Cognitive, Affective, & Behavioral Neuroscience (2023) 23:593–603
https://doi.org/10.3758/s13415-022-01034-8

SPECIAL ISSUE/UNCERTAINTY

Mental control of uncertainty

Samuel J. Gershman[1] · **Taylor Burke**[1]

Accepted: 1 September 2022 / Published online: 27 September 2022

Abstract

Can you reduce uncertainty by thinking? Intuition suggests that this happens through the elusive process of attention: if we expend mental effort, we can increase the reliability of our sensory data. Models based on "rational inattention" formalize this idea in terms of a trade-off between the costs and benefits of attention. This paper surveys the origin of these models in economics, their connection to rate-distortion theory, and some of their recent applications to psychology and neuroscience. We also report new data from a numerosity judgment task in which we manipulate performance incentives. Consistent with rational inattention, people are able to improve performance on this task when incentivized, in part by increasing the reliability of their sensory data.

Keywords Perception · Bayesian inference · Information theory · Numerosity · Rational inattention

Introduction

Because our senses only give us partial and unreliable information about the environment, uncertainty is ubiquitous. Bayesian models of perception have formalized how uncertainty is represented, computed, and used by downstream processes such as decision-making and confidence judgments (Knill & Richards, 1996; Kersten et al., 2004; Meyniel et al., 2015). According to these models, percepts are probabilistic beliefs about hidden features of the environment, updated based on sensory data according to Bayes' rule. A typical and often overlooked property of these models is the assumption that uncertainty arises from noise sources that are "exogeneous"—the brain cannot directly control the reliability of sensory data, and by extension it cannot directly control the degree of subjective uncertainty. Thus, it is common practice to treat sensory reliability either as a free parameter (estimated from psychophysical data) or as dependent on an experimental parameter (e.g., stimulus contrast).

In this paper, we survey recent work that endogenizes sensory reliability, thereby placing uncertainty at least partially under mental control. The key idea is that sensory reliability is a kind of internal action—attention—that can be optimized based on a cost–benefit analysis. The benefit of attention is better task performance. The cost can take many forms; an influential line of work, originating in economics (Sims, 2003; Woodford, 2009; Mackowiak & Wiederholt, 2009; Mackowiak et al., 2018), adopts an information-theoretic cost function. We will derive this cost function from first principles, starting from the assumption that sensory channels have an upper bound on the average number of bits that can be transmitted per signal. The level of attention that maximizes utility (net benefit minus cost) can be understood as rational in the classical economic sense of rational choice theory. By the same token, neglecting some information (i.e., inattention) is rational when the costs are sufficiently high. Accordingly, the framework was dubbed "rational inattention" by economists seeking to explain why and when people or firms fail to make use of all available information.

The rest of this paper is organized as follows. We will first provide a non-technical overview of rational inattention theory, beginning with its origins in economics, followed by a survey of applications in psychology and neuroscience. We will then present a technical treatment specialized to an analytically tractable model class, which will allow us to concretely expose the key features of rational inattention. As a quantitative test of this model, we report new experimental data from a numerosity judgment task in which we manipulated incentives. Our analyses of this data

✉ Samuel J. Gershman
gershman@fas.harvard.edu

[1] Department of Psychology and Center for Brain Science, Harvard University, MA, Cambridge, USA

set highlight several implications of rational inattention for simple perceptual judgments.

Economic origins of rational inattention theory

The idea of rational inattention emerged from attempts to solve a long-standing macroeconomic puzzle: why do variables such as wages and prices respond slowly and coarsely to economic shocks? A closely related microeconomic puzzle is the tendency of individuals to value their wealth or income in nominal (e.g., dollar) rather than real (purchasing power) terms, a phenomenon known as the "money illusion" (Shafir et al., 1997). Even a small amount of individual-level money illusion can give rise to substantial aggregate inertia in response to shocks (Fehr & Tyran, 2001). Intuitively, if individuals don't notice changes in their purchasing power, then the economic pressure to adjust wages or prices will be blunted. But why, given the potentially large financial stakes, would individuals fail to notice such changes?

If we think of the economy as comprised of rational agents who respond instantaneously to market conditions, then we would expect no delay between shocks and adjustments, and the adjustments should vary predictably with the shock. Keynes (1936) incorporated a form of "stickiness" into an equilibrium model of the economy, but a satisfying explanation of how such stickiness arises did not appear until Sims (1998) located it in the limited attentional capacities of agents. This initial insight was subsequently developed by Sims (2003), Woodford (2009), and Mackowiak and Wiederholt (2009). The central feature of these models is that agents ignore some economic shocks, which explains why they adjust slowly and coarsely. Importantly, agents allocate their attention so as to optimize profit subject to their attentional capacity limit.

The concept of an attentional capacity limit can be formalized using information theory. In this section, we will provide a non-technical sketch of these ideas; later we will develop a more detailed technical treatment. The brain, like all physical information processing systems, has an upper bound (the channel capacity) on how much information about the external world it can transmit across sensory channels. Transmitting a signal across a channel requires a number of bits equal (on average) to the entropy of the signal (Shannon, 1948). If the average number of bits per signal (the information rate) exceeds the channel capacity, then reliable transmission is impossible. However, it was also recognized by Shannon that not all errors are equally important, and therefore it may be acceptable to unreliably transmit some signals in order to optimize a distortion function that characterizes the costs of particular errors (Shannon, 1959). This insight was the foundation of rate-distortion theory (Berger, 1971), which marries information theory and decision theory.

Economists, beginning with Sims (2003), used these ideas to define attention as the sensitivity of an information processing system (which could be an individual, firm, or other economic entity) to external signals (e.g., prices, wages, etc.). Optimal attentional allocation depends both on the incentive structure and the capacity limit of the system. When incentives are larger, more attention can be allocated, but only up to a point.

Before proceeding, a terminological remark is needed. What economists call rational inattention is mathematically equivalent to what information theorists call rate-distortion theory: they are "duals" of one another, in the sense that they define different optimization problems with the same solution (Denti et al., 2020).[1] Therefore, it doesn't really make sense to discuss the two areas of research separately. Rate-distortion theory has recently become fertile ground for thinking about perception, memory, and decision-making (see Sims 2016; Lai and Gershman2021, for reviews). In this paper, we will refer to rational inattention because we feel that the emphasis on signal sensitivity provides a natural way of talking about particular phenomena. However, we want to make clear that this is a matter of emphasis rather than a deep theoretical distinction.

Applications to perception, memory, and decision-making

Psychologists studying memory and perception have long been interested in capacity limits, and indeed much of the early work on these topics was directly influenced by information theory (Attneave, 1954; Miller, 1956). What jumped out at Miller in his famous 1956 paper was the fact that the same capacity limit (measured in bits) seemed to appear across many different experiments with different stimuli. He suggested that items or sets of items could be compressed to make more efficient use of the fixed capacity limit. Compression also played an important role in Attneave's discussion of perceptual organization such as grouping, where redundant stimulus elements are compressed to produce a single Gestalt figure, thereby making more efficient use of a fixed perceptual capacity.

A stumbling block in these efforts was the fact that information theory by itself could not comprehensively explain why some stimuli were more memorable or perceptible than others. For example, if a subject receives larger payoffs for remembering some items than others, memory is better for the high-payoff items (Taub, 1965; Christ, 1969; Tolkmitt

[1] Some rational inattention models adopt cost functions that do not have an information-theoretic interpretation, and in these cases the two approaches diverge.

& Christ, 1970; Martin & Richards, 1972; Brissenden et al., 2021). A purely information-theoretic treatment of capacity only constrains memory based on the signal statistics and channel noise; it has no means of prioritizing some signals over others based on their differential payoffs. Rational inattention, in contrast, provides a natural explanation, since high-payoff items should be preferentially allocated bits in memory. One implication of this explanation, supported by the data, is that memory for low-payoff items should be *worse* compared to a condition in which there are no differential payoffs. This follows from the fact that extra bits for high-payoff items can only be obtained by reducing the bit allocation for low-payoff items.

Payoff history for different stimuli can affect prioritization in memory even when memory performance itself is not differentially rewarded. For example, Gong and Li (2014) associated some stimuli with low rewards and others with high rewards in a visual search task; they then used the same stimuli in a change detection task, finding that detection accuracy was better for the stimuli associated with high rewards, despite the fact that reward on the change detection task did not depend on these associations (see also Wallis et al., 2015; Thomas et al., 2016). In a similar vein, Bates et al., (2019) showed that changes are easier to detect along stimulus dimensions that are relevant to a previously learned category distinction. These findings suggest that the allocation of attentional capacity is adapted to the distribution of stimuli or tasks.

The data discussed so far come from studies with humans, but similar conclusions can be drawn from studies with animals. Using rat and pigeon subjects, Nevin et al., (1975) reported that signal sensitivity in a two-alternative forced choice task increased in proportion to the difference in payoff for correct and incorrect responses (see also Davison and McCarthy 1980). They pointed out that this finding appears to contradict classical signal detection theory, which assumes that differential payoffs should affect response bias but not sensitivity. More recently, Grujic et al., (2022) explicitly applied a rational inattention model to data from an orientation discrimination task in mice. By manipulating the mapping between stimuli and payoffs for correct responses, they were able to show that sensitivity was higher for orientations receiving larger payoffs. This pattern was quantitatively matched by a model that fully endogenized sensitivity based on the stimulus–reward mapping.[2]

Rational inattention theory can also be applied to more cognitive domains, where the signals are internally generated. For example, Gabaix and Laibson (2017) proposed that intertemporal decisions may involve mental simulation of the future. Intuitively, the value of a delayed payoff depends on the conditions at the time of receipt. Mental simulation is an inherently noisy way of prospecting about these conditions due to the fact that each simulation is a random sample path through a hypothetical future. The decision-maker can interpret the prospective values constructed by these simulations as signals, and then combine them with prior beliefs (via Bayes' rule) to estimate payoff value. Gabaix and Laibson showed that under certain distributional assumptions they could derive a hyperbolic temporal discounting function, where discounting of the future (myopia) increases with the simulation noise variance. This captures the idea that the decision-maker should rely less on simulations when they are less reliable indicators of future value. Gabaix and Laibson treated the simulation noise variance as exogenous, but evidence suggests that it might be adaptively calibrated. In particular, the fact that myopia diminishes for larger magnitudes (Thaler, 1981) suggests that the noise variance may diminish for larger magnitudes. Gershman and Bhui (2020) showed how this arises naturally from a rational inattention treatment of the Bayesian discounting model. The key innovation is to endogenize the noise variance so that it becomes sensitive to payoff magnitude. A novel prediction of this treatment, confirmed empirically, was that response variability decreases with magnitude, consistent with reduced simulation noise.

Links to the study of selective attention

Given that rational inattention theory invokes the psychological concept of attention, it is reasonable to expect that it should have something to say about the voluminous literature on attention. Here, we take a brisk tour of some key findings and ideas from this literature, drawing out some connections to rational inattention.

The earliest systematic studies of attention, exemplified by Broadbent (1958), suggested a "filter" theory of attention, whereby instructions to focus on one auditory source (presented to one ear) obliterated comprehension of another simultaneous auditory source in the other ear. The unattended source was apparently filtered out at an early stage of perception, rendering it completely unavailable for downstream computation. It turned out, however, that some information from the unattended source *is* available downstream, albeit in an attenuated form (Treisman 1960). The degree of attenuation is greater under conditions of high perceptual load (e.g., a large number of distractors; Lavie and Tsal 1994; Lavie 1995). These findings suggest that the amount of unattended information available to downstream computation depends on a capacity-limited perceptual channel:

[2] Note that Grujic et al., (2022) did not use the information-theoretic cost function, so it is somewhat different from the other models in this literature. Their cost function is motivated by neurobiological constraints, and is in that sense more similar in spirit to the work of Van den Berg and Ma (2018).

unattended signals are propagated only when capacity is not saturated. The idea that attentional resources can be flexibly allocated based on available capacity and priority is broadly consistent with rational inattention theory.

Subsequent work attempted to unpack the mechanisms underlying attentional enhancement of perceptual processing (see Carrasco 2011, for a review). Studies indicated that cues can affect both response bias and sensitivity, but these can be dissociated by different experimental manipulations (Wyart et al., 2012; Luo & Maunsell, 2015). For example, Wyart et al., (2012) showed, using a spatial cueing paradigm, that information about where a stimulus will appear affects response bias but not sensitivity, whereas information about whether a particular location will be queried (i.e., the location's task relevance) affects sensitivity but not response bias. This result can be understood in terms of rational inattention: sensitivity is predicted to increase when it pays off to decrease internal noise, as in the case of cues indicating task relevance (see also Engelmann and Pessoa 2007). Cues indicating stimulus location, on the other hand, do not dictate that a reduction of internal noise is worth the cost, but they do indicate how to set the optimal decision criterion.

A final link that we want to touch upon is the literature on enhancement of attention by reward associations (see Bourgeois et al., 2016, for a review). For example, Anderson et al., (2011) showed that stimuli previously associated with monetary rewards capture attention during visual search. This reward-induced enhancement can last for several days (Della Libera and Chelazzi, 2009), and is accompanied by changes in early visual cortex tuning (Itthipuripat et al., 2019). While rational inattention theory does not directly address associative learning, these effects are broadly consistent with the notion that reward shapes selective attention, in some cases by modulating activity in early perceptual areas (Serences, 2008; Luo & Maunsell, 2015).

Case study: Gaussian magnitude estimation

In this section, we will develop the mathematical details for a special case of rational inattention theory. This case will serve as the model for the behavioral experiment that we describe later.

We consider an agent that receives a signal $m \in \mathbb{R}$ representing a scalar magnitude (e.g., length, size, numerosity, etc.), drawn from some distribution $p(m)$. The goal of the agent is to estimate the expected value of m, denoted $\mu = \mathbb{E}[m]$. For analytical tractability, we will specialize this setup to the assumption that the signal distribution is Gaussian, $m \sim \mathcal{N}(\mu, \lambda^{-1})$, with precision (inverse variance) λ.

Bayesian estimation

Bayes' rule prescribes the normative solution to the estimation problem. The posterior distribution over μ is given by:

$$p(\mu|m) \propto p(m|\mu)p(\mu), \tag{1}$$

where $p(m|\mu) = \mathcal{N}(m;\mu, \lambda^{-1})$ is the *likelihood* and $p(\mu)$ is the *prior*. If we assume that the prior is Gaussian, $\mu \sim \mathcal{N}(\mu_0, \lambda_0^{-1})$, then the posterior is also Gaussian, with posterior mean $\hat{\mu}$:

$$\hat{\mu} = wm + (1-w)\mu_0, \tag{2}$$

where

$$w = \frac{\lambda}{\lambda+\lambda_0} \tag{3}$$

is the *sensitivity* to the signal. The sensitivity w is determined by the relative precision of the likelihood and prior. In particular, the sensitivity is high when signal precision (λ) is high relative to prior precision (λ_0).

Equation 2 offers a simple account of central tendency (also known as regression) effects, which are ubiquitous in magnitude estimation tasks (Petzschner et al., 2015): magnitude estimates tend to be pulled towards the prior mean, such that magnitudes less than the mean are overestimated and magnitudes greater than the mean are underestimated. Consistent with the Bayesian account, these effects are greater under conditions in which sensory noise is putatively higher (and hence signal sensitivity w is lower). For example, Xiang et al., (2021) showed that briefer stimulus presentations induce a stronger central tendency effect. The Bayesian account also explains why central tendency effects tend to be larger when the range of stimuli is wider (Teghtsoonian and Teghtsoonian, 1978; Petzschner & Glasauer, 2011): w decreases with λ_0, which is smaller when the range is wider.

Rational inattention

Intuition suggests that when signals are more important, the agent should pay more attention to them. This intuition can be formalized using the framework of rational inattention. The agent has a limited attentional capacity that can be allocated to the signal estimation problem. In information-theoretic terms, this capacity limit C is the maximum feasible information rate of the channel. The rate is defined as the mutual information between μ and m:

$$I(\mu;m) = H(\mu) - H(\mu|m), \tag{4}$$

where $H(\mu)$ is the entropy of the prior $p(\mu)$, expressing the amount of uncertainty about μ prior to observing the signal, and $H(\mu|m)$ is the conditional entropy, expressing the amount of uncertainty about μ after observing the signal. Thus, mutual information expresses the degree of uncertainty reduction due to the signal—a formal conceptualization of attention (see also Itti & Baldi 2009; Feldman & Friston 2010). For the Gaussian generative model described in the previous section, the mutual information is given by:

$$I(\mu;m) = \frac{1}{2}\log\left(1 + \frac{\lambda}{\lambda_0}\right). \tag{5}$$

Shannon's noisy channel theorem states that the minimum expected number of bits needed to communicate μ across a noisy channel without error is equal to $I(\mu;m)$. A corollary is that errorless communication is not possible if the agent's capacity is less than $I(\mu;m)$.

If the agent is operating at its attentional capacity limit, then the optimization problem is to maximize expected reward subject to the constraint that $I(\mu;m)$ cannot exceed the capacity C. In our setup, the parameter being optimized is the signal precision λ, which allows us to capture the idea that the agent can attend more to the signal (by increasing λ) when it is more important, and hence worth paying a higher information cost. We will first describe the reward structure and then show how this is integrated into a capacity-constrained optimization problem.

We assume that the reward an agent collects on a particular task is a monotonically decreasing and differentiable function $u(\epsilon)$ of the squared error $\epsilon = (\mu - \hat{\mu})^2$. Taking a first-order Taylor series approximation around $\epsilon = 0$ gives:

$$u(\epsilon) \approx u(0) - \beta(\mu - \hat{\mu})^2, \tag{6}$$

where $\beta > 0$ is the negative slope of $u(\epsilon)$ at $\epsilon = 0$. Intuitively, β expresses how rapidly reward decreases with estimation error for a given task. We will interpret β as an *attentional incentive* parameter to capture the idea that agents are motivated to pay more attention to the signal when reward is contingent on error. Evidence for reward contingency-dependent increases in cognitive effort are well documented (Kool et al., 2016; Kool et al., 2017; Manohar et al., 2017; Frömer et al., 2021).

For the Gaussian generative model, the expected reward is given by:

$$U = \mathbb{E}[u(\epsilon)] \approx u(0) - \frac{\beta}{\lambda + \lambda_0}, \tag{7}$$

where we have used the Taylor series expansion in the second expression. The denominator in the second term ($\lambda + \lambda_0$) is equal to the posterior precision; thus, the expected reward is inversely related to the posterior uncertainty (variance) scaled by the attentional incentive (β). Another way to understand this expression is by noting that the expected squared error is given by $\mathbb{E}[\epsilon] = 1/(\lambda + \lambda_0)$. Thus, $U \approx u(0) - \beta\mathbb{E}[\epsilon]$, meaning that β determines the first-order effect of squared error on utility.

We can now write down the optimization problem facing the agent. We formulate it as a Lagrangian "dual" function (the standard formulation in rational inattention models), which is equivalent to maximizing reward under a capacity constraint:

$$\lambda^* = \underset{\lambda}{argmax} U - \kappa I(\mu;m), \tag{8}$$

where $\kappa \geq 0$ is a Lagrange multiplier that can be interpreted as the attentional cost. More precisely, κ is the "exchange rate" between reward and information: one unit of reward can be "bought" for κ units of information (e.g., bits, if using the base 2 logarithm in the definition of mutual information). In general, κ decreases with the agent's capacity. Note that we no longer have the capacity limit C in this formulation; it is implicitly expressed by κ.

Using the Taylor series approximation, the optimal signal precision is given by:

$$\lambda^* = \max(0, 2\beta/\kappa - \lambda_0). \tag{9}$$

From this result, we can deduce that the optimal sensitivity is given by:

$$w^* = 1 - \frac{\lambda_0 \kappa}{2\beta}, \tag{10}$$

subject to the constraint that $w^* \in [0,1]$. Summarizing these results:

- As the attentional incentive β increases, optimal signal precision λ^* and sensitivity w^* increase.
- As the attentional cost κ increases, optimal signal precision λ^* and sensitivity w^* decrease.
- As the prior precision λ_0 increases, optimal signal precision λ^* and sensitivity w^* decrease.

Note that $\frac{d\lambda^*}{d\kappa}$ is negative and monotonically decreasing as a function of β; thus, β amplifies the effect of κ on optimal signal precision.

Plugging the optimal signal precision λ^* into Eq. 7 yields:

$$U^* = u(0) - \frac{\kappa}{2}. \tag{11}$$

Thus, when the agent is operating at its capacity limit, its expected reward is inversely proportional to the attentional cost κ. Mikhael et al., (2021), following earlier work (Niv et al., 2007; Beierholm et al., 2013; Hamid et al., 2016), proposed that tonic dopamine encodes average reward. Equation 11 means that the average reward interpretation of dopamine is equivalent (under the rational inattention analysis) to an interpretation of dopamine in terms of the exchange rate between reward and information. This equivalence also

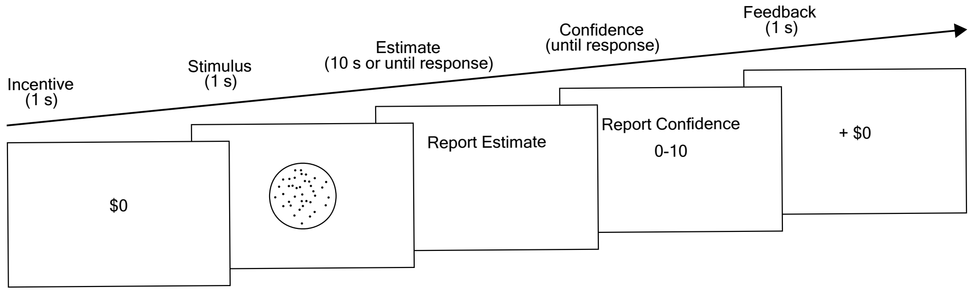

Fig. 1 Illustration of a trial. After viewing the trial-specific performance bonus, subjects viewed the stimulus, reported their numerosity estimate, and then rated their confidence on a ten-point scale. Lastly, they received feedback about whether they received the incentivized performance bonus

allows us to draw a connection to the interpretation of dopamine as posterior precision (Friston et al., 2012; FitzGerald et al., 2015; Tomassini et al., 2016; Shi et al., 2013): Eq. 7 shows that expected reward is increasing in posterior precision, and the rational inattention analysis implies that posterior precision is equal to $2\beta/\kappa$. These expressions show that interpretations of dopamine as information–reward exchange rate, average reward, and posterior precision are in a sense interchangeable.

We can also re-express the optimal sensitivity in terms of the average reward:

$$w^* = 1 + \frac{\lambda_0 U^*}{\beta}, \qquad (12)$$

which shows that optimal sensitivity increases with average reward, scaled inversely by β. One way to understand this is in terms of the task complexity interpretation of β: more complex tasks require a larger reward to induce the same degree of sensitivity.

Experiment: numerosity judgment with variable incentives

To test the predictions of the model described in the previous section, we collected a new data set from subjects performing a numerosity estimation task. We manipulated incentives on a trial-by-trial basis in order to determine whether errors, variability, confidence, and central tendency change in accordance with the principles of rational inattention.

Subjects

We recruited 377 subjects from Amazon Mechanical Turk. All subjects gave informed consent prior to testing. To ensure that subjects fully understood the experiment, they completed a three-question comprehension check after the instructions. No subjects were excluded for failing the comprehension check. All subjects received a $6 base payment. As a performance bonus, subjects were awarded the incentive ($1 or $5) on a randomly selected trial if their estimate on that trial was within three of the true numerosity. We excluded seven subjects who were either older than 69 or made more than 30 invalid responses (outside the range of possible numerosities, or more than 40 away from the true numerosity). All invalid responses for the remaining subjects were excluded from subsequent analysis. The study was approved by the Harvard Institutional Review Board.

Stimuli and procedure

The stimuli were black dots in a random spatial arrangement within an aperture. The number of dots ranged between 15 and 65.

To gain familiarity with the task, subjects were trained on four trials in which the true numerosity was revealed during feedback (in the rest of the task, this feedback was not provided). After training, subjects completed six blocks with 40 trials per block. Each block's average numerosity was randomly drawn from a uniform distribution (30 to 50, sampled without replacement). The average numerosity was revealed to subjects at block onset.

On each trial, a numerosity was sampled from a uniform distribution (block average stimulus magnitude ± 15, sampled with replacement). Trials were randomly assigned to either the high or low incentive condition, such that every condition was evenly distributed within and across blocks. High incentive trials were color-coded by a green border on the aperture and offered a potential performance bonus of $5. Low incentive trials were color-coded by a grey border and offered a potential performance bonus of $1.

As illustrated in Fig. 1, each trial began with the presentation of a green or grey color-coded incentive value at the center of the screen for 1 s. Then, a random dot array, bordered with the same color as the incentive, was presented for 1 s. Subjects then had 10 s to report their numerosity estimate using the number pad on their keyboard. Subjects were then prompted to rate the confidence in their estimate using a discrete slider ranging from 0 (random guessing) to 10 (very confident). Finally, subjects were given feedback ($0 for inaccurate performance, or the indicated incentive for accurate performance) for 1 s. Accurate performance was defined as being with ± 3 of the true numerosity.

Parameter estimation and model comparison

To model the behavioral data (subjective magnitude reports, denoted by y), we assume that the sensory signal m is generated from mean $\mu = \log x$, where x is the true numerosity. The log transformation captures diminishing sensitivity for larger magnitudes (i.e., the Weber-Fechner law; Dehaene 2003). Likewise, the prior is represented on the log scale: $\mu_0 = \mathbb{E}[\log \mu]$.

We model subjective magnitude reports as exponential transformations of the logarithmic posterior mean estimates, corrupted by zero-mean Gaussian response noise with variance τ:

$$\log y|m \sim \mathcal{N}(\hat{\mu}, \tau). \tag{13}$$

Marginalizing over the latent sensory signal m yields:

$$\log y \sim \mathcal{N}(w\mu + (1-w)\mu_0, w^2/\lambda + \tau). \tag{14}$$

We use the reduced-form parametrization $\alpha = \beta/\kappa$ because β and κ only influence the signal sensitivity via their ratio. When incentive is varied across discrete conditions, we fit a separate value of α for each condition.[3] Thus, for condition n, the optimal precision is $\lambda_n^* = 2\alpha_n - \lambda_0$, with the constraint $\alpha_n \geq \lambda_0/2$. We set μ_0 and λ_0 to match the mean and precision (respectively) of the log-transformed numerosities for each block of trials.

We will refer to the model described above as the *Rational Inattention* (RI) model. Its free parameters are $\theta = (\tau, \alpha_{1:N})$, where $\alpha_{1:N}$ is the set of attentional parameters for N different incentive conditions (in our experiment, $N = 2$). We compared it to the following alternative models:

- Fixed Precision (FP): the Bayesian estimator with fixed sensory precision (λ treated as a free parameter rather than endogenized based on α) and leak. Free parameters are $\theta = (\tau, \lambda)$.
- Non-Bayesian (NB): numerosity judgments are noisy linear functions of the sensory signal m, $\log y|m \sim \mathcal{N}(am, \tau)$, where $a \in \mathbb{R}$. In this model, the prior mean plays no role. Marginalizing over m yields: $\log y \sim \mathcal{N}(a_n\mu, a_n^2/\lambda + \tau)$, where we allow different signal weights for each incentive condition. Free parameters are $\theta = (a_{1:N}, \tau, \lambda)$.

To fit the model to data, we searched for parameters θ that maximize the log-likelihood of the data for each subject separately. The maximum likelihood estimates were obtained using a non-linear optimization routine with 5 random starting points. To compare models, we used random-effects Bayesian model selection (Rigoux et al., 2014), which estimates the frequency of each model in the population based on the log model evidence (marginal likelihood), which we approximated using the Bayesian information criterion (BIC; Bishop, 2006):

$$\log P(\mathbf{y}) = \int_\theta P(\mathbf{y}|\theta)P(\theta)d\theta \approx -\tfrac{1}{2}\mathrm{BIC}, \tag{15}$$

where $\mathbf{y}$ is the collection of numerosity estimates. This approximation is asymptotically correct in the limit of many samples, where the posterior concentrates on the maximum likelihood estimate.

Results

Rational inattention makes several predictions, which we tested using our data set. First, performance should be better when incentives are high compared to when they are low. The expected squared error is given by:

$$\mathbb{E}[(\mu - \hat{\mu})^2] = \frac{1}{\lambda + \lambda_0} = \frac{1}{2\alpha}. \tag{16}$$

Recall that α is the ratio of the attentional incentive (β) to the attentional cost (κ). We assume that α increases with incentive; we directly verify this hypothesis below. Since expected error is monotonically decreasing in α, it is also monotonically decreasing in incentive. This prediction was confirmed in our data set, as shown in Fig. 2a: the squared error was significantly larger in the low incentive condition [$t(369) = 4.02, p < 0.0001$].

Second, rational inattention predicts that variance of (signed) errors should be higher for lower incentives, due to the fact that mental precision (λ) is lower. This prediction was confirmed in our data set, as shown in Fig. 2b: the error variance was significantly larger in the low incentive condition [$t(369) = 3.04, p < 0.005$].

Third, rational inattention predicts that confidence should be higher under high incentives. There are a number of ways to formalize confidence (for an application to numerosity estimation, see Xiang et al., 2021), but in general confidence will increase in mental precision (λ), which increases with incentive. This prediction was confirmed in our data set, as shown in Fig. 2c: confidence was significantly larger in the high incentive condition [$t(369) = 8.55, p < 0.0001$].

Finally, rational inattention predicts that there should be a weaker central tendency effect (influence of the prior mean) under high incentives. Following Xiang et al., (2021), we quantified the central tendency effect using a linear mixed-effects model with terms for the true stimulus on each trial, prior

[3] Note that because payoffs depend on a hard distance threshold, our experimental design is not completely compatible with our assumption that reward is a differentiable function of squared error. However, we chose to overlook this nuance in order to retain analytical tractability.

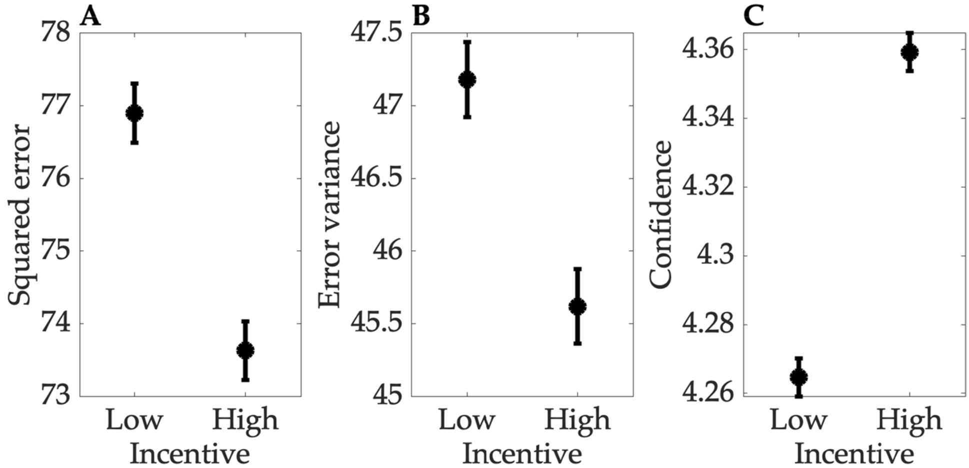

Fig. 2 Behavioral results in the numerosity estimation task. **A** Squared error. **B** Error variance. **C** Confidence. *Error bars* show within-subject standard error of the mean

mean on each block, and the interaction between prior mean and incentive (coded as −1 for low and +1 for high). This analysis revealed significantly positive coefficients for stimulus [$F(1,87249) = 4538.8$, $p < 0.0001$] and prior mean [$F(1,87249) = 8.6678$, $p < 0.005$], indicating an overall central tendency effect. Critically, the interaction coefficient was significantly negative [$F(1,87249) = 26.47$, $p < 0.0001$], indicating that the central tendency effect diminished when incentives were higher.

To more quantitatively evaluate the Rational Inattention (RI) model, we compared it to two other models described above: the Fixed Precision model (FP) and the Non-Bayesian model (NB). The FP model is identical to the RI model, except that it assumes a fixed mental precision rather than adaptive precision. The NB model has a similar functional form to the FP model but ignores the prior and does not use the posterior to determine the signal weight. The log Bayes factors (differences in log model evidence for two models) decisively favored the RI model (Fig. 3a). Random-effects Bayesian model comparison produced a protected exceedance probability indistinguishable from 1, indicating a very high posterior probability that the RI

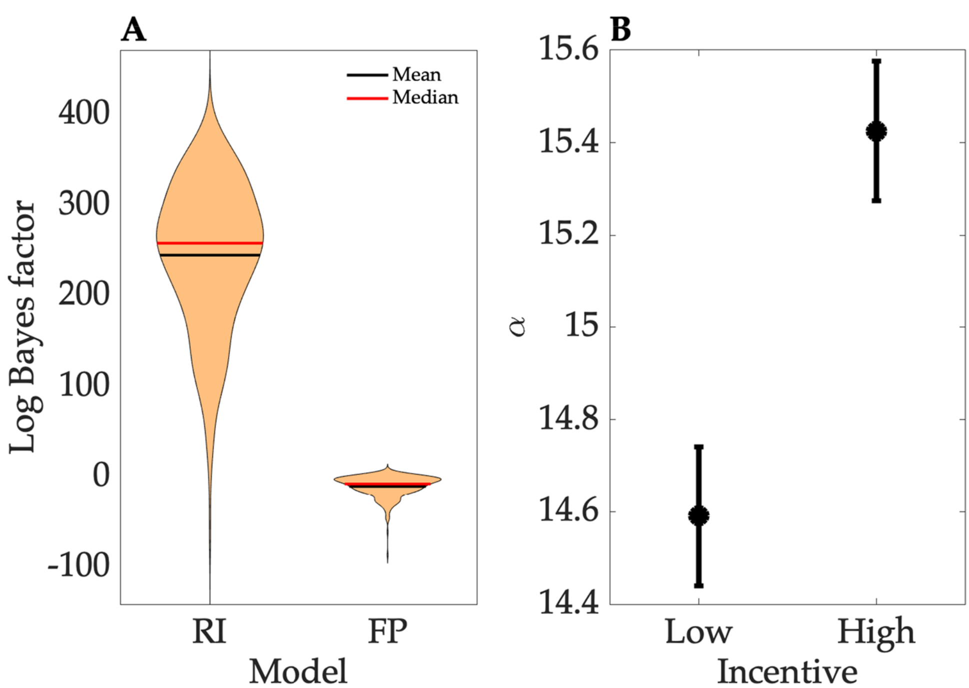

Fig. 3 Modeling results. **A** Log Bayes factors for the Rational Inattention (RI) and Fixed Precision (FP) models compared to the Non-Bayesian (NB) model. **B** Fitted α parameter for each incentive condition. *Error bars* show within-subject standard error of the mean

model is the most frequent model in the population. As shown in Fig. 3b, the fitted α parameter in the RI model was significantly larger in the high incentive condition compared to the low incentive condition [$t(369) = 2.77$, $p < 0.01$], supporting our claim that the benefit/cost ratio is sensitive to our incentive manipulation, putatively driving the effects of squared error, error variance, and confidence detailed above.

Discussion

The idea that uncertainty may be under mental control unifies many disparate observations from economics, psychology, and neuroscience. Rational inattention provides a formal framework for thinking about uncertainty endogenously rather than exogenously (the standard approach in cognitive science). Using magnitude estimation as a tractable case study, we developed a detailed mathematical model and tested its predictions experimentally with a numerosity estimation task. In particular, we found that subjects were more accurate, reliable, and confident when performance incentives were higher, supporting the view that mental uncertainty can be endogenously reduced when potential payoffs outweigh the attentional costs. Quantitative model fitting and comparison supported these conclusions.

Our findings from the numerosity estimation task are consistent with a recent study of numerosity discrimination, where subjects had to judge which display contained more elements (Dix and Li, 2020). The researchers found that accuracy was increased when incentives were larger, consistent with the rational inattention account. One advantage of our estimation task is that it is more straightforward to formalize using the rational inattention model (see Hébert & Woodford 2019, for an application to two-alternative forced choice tasks).

More broadly, rational inattention fits with a constellation of hypotheses that suggest how mental effort can be adaptively allocated based on task demands and incentives (Shenhav et al., 2017; Kool & Botvinick, 2018). For example, people will engage in more effortful thinking—such as planning (Kool et al., 2016; Kool et al., 2017) and storing information in memory (Kool & Botvinick, 2012; Westbrook & Braver, 2015)—when incentives are larger. Uncertainty may thus be one among many dials in the brain that are managed by control processes based on a cost-benefit analysis.

Acknowledgments This work was supported by a Bipolar Disorder seed grant from the Harvard Brain Initiative and by the Center for Brains, Minds and Machines (CBMM), funded by NSF STC award CCF-1231216.

References

Anderson, B.A., Laurent, P.A., & Yantis, S. (2011). Value-driven attentional capture. *Proceedings of the National Academy of Sciences*, *108*, 10367–10371.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*, 183–193.

Bates, C.J., Lerch, R.A., Sims, C.R., & Jacobs, R.A. (2019). Adaptive allocation of human visual working memory capacity during statistical and categorical learning. *Journal of Vision*, *19*, 11.

Beierholm, U., Guitart-Masip, M., Economides, M., Chowdhury, R., Düzel, E., Dolan, R., & Dayan, P. (2013). Dopamine modulates reward-related vigor. *Neuropsychopharmacology*, *38*, 1495–1503.

Berger, T. (1971). Rate Distortion Theory: A Mathematical Basis for Data Compression. NJ: Prentice-Hall.

Bishop, C.M. (2006). Pattern Recognition and Machine Learning. Springer.

Bourgeois, A., Chelazzi, L., & Vuilleumier, P. (2016). How motivation and reward learning modulate selective attention. *Progress in Brain Research*, *229*, 325–342.

Brissenden, J.A., Adkins, T.J., Hsu, Y.T., & Lee, T.G. (2021). Reward influences the allocation but not the availability of resources in visual working memory. bioRxiv.

Broadbent, D.E. (1958). Perception and Communication. Oxford University Press.

Carrasco, M. (2011). Visual attention: the past 25 years. *Vision Research*, *51*, 1484–1525.

Christ, R.E. (1969). Effects of payoff and probability on recall of multisymbol displays. *The Journal of General Psychology*, *80*, 81–92.

Davison, M., & McCarthy, D. (1980). Reinforcement for errors in a signal-detection procedure. *Journal of the Experimental Analysis of Behavior*, *34*, 35–47.

Dehaene, S. (2003). The neural basis of the Weber-Fechner law: a logarithmic mental number line. *Trends in Cognitive Sciences*, *7*, 145–147.

Della Libera, C., & Chelazzi, L. (2009). Learning to attend and to ignore is a matter of gains and losses. *Psychological Science*, *20*, 778–784.

Denti, T., Marinacci, M., & Montrucchio, L. (2020). A note on rational inattention and rate distortion theory. *Decisions in Economics and Finance*, *43*, 75–89.

Dix, A., & Li, S.-C. (2020). Incentive motivation improves numerosity discrimination: Insights from pupillometry combined with drift-diffusion modelling. *Scientific Reports*, *10*, 1–11.

Engelmann, J. B., & Pessoa, L. (2007). Motivation Sharpens Exogenous Spatial Attention. *Emotion*, *7*, 668–674.

Fehr, E., & Tyran, J.-R. (2001). Does money illusion matter? *American Economic Review*, *91*, 1239–1262.

Feldman, H., & Friston, K. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*, 215.

FitzGerald, T.H., Dolan, R.J., & Friston, K. (2015). Dopamine, reward learning, and active inference. *Frontiers in Computational Neuroscience*, *9*, 136.

Friston, K.J., Shiner, T., FitzGerald, T., Galea, J.M., Adams, R., Brown, H., ..., Bestmann, S. (2012). Dopamine, affordance and active inference. *PLOS Computational Biology*, *8*, e1002327.

Frömer, R., Lin, H., Wolf, C.D., Inzlicht, M., & Shenhav, A. (2021). Expectations of reward and efficacy guide cognitive control allocation. *Nature Communications*, *12*, 1–11.

Gabaix, X., & Laibson, D. (2017). Myopia and discounting. *Technical report, National bureau of economic research*.

Gershman, S.J., & Bhui, R. (2020). Rationally inattentive intertemporal choice. *Nature Communications*, *11*, 1–8.

Gong, M., & Li, S. (2014). Learned reward association improves visual working memory. *Journal of Experimental psychology. Human Perception and Performance*, *40*, 841–856.

Grujic, N., Brus, J., Burdakov, D., & Polania, R. (2022). Rational inattention in mice. *Science Advances*, *8*, eabj8935.

Hamid, A.A., Pettibone, J.R., Mabrouk, O.S., Hetrick, V.L., Schmidt, R., Vander Weele, C.M., ..., Berke, J.D. (2016). Mesolimbic dopamine signals the value of work. *Nature Neuroscience*, *19*, 117–126.

Hébert, B.M., & Woodford, M. (2019). Rational inattention when decisions take time. *Technical report, National Bureau of Economic Research*.

Itthipuripat, S., Vo, V.A., Sprague, T.C., & Serences, J.T. (2019). Value-driven attentional capture enhances distractor representations in early visual cortex. *PLoS Biology*, *17*, e3000186.

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, *49*, 1295–1306.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, *55*, 271–304.

Keynes, J.M. (1936) *The general theory of employment, interest and money*. Harcourt: Brace, & World.

Knill, D.C., & Richards, W. (1996). Perception as Bayesian Inference. Cambridge University Press.

Kool, W., & Botvinick, M. (2012). A labor/leisure tradeoff in cognitive control. *Journal of Experimental psychology. General*, *143*, 131–141.

Kool, W., & Botvinick, M. (2018). Mental labour. *Nature Human Behaviour*, *2*, 899–908.

Kool, W., Cushman, F.A., & Gershman, S.J. (2016). When does model-based control pay off? *PLoS Computational Biology*, *12*, e1005090.

Kool, W., Gershman, S.J., & Cushman, F.A. (2017). Cost-benefit arbitration between multiple reinforcement-learning systems. *Psychological Science*, *28*, 1321–1333.

Lai, L., & Gershman, S.J. (2021). Policy compression: An information bottleneck in action selection. In *Psychology of Learning and Motivation*, (Vol. 74 pp. 195–232): Elsevier.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology*, *21*, 451–468.

Lavie, N., & Tsal, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & Psychophysics*, *56*, 183–197.

Luo, T.Z., & Maunsell, J.H. (2015). Neuronal modulations in visual cortex are associated with only one of multiple components of attention. *Neuron*, *86*, 1182–1188.

Mackowiak, B., Matejka, F., & Wiederholt, M. (2018). Rational inattention: a disciplined behavioral model. CEPR Discussion Papers, 13243.

Mackowiak, B., & Wiederholt, M. (2009). Optimal sticky prices under rational inattention. *American Economic Review*, *99*, 769–803.

Manohar, S.G., Finzi, R.D., Drew, D., & Husain, M. (2017). Distinct motivational effects of contingent and noncontingent rewards. *Psychological Science*, *28*, 1016–1026.

Martin, D.W., & Richards, E. (1972). Locus of selectivity in a visual memory task using a color value indicator. *Perception & Psychophysics*, *12*, 65–68.

Meyniel, F., Sigman, M., & Mainen, Z.F. (2015). Confidence as Bayesian probability: From neural origins to behavior. *Neuron*, *88*, 78–92.

Mikhael, J.G., Lai, L., & Gershman, S.J. (2021). Rational inattention and tonic dopamine. *PLOS Computational Biology*.

Miller, G.A. (1956). The magic number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 91–97.

Nevin, J.A., Olson, K., Mandell, C., & Yarensky, P. (1975). Differential reinforcement and signal detection. *Journal of the Experimental Analysis of Behavior*, *24*, 355–367.

Niv, Y., Daw, N.D., Joel, D., & Dayan, P. (2007). Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology*, *191*, 507–520.

Petzschner, F.H., & Glasauer, S. (2011). Iterative Bayesian estimation as an explanation for range and regression effects: a study on human path integration. *Journal of Neuroscience*, *31*, 17220–17229.

Petzschner, F.H., Glasauer, S., & Stephan, K.E. (2015). A Bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences*, *19*, 285–293.

Rigoux, L., Stephan, K.E., Friston, K.J., & Daunizeau, J. (2014). Bayesian model selection for group studies—revisited. *NeuroImage*, *84*, 971–985.

Serences, J.T. (2008). Value-based modulations in human visual cortex. *Neuron*, *60*, 1169–1181.

Shafir, E., Diamond, P., & Tversky, A. (1997). Money illusion. *The Quarterly Journal of Economics*, *112*, 341–374.

Shannon, C. (1959). Coding theorems for a discrete source with a fidelity criterion. *IRE National Convention Record*, *4*, 142–163.

Shannon, C.E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*, 379–423.

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T.L., Cohen, J.D., & Botvinick, M.M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, *40*, 99–124.

Shi, Z., Church, R.M., & Meck, W.H. (2013). Bayesian optimization of time perception. *Trends in Cognitive Sciences*, *17*, 556–564.

Sims, C.A. (1998). Stickiness. In *Carnegie-Rochester Conference Series on Public Policy*, (Vol. 49 pp. 317–356): Elsevier.

Sims, C.A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, *50*, 665–690.

Sims, C.R. (2016). Rate-distortion theory and human perception. *Cognition*, *152*, 181–198.

Taub, H.A. (1965). Effects of differential value on recall of visual symbols1. *Journal of Experimental Psychology*, *69*, 135–143.

Teghtsoonian, R., & Teghtsoonian, M. (1978). Range and regression effects in magnitude scaling. *Perception & Psychophysics*, *24*, 305–314.

Thaler, R. (1981). Some empirical evidence on dynamic inconsistency. *Economics Letters*, *8*, 201–207.

Thomas, P.M., FitzGibbon, L., & Raymond, J.E. (2016). Value conditioning modulates visual working memory processes. *Journal of experimental psychology. Human perception and performance*, *42*, 6–10.

Tolkmitt, F., & Christ, R.E. (1970). Differential payoff and locus of selectivity in a visual memory task. *Perception & Psychophysics*, *7*, 47–50.

Tomassini, A., Ruge, D., Galea, J.M., Penny, W., & Bestmann, S. (2016). The role of dopamine in temporal uncertainty. *Journal of Cognitive Neuroscience*, *28*, 96–110.

Treisman, A.M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, *12*, 242–248.

Van den Berg, R., & Ma, W.J. (2018). A resource-rational theory of set size effects in human visual working memory. *ELife*, *7*, e34963.

Wallis, G., Stokes, M.G., Arnold, C., & Nobre, A.C. (2015). Reward boosts working memory encoding over a brief temporal window. *Visual Cognition*, *23*, 291–312.

Westbrook, A., & Braver, T.S. (2015). Cognitive effort: a neuroeconomic approach. *Cognitive, Affective, & Behavioral Neuroscience*, *15*, 395–415.

Woodford, M. (2009). Information-constrained state-dependent pricing. *Journal of Monetary Economics*, *56*, S100–S124.

Wyart, V., Nobre, A.C., & Summerfield, C. (2012). Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. *Proceedings of the National Academy of Sciences*, *109*, 3593–3598.

Xiang, Y., Graeber, T., Enke, B., & Gershman, S.J. (2021). Confidence and central tendency in perceptual judgment. *Attention, Perception, & Psychophysics*, *83*, 3024–3034.

Open practices statement All of the data and code are available at https://github.com/sjgershm/uncertainty-control. The experiment was not preregistered.