Optimal Evidence Accumulation and Stochastic Choice*

Michael Woodford
Columbia University

August 11, 2016

Abstract

A model is proposed in which stochastic choice results from noise in cognitive processing rather than random variation in preferences. The mental process used to make a choice is nonetheless optimal, subject to a constraint on available information-processing capacity that is partially motivated by neurophysiological evidence. The optimal information-constrained model is found to offer a better fit to experimental data on choice frequencies and reaction times than either a purely mechanical process model of choice (the drift-diffusion model) or an optimizing model with fewer constraints on feasible choice processes (the rational inattention model).

*Revision of a paper previously circulated under the title "An Optimizing Neuroeconomic Model of Discrete Choice." I would like to thank Ian Krajbich for sharing the data from Krajbich *et al.* (2010); Nico Foley, Ben Hébert, Ian Krajbich, Stephen Morris, Antonio Rangel, Aldo Rustichini, Michael Shadlen, and Jón Steinsson for helpful discussions; Stéphane Dupraz, Kyle Jurado, and Danyan Zha for research assistance; and the Institute for New Economic Thinking, the National Science Foundation, and the Kumho Visiting Professorship, Yale University, for supporting this research.

1 Modeling Stochastic Choice

Since at least the work of Mosteller and Nogee (1951), it has been noted in experimental studies of choice behavior that choices have an apparently random element. This is true not only in the sense that less than 100 percent of the variation in choices across trials can be accounted for by a deterministic function of a particular list of measured characteristics, but more tellingly, in the sense that a given subject does not always make the same choice when presented again with exactly the same set of alternatives, sometimes only a few minutes later (Hey, 1995, 2001; Ballinger and Wilcox, 1997; Cheremukhin *et al.*, 2011). This raises the question of how such randomness can be understood, and in particular, whether it can be reconciled with a theory according to which people's choices reflect well-defined and stable preferences, which one should (at least in principle) be able to infer from those choices.

The approach proposed here views stochastic choice as representing random errors in cognitive processing, akin to random errors in perception. A long literature in experimental psychology and neuroscience has documented the randomness of the responses that subjects give when asked to make perceptual judgments, in a variety of sensory domains. The probability of the subject's giving one response instead of the other (when required to choose between two possible responses) is found to vary continuously with continuous changes in the characteristics of the stimuli, rather than jumping discontinuously as the correct answer (say, to the question which of two tones is louder) would; and the "psychometric functions" that are plotted showing the experimentally determined relationship[1] are similar in form to plots showing how probabilities of choice among lotteries, say, vary with variation in the promised payoffs.[2]

Moreover, in economic choices just as in the case of perceptual discrimination tasks, there is observed to be a systematic relation between the *time* required for experimental subjects to make a decision and the characteristics of the alternatives; specifically, the average response time is shorter in the case of "easier" choices, which are also the ones for which repeated trials yield the same response a greater fraction of the time.[3] Clithero and Rangel (2013) and Krajbich *et al.* (2013) both find that the relationship between response time and the frequencies with which different choices are made is sufficiently systematic that data on a given subject's response times

[1] See, e.g., Green and Swets (1966) or Gabbiani and Cox (2010), chap. 25.

[2] See, e.g., Figure 2 from Mosteller and Nogee (1951).

[3] See Figure 1 below, for an example of choice between alternative consumption goods.

when choosing between particular pairs on goods on past information improves one's ability to predict the subject's future choices (even if response times themselves are not of intrinsic interest). The similarity in the association between response time and the degree of predictability of the subject's eventual choice that is observed both in economic choices and in perceptual tasks suggests the underlying cognitive processes involved in both kinds of judgments may be similar.

The present study takes this analogy seriously, by proposing a model of stochastic choice between discrete alternatives in which the randomness of choice is explained in the same way as randomness in perceptual judgments (in the literature discussed further below) — that is, as resulting from decisions being based on random evidence about the nature of the decision situation on a given occasion, rather than a precise description of it. In perceptual contexts, the randomness of the available evidence is assumed to result from randomness in the processing of the evidence received by the senses. In the case of stochastic choice, the randomness does not generally result from randomness in perceptual processing (for example, inaccuracy in visual estimation of the size of a piece of cake that one is offered). But the value of the available options in a choice situation depends not only on characteristics that can be directly observed or that are described to the decisionmaker (the type of cake offered), but also on associations that will have to be accessed from memory (the degree to which one enjoys that type of cake); and one may suppose that access to these contents of memory at the time of choice is subject to corruption by noise in a manner analogous to the noise in observations of the world through the senses.

The content of such a theory depends, of course, on the assumed nature of the statistical relationship between the true features of a given choice situation and the random evidence upon which choice must be based. We discipline our assumptions about this in two ways. On the one hand, we assume that decisions are made using a stochastic algorithm, the general structure of which derives from studies of the neural processing underlying certain types of perceptual judgments. At the same time, we assume that the kind of random evidence supplied as an input to this algorithm is optimal from the standpoint of maximizing expected utility, subject to a constraint on the information-processing capacity of the channels that supply the information about the current choice situation on the basis of which the decision is made. The kind of choice algorithm that is optimal in this constrained sense involves random errors, not because they are assumed to be inevitable but because such an algorithm reduces the amount of information about the variation in choice situations that must be obtained; and it makes specific (and empirically testable) predictions about the way in which the rate at which random errors are made should very with the choice situation. We briefly discuss each of these aspects of the motivation for our problem formulation.

1.1 Process Models of Perceptual Judgment and Stochastic Choice

At least since Thurstone (1927), a standard interpretation of the "psychometric function" showing how the probability of a particular perceptual judgment varies continuously with variation in stimulus characteristics has been to assume that the subject's responses are based on sampling from a probability distribution of possible subjective perceptions corresponding to each objective stimulus.[4] A further development of the approach, that seeks to account for measured response times as well as the frequencies with which different responses are given, replaces the hypothesis of a single sample from a distribution of possible perceptions by a model in which the subjective perception is the realization of stochastic process, unfolding over time as the subject observes the stimulus and contemplates the appropriate response.

The best-known of such stochastic-process models of binary perceptual discrimination tasks is the "drift-diffusion model" (DDM), a member of the broader family of "diffusion-to-bound" models.[5] In this model, a continuous-valued subjective state variable, that one may think of as an evolving perception of the weight of sensory evidence in favor of one response relative to the other, follows a random walk with drift on a bounded interval; a decision is made in favor of one or the other of the two possible responses when the bound corresponding to that decision is reached. The instantaneous variance of the diffusion process is assumed to be independent of the stimulus presented, while the drift of the process is assumed to depend on the stimulus (specifically, the degree to which it possesses the property that implies that one response rather than the other is correct). Under a particular (typically linear)

[4] See Green and Swets (1966) for a canonical exposition of this approach.

[5] Early proponents of models of this type include Stone (1960), Laming (1968), Link and Heath (1975), and Ratcliff (1978). Reviews of the mathematical properties of models in this family can be found in Smith (2000), Bogacz *et al.* (2006), and Shadlen *et al.* (2007).

specification of how the drift of the process depends on stimulus characteristics, the model has been found to give a fairly accurate quantitative account of the effects of variation in the stimulus on both response frequencies and reaction times.

Further evidence for the realism of this account of the mental processing involved in perceptual discriminations comes from measurement of electrical activity in parts of the brain involved in recognizing the direction of motion of visual stimuli while monkeys perform a motion-discrimination task. In a fascinating series of experiments by William Newsome, Michael Shadlen, and their collaborators, beginning in the late 1980s, monkeys were trained to indicate the perceived dominant direction of motion in "moving-dot" stimuli. In one of these experiments, the monkey is trained to focus its attention on a screen where several hundred dots are moving; at any point in time, a certain fraction of these are moving in a particular direction with a particular speed, while each of the other dots is moving in a randomly selected direction with a randomly selected speed. The fraction of dots moving coherently define the "motion strength" in the particular direction. The monkey is rewarded (by juice) if it correctly indicates (typically, by one of two possible eye movements) whether the dominant direction of motion of the dots is in one direction (say, to the right) or the opposite direction (i.e., to the left).

The DDM fits well the observed variation both in the frequency with which the monkey indicates motion to the right and in the average time that the monkey takes to respond, as functions of the degree of actual (net) "motion strength" to the right.[6] Moreover, the neurophysiological measurements uncover signatures of processes similar to those postulated by the model.[7] For example, Britten *et al.* (1993) found that the rate of firing of neurons in the middle-temporal area (MT) of the visual cortex, an area already known to be involved in recognizing visual motion, was affected by the motion of the dots; specifically, the average firing rate of "right-preferring" MT neurons increased by more the greater the degree of motion strength to the right. They further measured the random variation in the number of times that these neurons would fire over a given time interval in the case of each type of stimulus, and showed that the randomness of the monkey's responses in the case of varying degrees of motion strength (i.e., the "psychometric function") could be accounted for by the degree of random variation in the difference between the number of times that right-

[6]See Figure 10.1 in Shadlen *et al.* (2007).

[7]See section 10.4 of Shadlen *et al.* (2007) for a more detailed discussion.

preferring and left-preferring neurons would fire (Britten *et al.*, 1992), suggesting that the difference between the firing rates of the two groups of neurons behaves like the random innovations in the diffusion process postulated by the DDM.

Roitman and Shadlen (2002) further found that the rate of firing of direction-sensitive neurons in the lateral intraparietal area (LIP) appeared to integrate the cumulative evidence in favor of one response or the other, as is postulated to occur in the DDM. Unlike the MT neurons, the firing rate of which fairly quickly reaches a steady rate (around which there are only white-noise variations) that depends on the degree of motion strength in the stimulus, the firing rates of the LIP neurons follow path more like random walks with drift, until the decision is made; and as postulated by the DDM, the drifts vary systematically with the motion strength of the stimulus. Random variation in the degree to which the activation level of particular LIP neurons drifts up on a particular trial correlates with whether a particular decision is made on that trial, as shown by the difference in average sample paths for the cases in which the monkey eventually chooses one response or the other. And finally, as also postulated by the DDM, the timing of the monkey's decision appears to be determined by the time when the firing rate of certain LIP neurons reaches a particular threshold, that is independent of the stimulus characteristics and the time taken to reach it. All of these observations suggest that some form of diffusion-to-bound model is implemented by the brain in making perceptual decisions of this kind.

Authors such as Milosavljevic *et al.* (2010) and Krajbich *et al.* (2010) have proposed the DDM as a model of the way in which economic choices are made as well.[8] These authors show that also in this context, the DDM can successfully account for the way in which variation in the relative value of the options presented affects both the frequency with which one is option is chosen over the other, and the average time required to make a choice.[9] And while the neural implementation of such a mechanism would have to be different in the case of choices among economic options

[8]See Fehr and Rangel (2011) for a review of this literature. Busemeyer and Townsend (1993) had earlier proposed a variant form of accumulation-to-bound model to account for experimentally observed choice behavior; see Busemeyer and Johnson (2004) for a survey of the earlier literature, and Webb (2016) for a synthetic treatment of a family of accumulation-to-bound models of stochastic choice that generalizes the DDM.

[9]See Figure 1 below, plotting data from Krajbich *et al.* (2010) for choice between pairs of food items that the subject may eat. The figure is presented in the same format as Figure 10.1 of Shadlen *et al.* (2007), showing the similarity of the phenomena to be explained in the two domains.

— the particular areas of the cortex specialized in the perception of visual motion will obviously not be the ones involved in judgments of the relative value of consumption goods — one might well suppose that the underlying principles that result in this form of neural computation being used by the brain in one area should result in similar computations being used to make comparative judgments of other types as well.

Here we suppose that choices in economic contexts are indeed made as a result of computations similar to those depicted by the DDM. In particular, we take it as given that decisions will be made by moving backward or forward through an ordered sequence of possible internal states, in response to successive increments of evidence about the relative value of the two possible actions, until a sufficiently extreme internal state is reached for one decision or the other to be made. However, rather than also taking as given the way in which the stochastic process for the incremental evidence depends on the true values of the available options assumed by authors such as Milosavljevic *et al.* (2010) and Krajbich *et al.* (2010), we derive the evidence-accumulation process that would satisfy a particular constrained optimality property. In this way, we seek to provide a functional interpretation for the evidence accumulation process, that should be useful in suggesting the way that one might expect the model to generalize to more complex types of decision problems. The resulting dynamics, for the kind of binary choice problem considered here, are similar to those implied by the DDM, but not identical. Moreover, the way in which they differ from those of the DDM seems to allow an improvement in the empirical fit of the model.

1.2 Optimal Evidence Accumulation

Stochastic choice has also been modeled as resulting from a process of accumulation of noisy evidence by Natenzon (2013), Lu (2013), and Fudenberg *et al.* (2015). These authors take as given the kind of noisy evidence that is obtained through further consideration of the available options, but consider optimal choice on the basis of such information. In Natenzon (2013) and Lu (2013), the amount of noisy evidence that is collected is also exogenously given, but the effects of exogenous variation in the length of time allowed for additional observation are analyzed. In Fudenberg *et al.* (2015), the time at which further information ceases to be collected and a decision is made is also chosen optimally, given an assumed cost of further delaying the decision. The problem treated by Fudenberg *et al.* generalizes a classic problem in statistical decision theory (to which seminal contributions were made by Wald, 1947, and Arrow *et al.*, 1949), in which the statistical properties of an available test or experiment is specified and the question is how long to continue testing before making a decision.

The question posed here is different. Rather than asking how many times it is optimal to repeat the one type of test that is assumed to be possible, we ask *which type* of test it is optimal to perform at each stage in the decision process, from among a flexibly specified class of tests.[10] Essentially, we allow any kind of test with a binary outcome to be performed at each stage in the sequential decision process, and any kind of dependence of the test upon the history of outcomes of previous tests, subject only to a constraint on the average informativeness of the tests. Instead it is the rule for deciding when to stop testing that is exogenously specified in our case, rather than being chosen optimally, though we do not fix the length of time for which evidence is collected, as in models like those of Natenzon and Lu, and instead specify an endogenous stopping criterion like the one assumed in the DDM.[11] Our

[10]Che and Mierendorff (2016) similarly seek to endogenize the type of test performed at each stage, but consider a special family of potential tests, in which signals arrive as a Poisson process, with no information other than the passing of time being received over most small intervals, while a large "lump" of information, resulting in an immediate decision, will arrive on some discrete occasion. We instead consider here evidence accumulation processes in which slightly informative signals are received continually, as in psychological models like the DDM or decision field theory.

[11]The stopping criterion assumed in the DDM is known to be optimal (for a constant cost of additional testing, and the particular kind of information from testing assumed in the DDM) in the case of a problem in which only two possible states of the world (decision situations) are *ex ante* possible, as shown by the normative justifications for a sequential probability ratio test derived by Wald (1947) and Arrow *et al.* (1949). However, as discussed in Fudenberg *et al.* (2015), even under these assumptions about the cost of testing and the information from testing, it is not generally optimal if there are more than two possible states of the world, since the posterior after a given history of random evidence will not generally be a function solely of a single statistic. It also need not be optimal for the kind of evidence accumulation process derived here. Here we treat the stopping criterion as a constraint on the class of decision algorithms considered, and leave for future work the question whether this criterion can be justified as optimal, possibly under a constraint on the complexity of the criterion that can be implemented. Hébert and Woodford (2016) offer some preliminary analysis of a generalization of the problem treated here, in which both the information sampled at each stage in the evidence accumulation process and the stopping criterion are endogenized.

inquiry, which endogenizes information choice at each stage while taking as given the stopping criterion, can be viewed as complementary to that of Fudenberg *et al.*, which endogenizes the stopping criterion while taking as given the information available at each stage.

In proposing that that the accuracy of the information that can be obtained through additional evidence accumulation at each stage is limited by an information-theoretic constraint, the model here can viewed as an application of Sims' (2003, 2010) theory of "rational inattention."[12] However, unlike Sims' theory, the model proposed here imposes additional constraints on the class of choice algorithms that are assumed to be feasible, motivated by both behavioral and neurophysiological data.[13]

2 A Information-Constrained Dynamic Model of Discrete Choice

We now present a simple example of a process model of choice that can be derived from an optimization problem, subject to a constraint on the information-processing capacity of the structures through which the DM becomes aware of the nature of the choices faced at a given point in time. We then compare the properties of the optimal choice mechanism according to this theory with the predictions of the DDM, with empirical data, and with some other possible normative theories of stochastic choice.

2.1 Choice Algorithms

Suppose that choice among a discrete set of alternatives occurs using a mechanism with two elements.[14] First, there is a *sensor* that produces a signal about the nature of the current choice situation each time it is employed; and second, there is a *decoder*

[12]Previous discussions of rational inattention as a source of randomness in discrete choices include Woodford (2008, 2009), Pinkovskiy (2009), Cheremukhin *et al.* (2011), Matějka and MacKay (2013), and Caplin and Dean (2013). See further discussion of this theory in section 3.

[13]Webb (2015) also seeks to discipline the specification of a model of stochastic choice on the basis of both behavioral and neurophysiological evidence, but does not seek to provide optimizing foundations for the specifications that he considers.

[14]Note that the optimizing models of Natenzon (2013) and Fudenberg *et al.* (2015), as well as purely descriptive models like the DDM, are examples of decision algorithms with the general structure assumed here.

that receives the stream of signals from the sensor, and indicates the decision to be taken, based only on the signals from the sensor. Suppose furthermore that the signal each time the sensor is activated is a letter s in some finite alphabet S, and let H be the set of all possible finite sequences of letters of this alphabet (the set of possible finite signal histories), including an element $\emptyset$ corresponding to a situation in which no signals have been received. The operation of the sensor can then be specified by a function $\sigma : X \times H \to \Delta(S)$ indicating the probability with which a given signal will be sent by the sensor, as a function of the state of the world $x \in X$ at the time that the sensor is operated, and the history $h \in H$ of signals previously sent.[15]

The decoder decides, after each additional signal from the sensor is received, whether to announce a decision (an action a chosen from among a finite set of possible actions A) or to request another signal from the sensor. The algorithm begins with prior signal history $\emptyset$, and terminates as soon as a signal history h is reached for which the decoder announces a decision; once it terminates, the DM receives a reward specified by a function $U : A \times X \to \mathbb{R}$, that depends on both the action chosen and the state of the world at the time of decision. The operation of the decoder can then be specified by a set $C \subset H$ of "codewords" for which the decoder will announce a decision, and a decision rule $\delta : C \to A$ indicating the decision made when each of these codewords is encountered. Since we assume that the algorithm terminates as soon as a signal history $c \in C$ is encountered, the set C must constitute a "prefix code."[16]

The decision problem, for which we wish to design a choice algorithm, can then be defined by a prior over possible histories of states of the world (assumed to evolve independently of both observations and actions of the DM), the action set A and the reward function U. A specification of the sets S and C and the functions σ and δ completely describes a choice algorithm (that may or may not reach a decision in finite time) for this decision problem. For any specification of a sequence of states of the world associated with a given experimental trial, the algorithm yields a probability

[15]The assumption that the operation of the sensor may depend on the history of previous *signals*, but not the history of previous *states of the world* reflects an assumption that the sensor itself is memoryless, but that it may be influenced by the current state of the decoder unit. The decoder, in turn, has no direct access to the state of the world, and instead observes only the history of signals generated by the sensor.

[16]That is, it has the property that no codeword $c \in C$ is also the prefix (the initial sequence of signals) of some longer codeword in C (Cover and Thomas, 2006, chap. 5).

distribution of possible outcomes, where each outcome (in which a decision is reached) is a specification both of an action $a \in A$ that is chosen and a number of time steps T that are required for the decision.[17]

The DDM model of binary choice is an example of an algorithm of this kind. In this kind of decision problem, it is assumed that the action set A consists of two possible choices, which we may call "left" (L) and "right" (R), and in typical applications the state of the world x (i.e., the specification of the values of the two possible actions, U_L and U_R) is assumed to remain constant over a given trial, until a decision is reached. The prior over possible histories of states of the world then reduces to a simple specification of a prior $\pi \in \Delta(X)$ over possible constant states on a given trial. A discrete version of the DDM (which is more commonly specified as a continuous diffusion,[18] as in section 2.4) assumes that at each step of the algorithm prior to termination (i.e., a decision), the sensor produces one of two possible signals, that we may also denote L and R — the idea being that the signal L is additional evidence in support of the desirability of action L, while the signal R is additional evidence in support of action R.

In the neural implementation of the DDM proposed by Shadlen *et al.* (2007), the momentary evidence that causes the decision variable to drift in one direction or the other corresponds to the difference in firing rates between "right-preferring" and "left-preferring" neurons in region MT, where the firing rate of each type of neuron is a function of the state x (strength of motion to the right, in the experiments discussed by these authors). The neurons produce discrete spikes with random timing (often approximated as a Poisson process), rather than being continuously active at some level, and it is this randomness of the production of spikes that results in the random variation in the momentary evidence associated with a given state in the neural model. If we think of the successive steps in our algorithm as the instants in time at which MT neurons spike, then at each step the signal received will be of one of two types: either another spike from a "right-preferring" neuron (which we shall call the signal R) or

[17]Below, the framework is extended to allow as well for termination of the algorithm due to an exogenous event, "disappearance of the choice opportunity." In this extension of the model, the functions σ and δ continue to describe the functioning of the algorithm, conditional on an exogenous termination not having occurred.

[18]See, however, Shadlen *et al.* (2007) for a discrete-time presentation of the model, that also allows for the possibility that the momentary evidence is a discrete random variable, as assumed here.

another spike from a "left-preferring" neuron (the signal L). Thus our restriction to a binary signal space is not necessarily unrealistic. (In addition, as discussed further below, the Brownian motion with drift assumed in classic expositions of the DDM can be viewed as a continuous-time limit of this binary-signal model.)

At each step, the signal R is produced with some probability $0 < \lambda(x) < 1$ that depends on the (constant) state x, while L is produced with probability $1 - \lambda(x)$. The specific version of the DDM advocated by Fehr and Rangel (2011)[19] as a model of value-based choice assumes not only that these probabilities are independent of the history of previous signals, but that the log odds of the two signals are a linear function of the difference in value between the two options,

$$\ln \frac{\lambda(x)}{1 - \lambda(x)} = \frac{\alpha}{N} \left[U_R(x) - U_L(x)\right], \tag{2.1}$$

where $\alpha > 0$ is a parameter indicating the sensitivity of the momentary evidence to the degree of difference between the two available choices, and $N \geq 1$ is a parameter of the decoder discussed below.[20] This specifies a function σ for the sensor, in the terminology proposed here.

The classic DDM assumes that a decision is made as soon as the net evidence in favor of one alternative over the other reaches a certain threshold. For any possible signal history $h \in H$, let $n(h)$ denote the number of occurrences of the signal R minus the number of occurrences of the signal L. The algorithm is assumed not to decide, and so to continue collecting evidence, in the case of any finite signal history for which $-N < n(h) < N$, where $N \geq 1$ defines the required evidentiary threshold; it halts, and chooses the action R, if $n(h) = N$; and it halts, but chooses the action L, if $n(h) = -N$. This specifies a decision function δ for the decoder, where the codebook C consists of all finite sequences h such that $n(h)$ is equal to either N or $-N$, and such that $-N < n < N$ for all prefixes (truncations) of h.

[19] Fehr and Rangel actually discuss the continuous-time version of this process, presented in section 2.4 below, rather than the discrete version reflected in equation (2.1). They also discuss an extension of the model, proposed by Krajbich *et al.* (2010), in which the rate of drift (corresponding to the log odds here) depends not only on the relative value of the options, but also on the current visual fixation of the subject. We abstract from the latter complication in the analysis here.

[20] Writing α/N rather than a simple positive coefficient in (2.1) has the advantage that the model's predicted choice frequencies, given by equation (3.4) below, are then a function of α, and independent of the value assumed for N. This form also has the advantage that in this case α corresponds directly to a parameter of the continuous-time version of the model.

Alternatively, the operation of the decoder can be described by a finite-state automaton, with states $\{-N, -N+1, \ldots, N-1, N\}$. The automation begins in state 0 (corresponding to $n(\emptyset)$). Whenever the automaton is in state n, for any $-N < n < N$, it collects another signal from the sensor, and if signal s is received the state moves to $n' = T_s(n)$, where the transition function is given by

$$T_R(n) = n + 1, \qquad T_L(n) = n - 1. \tag{2.2}$$

If the automaton reaches either of the states $\{-N, N\}$, it halts, and announces the decision

$$a(-N) = L, \qquad a(N) = R. \tag{2.3}$$

It can be shown that such an algorithm produces a decision in finite time with probability 1, for any values U_L, U_R for the choices.

2.2 The Optimal Sensor Problem

We are interested in evaluating choice algorithms of this kind according to their implications for the DM's expected reward $\mathrm{E}[U]$, where $U(a, x)$ is evaluated at the time of choice[21] and the expectation is over all possible sequences of states of the world (using the prior) and, given a sequence of states of the world, over all possible signal histories (using the conditional probabilities defined by σ). For simplicity, we here consider only binary choice problems in which in each possible state of the world, $U_R = V(x), U_L = -V(x)$, for some real quantity $V(x)$ (that may be either positive or negative) that depends on the state.[22] The prior used to evaluate the expected reward implied by a given choice algorithm is then a prior over possible sequences of values for V.

This prior need not be identified with the rule used by an experimenter to select the decision situations presented to an experimental subject on successive trials; rather, it should reflect the probability with which the DM can expect to encounter different

[21]We may restrict attention to algorithms that imply that a decision is reached with probability 1. In fact, below we mainly emphasize a case in which the opportunity to choose disappears in finite time with probability 1 if the decision is delayed too long, and in this case there is no difficulty defining the DM's reward in the event of an algorithm that never reaches a decision: eventually the opportunity disappears, resulting in a zero reward.

[22]This is without loss of generality, since the optimal algorithm remains the same if the constant $(U_L + U_R)/2$ is subtracted from all rewards.

situations, given the environment to which its choice procedure has been adapted; the particular situations created by the experimenter may be quite atypical, relative to this prior. This is true even in the case where the experimenter tells the subject the probability distribution from which the experimental situations will be drawn. We may suppose that the DM's deliberations make use of two mental systems, as in Cunningham's (2015) refinement of the dual-systems theory of Kahneman (2011): a "system 1" that produces a quick recommendation a about the action to take, on the basis of fine-grained "low-level" information x about the characteristics of the options presented on the current occasion, and a more reflective "system 2" that can also draw inferences from "high-level" information about the way in which the experimental situation has been designed, and that observes the recommendation a of system 1 without having access to all of the information x used to produce it.

The model presented here is intended as a theoretical account of the automatic processing of the choice situation by system 1; the prior for which the algorithm used by system 1 has been optimized derives from some considerable body of previous experience by the organism, but may not be able to take into account "high-level" information about the experimental design, even when the DM's system 2 is aware of it. As long as the high-level information about the experimental design does not give system 2 a reason to change its expectation that the action recommended by system 1 is more likely to be the correct one than the reverse, the two-system DM will continue to take the action a recommended by system 1, even if the information processing by system 1 has been optimized for a prior that does not correspond precisely to system 2's probability beliefs. Note that in this respect the theory proposed here differs from Sims' (2003, 2010) theory of rational inattention. In order to minimize the number of free parameters used to fit the experimental data discussed below, we shall make a fairly simple assumption about the prior, and one under which the situations actually presented in the experiment would not have been judged extremely unlikely ex ante. However, we shall *not* assume that whatever state x is observed at a given point in time is expected to persist indefinitely, regardless of the time that the DM may take to decide between the two currently available options; it is reasonable to assume that the more typical situation is one in which opportunities for action appear, but will disappear again if a choice is not made sufficiently quickly, and that the choice algorithm is optimized for an environment with this characteristic.

Specifically, we shall assume that some measure π over the real line describes the prior probability for the value of V for the options that are initially available when a choice situation is presented and the algorithm is initiated. Then, each time that the algorithm fails to make a choice between the two options, there is a probability $0 < \rho < 1$ that the choice opportunity will continue to be available, in which case the sensor can be used to produce another signal about the relative value of the options, and another opportunity to choose or to defer choice will be presented. But with probability $1 - \rho$, an action is forced and the algorithm necessarily terminates; there is an equal probability of either action being selected (with no input from the decoder) in this case, so that the expected reward is zero.[23]

A given choice algorithm (S, C, σ, δ) and anticipated persistence of opportunities ρ then define probabilities $\{p(h|V)\}$ of reaching each possible finite history h and engaging the sensor, if the initially presented options are such that $U_R = V, U_L = -V$. We can similarly let $p^*(h|V)$ be the probability of reaching any history (of length greater than or equal to 1), whether the algorithm terminates at this history or not. Let $D \subset H$ be the set of histories such that neither h nor any prefix of h belongs to the codebook C. (That is, D is the set of histories for which the algorithm will not yet have halted.) Then these probabilities are defined recursively by the relations

$$p^*(hs|V) = p(h|V)\sigma(s|V, h)$$

for any $h \in D$ and $s \in S$, where hs denotes the signal history obtained by adjoining s to the previous signal history h, and $\sigma(s|h, V)$ is the probability of observing signal s if the sensor is activated when the value of option R is V and the prior signal history is h; and

$$p(\emptyset|V) = 1$$

for all V, while

$$p(h|V) = \rho p^*(h|V)$$

for any $h \in \bar{D}$, the set of histories in D of length greater than or equal to 1. We define $p(h|V) = 0$ for any $h \notin D$ and correspondingly $p^*(hs|V) = 0$ for all of the successors of any $h \notin D$. Given a prior π over the possible initial values of V, the

[23] More generally, the expected reward would be $(U_L + U_R)/2$ in this case.

expected reward from using the algorithm will then be[24]

$$\mathrm{E}[U] = \sum_{V} \pi(V) \sum_{c \in C} p^*(c|V) U(\delta(c), V), \tag{2.4}$$

where we now write $U(a, V)$ for the reward from choosing a if V takes this value and a decision is made before an action is forced.

In the present analysis, we shall furthermore take as given the rule used by the decoder, and consider only the optimal design of the sensor for a given class of potential decision problems. Specifically, we shall assume that the rule used by the decoder is given by (2.2)–(2.3), for some value of N, as in the case of the DDM, and consider the optimal choice of the signalling function σ. Note that while the assumption that the decoder uses the rule (2.2)–(2.3) implies something about the consequences of sending a signal R as opposed to L at any stage of the algorithm, we make no *a priori* assumptions about which signal must be sent under any particular conditions, so that the dynamics of the decision process remain relatively unconstrained. One advantage of focusing on the optimal sensor design problem is that, as we shall see, it is possible to characterize the optimal signalling function σ under only a relatively weak assumption about the prior π over possible relative values V; the optimal design of the decoder will instead generally be sensitive to further details of the prior, since a single choice of the decoder must be made to apply regardless of the value of V on a particular trial.

If we simply consider the problem of choosing a sensor function to maximize $\mathrm{E}[U]$, under no constraints beyond those already mentioned, the solution is simple: the sensor should always indicate the value R in a state in which $V > 0$, and the value L in a state in which $V < 0$. Such a rule implies that the correct (reward-maximizing) choice will always be made, assuming that an action is not forced before a choice can be made (i.e., before N observations of the signal have been completed); and it will maximize the probability of getting to choose before an action is forced (by always generating the shortest possible codeword among those leading to the correct choice). Such an analysis cannot, however, explain the experimental data reported in studies such as Krajbich *et al.* (2010); it implies that, in an experiment where the choice opportunity is maintained long enough for the subject to reach a decision,

[24]Here we assume that the set X is countable. This allows us to write sums rather than integrals, though the formalism is easily extended to deal with the case of a continuum of possible values V.

there should be no random variation in the choice made between any two options (except in the case of exactly identical values), and it implies that the time taken to make the decision should be the same for "hard" choices as for "easy" ones.

According to the hypothesis proposed here, this does not occur because it would require the output of the sensor to be too precisely coordinated with the current state of the world. We shall suppose that there is a limit to the information-processing capacity of the sensor, and model this (as in Sims, 2003, 2010) by a bound on the *mutual information* between the output of the sensor and the state of the world.[25] Following Shannon, we can define the rate of information transmission by the sensor by the average amount by which receipt of a signal s reduces an observer's uncertainty (measured by entropy) about the conditions z that have influenced the signal (Cover and Thomas, 2006, chap. 2). In our case, the operation of the sensor is defined by conditional probabilities $\lambda(z)$ of producing the signal R when the sensor receives inputs $z \equiv (V, h)$. The prior over states of the world together with the specification of the algorithm imply a set of unconditional probabilities $\pi(z)$ of being in a given state z when the sensor is engaged, given by

$$\pi(V, h) = \frac{\pi(V)p(h|V)}{\sum_{\tilde{V}} \sum_{\tilde{h} \in H} \pi(\tilde{V})p(\tilde{h}|\tilde{V})} = \frac{\pi(V)p(h|V)}{\mathrm{E}[T]},$$

where

$$\mathrm{E}[T] \equiv \sum_{V} \sum_{h \in H} \pi(V)p(h|V)$$

is the expected value of T, the (random) number of times the sensor is employed (and hence the number of steps in the algorithm) before a decision is reached.

The ex ante uncertainty about the value of z, for someone who observes only that the sensor has been engaged, is then given by the (unconditional) entropy[26]

$$H \equiv -\sum_{z} \pi(z) \ln \pi(z).$$

Observation of a signal s implies a set of posterior probabilities $\pi(z|s)$ given by

$$\pi(z|R) = \pi(z)\lambda(z)/\bar{\lambda}, \qquad \pi(z|L) = \pi(z)(1 - \lambda(z))/(1 - \bar{\lambda}),$$

[25] See also Wolpert and Leslie (2012) for use of a similar constraint.

[26] The use of the natural logarithm rather than a logarithm of base 2, as in classic expositions of information theory, is simply a change in the units in which the rate of information flow is measured: from "bits" (binary digits) per signal to "nats" per signal.

where

$$\bar{\lambda} = \sum_z \pi(z)\lambda(z) \tag{2.5}$$

is the ex ante probability of receiving signal R when the sensor is engaged. The entropy of the posterior distribution is therefore

$$H(s) \equiv -\sum_z \pi(z|s)\; \ln \pi(z|s),$$

and the mutual information I is then defined as the average entropy reduction per use of the sensor,

$$I \equiv H - [\bar{\lambda}H(R) + (1-\bar{\lambda})H(L)]$$

We assume that the only feasible sensors are ones for which $I \leq \bar{I}$, where the upper bound $\bar{I}$ is assumed to be considerably less than $\ln 2$ nats (i.e., one binary digit) per signal.[27]

Substitution of the above definitions allows us to alternatively write

$$\begin{aligned}
I &= \bar{\lambda}\sum_z \pi(z|R)\; \ln \pi(z|R) + (1-\bar{\lambda})\sum_z \pi(z|L)\; \ln \pi(z|L) - \sum_z \pi(z)\; \ln \pi(z) \\
&= \sum_z \pi(z)\lambda(z)\,[\ln \pi(z) + \ln \lambda(z) - \ln \bar{\lambda}] \\
&\qquad + \sum_z \pi(z)(1-\lambda(z))\,[\ln \pi(z) + \ln(1-\lambda(z)) - \ln(1-\bar{\lambda})] - \sum_z \pi(z)\; \ln \pi(z) \\
&= \sum_z \pi(z)\, D(\lambda(z)||\bar{\lambda})
\end{aligned}$$

for the average rate of information transmission per use of the sensor, where

$$D(\lambda||\bar{\lambda}) \equiv \lambda \ln(\lambda/\bar{\lambda}) + (1-\lambda)\ln(1-\lambda/1-\bar{\lambda})$$

is the relative entropy of a distribution that assigns probability λ to the receipt of signal R relative to one that assigns probability $\bar{\lambda}$ to that outcome.[28] We assume that

[27] The bound $I \leq \ln 2$ is satisfied by any binary signalling mechanism, including the (deterministic) optimal reporting rule characterized above.

[28] For the properties of relative entropy as a measure of the distance of one probability distribution from another, see, e.g., Cover and Thomas (2006), chaps. 2 and 11. This formulation makes it clear that a limit on information flow makes it difficult to arrange for the sensor to produce L and R signals with frequencies that *differ* systematically with the conditions under which the sensor is operating.

the set of feasible choice algorithms is constrained by an upper bound on the average cumulative information flow through the sensor per trial, summing the information transmitted each time the sensor is used. Thus the proposed information constraint requires that the sensor be operated in a way that satisfies the bound $I \cdot \mathrm{E}[T] \leq K$, for some finite value of K, or alternatively,

$$\sum_V \pi(V) \sum_{h \in D} p(h|V)\, D(\lambda(V,h)||\bar{\lambda}) \leq K, \tag{2.6}$$

where $\bar{\lambda}$ is given by (2.5).

Note furthermore that if $\bar{\lambda}$ is treated as a separate design parameter (rather than being defined by (2.5), then for any choice algorithm (implying values for the $\pi(z)$ and $\lambda(z)$), the value of $\bar{\lambda}$ that minimizes the left-hand side of (2.6) is given by (2.5). We can accordingly restate our problem as the choice of signal probabilities $\{\lambda(z)\}$ for each of the states that can possibly be reached, and a constant $\bar{\lambda}$, so as to maximize $\mathrm{E}[U]$ subject to constraint (2.6). Alternatively, we can state the problem as the choice of $\{\lambda(z)\}$ and $\bar{\lambda}$ to maximize a Lagrangian of the form $\mathcal{L} \equiv \mathrm{E}[U] - \theta I \cdot \mathrm{E}[T]$, or

$$\mathcal{L} = \sum_V \pi(V) \left\{ \sum_{c \in C} p^*(c|V) U(\delta(c), V) - \theta \sum_{h \in D} p(h|V)\, D(\lambda(V,h)||\bar{\lambda}) \right\}, \tag{2.7}$$

where θ is a non-negative Lagrange multiplier indicating the value of relaxing the information constraint (2.6).

2.3 Optimal Evidence Accumulation Dynamics

We now characterize the signaling function that solves problem (2.7), for a given value of the shadow cost of processing capacity $\theta > 0$. We note that for any V, the final term in (2.7) can be written in the form

$$\begin{aligned}
\sum_{h \in D} p(h|V)\, D(\lambda(V,h)||\bar{\lambda}) &= \sum_{h \in D} \sum_{s \in S} p^*(hs|V) \ln(p^*(hs|V)/p(h|V)\bar{p}(s) \\
&= \sum_{c \in C} \mathrm{var}\phi(p^*(c|V)) + \left(\frac{1-\rho}{\rho}\right) \sum_{h \in \bar{D}} \varphi(p^*(h|V) - \sum_{h \in D, s \in S} p^*(hs|V) \ln \bar{p}(s),
\end{aligned}$$

where we define $\bar{p}(R) \equiv \bar{\lambda}, \bar{p}(L) \equiv 1 - \bar{\lambda}$, and $\varphi(p) \equiv p \ln p$ for any $0 < p < 1$, with $\varphi(0) \equiv \varphi(1) \equiv 0$. Thus this term is a strictly convex function of the probabilities $\{p^*(h|V)\}$ and a strictly convex function of the probabilities $\{\bar{p}(s)\}$. It follows that for any $\theta > 0$, (2.7) is a strictly concave function of the probabilities $\{p^*(h|V)\}$ (implied by the probabilities $\{\lambda(z)\}$) and a strictly concave function of $\bar{\lambda}$.

If for any probabilities $\{p^*(h|V)\}$ defined over signal histories that are immediate successors of histories $h \in D$, we let $\{p^{*\dagger}(h|V)\}$ be the probabilities that reverse the roles of L and R (that is, $p^{*\dagger}(h|V) \equiv p^*(h^\dagger|V)$, where if $h = LLRL$, $h^\dagger = RRLR$, and so on), then (2.7) can be written in the form

$$\mathcal{L} = \sum_V \pi(V) L(V)$$

where for each value of V, $L(V)$ is a function of the probabilities $p^*(\cdot|V)$ and of $\bar{\lambda}$ such that

$$L(V, p^*(\cdot|V), \bar{\lambda}) = L(-V, p^{*\dagger}(\cdot|-V), 1-\bar{\lambda}).$$

Let us now assume further that the prior π over possible values of V is *symmetric:* that is, the probability of a decision situation in which $U_R - U_L = V$ is the same as the probability of one in which $U_R - U_L = -V$, for any V. (This simply means that the assignment of which of the options will be the "right" option on a given trial is independent of the options to be compared.) It then follows that

$$\mathcal{L}(p^*, \bar{\lambda}) = \mathcal{L}(p^{*\dagger}, 1-\bar{\lambda}),$$

if the Lagrangian is evaluated for an arbitrary signalling function σ (defined for all possible values of V) and for an arbitrary value of $\bar{\lambda}$.

Then for any signalling function giving rise to probabilities $\{p^*(h|V)\}$ (defined for all V and for all $h \in D$), the reversed probabilities $\{p^{*\dagger}(h|-V)\}$ define an alternative feasible signalling function, and the convex combination

$$\bar{p}^*(h|V) \equiv \frac{1}{2} p^*(h|V) + \frac{1}{2} p^{*\dagger}(h|-V)$$

defines yet another feasible signalling function. The strict concavity of $\mathcal{L}$ then implies that

$$\mathcal{L}(\bar{p}^*, 1/2) \geq \frac{1}{2}\mathcal{L}(p^*, \bar{\lambda}) + \frac{1}{2}\mathcal{L}(p^{*\dagger}, 1-\bar{\lambda}) = \mathcal{L}(p^*, \bar{\lambda}),$$

where the inequality is strict unless $p^* = p^{*\dagger}$ and $\bar{\lambda} = 1/2$. From this we can conclude that the optimal value of $\bar{\lambda}$ must be 1/2, and consider the problem of maximizing (2.7) over choices of the $\{\lambda(z)\}$, taking the value $\bar{\lambda} = 1/2$ as given.

It is now possible to consider the problem of choosing the $\{\lambda(V, h)\}$ for a given value of V so as to maximize the term $L(V)$ in the Lagrangian, independently of the signalling function that is to be used in any other states. In particular, the prior π over possible values of V plays no role in this problem. Thus we can derive predictions about the probability of choosing one action over the other and about the probability distribution of reaction times for alternative values of V, without requiring any assumption about the prior distribution over possible values of V, apart from the symmetry assumption just invoked.

Let us consider this problem for a particular value of V. Suppressing the index V, our goal is to choose transition probabilities $\{\lambda(h)\}$ for the finite histories $h \in D$ so as to maximize

$$\sum_{c\in C} p^*(c)U(\delta(c)) - \theta \sum_{h\in D} p(h)\, D(\lambda(h)), \tag{2.8}$$

where we now adopt the shorthand $D(\lambda) \equiv D(\lambda||1/2)$. This can be solved using the method of dynamic programming. For any history $h \in D$, let $W(h)$ denote the maximum achievable value of the continuation objective

$$\sum_{c\in C_h} p^*(c|h)U(\delta(c)) - \theta \sum_{\tilde{h}\in D_h} p(\tilde{h}|h)\, D(\lambda(\tilde{h})),$$

where $C_h \subset C$ is the set of codewords for which h is a prefix, and $D_h \subset D$ is the set consisting of h and the other non-terminating histories for which h is a prefix; $p^*(c|h)$ is the probability of reaching codeword c, conditional on having received the initial signal history h; and $p(\tilde{h}|h)$ is the probability of reaching $\tilde{h}$ and activating the sensor, conditional on the initial history h. In this definition, the maximization is over possible choices of $\{\lambda(\tilde{h})\}$ for the $\tilde{h} \in D_h$. This value function must satisfy a Bellman equation of the form

$$W(h) = \max_{\lambda(h)} \{\rho[\lambda(h)W(hR) + (1 - \lambda(h))W(hL)] - \theta D(\lambda(h))\},$$

where for histories $c \in C$ we define $W(c) \equiv U(c)$.

One can furthermore observe that the set of continuation plans and the payoffs associated with them are exactly the same in the case of all histories $h \in D$ for which $n(h)$, the number of occurrences of the signal R minus the number of occurrences of the signal L, has the same value n. Thus we can simply write $W(n)$ for the common value of $W(h)$ in the case of all such histories, and write the Bellman equation in the form

$$W(n) = \max_{\lambda(n)} \{\rho[\lambda(n)W(n+1) + (1-\lambda(n))W(n-1)] - \theta D(\lambda(n))\} \tag{2.9}$$

for any $-N < n < N$. The value function (for any value V of the action R) is then given by a sequence of values $\{W(n)\}$ for $-N \leq n \leq N$ that satisfy (2.9) for each $-N < n < N$, together with the boundary conditions

$$W(-N) = -V, \qquad W(N) = V. \tag{2.10}$$

The solution to the optimization problem in (2.9) is easily seen to be given by the $\lambda(n)$ corresponding to the log odds

$$\ln \frac{\lambda(n)}{1-\lambda(n)} = \frac{\rho}{\theta}[W(n+1) - W(n-1)]. \tag{2.11}$$

One observes that for any finite values $\{W(n)\}$, the solution to (2.11) will imply that $0 < \lambda(n) < 1$ for all $-N < n < N$; thus regardless of the value of V, and regardless of the previous signal history, there will at each point in time be positive probabilities of receiving either L or R as the next signal, so that the value of n will fluctuate randomly, drifting up over some time intervals and down over others, as in the DDM. Note that this is not an assumption, but rather a conclusion about the kind of process that economizes on information: for our model assumes that it is feasible, at finite cost, to arrange for the transition probability $\lambda(h)$ to equal either 0 or 1, though it turns out never to be efficient for the decision process to evolve so predictably.

Substituting the solution (2.11) into (2.9) we obtain

$$W(n) = \theta \ln \left[\frac{1}{2} \exp\left(\frac{\rho}{\theta} W(n-1)\right) + \frac{1}{2} \exp\left(\frac{\rho}{\theta} W(n+1)\right)\right] \tag{2.12}$$

for each $-N < n < N$. The value function can then be characterized as the solution to this nonlinear difference equation consistent with boundary conditions (2.10).

As with the DDM, the present model implies that the bias of this process toward one of the absorbing barriers or the other depends on the sign and magnitude of V (i.e., of the difference in value of the two options). Condition (2.11) implies that the log odds are an increasing function of the gradient of the value function, and the boundary conditions (2.10) imply that the average value of this gradient must be proportional to V. However, this model implies that the log odds of obtaining a signal R as opposed to L will also depend on the current balance of the accumulated evidence (i.e., the value of n at a given point in the process), and not only on the value of V. As with the DDM, the decision dynamics can be described by a Markov chain on the set of subjective states $\{-N, -N+1, \ldots, N-1, N\}$, but the transition law is no longer precisely a random walk with drift.

In the special case $\rho = 1$ (so that opportunities are expected to persist as long as may be necessary for a choice to be made), the difference equation (2.12) has a closed-form solution,

$$W(n) = \theta[a + \ln(n+b)], \qquad \text{where} \quad a = \ln \frac{e^{V/\theta} - e^{-V/\theta}}{2N}, \quad \frac{b}{N} = \frac{e^{V/\theta} + e^{-V/\theta}}{e^{V/\theta} - e^{-V/\theta}} > 1,$$

if $V > 0$, or

$$W(n) = \theta[a + \ln(-b-n)], \qquad \text{where} \quad a = \ln \frac{e^{-V/\theta} - e^{V/\theta}}{2N}, \quad \frac{b}{N} = \frac{e^{V/\theta} + e^{-V/\theta}}{e^{V/\theta} - e^{-V/\theta}} < -1,$$

if $V < 0$. (In the case that $V = 0$, the solution is simply $W(n) = 0$ for all n.) It then follows from (2.11) that

$$\lambda(n) = \frac{1}{2} + \frac{1}{2(n+b)}. \tag{2.13}$$

It follows that for any value of n, $\lambda(n)$ is a monotonically increasing function of V/θ (approaching 0 as $V/\theta \to -\infty$ and approaching 1 as $V/\theta \to +\infty$). However, the probability of receiving a signal R (and hence the average rate of drift of the accumulated evidence) depends on the current value of n: for any $V \neq 0$, $\lambda(n)$ is a monotonically decreasing function of n.

Under the more realistic assumption that $\rho < 1$, a closed-form solution is unavailable, but the difference equation (2.12) can be solved numerically for each possible specification of the boundary conditions. One finds in general that λ is a function of the values of V/θ and n.

2.4 The Continuous-Time Limit

Suppose that we let the value of N become large, but shrink the length of time Δ required for each additional use of the sensor in proportion to N^{-2}, so that $N^2\Delta$ is held constant as N is increased. Then if we let $\tau \equiv t\Delta$ be the amount of clock time that has passed after t uses of the sensor, the model approximates one in which the rescaled state variable $z \equiv \Delta^{1/2} n$ evolves with a continuous sample path as τ

increases, The range over which this variable varies will be the interval $[-B, B]$, where the decision threshold $B \equiv (N^2\Delta)^{1/2}$ is independent of N.

Furthermore, over any time interval that is sufficiently small for z (and hence the probability $\lambda(z)$ of receiving a signal R on each use of the sensor) to change little over the interval, the cumulative change in the value of z over the interval is approximately the sum of a large number of independent draws of a bounded random variable, so that the distribution approaches a Gaussian as N is made large. If for any $-B < z < B$, the log odds $\ln(\lambda(z)/1-\lambda(z))$ become small at the rate N^{-1} as N is made large (as specified for example in (2.1)), then

$$\mu(z) \equiv \frac{N}{2B} \ln \frac{\lambda(z)}{1-\lambda(z)} \tag{2.14}$$

has a well-defined limit as N is made large, and the mean increment in z over a small time interval will equal $\mu(z)$ times the length of the time interval, neglecting an error of order N^{-1} in the estimate of the drift. Under the same assumption, the variance of the increment in z will equal the length of the time interval,[29] where we again neglect an error of order N^{-1} in this estimate of the variance per unit time. Hence in the limit of large enough N, the process approximates the trajectories of a Wiener process

$$dz_\tau = \mu(z_\tau)d\tau + dW_\tau, \tag{2.15}$$

where W_τ is a standard Brownian motion (with zero drift and a unit instantaneous variance).

The continuous-time version of the DDM is then given by equation (2.15), in which μ is a constant (i.e., a number independent of z) given by

$$\mu = \frac{\alpha}{2B}(U_R - U_L),$$

as a consequence of (2.1) and (2.14). This is the version of the DDM presented, for example, in Fehr and Rangel, 2011. Note that the testable predictions of the model about a given choice situation are entirely functions of two parameters, the value of B (that is the same for all choice situations) and the value of $\alpha(U_R - U_L)$ for the particular options available on this occasion.

[29]The normalization used in the definition of z has been chosen to imply this unit instantaneous variance.

In the case of the optimal information-constrained model (OICM), instead, the drift will depend on the current value of z. Let us suppose that the arrival rate of the event that forces an action is fixed per unit of time (rather than per use of the sensor) as the time required for each use of the sensor shrinks, so that

$$\delta \equiv (-\ln \rho)/\Delta > 0$$

is held constant as N increases. Then (2.11) implies that

$$\mu(z) = \frac{w'(z)}{\theta} \tag{2.16}$$

for all $-B < z < B$, where $w(z) \equiv W(\Delta^{-1/2}z)$ expresses the value function in terms of the rescaled state variable.

In this continuous-time limit, the Bellman equation (2.9) becomes

$$\delta w(z) = \max_{\mu} \left\{ w'(z)\mu + \frac{1}{2}w''(z) - \frac{\theta}{2}\mu^2 \right\}$$

for all $-B < z < B$. The first-order condition for the inner problem is easily seen to be (2.16), and substitution of this for $\mu(z)$ yields the differential equation

$$\delta w(z) = \frac{1}{2\theta}w'(z)^2 + \frac{1}{2}w''(z)$$

as the continuous analog of (2.12) above. If we alternatively define the normalized value function $v(z) \equiv w(z)/\theta$, we can characterize it by the differential equation

$$2\delta v(z) = v'(z)^2 + v''(z) \tag{2.17}$$

for all $-B < z < B$, together with the boundary conditions

$$v(-B) = -\nu, \qquad v(B) = \nu, \tag{2.18}$$

where $\nu \equiv V/\theta$. We observe from (2.16) that

$$\mu(z) = v'(z),$$

so that computation of the normalized value function suffices to determine the drift of the process (2.15).

We observe from the form of the equations (2.17) and (2.18) that the solution for the normalized value function (and hence for the optimal drift $\mu(z)$ at all points in the interval $-B < z < B$) for any given decision problem depends only on the values of the parameters B, δ, and ν. Accordingly, the predictions of the OICM with regard to both the probability of choosing each of the two options and the probability distribution of response times depend only on these three parameters. It follows that in the large-N limit, it is only these parameters that can be identified from observations of choice and reaction time; we should not expect to be able to identify numerical values for N, ρ or Δ. Alternatively (as in the discussion of parameter estimates below), we can arbitrarily fix a large value of N, and use the data to identify the implied values of ρ, Δ, and θ; but it should be understood that the numerical values of ρ and Δ obtained in this way are only meaningful in terms of the values that they imply for δ and B. It should also be noted that arbitrarily fixing an assumed value for N does not imply that one is making an assumption about the distance between the decision thresholds, rather than determining this empirically; for the value of B (which determines the distance between the thresholds, in units of the instantaneous standard deviation of the Brownian motion) is not implied by a given choice of N.

3 Comparison with Experimental Evidence

We turn now to a discussion of the degree to which the OICM succeeds as an explanation of observed behavior.

3.1 Explaining Logistic Choice

The model can be used to predict the probability of choosing each of the two options, as functions of the relative value of the two options to the DM. In using the model to predict the outcomes that should be observed in experimental settings where the experimenter allows the subject to take as much time as desired to decide between the options, we must now consider the probability distribution of outcomes predicted by the choice algorithm derived above, for a given value of V and under the assumption that the opportunity to choose *persists indefinitely*, rather than a decision being forced with probability $1 - \rho$ after each additional use of the sensor.[30]

[30]This does not require that we solve the equations of the previous section under the assumption $\rho = 1$. The value of ρ indicates the degree of persistence that has been typical of the environment to

Suppose that the value of the opportunity presented on a particular occasion is V (that is, that $U_R = V, U_L = -V$), and that the opportunity persists until a choice is made. Under this assumption, the probability that the algorithm will eventually terminate in a choice of R, conditional upon the signal history received thus far, will depend only on the quantity n by which the number of R signals exceeds the number of L signals. Furthermore, this probability $\Lambda(n)$ must satisfy

$$\Lambda(n) = \lambda(n)\Lambda(n+1) + (1-\lambda(n))\Lambda(n-1) \tag{3.1}$$

for all $-N < n < N$, together with the boundary conditions

$$\Lambda(-N) = 0, \qquad \Lambda(N) = 1, \tag{3.2}$$

where $\{\lambda(n)\}$ is the sequence determined in the way discussed above. (Here the dependence of both $\lambda(n)$ and $\Lambda(n)$ on the value of V/θ is not made explicit.)

3.1.1 The Undiscounted Case

In the special case $\rho = 1$, the sequence $\{\lambda(n)\}$ is given by (2.13), and (3.1) has a closed-form solution,

$$\Lambda(n) = \frac{b+N}{2N}\frac{n+N}{n+b},$$

so that the initial probability of an eventual choice of R is equal to

$$\Lambda(0) = \frac{b+N}{2b} = \frac{e^{2V/\theta}}{1+e^{2V/\theta}}.$$

Thus in this case the model predicts a logistic relation between the difference in value of the two options and the frequency with which they will each be chosen,

$$\mathrm{Prob}(R) = \frac{e^{(U_R-U_L)/\theta}}{1+e^{(U_R-U_L)/\theta}}. \tag{3.3}$$

This kind of logistic relationship is very commonly fit to data on binary choices (both from laboratory experiments and the field); if the difference in the values of the two alternatives is assumed to be a linear function of some vector of measured

which the subject's cognitive system has adapted, but this may differ from the degree of persistence for the particular set of experiments for which we now wish to derive a probability distribution of predicted outcomes.

characteristics, one obtains the familiar logistic regression model.[31] The standard interpretation given to this statistical specification is in terms of a random-utility model (RUM), in which R is chosen over L if and only if $U_R + \epsilon_R > U_L + \epsilon_L$, where ϵ_L, ϵ_R are two independent draws from an extreme value distribution of type I (McFadden, 1974). But while the RUM provides a possible justification for the econometric practice, there is no obvious reason to expect that the additive random terms in people's valuations (even supposing that they vary randomly from each occasion of choice to the next) should be drawn from this particular type of distribution, so that there is little reason to expect logistic regressions to be correctly specified, under this interpretation.

3.1.2 Comparison with the DDM

The DDM provides an alternative interpretation of logistic choice. Substitution of (2.1) into (3.1) yields a difference equation with the solution

$$\Lambda(n) = \frac{e^{\alpha(U_R - U_L)} - e^{-\alpha(U_R - U_L)n/N}}{e^{\alpha(U_R - U_L)} - e^{-\alpha(U_R - U_L)}},$$

so that

$$\Lambda(0) = \frac{e^{\alpha(U_R - U_L)}}{1 + e^{\alpha(U_R - U_L)}}, \tag{3.4}$$

again a logistic function of $U_R - U_L$.[32] This makes the logistic outcome depend on assumptions that seem less arbitrary; but it still depends on the particular functional form (2.1) for the relation between the relative value of the two options and the drift of the diffusion process, which is not motivated by any considerations deeper than analytical convenience.

The OICM instead provides an explanation for the empirical fit of a logistic relationship (or something close to it) that does not depend on the *a priori* assumption of either a special probability distribution or any special functional forms (apart from the information-theoretic measure of the information required by different sensors). The exact logistic form (3.3) is predicted only in the case that $\rho = 1$; however, even in

[31] For a variety of applications, see Cramer (2003).

[32] Thus we observe that the parameter θ of the OICM has the same consequences for the predicted relationship between relative value and choice probabilities as the parameter α^{-1} in the DDM. Note however that the decision dynamics implied by the two models are not equivalent under this identification: for example, the predicted values of $\Lambda(n)$ are not the same, except when $n = 0$.

the case of a value of ρ slightly less than 1 (argued below to be the empirically realistic case), the predicted relationship is quite similar to a logistic curve, as illustrated by Figure 1(a) below.

Comparison of (3.4) with (3.3) shows that the parameter θ of the OICM has the same consequences for the predicted relationship between relative value and choice probabilities as the parameter α^{-1} in the DDM. Nonetheless, the decision dynamics implied by the two models are not equivalent under this identification; for example, the predicted values of $\Lambda(n)$ are not the same, except when $n = 0$. The differences between the implications of the two models are particularly evident when we consider what they imply about the time that should be required to make a decision.

3.2 Explaining Variation in Response Times

A key argument in favor of the DDM (or related evidence-accumulation models of decisionmaking) is the fact that it can explain not only the way choice probability varies with the alternatives presented, but also the way that the average time required for subjects to announce their decision varies with these alternatives. We now consider the predictions of the OICM in this regard as well.

At any stage of the decision procedure, the expected number of additional operations of the sensor that will be required in order for a decision to be made will depend only on the current value of n. Let this expected value be denoted $T(n)$, where we suppress the dependence on the value of V. It is easy to see that the quantity $T(n)$ must satisfy a recursion of the form

$$T(n) = \lambda(n)T(n+1) + (1-\lambda(n))T(n-1) \tag{3.5}$$

for all $-N < n < N$, together with the boundary conditions

$$T(-N) \;=\; T(N) \;=\; 0, \tag{3.6}$$

where $\{\lambda(n)\}$ is the sequence indicating the probability of a signal R (conditional on the given value of V) conditional on being in decoder state n. This recursive formulation implies in the case of both the DDM and the OICM; only the sequences $\{\lambda(n)\}$ are different for the two models, as explained above.

In the case of the DDM, $\lambda(n) = \lambda$ for all n, where the value of λ depends on V through (2.1). In this case, (3.5) becomes a difference equation with a closed-form solution,

$$T(n) = \frac{(N+n)(1-\lambda)^{2N} + (N-n)\lambda^{2N} - 2N(1-\lambda)^{N+n}\lambda^{N-n}}{[\lambda-(1-\lambda)][\lambda^{2N}-(1-\lambda)^{2N}]}.$$

The expected number of uses of the sensor required for a decision, when the problem is initially presented (and thus before any evidence has been accumulated), is thus given by

$$T(0) = \frac{N\,[\lambda^N-(1-\lambda)^N]}{[\lambda-(1-\lambda)]\,[\lambda^N+(1-\lambda)^N]}. \tag{3.7}$$

One observes that $T(0)$ is an increasing function of λ for all $0 < \lambda < 1/2$, and decreasing for all $1/2 < \lambda < 1$. (The function is also symmetric, in the sense that it takes the same value for λ and for $1-\lambda$, for any λ. Since (2.1) implies that λ is an increasing function of V, with $\lambda(-V) = 1-\lambda(V)$ for all V and $\lambda = 1/2$ when $V = 0$, it follows that $T(0)$ is a monotonically decreasing function of $|V|$, so that the mean response time is shorter the greater the difference in value of the two alternatives. This prediction agrees with what is observed experimentally, as illustrated by panel (b) of Figure 1 below.

3.2.1 The OICM in the Undiscounted Case

In the special case $\rho = 1$, the sequence $\{\lambda(n)\}$ for the OICM is instead given by (2.13). Substituting this into (3.5), we obtain a difference equation that has a closed-form solution,

$$T(n) = \frac{b+\frac{n}{3}}{b+n}\,(N^2-n^2). \tag{3.8}$$

The expected number of uses of the sensor required for a decision when the problem is initially presented is then predicted to be

$$T(0) = N^2.$$

The way in which this scales with the distance of the thresholds $\pm N$ required for decision is the same as in the DDM.[33] However, the result is independent of the value of b, and hence of the value of V for a given decision problem. Thus this model predicts (contrary to the experimental evidence shown in Figure 1(b) below) that the

[33]Note that (3.7) implies that the mean response time when $V = 0$ is also equal to N^2 in the DDM, though decisions are faster if $V \neq 0$.

mean decision time should be the same for all values of V. The undiscounted version of the OICM is much less successful than the DDM in explaining this aspect of the experimental data.

Allowing for values $\rho < 1$ makes a considerable difference for this prediction of the model, though, while affecting the predictions about how choice probabilities vary with the relative value of the two options to a much smaller extent. As analytical results are unavailable in the more general case, we provide a numerical illustration of the effects of allowing $\rho < 1$. We consider the degree to which the model can fit a particular set of experimental data, from one of the leading studies finding support for the DDM as an account of the mental processing underlying observed choice behavior, namely that of Krajbich *et al.* (2010).

3.2.2 Relative Value and Response Times in Krajbich *et al.* (2010)

Krajbich *et al.* ask subjects to rate how much they would like to eat each of 70 possible food items at the end of the experiment, on an integer scale between -10 and 10. (These rankings provide a measure of the value of each of the possible outcomes to the individual subject, independent of any comparison between the goods and any particular alternative.) In a second stage of the experiment, a subject is then asked to choose between pairs of food items that have previously been ranked between 0 and 10; the experimenters record both the subject's choice in the case of each pair and the time taken to decide. After making 100 such decisions, the subject is allowed to actually eat the item chosen in one of the binary choices, selected at random.

Figure 1 plots certain summary statistics of the choices of 39 subjects, each of whom made 100 binary choices. In each panel, trials are grouped according to the extent to which the right option was ranked higher than the left option by that particular subject; thus a "relative value of R" equal to 3 might mean that the right option was ranked 8 by that subject while the left option was ranked 5, or that the right option was ranked 3 while the left option was ranked 0. Panel (a) shows the fraction of trials on which the option R was chosen, as a function of the relative rank of option R; panel (b) shows the average time taken to decide (in milliseconds), again as a function of the relative value. (In each panel, the height of the bar indicates the value of the statistic for a given bin, while the width of the bar is proportional to the

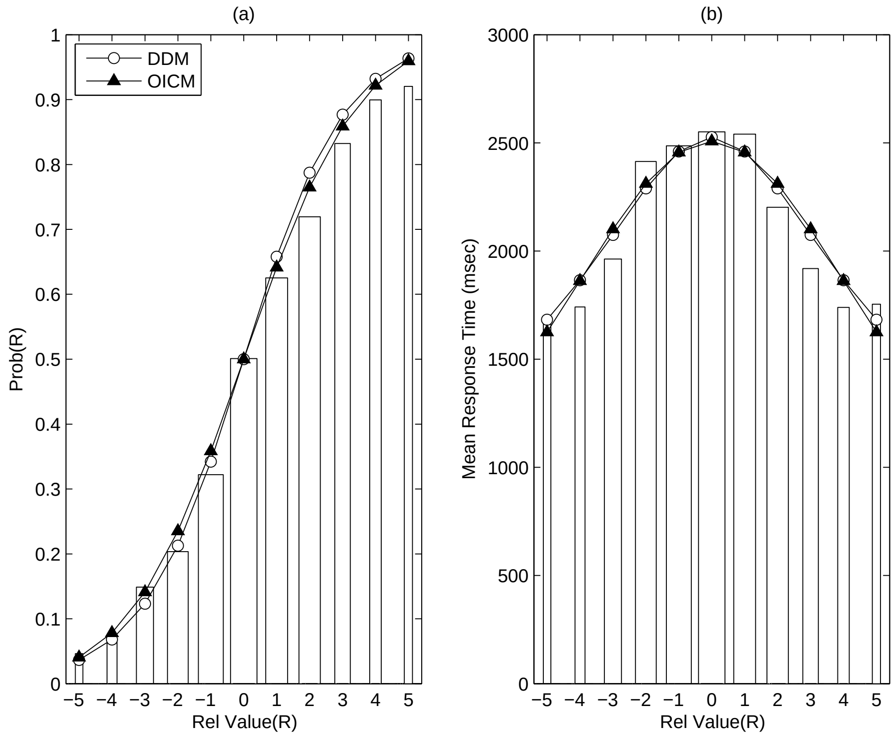

Figure 1: Experimental data and fitted theoretical predictions for (a) probability of choosing R and (b) mean time required to make a decision, each plotted as a function of the amount by which the subject's reported valuation of option R exceeds that of option L. Bars plot the experimental data of Krajbich *et al.* (2010), circles the predictions of the DDM, and triangles the predictions of the present model.

number of trials assigned to that bin.[34])

In the same figure, the open circles indicate the predictions of the DDM for these two statistics, if the relative reward $U_R - U_L$ is assumed to be proportional to the subject's reported relative ranking of the two options, with the same constant of proportionality for all subjects, and if the other parameters of the model are assumed to be identical for all subjects as well. The assumption that $U_R - U_L$ is proportional to the relative rank amounts to an assumption that each subject's utility from consuming

[34]There are also a small number of trials for which the relative rank has an absolute value greater than five. But because the number of trials in these bins are small, and the estimated statistics are correspondingly inaccurate, these bins have not been shown in the figures, though all data are used in the model evaluation exercise below. The truncation of the figures follows Krajbich *et al.* (2010).

the various food items is a roughly linear function of the rank assigned to them on the scale from -10 to 10. Let the slope of this relationship be called the subject's "marginal utility of rank," *mur*. Since the model predictions depend only on the ratio $(U_R - U_L)/\theta$ for each pair of goods (or $\alpha(U_R - U_L)$ in the notation used above for the DDM), a crucial additional assumption of this test of the model is that the ratio mur/θ (or $\alpha \cdot mur$) is the same (or sufficiently similar) for all subjects.[35]

In the case of the DDM, it follows from our discussion above of the large-N limit that there are two parameters that should be identifiable from data on choice and response times, the parameter α introduced in equation (2.1) — or more precisely, the value of αmur[36] — and the parameter $B \equiv \Delta^{1/2} N$ introduced in section 2.4. Alternatively, since N is not identified (any sufficiently large value should lead to an equivalent value), we may arbitrarily fix a value for N (set equal to 100 in the numerical work reported here[37]) and then identify the values of α and Δ that best fit the data. In the fit shown in Figure 1(b), we also introduce a third free parameter, a constant intercept A in the prediction

$$\text{Response Time} = A + T\Delta$$

for measured response time, where T is the random number of times that the sensor is employed in making the decision. The allowance for an intercept reflects an assumption that the observed response time may include an additional time interval (required to engage the choice mechanism and/or to communicate the choice once it has been made) that is independent of the choice options presented on a given trial, in addition to the time required to execute the choice algorithm modeled above.[38]) In

[35]In the case of the OICM, it is a prediction of the model (assuming that each subject's classification system has been adapted to the same frequency distribution of choice situations, and that each subject's system is subject to the same information constraint $\bar{I}$ in nats per millisecond) that the Lagrange multiplier θ for each subject will be proportional to that subject's *mur*. In the case of the DDM, there is no underlying theory of the determinants of the parameter α, so that it would be a coincidence for this to be true, unless we assume that both *mur* and α are the same across subjects.

[36]We call this parameter α, measuring utility in units of the numerical rank supplied by the subjects, so that *mur* is assumed to equal 1. This is without loss of generality, since only $\alpha \cdot mur$ is identified.

[37]We verify numerically that choosing $N = 50$ or 200 instead makes only a negligible difference for the reported results.

[38]This is standard in empirical tests of the DDM; see, for example, the curve-fitting exercise shown in Figure 1 of Shadlen *et al.* (2007), in which a 3-parameter version of the DDM is fit to two curves

the figure, these parameters have been chosen to minimize a weighted sum of squared prediction errors.[39]

The solid triangles instead indicate the predictions of the OICM, if the parameters of the model are chosen to minimize the same weighted-least-squares criterion. In this case, there are again three parameters that are adjusted to fit the data. It follows from our discussion above that in the large-N limit, the only identified parameters of this model are θ (actually, θ/mur), δ and B. Alternatively, we can fix an arbitrary large value for N (again, $N = 100$ in the results reported here), and then determine the best-fitting values for θ, ρ and Δ.[40] One observes that with the same (fairly small) number of free parameters, the OICM does a reasonably similar job of accounting for these statistics; in panel (a), the weighted sum of squared errors is slightly smaller for the OICM than for the DDM, though the errors are somewhat larger for the OICM than for the DDM in panel (b).

The predictions of the two models are not identical, and one of the more noteworthy differences is shown in panel (a) of Figure 2. Here the mean response time is plotted not as a function of the relative value assigned to the R option, but rather as a function of the relative value of whichever option was actually chosen by the subject in that particular trial. The predictions of the DDM are the same in Figure 2(a) as in Figure 1(b): the model predicts that in the case of two options that differ in value by a given amount, the mean response time is *the same* in those cases in which the lower-valued option is chosen as in those in which the higher-valued option is chosen. The data, instead, show that "correct" choices (cases in which the higher-valued option is chosen) are on average made more quickly than "incorrect" choices, for a given degree of difficulty of the choice (based on the absolute difference in rankings). This is in fact a common feature of experimental data on response times in binary perceptual classification tasks as well, and is a famous empirical failing of the DDM

of the kind shown in panels (a) and (b) of Figure 1.

[39]See the appendix for a precise statement of the criterion that is minimized, and the best-fitting parameter values.

[40]No intercept A is allowed in the linear relationship between T and the measured response time in this case. In fact, allowing for a positive intercept would not reduce the sum of squared prediction errors in this case; allowing for a negative intercept would, but would have no interpretation in terms of the theoretical model. The best-fitting values of these parameters are again indicated in the appendix.

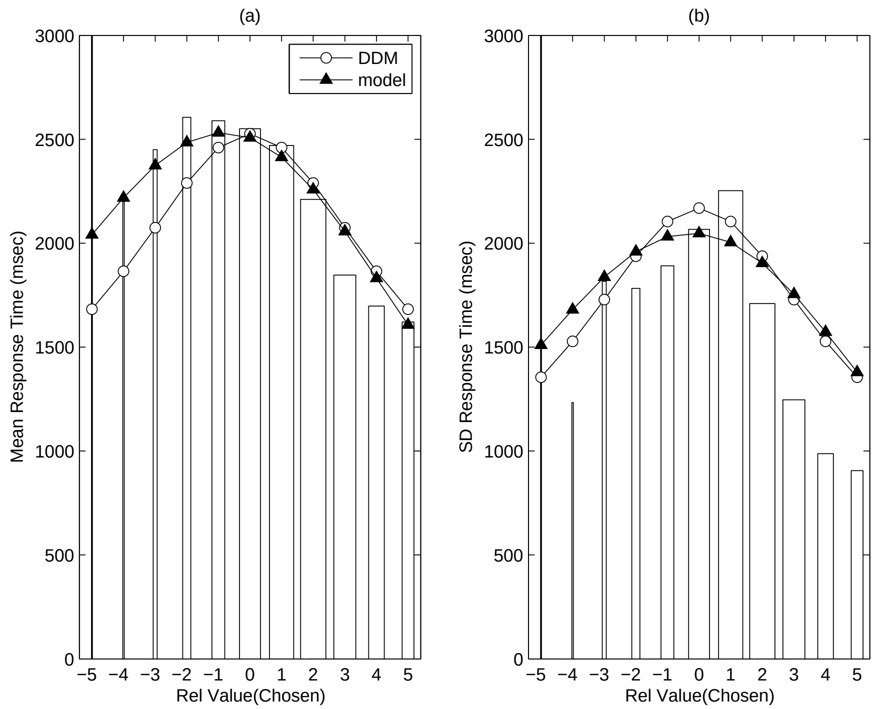

Figure 2: Empirical data and fitted theoretical predictions for (a) mean response time and (b) standard deviation of response times across trials, each now plotted as a function of the amount by which the subject's reported valuation of the option chosen exceeds that of the option not chosen. Experimental data are again from Krajbich *et al.* (2010).

(Luce, 1986).[41] The OICM instead correctly predicts that correct choices should be

[41] According to Shadlen and Kiani (2013, Box 2), this "apparent refutation" led many mathematical psychologists to abandon the DDM as an explanation of stochastic choice and reaction times, until the more recent discovery of neurophysiological evidence for a mechanism of a similar form. A number of variations on the basic DDM presented here have been proposed in the literature that can allow for shorter average response time in the case of correct responses (see, e.g., Link and Heath, 1975; Ratcliff and Rouder, 1998; or Ditterich, 2006), though often no motivation is given for these more general (purely descriptive) hypotheses. Drugowitsch *et al.* (2012) and Fudenberg *et al.* (2015) show that "correct" choices are also made more quickly on average under optimal Bayesian decisionmaking, when the random evidence at each stage is the kind assumed in the DDM but the stopping criterion is optimal, if a large number of different relative values for the two options are possible *ex ante.* The latter models can also be viewed as generalizations of the classic DDM, as they differ only in proposing a more sophisticated stopping criterion.

made more quickly, and by roughly the amount by which the mean response times are different in the data.[42] This prediction results from the way in which the probability of receiving an R signal varies with the net accumulated evidence in favor of the R choice in the OICM; a departure of this sort from the assumption of a constant-drift diffusion process thus has some empirical plausibility.

4 Comparison with the Basic Rational Inattention Model

The explanation given by the OICM for the stochasticity of discrete choice, and even for logistic choice (or at least a nearly-logistic functional relationship) could also be obtained more simply from the theory of "rational inattention" (RI) as formulated by Sims (2003, 2010), without any need to discuss the specific class of dynamic decision processes proposed here.[43] Sims postulates that decisions must be based, as in the model proposed here, on a signal s that is a stochastic function of the true state x (that determines the reward $U(a, x)$ from alternative possible actions a); but he supposes that the set S of possible signals, the function $\sigma : X \to \Delta(S)$ describing the conditional probability of different signals being received in any given state, and the decision rule $\delta : S \to A$ are all chosen so as to maximize $\mathrm{E}[U]$, subject only to an upper bound on the mutual information between x and s.[44]

One can easily show that the maximum achievable value of $\mathrm{E}[U]$ is attained in the case of a signal that simply indicates which action to take. In this case, the information structure is specified by a function $\lambda : X \to [0, 1]$ indicating the probability of

[42] The model prediction differs substantially from the same mean response time in the data only in the case when the relative value of the chosen option equals -5. There are, however, only a few trials in this bin (as shown by the narrowness of the bar), since subjects seldom chose the less-valued option in the case of such an extreme difference, and the response times were quite variable across the trials in this bin (as shown in panel (b)), so that the mean response time for this case should be assigned a wide confidence interval.

[43] See Woodford (2008), Cheremukhin *et al.* (2011), and Matějka and McKay (2015) for previous discussions of this implication of RI. Ortega and Braun (2013) also derive logistic choice from a theory that is essentially equivalent to RI.

[44] Here we use the notation s to refer to the complete set of signals on the basis of which the decision must be made, rather than to the signal obtained from a single use of the "sensor," as in the exposition above.

choosing R in any state, and the Lagrangian for the optimal information structure problem is

$$\mathcal{L} = \sum_x \pi(x) \left\{ (1 - \lambda(x))U_L(x) + \lambda(x)U_R(x) - \theta D(\lambda(x)||\bar{\lambda}) \right\},$$

where $\bar{\lambda} \equiv \sum_x \pi(x)\lambda(x)$ is the average frequency of occurrence of the signal that results in the choice of R. This in turn is maximized by a function $\lambda(x)$ given by the right-hand side of (3.3).

This might seem a simpler explanation for the empirical success of the logistic specification. However, it fails to provide an explanation of the data on response times, to which we turn next. One might think that RI simply makes no prediction about response times, and so is neither confirmed nor disconfirmed by such evidence, but this is not true. If an algorithm obtains no information about the state x other than a coarse signal s, then however the algorithm operates, then joint distribution of *both* action choices a and stopping times T must depend only on the information revealed by s. Hence to the extent that response time is informative about the decision problem faced on a given occasion, the algorithm that produces the decision in that period of time must have access to (at least) the information about the state that can be inferred from the time taken for the algorithm to stop.

If we treat the pair (a, T) as the output of the choice algorithm, and ask what stochastic algorithm will maximize $\mathrm{E}[U]$ subject to an upper bound on the mutual information between the vector (a, T) and the state x, under the assumption that the reward depends only on the initial state x and the action chosen, we obtain a simple prediction: (i) the probability of choosing option R in any state x should be given by (3.3), for some Lagrange multiplier $\theta \geq 0$ on the information constraint; and (ii) conditional on the action a that is chosen, the response time T should convey no additional information about the state x. This requires that the probability distribution of response times on those trials in which R is chosen should be independent of x, and similarly for the distribution of response times when L is chosen. Hence the average response time would have to be an affine transformation of $\lambda(x)$,

$$\mathrm{E}[T](x) = (1 - \lambda(x))T_L + \lambda(x)T_R,$$

where T_a is the mean response time conditional on action a.[45] But this is clearly

[45]The RI hypothesis would provide no reason for T_L to differ from T_R, but it would not preclude such an asymmetry, either.

not consistent with a comparison between the two panels of Figure 1. If we add the additional postulate (roughly consistent with the data shown in Figure 1(b)) that the average response time for a given choice is independent of which option is presented as the left option as opposed to the right option, then T_L would have to equal T_R, and RI would require average response time to be completely independent of x.

Even without this last assumption, if the frequency distribution of value differences $U_R - U_L$ is symmetric around zero, the bins in Figure 2 should each contain as many cases in which R is chosen as cases in which L is chosen, and so the distribution of response times for trials in each bin should be the same (an equally weighted mixture of the distribution for trials in which R is chosen and the distribution for trials in which L is chosen). It follows that in Figure 2(a), the mean response time should be the same for each relative value. But again, this is clearly not the case.

These predictions of RI are obtained under the assumption that response time has no consequences for reward. If instead we assume, as in our derivation of the OICM, that the choice algorithm is optimized for a prior under which choice may be forced with a certain probability if the decision is delayed, delay influences expected reward. In this case, however, in the absence of any constraints other than the information constraint, it would be optimal to always decide immediately (or in the minimum feasible time). Then average response time should be independent of the values of both of the options presented; but this is again inconsistent with both Figures 1(b) and 2(a). Hence an additional constraint on possible algorithms, beyond the information constraint alone, is necessary in order to account simultaneously for choice behavior and response times. The OICM represents a relatively simple example of such an additional constraint.

This does not mean, however, that our (still very simple) model successfully explains all aspects of observed response times with only three parameters. For example, as shown in panel (b) of Figure 2, when the parameters of the model are chosen to fit the data on choice frequencies and mean response times, it generally over-predicts the variability of response times. On the other hand, this is also a problem for the classic DDM as well.[46] Quite possibly it results from the fact that the version of

[46]Figure 2(b) is even more problematic for the RI theory. As explained above, RI would imply that the probability distribution of response times should be the same for each of the bins in Figure 2, so that along with the mean response time (discussed above), the variance of the response time should be the same in each bin; but this is not at all what Figure 2(b) shows.

the OICM analyzed here, like the DDM, assumes that the choice algorithm takes no account of the passage of time, but only the net accumulation of evidence in favor of one option over the other.

This feature of the OICM results from our adoption of the simple form of "decoder" assumed by the DDM, rather than any intrinsic consequence of the hypothesis of optimal information-constrained classification. Drugowitsch *et al.* (2012) and Fudenberg *et al.* (2015) consider instead algorithms in which the threshold for a decision decreases with the passage of time, making very long response times less likely than in the models analyzed here, and show that this is a property of an optimal stopping criterion if, as in the DDM, one assumes an evidence flow given a Brownian motion with a constant drift that is exogenously specified as a linear function of the value difference $U_R - U_L$. Even in a model with optimal endogenous variation in the nature of the evidence sampled over the course of the decision process, as allowed for here, such time-dependence in the stopping criterion may well be optimal.[47] It will be interesting to consider, in future work, whether a more sophisticated form of decoder might be both more fully in conformity with the hypothesis of economizing on information-processing capacity and more consistent with empirical evidence.

[47] In a model of endogenous information sampling related to the one proposed here, Hébert and Woodford (2016) show that the optimal stopping criterion should depend on the posterior reached after the history of signals received to that point; but this posterior will generally not be a function of a single real-valued summary statistic, such as the cumulative excess of R signals over L signals, so that there is no reason to expect the optimal stopping criterion to be a function of a single summary statistic.

A Appendix: Numerical Parameter Values

Here we provide additional details of the parameters that are used in the numerical predictions shown in Figures 1 and 2. In the case of both the DDM and the OICM, parameters are chosen to minimize a weighted sum of prediction errors,

$$N_{tot}^{-1} \sum_j N_j (a_j - \hat{a}_j)^2 + \frac{1}{2} N_{tot}^{-1} \sum_j N_j (b_j - \hat{b}_j)^2 \frac{1}{2} N_{tot}^{-1} \sum_j \tilde{N}_j (c_j - \hat{c}_j)^2, \tag{A.1}$$

taking into account the model predictions for the quantities plotted in Figures 1(a), 1(b), and 2(a). (In focusing on the model predictions for choice frequencies and mean reaction times, we follow much of the literature on the empirical support for the DDM; see for example the plots in Shadlen *et al.* (2007) and Krajbich *et al.* (2010), focusing on the same statistics as in Figures 1(a) and 1(b) here.)

In this expression, for each integer $10 \leq j \leq 10$, N_j is the number of trials on which the subject's ranking of the R option exceeds the ranking of the L option by j, and $\tilde{N}_j$ is the number of trials on which the ranking of the option that is chosen exceeds the ranking of the other option by j. N_{tot} is the total number of trials ($\sum_j N_j = \sum_j \tilde{N}_j = N_{tot}$). Note that in computing (A.1), we sum over all of the bins $-10 \leq j \leq 10$ for each of the two ways of classifying the data, and not just the bins $-5 \leq j \leq 5$ shown in Figures 1 and 2. (Most of the data fall in the bins $-5 \leq j \leq 5$, however, and given the weights in the goodness-of-fit criterion, these bins are largely responsible for the conclusions about the best-fitting parameter values. This is why only the central bins are shown in the figures.)

The quantities a_j, b_j and c_j are the quantities plotted in Figures 1(a), 1(b) and 2(a), respectively, while the corresponding hatted variables are the predictions of the model for these same quantities. Thus the model parameters are chosen to minimize a weighted sum of squared prediction errors, where the weights applied to the squared prediction errors for each bin are proportional to the number of trials in that bin. The factors $(1, 1/2, 1/2)$ that pre-multiply the three terms are chosen so as to put equal weight on fitting the average frequency of R choices and on fitting average response times; and then, with regard to the goal of fitting the evidence on average response times, putting equal weight on fitting the evidence in Figure 1(b) and fitting the evidence in Figure 2(a).

The quantity a_j is the fraction of the trials on which the relative rank is j in which the subject chose R; it varies over a theoretical range from 0 to 1. The quantities b_j and c_j are instead mean response times in milliseconds, divided by 1600, so that the range of variation in these variables is of a similar order of magnitude to the range of variation in the variable a_j.

In the case of both models, the value $N = 100$ is fixed arbitrarily, as discussed in the text. In the case of the DDM, the parameters α, Δ, and A are then chosen to

Table 1: Parameter values used to fit the experimental data, for each of the two models.

	θ (α^{-1})	ρ	Δ (msec)	A (msec)
DDM	0.76	—	0.19	578
OICM	1.04	0.9996	0.25	0

Table 2: Goodness-of-fit statistics, for each of the two models.

	$\langle(a-\hat{a})^2\rangle$	$\frac{1}{2}\langle(b-\hat{b})^2\rangle$	$\frac{1}{2}\langle(c-\hat{c})^2\rangle$	*Loss*
DDM	0.0019369	0.0033134	0.0092264	0.014477
OICM	0.0015782	0.0048181	0.0059117	0.012308

minimize the criterion (A.1); in the case of the OICM, the parameters θ, ρ, and Δ are chosen to minimize the same criterion, fixing $A = 0$ in this case. The best-fitting parameter values for each model are reported in Table 1.

The minimized values of the criterion (A.1) in each case are reported in the final column of Table 2. Table 2 also reports the values of each of the three terms in (A.1) separately, in the first three columns of the table, evaluated at the parameter values reported in Table 1. (Thus the final column in Table 2 is the sum of the other three columns.) On this criterion, the OICM fits somewhat better overall; but admittedly, such a conclusion is sensitive to the relative weights chosen for the different parts of the goodness-of-fit criterion.

The best-fitting parameterizations of the two models imply fairly similar dynamics. Apart from the similarity of the predictions for the the statistics plotted in the two panels of Figure 1, one observes from Table 1 that the parameter indicating the average sensitivity of the sensor log odds to variations in the value gradient (θ in the case of the OICM, or α^{-1} in the case of the DDM) is of similar magnitude in the two cases. The best-fitting value of Δ is similar in magnitude in both cases as well; this means that the implied value of B (the distance to one of the decision barriers, in units of the instantaneous standard deviation of the diffusion process in the continuous-time limiting model) is similar for the two models.

The modest differences in the best-fitting numerical parameter values do point to some differences in the dynamics implied by the two models, however. The fact that the best-fitting θ for the OICM is larger than α^{-1} for the DDM accounts for the fact that the predicted curve in Figure 1(a) is slightly steeper for the DDM. In fact (as shown by the first column of Table 2), the fit of the OICM to Figure 1(a) is slightly better than that of the DDM. This indicates that in the case of the DDM, there is more tension between the value of the sensitivity parameter α needed to fit the choice frequency data and the value that allows the model to better fit the data on average response times.

The best-fitting value of Δ is also about 25 percent shorter in the case of the DDM than that used to fit the OICM. This means that the DDM is only able to account as well as the OICM for the average response time by assuming a substantial fixed time requirement A (a value that accounts for more than a quarter of the average response time), in addition to the time required for the stochastic algorithm to reach a decision threshold. Thus in the case of the DDM, there is a tension between the value of Δ needed to account for the average response time and the value needed to account for the difference in response times between "easy" and "hard" choices, that is minimized through the introduction of an additional free parameter, A, while this is not needed in the case of the OICM. (On the other hand, the OICM has a free parameter, ρ, with no analog in the case of the DDM, and this parameter is also crucial for allowing the model to match the observed difference in average response times between "easy" and "hard" choices.)

References

[1] Arrow, K.J., D. Blackwell, and M.A. Girshick, "Bayes and Minimax Solutions of Sequential Decision Problems," *Econometrica* 17: 213-244 (1949).

[2] Ballinger, T. Parker, and Nathaniel T. Wilcox, "Decisions, Error and Heterogeneity," *Economic Journal* 107: 1090-1105 (1997).

[3] Bogacz, Rafal, Eric Brown, Jeff Moehlis, Philip Holmes, and Jonathan D. Cohen, "The Physics of Optimal Decision Making: A Formal Analysis of Models of Performance in Two-Alternative Forced-Choice Tasks," *Psychological Review* 113: 700-765 (2006).

[4] Britten, Kenneth H., Michael N. Shadlen, William T. Newsome, and J. Anthony Movshon, "The Analysis of Visual Motion: A Comparison of Neuronal and Psychophysical Performance," *Journal of Neuroscience* 12: 4745-4765 (1992).

[5] Britten, Kenneth H., Michael N. Shadlen, William T. Newsome, and J. Anthony Movshon, "Responses of Neurons in Macaque MT to Stochastic Motion Signals," *Visual Neuroscience* 10: 1157-1169 (1993).

[6] Busemeyer, Jerome R., and J.T. Townsend, "Decision Field Theory: A Dynamic-Cognitive Approach to Decision Making in an Uncertain Environment," *Psychological Review* 100: 432-459 (1993).

[7] Busemeyer, Jerome R., and J.G. Johnson, "Computational Models of Decision-making," in Derek J. Koehler and Nigel Harvey, eds., *Blackwell Handbook of Judgment and Decisionmaking,* Oxford: Blackwell Publishing, 2004, pp. 133-154.

[8] Caplin, Andrew, and Mark Dean, "Revealed Preference, Rational Inattention, and Costly Information Acquisition," *American Economic Review* 105: 2183-2203 (2015).

[9] Che, Yeon-Koo, and Konrad Mierendorff, "Optimal Sequential Decision with Limited Attention," unpublished, Columbia University, May 2016.

[10] Cheremukhin, Anton, Anna Popova, and Antonella Tutino, "Experimental Evidence on Rational Inattention," Federal Reserve Bank of Dallas Research Department Working Paper no. 1112, December 2011.

[11] Clithero, John A., and Antonio Rangel, "Combining Response Times and Choice Data Using a Neuroeconomic Model of the Decision Process Improves Out-of-Sample Predictions," unpublished, California Institute of Technology, November 2013.

[12] Cover, Thomas M., and Joy A. Thomas, *Elements of Information Theory,* New York: Wiley-Interscience, 2d ed., 2006.

[13] Cramer, J.S., *Logit Models from Economics and Other Fields,* Cambridge: Cambridge University Press, 2003.

[14] Cunningham, Tom, "Hierarchical Aggregation of Information and Decision-Making," unpublished, Institute for International Economic Studies, Stockholm University, February 2015.

[15] Ditterich, Jochen, "Stochastic Models of Decisions about Motion Direction: Behavior and Physiology," *Neural Networks* 19: 981-1012 (2006).

[16] Drugowitsch, Jan, Rubén Moreno-Bote, Anne K. Churchland, Michael N. Shadlen, and Alexandre Pouget, "The Cost of Accumulating Evidence in Perceptual Decision Making," *Journal of Neuroscience* 32: 3612-3628 (2012).

[17] Fehr, Ernst, and Antonio Rangel, "Neuroeconomic Foundations of Economic Choice — Recent Advances," *Journal of Economic Perspectives* 25(4): 3-30 (2011).

[18] Fudenberg, Drew, Philipp Strack, and Tomasz Strzalecki, "Stochastic Choice and Optimal Sequential Sampling," unpublished, Harvard University, June 2015.

[19] Gabbiani, Fabrizio, and Steven J. Cox, *Mathematics for Neuroscientists,* Amsterdam: Academic Press, 2010.

[20] Green, David M., and John A. Swets, *Signal Detection Theory and Psychophysics,* New York: Wiley, 1966.

[21] Hébert, Benjamin, and Michael Woodford, "Rational Inattention with Sequential Information Sampling," unpublished, Stanford University, July 2016.

[22] Hey, John D., "Experimental Investigations of Errors in Decision Making under Risk," *European Economic Review* 39: 633-640 (1995).

[23] Hey, John D., "Does Repetition Improve Consistency?" *Experimental Economics* 4: 5-54 (2001).

[24] Kahneman, Daniel, *Thinking, Fast and Slow,* New York: Farrar, Straus and Giroux, 2011.

[25] Krajbich, Ian, Carrie Armel, and Antonio Rangel, "Visual Fixations and the Computation and Comparison of Value in Goal-Directed Choice," *Nature Neuroscience* 13: 1292-1298 (2010).

[26] Krajbich, Ian, Bastiaan Oud, and Ernst Fehr, "Benefits of Neuroeconomics Modeling: New Policy Interventions and Predictors of Preference," unpublished, Ohio State University, September 2013.

[27] Laming, D.R.J., *Information Theory of Choice-Reaction Times,* London: Academic Press, 1968.

[28] Link, S.W., and R.A. Heath, "A Sequential Theory of Psychological Discrimination," *Psychometrika* 40: 77-105 (1975).

[29] Lu, Jay, "Random Choice and Private Information," unpublished, Princeton University, October 2013.

[30] Luce, R. Duncan, *Response Times,* New York: Oxford University Press, 1986.

[31] Matějka, Filip, and Alisdair McKay, "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," *American Economic Review* 105: 272-298 (2015).

[32] McFadden, Daniel, "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, ed., *Frontiers in Econometrics,* New York: Academic Press, 1974.

[33] Milosavljevic, Milica, Jonathan Malmaud, Alex Huth, Christof Koch, and Antonio Rangel, "The Drift Diffusion Model Can Account for the Accuracy and Reaction Times of Value-Based Choice under High and Low Time Pressure," *Judgment and Decision Making* 5: 437-449 (2010).

[34] Mosteller, Frederick, and Philip Nogee, "An Experimental Measurement of Utility," *Journal of Political Economy* 59: 371-404 (1951).

[35] Natenzon, Paulo, "Random Choice and Learning," unpublished, Washington University, September 2013.

[36] Ortega, Pedro A., and Daniel A. Braun, "Thermodynamics as a Theory of Decision-Making with Information-Processing Costs," *Proceedings of the Royal Society A* 469: 20120683 (2013).

[37] Pinkovskiy, Maxim, "Rational Inattention and Choice under Risk: Explaining Violations of Expected Utility through a Shannon Entropy Formulation of the Costs of Rationality," *Atlantic Economic Journal* 37: 99-112 (2009).

[38] Ratcliff, Roger, "A Theory of Memory Retrieval," *Psychological Review* 85: 59-108 (1978).

[39] Ratcliff, Roger, and Jeffrey N. Rouder, "Modeling Response Times for Two-Choice Decisions," *Psychological Science* 9: 347-356 (1998).

[40] Roitman, Jamie D., and Michael N. Shadlen, "Response of Neurons in the Lateral Intraparietal Area During a Combined Visual Discrimination Reaction Time Task," *Journal of Neuroscience* 22: 9475-9489 (2002).

[41] Shadlen, Michael N., Timothy D. Hanks, Anne K. Churchland, Roozbeh Kiani, and Tianming Yang, "The Speed and Accuracy of a Simple Perceptual Decision: A Mathematical Primer," in K. Doya, S. Ishii, A. Pouget, and R.P.N. Rao, eds., *Bayesian Brain: Probabilistic Approaches to Neural Coding,* Cambridge, MA: MIT Press, 2007.

[42] Shadlen, Michael N., and Roozbeh Kiani, "Decision Making as a Window on Cognition," *Neuron* 80: 791-806 (2013).

[43] Sims, Christopher A., "Implications of Rational Inattention," *Journal of Monetary Economics* 50: 665-690 (2003).

[44] Sims, Christopher A., "Rational Inattention and Monetary Economics," in B.M. Friedman and M. Woodford, eds., *Handbook of Monetary Economics,* vol. 3A, Amsterdam: Elsevier, 2010.

[45] Smith, Philip L., "Stochastic Dynamic Models of Response Time and Accuracy: A Foundational Primer," *Journal of Mathematical Psychology* 44: 408-463 (2000).

[46] Stone, Mervyn, "Models for Choice-Reaction Time," *Psychometrika* 25: 251-260 (1960).

[47] Thurstone, Louis L., "A Law of Comparative Judgment," *Psychological Review* 34: 273-286 (1927).

[48] Wald, Abraham, *Sequential Analysis,* New York: John Wiley and Sons, 1947.

[49] Webb, Ryan, "The Dynamics of Stochastic Choice," unpublished, University of Toronto, March 2016.

[50] Wolpert, David, and David S. Leslie, "Information Theory and Observational Limitations in Decision Making," *The BE Journal of Theoretical Economics,* volume 12, issue 1, article 5, 2012.

[51] Woodford, Michael, "Inattentiveness as a Source of Randomized Discrete Adjustment," unpublished, Columbia University, April 2008.

[52] Woodford, Michael, "Information-Constrained State-Dependent Pricing," *Journal of Monetary Economics* 56(S): 100-124 (2009).