. The need for data enrichment is particularly clear in this case, as any pattern of play can be rationalized in any given game as a Nash equilibrium of a subjectively perceived distortion of the actual game.

With SDSC data, there is a natural relaxation of Nash play on a game-by-game basis, which corresponds to some form of Nash play in the perceived game rather than the actual game. In the case in which these perceptions are the result of Bayesian updating, it seems natural to consider a mutual version of the NIAS conditions to provide the corresponding restrictions on the data. Interestingly, Bergemann & Morris (2014) show mutual NIAS to characterize a robust formulation of Bayesian Nash equilibrium play in a given game of imperfect information. Of course, there is an important interpretational difference. In the Bergemann & Morris interpretation, players are involved in only one game, and the Bayesian Nash equilibrium concept summarizes the extent to which they learn this game. In the SDSC data, it is instead a reflection of the chosen level of attention to the game that is in fact being played. Here NIAS does not imply Nash play in any given game. Rather it is a cross-game restriction that specifies that, on average, when picking the first row, it must be that a wholesale switch to the second row is not beneficial in the entire class of games being played. In the specific game that the observer knows is being played in a particular instance, anything goes.

The general point is that once one allows for inattention, it is clear that almost any pattern of play can be rationalized in almost any objective game by assertions as to how the game was misperceived. Disciplining these assertions requires gathering data on the richer set of games that are hypothesized to underlie the misperception. SDSC data look particularly promising in this regard.

6.2. Time to Decide and Learning

The value of exploring data on decision time has been stressed by Wilcox (1993), Kocher & Sutter (2006), Rubinstein (2007), Chabris et al. (2009), Achtziger & Alos-Ferrer (2014), Spiliopoulos & Ortmann (2014), Geng (2016), and Caplin & Martin (2016), among others. In a typical use, it is correlated with the degree to which attention appears to have been given to the decision problem or game of interest. The broad finding is that longer decision times are indicative of greater attention to the details of the problem and hence better task performance.

As has generally been the case with attentional measurements, psychometricians focused on decision time well before economists. Indeed, they took an important step beyond this by specifically studying the relationship between decision time and decision quality in the literature on the DDM (see Ratcliff 1978, Ratcliff & McKoon 2008). Variants of this model are increasingly making their way into economics (see Fehr & Rangel 2011, Krajbich & Rangel 2011, Krajbich et al. 2012, Fudenberg et al. 2015), so that measurement and model enrichments are increasingly on the research agenda.

In the classical DDM experiment, a prize is available to a DM in one of two locations, and there is a flow of evidence indicative of which box contains the prize. The DM decides when to stop the flow of evidence by picking a box, and which box to pick. The data set that is produced is precisely SDSC data, with the particular feature that included among the actions is the possibility of delaying choice. Technically, a general version of the model allows for action set A^*, which comprises not only which of a finite number of standard actions $a \in A$ is taken, but also when it is taken:

$$A^* \equiv \{(a, t) | a \in A, t \geq 0\}.$$

The data reveal how this choice is linked to an underlying state of the world $\omega \in \Omega$, which specifies precisely which reward is in which location. The prior $\mu \in \Delta(\Omega)$ is also fixed in a typical experiment, and indicates how likely is each location to contain each reward. Hence, the data set is precisely in the form of SDSC data, $P \in \mathcal{P}(\mu, A^*)$.

The standard theory of choice in theoretical treatments of the DDM involves the gradual accretion of evidence to a decision threshold that triggers choice. Key to this is the level of state-dependent EU $u(a, \omega)$, which is generally assumed to correspond to some fixed underlying utility over the set of prizes that are available, independent of prize location. Another common assumption is that the DM discounts delay so that utility diminishes to $\beta^t u(a, \omega)$ if the choice is made at discrete time t.

The current approach to DDM experiments involves tightly specifying particular models of the learning process and then estimating parameters. Perceiving of the data set rather than this particular theory as primary, it becomes clear that it may be of value to introduce specification tests for general classes of theories in addition to parametric tests of particular models of learning. Included in this might be tests of (recursively) RI. Of course, the data may reject this theory, in which case it may be of value to characterize alternative models in these same data. For example, an interesting class of alternative models are those in which Bayesianism is a desire that is hard to achieve rather than being instinctive and free. This is much in the spirit of the fact that, on being told of violations of the EU axioms, many revise their decisions in the direction of consistency. It is hard not to believe that it may take effort to achieve the Bayesian ideal. An open question in this modeling is what bias shows up if one puts in less effort, and how precisely one can use SDSC data in identifying such failings.

Another open set of issues relates to the utility of learning, an element that is entirely missing in the simple RI model. One model amendment that would allow for this involves modeling for nonseparability between the prizes and the costs/utility of attention. This would enable one to capture spending more time thinking about good than bad outcomes. There are also more direct differences in utility: Some things are enjoyable to think and learn about; others are not. Research on the utility of thought is of particular interest in light of the ongoing movement to gamify learning, which is based on the reasonable idea that more will be learned if the experience itself is more enjoyable.

7. CONCLUDING REMARKS

The review stresses the value of suitably enriched data sets for attentional research, focusing in particular on state-dependent stochastic choice data. Nonstandard data of this form have the potential to provoke rich and mutually beneficial dialog between theory and measurement.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENT

I thank Joseph Briggs, Daniel Martin, and Ruth Wyatt for valuable guidance on how to organize and present this material.

LITERATURE CITED

Achtziger A, Alos-Ferrer C. 2014. Fast or rational? A response-times study of Bayesian updating. *Manag. Sci.* 60:923–38

Afriat S. 1967. The construction of a utility function from demand data. *Int. Econ. Rev.* 8:67–77

Agranov M, Caplin A, Tergiman C. 2015. Naive play and the process of choice in guessing games. *J. Econ. Sci. Assoc.* 1:1–12

Ahn DS, Sarver T. 2013. Preference for flexibility and random choice. *Econometrica* 81:341–61

Allcott H, Mullainathan S. 2010. Behavioral science and energy policy. *Science* 327:1204–5

Alonso R, Brocas I, Carrillo JD. 2014. Resource allocation in the brain. *Rev. Econ. Stud.* 81:501–34

Ameriks J, Zeldes S. 2000. *How do household portfolio shares vary with age?* Work. Pap., Columbia Univ., New York. **https://www0.gsb.columbia.edu/mygsb/faculty/research/pubfiles/16/Ameriks_Zeldes_age_Sept_2004d.pdf**

Bénabou R, Tirole J. 2002. Self-confidence and personal motivation. *Q. J. Econ.* 117:871–915

Bénabou R, Tirole J. 2011. Identity, morals, and taboos: beliefs as assets. *Q. J. Econ.* 126:805–55

Benoît J-P, Dubra J. 2011 Apparent overconfidence. *Econometrica* 79:1591–625

Bergemann D, Morris S. 2014. *Bayes correlated equilibrium and the comparison of information structures in games*. Discuss. Pap. 1909RR, Cowles Found., Yale Univ., New Haven, CT

Blackwell D. 1953. Equivalent comparisons of experiments. *Ann. Math. Stat.* 24:265–72

Block HD, Marschak J. 1960. Random orderings and stochastic theories of response. In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, ed. I Olkin, pp. 97–132. Stanford, CA: Stanford Univ. Press

Bordalo P, Gennaioli N, Shleifer A. 2013. Salience and consumer choice. *J. Polit. Econ.* 121:803–43

Brocas I, Carrillo JD. 2008. The brain as a hierarchical organization. *Am. Econ. Rev.* 98:1312–46

Campbell D. 1978. Realization of choice functions. *Econometrica* 46:171–80

Caplin A. 2003. Fear as a policy instrument. In *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice*, ed. G Loewenstein, D Read, RF Baumeister, pp. 441–58. New York: Russell Sage Found.

Caplin A, Dean M. 2011. Search, choice, and revealed preference. *Theor. Econ.* 6:19–48
Caplin A, Dean M. 2015. Revealed preference, rational inattention, and costly information acquisition. *Am. Econ. Rev.* 105:2183–203
Caplin A, Dean M, Leahy J. 2015a. *Consideration sets and rational inattention*. Work. Pap., New York Univ., New York
Caplin A, Dean M, Martin D. 2011. Search and satisficing. *Am. Econ. Rev.* 101:2899–922
Caplin A, Leahy J. 2001. Psychological expected utility theory and anticipatory feelings. *Q. J. Econ.* 116:55–79
Caplin A, Leahy J, Matějka F. 2015b. *Learning from market share when consumers are rationally inattentive*. Work. Pap., New York Univ., New York
Caplin A, Lowrance R. 2014. The mortgage mess, the press, and the politics of inattention. *Am. Econ. Rev.* 104(5):77–81
Caplin A, Martin D. 2015. A testable theory of imperfect perception. *Econ. J.* 125:184–202
Caplin A, Martin D. 2016. The dual-process drift diffusion model: evidence from response times. *Econ. Inq.* 54:1274–82
Caplin A, Spulber D. 1987. Menu costs and the neutrality of money. *Q. J. Econ.* 102:703–25
Chabris C, Morris C, Taubinsky D, Laibson D, Schuldt J. 2009. The allocation of time in decision-making. *J. Eur. Econ. Assoc.* 7:628–37
Chetty R, Looney A, Kroft K. 2009. Salience and taxation: theory and evidence. *Am. Econ. Rev.* 99:1145–77
Choi S, Fisman R, Gale D, Kariv S. 2007. Consistency and heterogeneity of individual behavior under uncertainty. *Am. Econ. Rev.* 97:1921–38
Colombo L, Femminis G, Pavan A. 2014. Information acquisition and welfare. *Rev. Econ. Stud.* 81:1438–83
Cover T, Thomas J. 2006. *Elements of Information Theory*. New York: Wiley. 2nd ed.
De Oliveira H, Denti T, Mihm M, Ozbek MK. 2013. *Rationally inattentive preferences*. SSRN Work. Pap. **http://ssrn.com/abstract=2274286**
Delavande A, Rohwedder S. 2008. Eliciting subjective probabilities in Internet surveys. *Public Opin. Q.* 72:866–91
DellaVigna S. 2007. *Psychology and economics: evidence from the field*. NBER Work. Pap. 13420
Dixit A. 2012. Paul Samuelson's legacy. *Annu. Rev. Econ.* 4:1–31
Dye RA. 1985. Costly contract contingencies. *Int. Econ. Rev.* 26:233–50
Eliaz K, de Clippel G, Rozen K. 2014. Competing for consumer inattention. *J. Polit. Econ.* 122:1203–34
Eliaz K, Spiegler R. 2011. Consideration sets and competitive marketing. *Rev. Econ. Stud.* 78:235–62
Ergin H, Sarver T. 2010. A unique costly contemplation representation. *Econometrica* 78:1285–339
Fehr E, Rangel A. 2011. Neuroeconomic foundations of economic choice—recent advances. *J. Econ. Perspect.* 25(4):3–30
Fudenberg D, Strack P, Strzalecki T. 2015. Stochastic choice and optimal sequential sampling. arXiv:1505.03342 [q-bio.NC]
Gabaix X. 2014. A sparsity-based model of bounded rationality. *Q. J. Econ.* 129:1661–710
Gabaix X, Laibson D. 2001. The 6D bias and the equity-premium puzzle. In *NBER Macroeconomics Annual 2001*, ed. BS Bernanke, K Rogoff, pp. 257–330. Cambridge, MA: MIT Press
Gabaix X, Laibson D. 2006. Shrouded attributes, consumer myopia, and information suppression in competitive markets. *Q. J. Econ.* 121:505–40
Gabaix X, Laibson D, Moloche G, Weinberg S. 2006. Costly information acquisition: experimental analysis of a boundedly rational model. *Am. Econ. Rev.* 96:1043–68
Geng S. 2016. Decision time, consideration time, and status quo bias. *Econ. Inq.* 54:433–49
Glimcher P. 2011. *Foundations of Neuroeconomic Analysis*. New York: Oxford Univ. Press
Golman R, Hagmann D, Loewenstein G. 2015. *Information avoidance*. SSRN Work. Pap. **http://ssrn.com/abstract=2633226**
Golosov M, Lucas R. 2007. Menu costs and Phillips curves. *J. Polit. Econ.* 115:171–99
Gul F, Pesendorfer W. 2001. Temptation and self-control. *Econometrica* 69:1403–35
Hayek FA. 1937. Economics and knowledge. *Economica* 4:33–54
Hayek FA. 1945. The use of knowledge in society. *Am. Econ. Rev.* 35:519–30
Hellwig C, Kohls S, Veldkamp L. 2012. Information choice technologies. *Am. Econ. Rev.* 102(3):35–40

Ho T, Camerer C, Weigelt K. 1998. Iterated dominance and iterated best-response in experimental "*p*-beauty contests." *Am. Econ. Rev.* 88:947–69
Hommes C. 2013. *Behavioral Rationality and Heterogeneous Expectations in Complex Economic Systems*. Cambridge, UK: Cambridge Univ. Press
Iyengar SS, Lepper MR. 2000. When choice is demotivating: Can one desire too much of a good thing? *J. Person. Soc. Psychol.* 79:995–1006
Jehiel P. 2005. Analogy-based expectation equilibrium. *J. Econ. Theory* 123:81–104
Jin G, Luca M, Martin D. 2015. *Is no news (perceived as) bad news? An experimental investigation of information disclosure*. NBER Work. Pap. 21099
Johnson EJ, Camerer C, Sen S, Rymon T. 2002. Detecting failures of backward induction: monitoring information search in sequential bargaining. *J. Econ. Theory* 104:16–47
Jovanovic B, MacDonald GM. 1994. Competitive diffusion. *J. Polit. Econ.* 102:24–52
Juster FT. 1966. Consumer buying intentions and purchase probability: an experiment in survey design. *J. Am. Stat. Assoc.* 61:658–96
Kamenica E, Gentzkow M. 2011. Bayesian persuasion. *Am. Econ. Rev.* 101:2590–615
Kézdi G, Willis RJ. 2003. *Who becomes a stockholder? Expectations, subjective uncertainty, and asset allocation*. Res. Pap. WP 39, Mich. Retire. Res. Cent., Ann Arbor
Kocher MG, Sutter M. 2006. Time is money—time pressure, incentives, and the quality of decision-making. *J. Econ. Behav. Organ.* 61:375–39
Koopmans TC. 1964. *On flexibility of future preference*. Discuss. Pap. 150, Cowles Found. Res. Econ., Yale Univ., New Haven, CT
Kőszegi B, Szeidl A. 2013. A model of focusing in economic choice. *Q. J. Econ.* 128:53–104
Krajbich I, Lu D, Camerer C, Rangel A. 2012. The attentional drift-diffusion model extends to simple purchasing decisions. *Front. Psychol.* 3:193
Krajbich I, Rangel A. 2011. Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *PNAS* 108:13852–57
Kreps DM. 1979. A representation theorem for preference for flexibility. *Econometrica* 47:565–77
Kreps DM, Porteus EL. 1978. Temporal resolution of uncertainty and dynamic choice theory. *Econometrica* 46:185–200
Kuhnen CM. 2015. Asymmetric learning from financial information. *J. Finance* 70:2029–62
Lichtenstein S, Slovic P. 1971. Reversals of preference between bids and choices in gambling decisions. *J. Exp. Psychol.* 89:46–55
Madrian BC, Shea DF. 2001. The power of suggestion: inertia in 401(k) participation and saving behavior. *Q. J. Econ.* 116:1149–87
Mankiw NG, Reis R. 2002. Sticky information versus sticky prices: a proposal to replace the New Keynesian Phillips curve. *Q. J. Econ.* 117:1295–328
Manski CF. 1990. The use of intentions data to predict behavior: a best-case analysis. *J. Am. Stat. Assoc.* 85:934–40
Manski CF. 2004. Measuring expectations. *Econometrica* 72:1329–76
Manzini P, Mariotti M. 2007. Sequentially rationalizable choice. *Am. Econ. Rev.* 97:1824–40
Marcet A, Sargent TJ. 1989. Convergence of least-squares learning in environments with hidden state variables and private information. *J. Polit. Econ.* 97:1306–22
Martin D. 2014. *Strategic pricing and rational inattention to quality*. Work. Pap., New York Univ., New York
Martin D. 2015. *Consumer theory with inattention to prices*. Unpublished manuscript, Manag. Econ. Decis. Sci., Northwestern Univ., Evanston, IL
Masatlioglu Y, Nakajima D, Ozbay E. 2011. Revealed attention. *Am. Econ. Rev.* 102:2183–205
Matějka F, McKay A. 2015. Rational inattention to discrete choices: a new foundation for the multinomial logit model. *Am. Econ. Rev.* 105:272–98
McCall JJ. 1970. Economics of information and job search. *Q. J. Econ.* 84:113–26
McFadden D. 1999. Rationality for economists? *J. Risk Uncertain.* 19:73–105
Milgrom PR. 1981. Good news and bad news: representation theorems and applications. *Bell J. Econ.* 12:380–91
Munshi K. 2003. Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution. *Rev. Econ. Stud.* 73:175–203

Nagel R. 1995. Unraveling in guessing games: an experimental study. *Am. Econ. Rev.* 85:1313–26
Natenzon P. 2016. *Random choice and learning*. Unpublished manuscript, Univ. Washington St. Louis. **https://pages.wustl.edu/files/pages/imce/pnatenzon/2016-06-13-rcl-natenzon.pdf**
Nyarko Y, Schotter A. 2002. An experimental study of belief learning using elicited beliefs. *Econometrica* 70:971–1005
Paciello L, Wiederholt M. 2014. Exogenous information, endogenous information, and optimal monetary policy. *Rev. Econ. Stud.* 81:356–88
Payne JW, Bettman JR, Johnson EJ. 1993. *The Adaptive Decision Maker*. Cambridge, UK: Cambridge Univ. Press
Ratcliff R. 1978. A theory of memory retrieval. *Psychol. Rev.* 85(2):59–108
Ratcliff R, McKoon G. 2008. The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput.* 20:873–922
Recalde M, Riedl A, Vesterlund L. 2013. *Intuitive generosity and error prone inference from response time*. Unpublished manuscript, Univ. Pittsburgh, PA
Reutskaja E, Pulst-Korenberg J, Nagel R, Camerer C, Rangel A. 2011. Economic decision-making under conditions of extreme time pressure and option overload: an eye-tracking study. *Am. Econ. Rev.* 101:900–26
Rubinstein A. 1988. Similarity and decision-making under risk (is there a utility theory resolution to the Allais paradox?). *J. Econ. Theory* 46:145–53
Rubinstein A. 2007. Instinctive and cognitive reasoning: a study of response times. *Econ. J.* 117:1243–59
Rubinstein A, Salant Y. 2006. A model of choice from lists. *Theor. Econ.* 1:3–17
Salant Y, Rubinstein A. 2008. (A, f): choice with frames. *Rev. Econ. Stud.* 75:1287–96
Samuelson PA. 1938. A note on the pure theory of consumer's choice. *Economica* 5:61–71
Samuelson W, Zeckhauser R. 1988. Status quo bias in decision making. *J. Risk Uncertain.* 1:7–59
Selten R. 1998. Aspiration adaptation theory. *J. Math. Psychol.* 42:191–214
Shannon CE. 1948. A mathematical theory of communication. *Bell Syst. Tech. J.* 27:379–423
Shaw ML, Shaw P. 1977. Optimal allocation of cognitive resources to spatial locations. *J. Exp. Psychol.* 3:201–11
Simon H. 1955. A behavioral model of rational choice. *Q. J. Econ.* 69:99–118
Sims CA. 1998. Stickiness. *Carnegie-Rochester Conf. Ser. Public Policy* 49:317–56
Sims CA. 2003. Implications of rational inattention. *J. Monet. Econ.* 50:665–90
Sims CA. 2010. Rational inattention and monetary economics. In *Handbook of Monetary Economics*, Vol. 3, ed. BM Friedman, M Woodford, pp. 155–81. Amsterdam: North-Holland
Spiegler R. 2011. *Bounded Rationality and Industrial Organization*. New York: Oxford Univ. Press
Spiliopoulos L, Ortmann A. 2014. *The BCD of response time analysis in experimental economics*. SSRN Work. Pap. **http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2401325**
Stigler GJ. 1961. The economics of information. *J. Polit. Econ.* 69:213–25
Thaler RH, Sunstein CR. 2008. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven, CT: Yale Univ. Press
Tversky A. 1969. Intransitivity of preferences. *Psychol. Rev.* 76:31–48
Tversky A. 1972. Elimination by aspects: a theory of choice. *Psychol. Rev.* 79:281–99
Van Nieuwerburgh S, Veldkamp L. 2009. Information immobility and the home bias puzzle. *J. Finance* 64:1187–215
Van Nieuwerburgh S, Veldkamp L. 2010. Information acquisition and under-diversification. *Rev. Econ. Stud.* 77:779–805
Varian HR. 1982. The nonparametric approach to demand analysis. *Econometrica* 50:945–73
Wald A. 1973. *Sequential Analysis*. New York: Dover
Weber EH. 1996 (1834). *On the Tactile Senses*, ed. HE Ross, DJ Murray. New York: Exp. Psychol. Soc. 2nd ed.
Wilcox NT. 1993. Lottery choice: incentives, complexity and decision time. *Econ. J.* 103:1397–417
Wiswall M, Zafar B. 2015. Determinants of college major choice: identification using an information experiment. *Rev. Econ. Stud.* 82:791–824
Woodford M. 2009. Information-constrained state-dependent pricing. *J. Monetary Econ.* 56:100–24
Woodford M. 2012. *Inattentive valuation and reference dependent choice*. Work. Pap., Columbia Univ., New York
Yang M. 2015. Coordination with flexible information acquisition. *J. Econ. Theory* 158:721–38

Annual Review of Economics

Volume 8, 2016

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Economics* articles may be found at http://www.annualreviews.org/errata/economics