

## **Reconciling Novelty and Complexity through a Rational Analysis of Curiosity**

Rachit Dubey

Department of Computer Science,  
Princeton University

Thomas L. Griffiths

Department of Psychology,  
Princeton University

accepted at *Psychological Review*

Corresponding author: Rachit Dubey Email: [rdubey@princeton.edu](mailto:rdubey@princeton.edu)

### Author Note

We thank Akshat Dave, Fred Callaway, Jordan Suchow, Mark Ho, Shaun O' Grady, and the anonymous reviewers for helpful discussions and comments. A preliminary version of the model and its evaluation (Experiment 1) was presented at the 39th Annual Meeting of the Cognitive Science Society and appeared in the conference proceedings. This work has been updated and extended for inclusion in the current article.

## Abstract

Curiosity is considered to be the essence of science and an integral component of cognition. What prompts curiosity in a learner? Previous theoretical accounts of curiosity remain divided – novelty-based theories propose that new and *highly* uncertain stimuli pique curiosity whereas complexity-based theories propose that stimuli with an *intermediate* degree of uncertainty stimulate curiosity. In this article, we present a rational analysis of curiosity by considering the computational problem underlying curiosity, which allows us to model these distinct accounts of curiosity in a common framework. Our approach posits that a rational agent should explore stimuli that maximally increase the usefulness of its knowledge and that curiosity is the mechanism by which humans approximate this rational behavior. Critically, our analysis shows that the causal structure of the environment can determine whether curiosity is driven by either highly uncertain or moderately uncertain stimuli. This suggests that previous theories need not be in contention but are special cases of a more general account of curiosity. Experimental results confirm our predictions and demonstrate that our theory explains a wide range of findings about human curiosity, including its subjectivity and malleability.

*Keywords:* curiosity, exploration, rational analysis, computational model

### **Reconciling Novelty and Complexity through a Rational Analysis of Curiosity**

In 1928, upon returning from a vacation, Alexander Fleming, who was a professor of Bacteriology at St. Mary's Hospital in London, noticed how a mold floating in one of his dirty petri dishes held the surrounding bacteria at bay. This peculiar event led him to develop a hypothesis that would be a prelude to the development of penicillin. The history of science abounds with incidents in which an event piqued the curiosity of a scientist thereby leading to important discoveries (other examples include Curie, Faraday, and Planck). For this reason, intellectual curiosity has long been recognized as the essence of science. In fact, Herbert Simon famously titled a 1992 talk given at Carnegie Mellon as 'The cat that curiosity couldn't kill' and described curiosity to be not only the beginning of all science, but also its end (Gobet, 2016). Curiosity is not just integral to scientists. It is a basic element of cognition from early on in life, and has been linked to better learning (Von Stumm et al., 2011; Gruber et al., 2014), memory (Kang et al., 2009), and decision-making (Pierce et al., 2005; Law et al., 2016). Thus, a better understanding of curiosity will have implications for science, pedagogy, and our understanding of human cognition.

Considering how important curiosity is to scientific discoveries and many other aspects of cognition, it is surprising that our understanding of curiosity as a psychological phenomenon remains quite limited (Gottlieb et al., 2013; Kidd & Hayden, 2015; Simon, 2001). Several theories have been proposed to describe the mechanisms behind curiosity and they generally fall under two different theoretical approaches. The first approach includes *novelty-based* theories, which link curiosity with novelty by hypothesizing that gaining information about novel stimuli is intrinsically rewarding (Berlyne, 1950; Lehman & Stanley, 2011). The second approach includes *complexity* theories, which posit that curiosity is driven by stimuli that are at an intermediate level of complexity i.e. neither overly simple nor overly complex (Berlyne, 1960; Loewenstein, 1994; Schmidhuber, 1991; Oudeyer et al., 2007). These theories predict that curiosity is a U-shaped pattern of preference for stimulus complexity (Kang et al., 2009; Baranes et al., 2014).

A large body of research has demonstrated the strengths of each of these theories and shown

evidence of curiosity being influenced by both novelty and stimulus complexity (Ranganath & Rainer, 2003; Wittmann et al., 2007; Kang et al., 2009). However, both these theories make different predictions about curiosity and it is not yet clear when and why is curiosity driven by novelty and not stimulus complexity (and vice-versa). Thus, to date, the field has not reached a consensus on a unified or an integrated view of curiosity (Kidd & Hayden, 2015; Coenen et al., 2017).

In this article, we present a formal framework that unifies previously distinct theories of curiosity and explains how human curiosity operates across a wide range of contexts. In contrast to previous theories, we do not provide a mechanistic explanation in terms of psychological processing steps or neural machinery. Rather, we provide a functional characterization of curiosity by performing a rational analysis. Our explanation takes the form of a computational-level theory, in the sense introduced by Marr (1982) and pursued by Anderson (1990) among others (Chater & Oaksford, 1999; Griffiths & Tenenbaum, 2005).

We start by identifying the abstract computational problem addressed by curiosity and derive an optimal solution to that problem. In other words, to begin with, we do not attempt to explain what curiosity does but rather first explain what curiosity *should* do. Our approach posits that a rational agent should explore stimuli that maximally add to the value of its knowledge. Depending on the structure of the environment, the stimuli that maximize this value can either be ones that are completely novel or that are of intermediate complexity. We then use this rational solution to explain human behavior, positing that curiosity may be a mechanism by which people come to act in accordance with this solution. This perspective helps to explain when people's curiosity should be influenced by novelty or stimulus complexity. As a consequence, our rational analysis serves as a useful *tool* to understand the relationship between previous distinct theories of curiosity that have dominated previous work and suggests a way to unify them into a single framework. From the perspective of our framework, these apparently disparate factors influencing curiosity are not discrete cases requiring separate explanations but rather points on a continuum.

The plan of this paper is as follows. First, we review different definitions of curiosity that

have existed within the literature, both to provide background and to establish our viewpoint. We then review past models of curiosity, dividing them into novelty-based and complexity-based approaches. Next, we introduce our rational model of curiosity and show how our model offers support to previous accounts of curiosity under different circumstances thereby unifying them in a single model. We then conduct behavioral experiments to test our model's predictions and evaluate how our model accords with human curiosity. We close with a discussion of the significance and implications of our results.

### **Background**

We begin by reviewing two strands of research that have primarily concerned researchers studying curiosity – defining curiosity and studying the factors that influence curiosity. For a comprehensive review, we refer the reader to Baldassarre and Mirolli (2013), Kidd and Hayden (2015), Gottlieb et al. (2013), and Gottlieb and Oudeyer (2018).

#### **What is curiosity?**

The task of defining curiosity has turned out to be a surprisingly difficult endeavor, as is evident from the lack of a single widely accepted definition of curiosity. One of the earliest and most influential accounts of curiosity was laid out by Daniel Berlyne who classified curiosity into four different forms (Berlyne, 1966). The first form, 'epistemic curiosity', is the desire for information and knowledge. The second form, 'perceptual curiosity', is curiosity driven by novel objects and stimuli. The third form, 'specific curiosity' is the desire for a particular piece of knowledge such as the final piece of a puzzle or knowledge about a particular question. Finally, the fourth form, 'diversive curiosity', is less directed and simply stimulation seeking to escape boredom. While these distinctions seem intuitive, researchers note that the definitions laid out by Berlyne are not very straightforward as they are rather convergent (Litman & Spielberger, 2003). For example, curiosity can be perceptual-specific (attending to novel features to aid problem solving) or perceptual-diversive (attending to novel stimuli to avoid boredom) or epistemic-specific (reading a specific book that one enjoys) and even epistemic-diversive (reading about random topics

to avoid boredom). Modern researchers define curiosity as information-seeking that is motivated by an intrinsic drive (Loewenstein, 1994; Oudeyer et al., 2007). This view of curiosity is distinct from traditional information-seeking which can be both internally (studying for the sake of pleasure or enjoyment) or externally (such as studying for an exam for sake of better grades) motivated . However, viewing curiosity in the light of intrinsic or extrinsic motivation leads us to a challenging philosophical problem of understanding a decision maker's mental states and delineating what is extrinsic and what is intrinsic (Kidd & Hayden, 2015; Mišćević, 2016).

Given the ambiguity and lack of any universally agreeable definition of curiosity, we here adopt a broader view of curiosity. Like prior researchers, we consider curiosity as information-seeking but we are agnostic to whether it is driven by intrinsic or extrinsic motivations. Some researchers have also advocated for adopting a similar viewpoint (Kidd & Hayden, 2015; Murayama et al., 2018). Taking this approach helps us to better understand and relate different strands of curiosity research into a common framework. Further, this is also consistent with our aim of providing a functional explanation of curiosity as drives (intrinsic or extrinsic) are mechanisms rather than functions.

### **What causes curiosity?**

A number of theories have been proposed in the past to describe the psychological processes involving curiosity. These theories can be broadly organized into two classes: *novelty-based* and *complexity-based*. In this section, we describe these theories in brief, and provide their individual strengths and weaknesses.

**Curiosity based on novelty.** Several psychological theories have linked curiosity with novelty by hypothesizing that gaining information about novel stimuli is intrinsically rewarding (Berlyne, 1950). Berlyne described this as a driving force that motivates an organism to seek out novel stimuli which diminishes with an increase in exposure. Novelty-seeking has also been supported by various neuroscientific studies that show that novel stimuli activate reward-responsive areas in the brain (Ranganath & Rainer, 2003; Wittmann et al., 2007, 2008; Düzel et al., 2010).

Recent computational models and reinforcement learning algorithms have also demonstrated the use of novelty as an efficient exploration strategy in various domains (Lehman & Stanley, 2011; Tang et al., 2017). However, a severe limitation of novelty-based theories is that they assume that it is optimal for an individual to explore novel stimuli in all environments. A novel stimulus doesn't necessarily mean that it contains information that is useful or generalizable to an individual. This is also pointed out by previous studies that show that exploration based only on novelty could lead agents to be trapped in unlearnable situations (Gottlieb et al., 2013). Further, as pointed by several researchers, novelty theories also cannot account for people's familiarity preferences and their tendency to sometimes avoid highly uncertain situations (Loewenstein, 1994; Kidd & Hayden, 2015).

**Curiosity based on stimulus complexity.** An alternative perspective on curiosity is provided by theories based on the idea that curiosity is a function of stimulus complexity, that people's curiosity is driven by stimuli that are neither too simple nor too complex. Thus, people's curiosity is drawn towards optimally incongruent events (Berlyne, 1960). One such theory is Loewenstein's *information-gap* hypothesis (Loewenstein, 1994) which proposes that curiosity arises whenever an individual has a gap in information prompting it to complete its knowledge and resolve the uncertainty. Thus, curiosity peaks when one has a small amount of knowledge but it diminishes when one knows too little or too much about the stimulus. A number of studies have supported this prediction and showed that curiosity is an inverted U-shaped function of confidence, with people showing the highest curiosity for topics that they are moderately confident about (Kang et al., 2009; Law et al., 2016). While the information-gap hypothesis has considerable strengths, it also faces several limitations. For example, the theory is constrained in that an individual can only be curious about stimuli in known contexts. Thus, if an individual has no prior knowledge about stimuli in the environment then it is not clear how curiosity will function in that environment (as one will not be curious about any stimuli in an environment where it has no prior knowledge). Further, this theory also can not explain why people's curiosity is often influenced by novel stimuli in the environment (as one has little confidence and larger gaps of information with respect to novel

stimuli).

Another recent theory based on the idea of stimulus complexity is the *learning-progress* hypothesis which proposes to model optimal incongruity via prediction errors. According to the learning progress hypothesis, curiosity is guided by the hypothesis that learning progress generates intrinsic reward (Schmidhuber, 1991, 2010; Oudeyer et al., 2007). This hypothesis proposes that the brain is intrinsically motivated to pursue tasks in which one's predictions are always improving. Thus, an individual will not be interested in tasks that are too easy or too difficult to predict but will rather focus on tasks that are learnable. A number of papers in developmental robotics and reinforcement learning have supported this idea by operationalizing learning progress using information-theoretic measures such as surprise, uncertainty, information-gain etc., and showed that based on learning progress, an agent can efficiently explore in an unknown environment (Oudeyer & Kaplan, 2006; Oudeyer et al., 2007; Sun et al., 2011; Pathak et al., 2017). However, despite its strength, this theory is again limited in certain situations. For example, at any given point of time, there exist many stimuli in the environment for which an agent can reduce its prediction error and make learning progress. In those cases, a purely learning progress driven agent may be sub-optimal as the agent will not prioritize stimuli based on usefulness or cost of learning.

### **Empirical findings**

A number of studies have been conducted in the past that lend empirical support to either novelty-based or complexity-based theories. In this section, we briefly review four major empirical findings on curiosity. Later, we will use these findings to evaluate our own approach to explaining and understanding curiosity.

**1) Curiosity is influenced by novelty.** Across a series of experiments, Berlyne showed that rats and humans alike have a tendency to be more curious about stimuli that are novel to them (Berlyne, 1950, 1966).<sup>1</sup> Smock and Holt (1962) confirmed Berlyne's findings in another series of experiments and showed that young children tend to explore and play with novel toys much longer

---

<sup>1</sup> Here, novelty is defined to be something unlike anything encountered before.



than familiar toys. We note that this finding is in line with novelty-based theories but incompatible with complexity-based theories.

**2) Curiosity is an inverted U-shape function of confidence.** A number of studies have explored the effect of confidence on curiosity and found that people's curiosity is highest for stimuli for which they are moderately confident about (Kang et al., 2009; Baranes et al., 2015). For example, Kang et. al. (2009), asked participants to rate their curiosity for knowing the answers to different trivia questions and found that people were most curious about trivia questions for which they had a moderate amount of confidence. Other studies have also demonstrated this effect of confidence on curiosity using similar experimental paradigms (Baranes et al., 2015; Marvin & Shohamy, 2016). While this finding is in alignment with complexity-based theories such as the information-gap hypothesis, novelty-based theories cannot explain this finding.

**3) Curiosity is highest for moderately difficult tasks.** In recent years, researchers have also studied the influence of task difficulty on curiosity (Baranes et al., 2014; Geana et al., 2016). Similar to the effect of uncertainty, these studies demonstrate that in the absence of any external rewards, people prefer to spend most time exploring tasks that are moderately difficult. This finding is in accordance with complexity-based theories such as the learning-progress hypothesis but not novelty-based theories.

**4) Curiosity increases progressively.** In a recent study, Baranes et. al., (2014) showed that in a free play task, people first explore easier tasks and then with an increase in time and exposure, people gradually switch from exploring easier tasks to progressively harder tasks. This finding is again in accordance with complexity-based theories such as the learning-progress hypothesis but not novelty-based theories.

### **Summary and prospectus**

*Novelty-based* and *complexity-based* theories make different commitments about the psychological factors that relate to human curiosity and make a compelling case for an influence of these different factors on curiosity. However, due to their orthogonal predictions, it is not yet clear

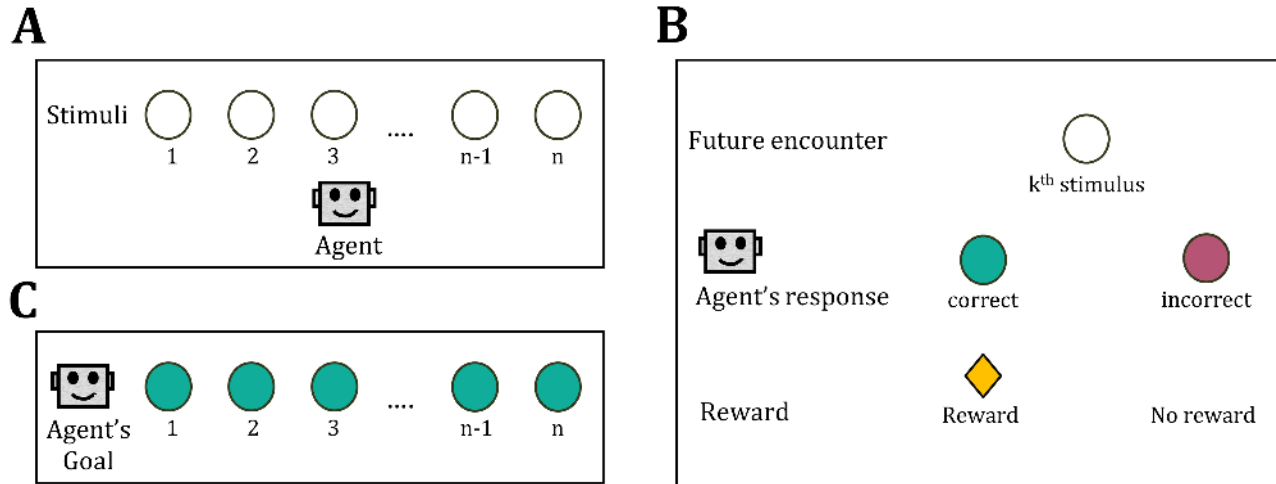
exactly how and when curiosity is influenced by one of these factors and not the other.

Despite differences in their predictions, a common theme of these models is that they all provide a mechanistic explanation of curiosity and describe *how* curiosity functions and how it relates to different psychological factors. However, none of these theories adequately explain *why* curiosity is influenced by these factors in the first place. In other words, what is the function of curiosity and why does curiosity work the way it does? Previous theories of curiosity are vague at this *ultimate* (Tinbergen, 1963) or *computational* (Marr, 1982) level of analysis.

To address this challenge, we take a fundamentally different view, focusing on the abstract problem of curiosity and the form that the optimal solution to that problem should take, rather than the process. Our approach supports past theories rather than supplanting them, and suggests that previous theories need not be in contention but are all special cases of a more general account of curiosity. As we will describe in the next section, our work provides a way to understand and unify novelty and complexity based approaches in one common framework.

### **A Rational Analysis of Curiosity**

We now provide a rational analysis of curiosity, in the spirit of Marr (1982) and Anderson (1990). Marr (1982) proposed three levels at which any information processing system should be understood: the levels of computational theory, representation and algorithm, and hardware implementation. Analyses at the computational level answer the question “What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?” (Marr (1982), p. 25). In the context of curiosity, a computational-level analysis seeks to identify the abstract problem being solved by curiosity and to understand the logic that makes it possible to solve this problem. Instead of explaining what curiosity does, our analysis here serve to explain what curiosity *should* do. However, in pursuing a rational analysis, we are not trying to make claims about the rationality of human curiosity, nor are we making commitments to the mechanisms or the psychological processes involved in curiosity. In fact, our framework is not strictly normative – it also suggests how people may become curious about seemingly sub-optimal



*Figure 1. The computational problem.* (A) An agent is in an environment with  $n$  stimuli. (B) Suppose the agent encounters the  $k$ th stimulus in the future (top row). If the agent produces the correct response to the stimulus (middle row, left), then it receives a reward (bottom row, left). If the agent produces the incorrect response (middle row, right), it receives no reward (bottom row, left). (C) Given this, to maximize rewards, the agent wants to learn the correct response for as many stimuli as possible.

tasks and we briefly outline some possible psychological mechanisms in the General Discussion. In this way, we expect that our approach of studying what a rational agent should do will be a valuable tool in understanding how people's curiosity actually works. We now begin our analysis by considering the abstract computational problem underlying curiosity and then formally deriving an optimal solution to this problem.

### Computational problem underlying curiosity

Consider the following problem. An agent is in an environment with  $n$  stimuli (refer to Figure 1(a)). Suppose that in the future, the agent encounters some of these  $n$  stimuli again. In that future encounter, if the agent produces the appropriate response to those stimuli, then the environment provides a reward to the agent (Figure 1(b)). Given this, the goal of the agent is to maximize future rewards by learning the correct response for every stimulus (Figure 1(c)). In other

words, the agent wants to know as much as possible about the reward structure of the world.<sup>2</sup>

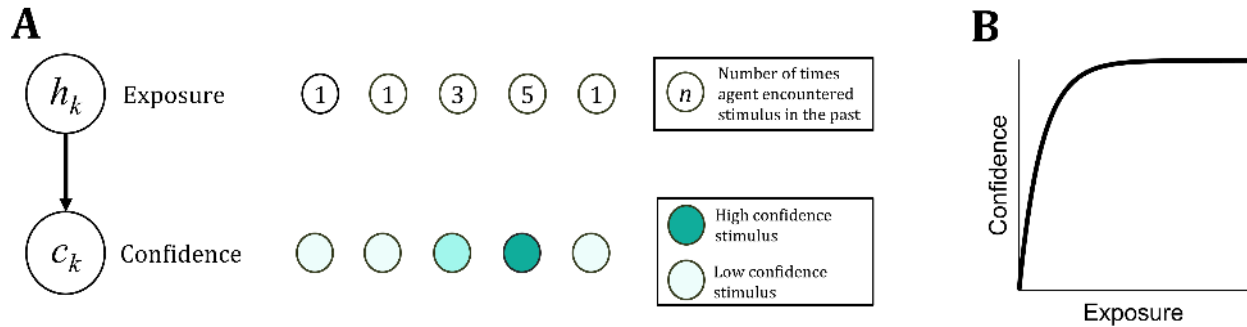
Unfortunately, due to resource and computational constraints, the agent cannot learn the correct response for every stimulus in the environment (Simon, 1990; Gabaix et al., 2006; Pierson & Goodman, 2014; Griffiths et al., 2015). Thus, the agent faces the problem of deciding what to explore in the environment in order to maximize its knowledge (and subsequent rewards) in the future.<sup>3</sup>

How should the agent solve this problem? A crude way to do so would be to explore various stimuli in the environment using random exploration. However, this approach is unlikely to be optimal in complex environments that contain a large number of stimuli. One way for the agent to improve its performance is to align its exploration with respect to future occurrence of various stimuli in the environment. For instance, if the agent will never encounter some stimuli in the future, then the agent should not explore those stimuli (as knowing about those stimuli will not help in maximizing rewards in future). To account for this, the agent computes the “probability of

---

<sup>2</sup> Note that for the sake of simplicity, we treat each stimulus as being independent (i.e. knowledge of one stimulus is not correlated with knowledge of other stimuli). However, our analysis can be easily extended to cases where the knowledge of one stimulus is correlated with other stimuli (i.e. when learning about one stimulus provides information about other stimuli). Furthermore, our analysis also treats different stimuli as having the same reward magnitude. Our analysis can again be extended to cases where different stimuli (and subsequently pieces of information) have varying reward values. While people undoubtedly have to solve more complex problems (for example, problems in which sequential dependencies hold between current and future actions) we focus on this case for its simplicity and its resemblance to the tasks typically used in experiments exploring curiosity.

<sup>3</sup> As an aside, we note that the information-seeking problem described here is similar to the multi-armed bandit problem (Daw et al., 2006; Cohen et al., 2007; Bubeck et al., 2012). However, multi-armed bandit problems face a tension between exploiting a familiar option for a known reward and exploring unfamiliar options for unknown rewards – the so called explore/exploit trade-off. In contrast, there is no such tension in our version of the problem – the agent’s goal is to simply explore, i.e. obtain information about the correct response for as many arms as possible. This information can then be exploited by the agent in the future to maximize rewards (as more information will help agent produce the appropriate response to various stimuli resulting in higher rewards in future). We refer the reader to Appendix A where we expand on the similarities and differences between these two problems in detail.



**Figure 2. Relationship between exposure  $h$  and confidence  $c$ .** (A) If the agent has encountered a stimulus often in the past (top row), then it would have a higher confidence value  $c$  (bottom row) for that stimulus (and vice-versa). (B) We assume that the confidence  $c$  of the agent increases at a decreasing rate with respect to exposure  $h$ .

occurrence”,  $p_k$  i.e. the probability that the  $k$ th stimulus will occur in the future (Anderson, 1990).<sup>4</sup> However, simply making decisions based on probability of occurrence will not be enough – the agent should also align its exploration with respect to its current knowledge. For example, the agent should avoid exploring stimuli for which the agent already knows the correct response (as the agent will definitely get rewards for those stimuli in future and it should aim to maximize knowledge for other stimuli). In line with this, for each stimulus  $k$  in the environment, the agent computes the probability it knows the correct response to that stimulus. The higher this probability is, the higher the chance of agent making the correct response to that stimulus and obtaining rewards in future. We call this probability the agent’s “confidence value”,  $c_k$ . In other words, confidence,  $c_k$  is simply the probability computed by the agent that if presented with the  $k$ th stimulus it will respond appropriately. Note that, the agent’s confidence  $c_k$  is directly influenced by the number of times the agent has encountered the  $k$ th stimulus in the past. That is, if the agent has encountered a stimulus often in the past, then it would have a higher confidence value for that stimulus (and vice-versa). We call this number the agent’s “exposure”,  $h_k$ . We here assume that confidence,  $c$  increases at a

<sup>4</sup> Note that the probability of occurrence  $p$ , is determined by the environment itself. Further,  $p$  is an *estimate* of the true probability computed by the agent and hence is prone to errors and can have individual variations in its computation – we expand on this further in the General Discussion.

decreasing rate with respect to exposure,  $h$  (refer to Figure 2).

For convenience, we describe the relationship between  $c$  and  $h$ , a kind of a learning rate, by a bounded growth function,

$$c_k = 1 - e^{-\gamma_k \cdot h_k}. \quad (1)$$

However, our qualitative predictions hold for any monotonically increasing function. Here,  $\gamma_k$  is a constant value that denotes the learning rate i.e. how quickly can the agent learn about the  $k$ th stimulus. This learning rate  $\gamma$  is different for each stimulus and thus confidence increases at different rates with respect to exposure for different stimuli. Stimuli with low learning rates would be harder to learn, as the confidence of the agent will increase slowly with respect to exposure. Stimuli with high learning rates are easier to learn, as the confidence of the agent will increase sharply with respect to exposure.<sup>5</sup>

**Value of knowledge.** Given estimation of  $p$  and  $c$  for every stimulus, the agent then computes the value of its overall knowledge. If each stimulus was equally likely to occur in the future, then the value of the agent's knowledge would simply be the average of its current knowledge i.e. confidence  $c$ . However, since some stimuli are more likely to occur in the future than others, the value, denoted as  $V$ , is a function of the probability of occurrence  $p$  and the confidence factor  $c$  and is given as follows:

$$V = \sum_k p_k \cdot c_k. \quad (2)$$

According to this equation, the value of an agent's knowledge is the sum over all stimuli of the probability that a stimulus will occur in the future multiplied by the probability that the agent knows how to react to that stimulus. This is not a mathematical assumption but simply a consequence of the structure of the problem – the value of knowledge is simply the chance of successfully responding to the next stimulus computed by summing over all stimuli in the environment.

**The problem.** Recall that the goal of the agent is to decide what to explore in the environment in order to maximize its knowledge and subsequently rewards in the future. In other

---

<sup>5</sup>  $\gamma_k$  is again an *estimate* of the true learning rate computed by the agent. This can be estimated using Bayesian inference and hence will be influenced by the prior on the learning rate as well as the kind of the data the agent receives.

words, the goal of the agent is to increase the value of its current knowledge  $V$  and the agent can do that by taking actions to increase  $h$  for the various stimuli in the environment (as that will increase  $c$  and in turn increase the value  $V$ ). Thus, the computational problem faced by the agent is choosing the appropriate action i.e. deciding which stimulus  $k$  to increase exposure  $h_k$ , in order to increase the value  $V$  which in turn will help the agent maximize future rewards.

**Deriving an optimal solution.** To solve the problem of choosing which stimulus to explore further, the agent can evaluate the change in  $V$  as it explores each stimulus in the environment. Thus, for every stimulus  $k$  in the environment, the agent can compute the change in its knowledge that would result from exploring that stimulus. The agent can then use that to select which stimulus to explore first – the stimulus that causes the largest increase in the overall value should be explored first. This is because exploring the stimulus that causes the largest increase in the overall value implies that the agent would end up exploring the stimulus that increases the reward magnitude maximally. Therefore, it would be optimal for the agent to explore the stimulus that causes the largest increase in the overall value. The agent can perform this computation simply by differentiating  $V$  with respect to  $h_k$ ,

$$\frac{dV}{dh_k} = p_k \cdot \frac{dc_k}{dh_k}. \quad (3)$$

**Optimal solution and curiosity.** An agent operating according to the above model will explore stimuli that have a high rate of change of value with respect to exposure associated with them. Maximizing this rate of change, i.e.  $\frac{dV}{dh_k}$ , is the *optimal* behavior under which the agent will explore stimuli to maximize its knowledge and subsequent rewards. This rational solution i.e. maximizing rate of change is the behavior which the agent *should* adapt. We postulate that the sense of curiosity is the mechanism that produces and drives this behavior in humans. Under this view, we take the rate of change  $\frac{dV}{dh_k}$  to be approximately the curiosity the agent has for knowing about the  $k$ th stimulus, which we denote as  $\Omega_k$ , assuming that the agent will explore the stimuli in the environment that it is most curious about.<sup>6</sup> Thus, according to our framework, the function of

<sup>6</sup> Since we expect curiosity  $\Omega$  to be the mechanism that approximates the optimal solution, human curiosity doesn't necessarily need to be always optimal and will be prone to errors – we expand on this in the General Discussion.

curiosity is to simply help the agent to achieve its goal of maximizing rewards in the future. Under the choice of the form of  $c_k$  given in Equation 1, we calculate this derivative as follows:

$$\Omega_k \approx p_k \cdot \frac{d(1 - e^{-\gamma_k \cdot h_k})}{dh_k}. \quad (4)$$

Upon differentiation, we get the relationship of curiosity  $\Omega_k$  with respect to probability of occurrence  $p_k$ , learning rate  $\gamma_k$ , and exposure  $h_k$  as follows:

$$\Omega_k \approx \gamma_k \cdot p_k \cdot e^{-\gamma_k \cdot h_k}. \quad (5)$$

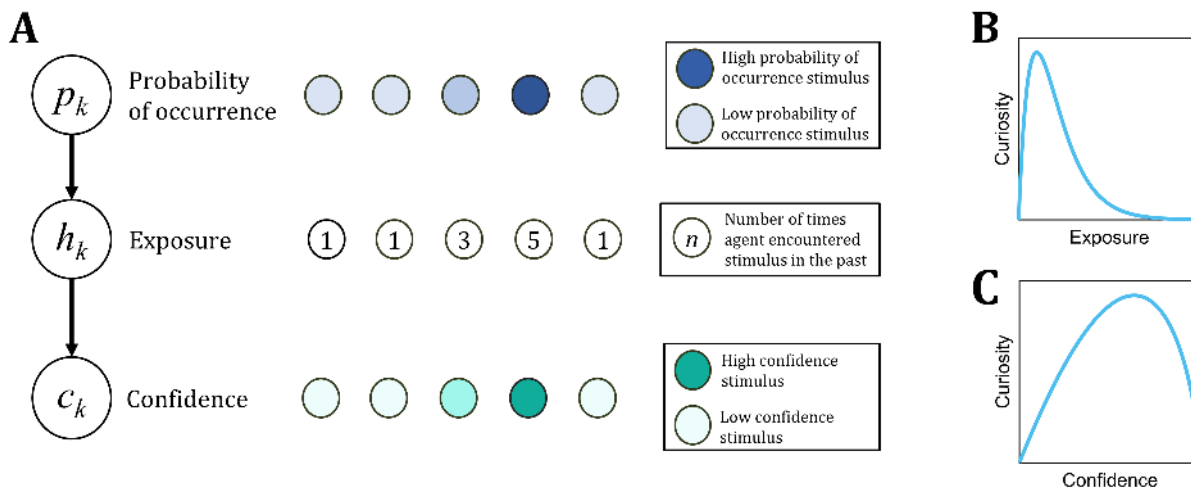
## Summary

In this section, we performed a rational analysis of curiosity in which we considered an agent whose goal was to know as much as possible about the reward structure of the world to maximize its expected rewards. We suggested that the agent should consider the future occurrence of various stimuli in the environment (which we termed ‘probability of occurrence’,  $p$ ) as well as its current knowledge about various stimuli in the environment (‘confidence’,  $c$ ), and use those estimates to compute the ‘value of knowledge’  $V$ . Note that the confidence  $c$  in turn was determined by number of times the agent encountered a stimulus in the past (‘exposure’,  $h$ ) as well as rate at which the agent could increase confidence of a stimulus (the ‘learning rate’,  $\gamma$ ). To maximize the expected rewards, according to our framework, the agent should simply maximize the value of knowledge  $V$ . The optimal solution is to compute the rate of change of  $V$  w.r.t exposure  $h$  and then increase exposure  $h$  for the stimulus that maximally increases  $V$ . This optimal solution is the behavior which a rational agent should adopt to maximize rewards. We suggest that curiosity is the mechanism by which people *approximate* this rational behavior. As we will show in the next section, studying curiosity under this lens can help to explain various empirical findings about human curiosity.

## Reconciling Novelty and Complexity-based Theories

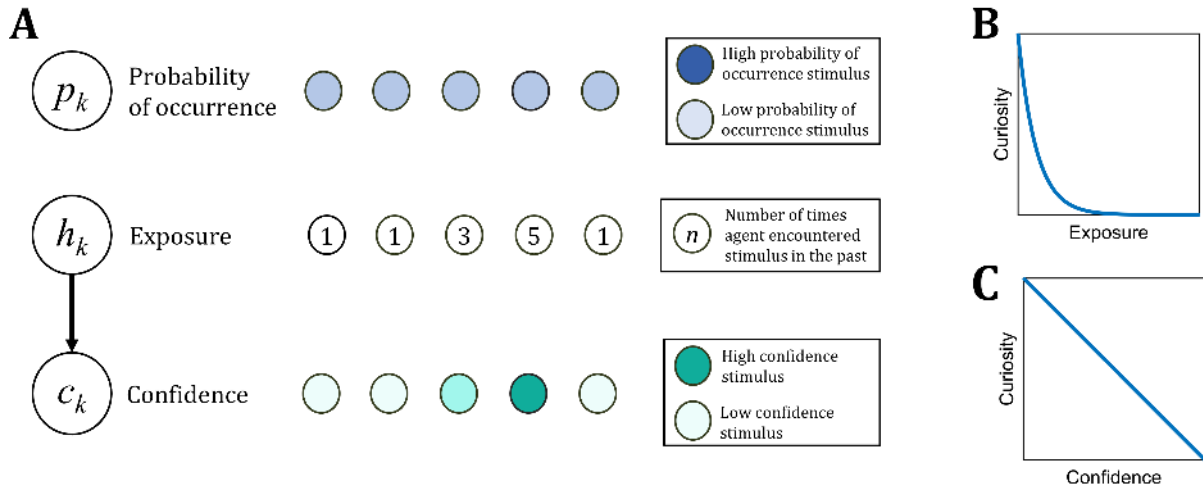
Having a formal account of curiosity, we now describe how our model relates to previous theories of curiosity. First, we note that in our framework, there is no fixed assumption about the





**Figure 3. When past and future are correlated, curiosity is highest for stimuli with moderate exposure and confidence.** (A) Graphical depiction of an environment where probability of occurrence  $p_k$  is related to exposure  $h_k$  and subsequently confidence  $c_k$ . Here, the probability that the agent will encounter a stimulus in the future (top row) determines the number of times the agent is exposed to the stimulus (middle row) which in turn determines the confidence of the agent about that stimulus (bottom row). (B) Model prediction about the relationship between curiosity and exposure in this environment. (C) Model prediction about the relationship between curiosity and confidence in this environment. In this environment, the agent's curiosity is an inverted U-shape function of exposure and confidence.

form of the environment. We consider two different forms of environmental structure. The first form is an environment where  $p_k$  is related to  $h_k$  (as described in Figure 3(a)). In this environment, stimuli frequently encountered by an agent are more likely to be encountered in the future. Thus, the probability that the agent will encounter a stimulus in the future determines the number of times the agent is exposed to the stimulus which in turn determines the confidence of the agent in knowing that stimulus. The second form is an environment where  $p_k$  and  $h_k$  are independent of each other (as described in Figure 4(a)). In this environment, the agent can encounter any stimulus in the future regardless of previous occurrence. These are obviously two extremes – perfect correlation of the past and the future and perfect independence – but they allow us to explore



*Figure 4. When past and future are independent, curiosity is highest for stimuli with lowest exposure and confidence.* (A) Graphical depiction of an environment where probability of occurrence  $p_k$  is independent of exposure  $h_k$  and confidence  $c_k$ . Here, the agent can encounter any stimulus in the future (top row) regardless of previous occurrence (middle row). (B) Model prediction about the relationship between curiosity and exposure in this environment. (C) Model prediction about the relationship between curiosity and confidence in this environment. In this environment, the agent's curiosity decreases with an increase in exposure and confidence.

correspondingly extreme forms of curiosity.

**Complexity-based theories.** When the agent is in an environment where  $p_k$  and  $h_k$  are related to each other (as in Figure 3(a)), then  $p_k$  is proportional to  $h_k$  and the relationship between curiosity and exposure given in Equation 5 reduces to

$$\Omega_k \propto \gamma_k \cdot h_k \cdot e^{-\gamma_k \cdot h_k}. \quad (6)$$

Subsequently, using Equation 1, confidence will be related to curiosity as

$$\Omega_k \propto [-\log(1 - c_k)] \cdot (1 - c_k). \quad (7)$$

Interestingly, this relationship between curiosity and confidence is very similar to the one described by the information-gap hypothesis (Loewenstein, 1994). Loewenstein used Shannon's (1948)

entropy formula to describe the relationship between curiosity and confidence:

$$\Omega_k = [-\log(c_k)].c_k. \quad (8)$$

Both the information-gap theory (Equation 8) and our model (Equation 7) predict an inverted U-shape relationship between curiosity and confidence (empirical finding 2). When an agent is in an environment where probability of occurrence is related to exposure, our rational model thus parallels the information-gap theory.

Intuitively, this makes sense. If the agent is in an environment where frequently encountered stimuli are more likely to appear in the future, then to maximize rewards, the agent should explore stimuli for which it has moderate exposure and confidence. This is simply because the stimuli for which the agent has little exposure are very unlikely to occur in the future, and while the stimuli for which the agent has high exposure have a high probability of occurring in the future, the agent already has high confidence for those stimuli.

Our model also relates to the learning progress theory (Oudeyer et al., 2007) in this kind of environment. According to the learning progress hypothesis, an agent is intrinsically motivated to pursue tasks in which predictions are constantly improving, thereby avoiding boring or extremely complicated tasks. An agent operating under this model ends up exploring stimuli of ‘intermediate complexity’. If we consider the case where various stimuli within the environment have different learning rates  $\gamma_k$ , then according to Equation 6, the agent will be most curious about stimuli that have a moderate learning rate (empirical finding 3). This is because the agent will first explore stimuli that have a high learning rate (i.e. are easier to learn), but with increased exposure, the agent’s curiosity would be primarily driven towards stimuli that have a moderate learning rate. Stimuli that are very hard to learn (i.e. low learning rate) would induce the lowest curiosity. Thus, in this environment, an agent that aims to maximize its knowledge behaves similarly to an agent whose curiosity is driven by learning progress.

**Novelty based curiosity.** Under theories that are based on curiosity driven by novelty, an agent is most curious about stimuli that it is least confident about:

$$\Omega_k = 1 - c_k. \quad (9)$$

According to our rational model, when the agent is in an environment where  $p_k$  and  $h_k$  are independent of each other (as described in Figure 4(a)), the relationship between curiosity and exposure will be the one described in Equation 5, where  $p_k$  is simply a constant value. Thus, curiosity is highest when exposure is lowest and it decreases as exposure increases i.e. curiosity is highest for novel stimuli (similar to empirical finding 1 discussed above). The relationship between curiosity and confidence can be rewritten as

$$\Omega_k \approx \gamma_k \cdot p_k \cdot (1 - c_k). \quad (10)$$

If  $p_k$  and  $\gamma_k$  are equal for all  $k$ , this reduces to Equation 9.<sup>7</sup> Thus, when the probability of occurrence and exposure are not related to each other, our rational model is similar to the previously proposed novelty-based theories of curiosity.

Intuitively, we can interpret this as follows. If the agent is in an environment where any stimuli can appear in the future regardless of past experience, then to maximize the value of its knowledge, the agent should explore novel stimuli, i.e. stimuli to which it has least exposure, as these stimuli are equally likely to appear in the future and the agent currently does not have any knowledge about them. The expected value of knowledge is increased most rapidly by learning more about the stimuli with least exposure and lowest confidence.

### **Intuitions and predictions**

Whereas previous theories associated curiosity with factors such as novelty and complexity our model shows that, depending on the structure of the environment, an agent's curiosity can be driven by any of these factors. In this way, our rational model allows to bridge previous theories related to curiosity in a single framework. In an environment where the probability of occurrence and exposure are directly related, our rational model makes predictions corresponding to complexity-based theories of curiosity. In an environment where probability of occurrence and exposure are not related, our model is akin to novelty-based theories of curiosity.

---

<sup>7</sup> This is also the case if the agent does not have information about  $\gamma_k$  and  $p_k$ , so the *expected* learning rate and probabilities are the same.

As an illustration, consider an individual living in San Francisco. Presumably, on a given day, this individual would be more likely to be curious about the weather of San Francisco compared to the weather of a novel city such as Kakamega. This is perhaps likely driven by an assumption being made by the individual that their past and future experiences are correlated with each other and therefore the probability of requiring information about the weather of Kakamega will be very low. Now, consider a scenario wherein the same individual has to move to Kakamega for a new job opportunity. In this setting, the individual might be more likely to be curious about the weather of this novel city. This is partly due to the fact that in this new scenario, the individual's past experience and future probability of occurrence are very weakly correlated with each other resulting in the individual to be more curious about the novel city (Kakamega). While the two environment structures considered here are obviously two extremes, this can be a useful tool to understand people's curiosity in different settings.

In the following experiments, we test our model predictions by evaluating how people's curiosity is driven by novelty and complexity in different environment structures. Our first set of predictions come in the setting where the learning rate  $\gamma$  is equal for all stimuli in the environment. In that setting, our model predicts that depending on the environment structure, curiosity can be driven by either novelty or stimulus complexity. We test these predictions in Experiment 1. Our next set of predictions come in the setting where the learning rate  $\gamma$  is unequal for various stimuli in the environment. In this setting, our model predicts that the agent will have highest curiosity for stimuli that have a moderate learning rate. However, an agent's curiosity towards harder stimuli can be modulated by either manipulating the probability of occurrence or exposure of those stimuli. We test these predictions in Experiment 2. We note that both experiments use a paradigm that has been widely employed in experiments on curiosity (Kang et al., 2009; Gruber et al., 2014; Marvin & Shohamy, 2016; Baranes et al., 2014; Geana et al., 2016), which parallels the set of assumptions used in defining our rational model.

### Experiment 1: Switching between Novelty and Complexity

We first begin by considering an environment, where the learning rate  $\gamma_k$  is equal for all stimuli  $k$ , i.e. the agent's confidence increases at the same rate with respect to exposure for all stimuli. In this case, our rational model makes two different empirical predictions.

**Prediction 1.** The first prediction arises when the agent is in an environment  $p_k$  and  $h_k$  are related to each other (as in Figure 3(a)). The relationship between curiosity and exposure can be described using Equation 6 and between curiosity and confidence using Equation 7. Thus, Equation 6 predicts that an inverted U-shape relationship will exist between curiosity and exposure and Equation 7 similarly predicts that curiosity will be highest when the agent is moderately confident about a stimulus (see Figure 3). We test this prediction in the *confidence* sampling condition of Experiment 1.

Note that this prediction fits the complexity-based theories such as the information gap and the learning progress hypothesis, which also predict an inverted U-shape curve between curiosity and confidence i.e., empirical finding 2. While several studies have supported the existence of this U-shaped relationship, our model also predicts how to make this effect go away as described in our second prediction.

**Prediction 2.** Our second prediction comes in when the agent is in an environment where the relationship between  $p_k$  and  $h_k$  no longer holds true (as described in Figure 4(a)). Then the relationship between curiosity and exposure will be the one described in Equation 5 and the relationship between curiosity and confidence will be that given in Equation 10. Equation 5 predicts that curiosity is highest when exposure is lowest and it decreases as exposure increases. Similarly, Equation 10 predicts that curiosity will be highest when confidence is the lowest (also shown in Figure 4). We test this prediction in the *uniform* sampling condition of Experiment 1.

This combination of predictions is unique to our account of curiosity. While the second prediction accords with the novelty based hypothesis and empirical finding 1, that hypothesis cannot reproduce our model's first prediction. On the other hand, while the complexity theories were in line with our model's first prediction, those theories fail to explain our model's second prediction.

We now conduct a behavioral experiment in order to test our model predictions. The experiment used two different scenarios – confidence sampling and uniform sampling – to assess whether people’s curiosity is affected by changes in the relationship between probability of occurrence and confidence. In the confidence sampling condition, we created an environment such that probability of occurrence was related to confidence (Figure 3(a)) and in the uniform sampling condition they were independent of each other (Figure 4(a)). Based on our model predictions, we hypothesize that an inverted U-shape relation will exist between confidence and curiosity in the confidence condition and a decreasing relation will exist in the uniform sampling condition.

## Methods

**Participants.** We recruited 303 participants from Amazon Mechanical Turk. They earned \$1.50 for participation with the option of earning an additional bonus of \$0.80. Participants in the experiment were randomly assigned to one of two conditions: confidence sampling condition (157 participants) and uniform sampling condition (146 participants). Informed consent was obtained using a consent form approved by the institutional review board at the University of California, Berkeley.

**Stimuli.** The stimuli used in the experiment were 40 trivia questions on various topics that were taken directly from Experiment 1 in Kang et al. (2009). According to the authors, these questions were designed to measure curiosity about semantic knowledge and evoke a range of curiosity levels.

**Procedure.** The experiment was divided into three phases – Phase 1, Phase 2, and Phase 3 (refer to Figure 5). In Phase 1, participants were shown 40 trivia questions one after another and were asked to rate their confidence (i.e., probability that they know the correct answer) and curiosity in knowing the correct answer. Curiosity ratings were on a scale from 1 to 7 and the confidence scale ranged from 0 to 100%. The order of trivia questions was randomized for each participant. Note that Phase 1 of our experiment followed the procedure of Kang et al.’s design closely. This part of the experiment took approximately 7-8 minutes to complete.

Phase 1 Judgments	Phase 2 Reveal Answers	Phase 3 Answer questions
<p><b>Question</b></p> <p>What book is the most shoplifted book in the world?</p> <p><b>Responses</b></p> <p>Please indicate how curious you are to know the correct answer:</p> <p style="text-align: center;"> <input type="text" value="not curious at all (1)"/> ... <input type="text" value="very curious (7)"/> </p> <p>How likely is it that you would answer this question correctly? (%)</p> <p style="text-align: center;"> <input type="text" value="not likely at all (0)"/> ... <input type="text" value="very likely (100)"/> </p>	<p><b>Question</b></p> <p>What book is the most shoplifted book in the world?</p> <p style="text-align: center;">Press V if you want view the answer</p> <p style="text-align: center;">Question 1/40</p>	<p>What book is the most shoplifted book in the world?</p> <p><input type="text"/></p> <p style="text-align: center;">⋮</p> <p style="text-align: center;">⋮</p> <p>Which is the only animal that can get sunburnt?</p> <p><input type="text"/></p>

**Figure 5. Design of Experiment 1.** The experiment was divided into three phases. In Phase 1, participants were presented with the 40 trivia questions and asked to rate their curiosity and confidence in knowing the correct answer. In Phase 2, participants had the choice to reveal the answers to the previously shown 40 questions at the cost of waiting for 10 seconds for the next question to appear. Following that, in Phase 3, participants had the chance to answer 10 of the previously shown questions to earn bonus money. Note that instructions were provided before each phase.

In Phase 2, all previous 40 questions from Phase 1 were again shown one after another and participants had the choice to reveal the answer to those questions. However, each time participants chose to reveal an answer, they had to wait an extra 10 seconds for the next question to appear. Findings from Experiment 3 of Kang et al., (2009) showed that participants were more likely to spend time, to wait longer, for the answers that they were more curious about. Thus, requiring participants to spend time to obtain information served as a proxy to measure their curiosity.

In Phase 3, participants were shown 10 out of the previously shown 40 questions and were given the chance to answer these 10 questions. For each correct answer, participants were given a bonus of \$0.08. To discourage participants from using Google or other search engines, they were only given 2 minutes in total to answer the questions.

At the beginning of the experiment, participants were randomly assigned to two conditions – the *confidence* and the *uniform* condition. These conditions differed in the way the 10 questions were sampled in Phase 3 and apart from that, the two conditions were exactly the same. In the



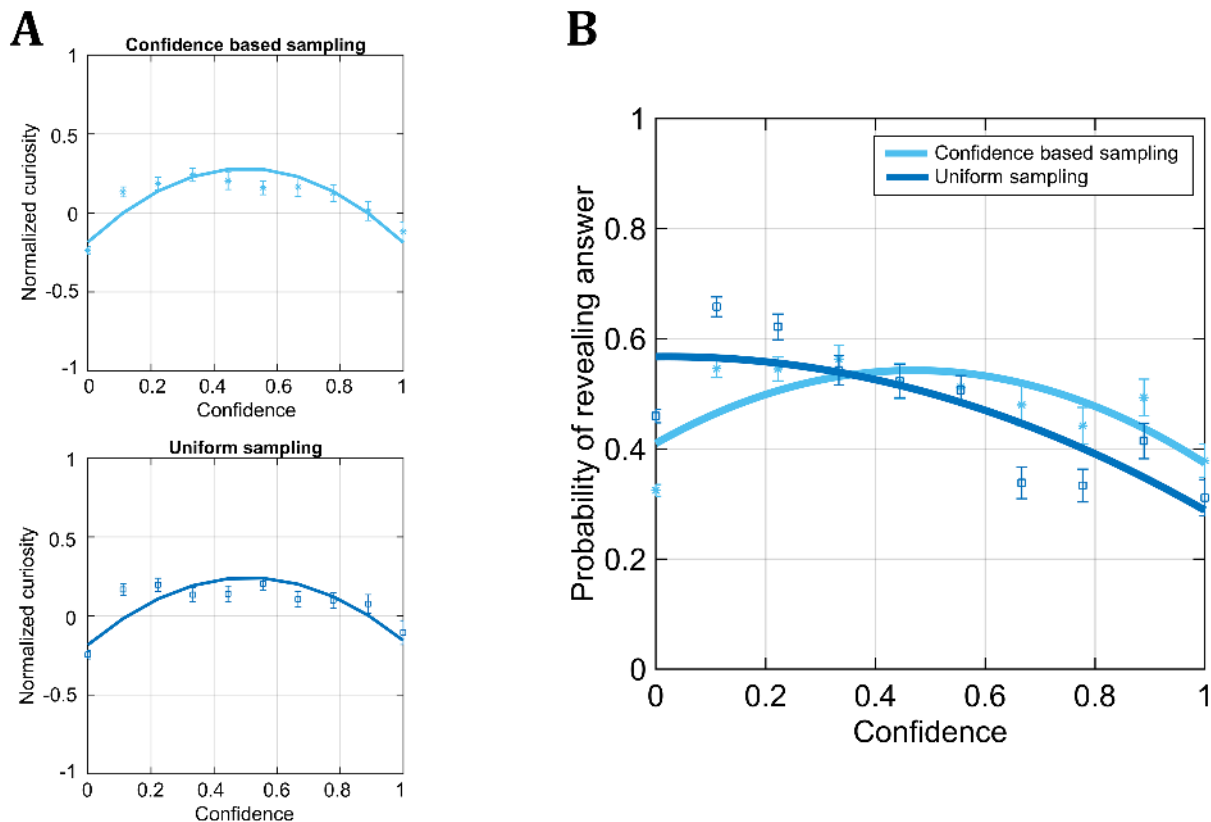
confidence condition, the sampling in Phase 3, was done based on the confidence ratings provided by the participants i.e. the questions for which participant's confidence rating was higher were more likely to appear in Phase 3. In the uniform condition, this sampling procedure was completely random i.e. each question was equally likely to appear in Phase 3. Importantly, participants were informed about the sampling procedure for their respective condition before the beginning of Phase 2. The confidence condition thus creates a situation in which confidence is related to probability of occurrence (Figure 3(a)) and the uniform condition breaks this relationship (Figure 4(a)). To ensure that participants understood the sampling procedure, participants had to answer a multiple-choice question about the sampling procedure after the instructions for Phase 2 were shown to them. If they gave an incorrect answer, participants were shown the instructions again and had to re-answer the question (this process was repeated until they answered the question correctly).

According to our model's predictions we should see an inverted U-shape between curiosity and confidence for both the conditions in Phase 1. However, the curiosity of participants (i.e. the answers they revealed) should be different for both conditions in Phase 2. This is because participants were informed about the sampling procedure only after Phase 1 (and before Phase 2) and therefore people's curiosity in Phase 2 should be affected by their knowledge of the structure of the environment. In particular, for the confidence condition, participants' probability of revealing an answer should be highest for questions which they were moderately confident about. On the other hand, in the uniform condition, participants should be most curious about questions for which they were least confident about.

## Results

For all analyses that follow, we removed participants that revealed either none (=0) or all of the answers (=40) in Phase 2. A total of 49 participants were excluded based on this criterion and our final data consisted of 253 participants (133 in the confidence condition and 120 in the uniform condition).

**Phase 1.** Following Kang et al.'s methodology, the raw curiosity ratings were individually



**Figure 6. Empirical results from Experiment 1.** (A) Results from Phase 1. Relationship of curiosity and confidence in Phase 1 for confidence sampling condition (top) and uniform condition (bottom). The markers indicate mean curiosity at each confidence level and the solid curve is the regression line. Curiosity is an inverted-U function of confidence for both conditions. (B) Results from Phase 2. The graph plots the probability of participants revealing an answer as a function of confidence in Phase 2. Consistent with our model’s prediction, an inverted U-shape exists between curiosity and confidence in the confidence based sampling condition and a decreasing relationship exists in the uniform sampling condition.

normalized and confidence was re-scaled to range from 0 to 1, and we fitted the data to the equation:  $\text{curiosity} = b_0 + b_1 \times c + b_2 \times c \times (1 - c)$  to both the conditions, where  $c$  was the re-scaled confidence score. Consistent with our prediction, an inverted U-shape exists between curiosity and confidence for both conditions (Figure 6(a)). For the confidence condition, the model provided  $r = 0.2$  and a significant coefficient for the quadratic coefficient i.e.  $c \times (1-c)$  ( $b_2$  estimate = 1.87,

$p < 0.05$ ) and an insignificant coefficient for confidence ( $p = 0.9$ ). For the uniform condition, the model provided similar results with  $r = 0.2$  and significant coefficient for the quadratic coefficient ( $b_2$  estimate = 1.65,  $p < 0.05$ ) and insignificant coefficient for confidence ( $p = 0.6$ ). We obtained equivalent results whether we used the raw or re-scaled confidence ratings. Thus, for both conditions, the model provided a significant quadratic coefficient, supporting the presence of an inverted U-shape between curiosity and confidence for both conditions.

**Phase 2.** We first computed the probability of participants revealing an answer conditioned on the confidence rating for both the conditions. As per our model's predictions, an inverted U-shape exists for the confidence condition and a decreasing relationship exists for the uniform condition (Figure 6(b)). Similar to the previous analysis, we fitted data to the equation:  $\text{curiosity} = b_0 + b_1 \times c + b_2 \times c \times (1 - c)$  to both the conditions. For the confidence condition, the model provided  $r = 0.75$  and a significant coefficient for the quadratic coefficient ( $b_2$  estimate = 0.6,  $p < 0.05$ ) and an insignificant coefficient for confidence ( $b_1$  estimate = -0.03,  $p = 0.5$ ) thereby implying a U-shape relationship between curiosity and confidence. For the uniform condition, the model provided  $r = 0.8$  but the coefficient for the quadratic coefficient was not significant ( $p = 0.34$ ). On the other hand, the coefficient for confidence was significant ( $b_1$  estimate = -0.27,  $p < 0.05$ ), implying a decreasing relationship of curiosity with confidence for the uniform condition. We also conducted individual-level analysis of the data that further supports these results and refer the reader to Appendix B for the details.

## Discussion

Experiment 1 tested and found support for two key predictions of our model. When participants were in an environment where probability of occurrence influenced exposure, then their curiosity was highest for stimuli for which they were moderately confident about. When participants were in an environment where probability of occurrence and exposure were independent of each other then curiosity was highest for novel stimuli, i.e. stimuli for which participants had little confidence. These results suggest that human curiosity is not only sensitive to

the properties of the stimuli but it is also influenced by the structure of the environment – depending on the environment structure, curiosity can be driven by either novelty or stimulus complexity.

### Experiment 2: Manipulating Learning Rate

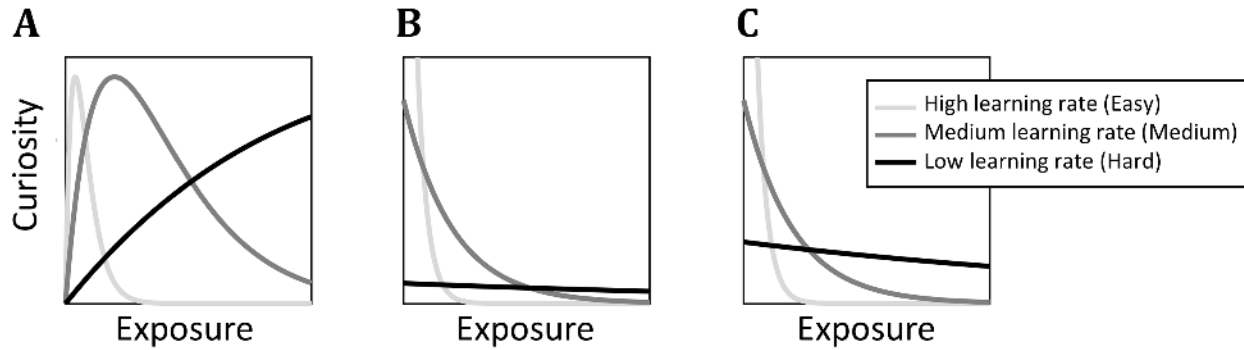
Until now, we have considered the case wherein all stimuli in the environment had equal learning rate  $\gamma$ . We now consider the setting where the learning rate  $\gamma$  is not the same for the various stimuli in the environment. Here, the agent would learn about different stimuli in the environment at a different rate (with respect to exposure) and thus some stimuli will be harder to learn than others.

Our model predictions are shown in Figure 7 (derived from Equations 5 and 6). We first begin by noting that regardless of the environment, the agent’s curiosity increases progressively i.e. it first starts by being curious about easy to learn stimuli and then gradually becomes curious about moderate stimuli and then eventually harder to learn stimuli. Second, regardless of the environment, the agent’s curiosity integrated over exposure is highest for stimuli with medium learning rates (refer to Figure 7(a) and Figure 7(b)). However, as also shown in Figure 7(a) and (b), with a further increase in exposure, the agent gradually becomes less curious about stimuli that have a medium learning rate (as its confidence eventually reaches 100%) and eventually becomes curious about stimuli that are hard to learn. In the second case i.e. Figure 7(b), we assumed equal probability of occurrence  $p_k$  for all  $k$ . However, recall that when probability of occurrence and exposure are independent of each other (Figure 4(a)), then  $p_k$  is simply a constant. Therefore, if the probability of occurrence of low learnability stimuli (i.e. harder to learn stimuli) is higher than other stimuli, then our model predicts that the agent will be most curious about low learnability stimuli (refer to Figure 7(c)).<sup>8</sup> We summarize our predictions below:

**Prediction 3.** People will be maximally curious about stimuli that have a moderate learning rate i.e., stimuli that are moderately difficult (consistent with empirical finding 3). We test this prediction in the *baseline* condition of Experiment 2.

---

<sup>8</sup> Intuitively, this is akin to implicitly modeling extrinsic motivation as greater probability of occurrence makes a stimulus more important thereby inducing higher curiosity.



**Figure 7. Model predictions about the relationship between curiosity and exposure when different stimuli have different learning rates.** (A) Environment where probability of occurrence is related to exposure, (B) Environment where they are independent of each other, and (C) Same environment setting as B but here the hard stimuli i.e. stimuli with low learning rate have a higher probability of occurrence. In all three cases, the agent's curiosity increases progressively from easier to harder stimuli. In A and B, the agent's curiosity integrated over exposure is highest for stimuli with medium learning rates. In C, the agent's curiosity for hard stimuli i.e. with low learning rates is higher (compared to A and B) due to their higher probability of occurrence.

**Prediction 4.** With an increase in exposure, people will eventually become more curious about stimuli that have a low learning rate i.e. stimuli that are harder to learn (consistent with empirical finding 4). We test this prediction in the *increase* condition of Experiment 2.

**Prediction 5.** People's curiosity towards harder to learn stimuli would also increase if the probability of occurrence of those stimuli is high. We test this prediction in the *probability* condition of Experiment 2.

We note that Predictions 3 and 4 are in line with predictions from previous complexity-based theories like the learning-progress hypothesis. In addition to these, our model also predicts curiosity can be modulated towards harder to learn stimuli by manipulating their probability of occurrence (Prediction 5). We now conduct another behavioral experiment to test these predictions. The experiment used three different conditions – *baseline* condition, *increase* condition, and *probability* condition – to assess whether people's curiosity towards harder to learn tasks is affected by increase

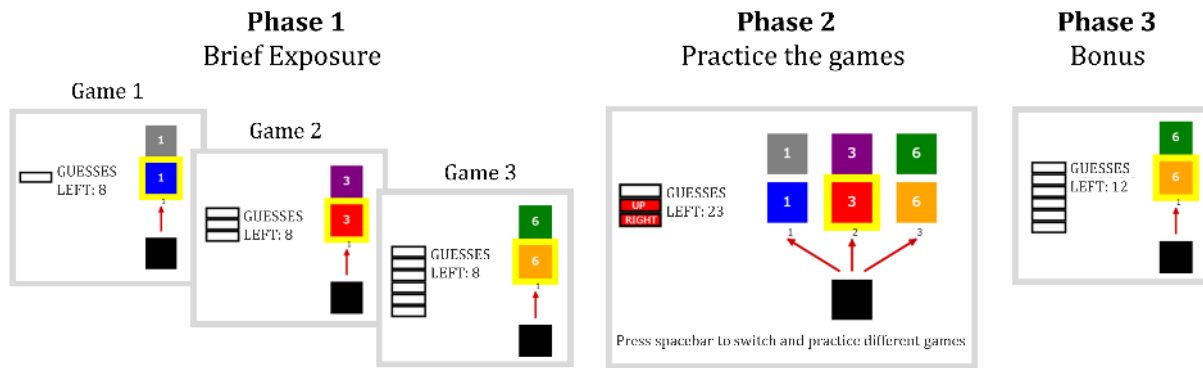
in exposure and probability of occurrence.

## Methods

**Participants.** We recruited 209 participants from Amazon Mechanical Turk. They earned \$0.75 for participating in an experiment that took approximately 5 minutes to complete with the option of earning an additional bonus of \$0.25. Participants in the experiment were randomly assigned to one of three conditions: *baseline* condition (78 participants), *increase* condition (64 participants), and *probability* condition (67 participants). Informed consent was obtained using a consent form approved by the institutional review board at Berkeley.

**Stimuli.** The stimuli used in the experiment were three games of varying difficulty – easy (Game 1), medium (Game 2), and hard (Game 3, also refer to Figure 8). Each game consisted of two uniquely colored square boxes that could be unlocked via a key sequence consisting of the four arrow keys (left, up, down, right). The participant’s goal was to learn the correct key sequence for the six square boxes. Importantly, the number of key sequences required to unlock the boxes were different for each Game (this length was indicated by a number printed in the center of each box). More specifically, the two boxes in Game 1 required only 1 key input to unlock them. The two boxes in Game 2 required 3 key inputs to be unlocked. Finally, the two boxes in Game 3 required 6 key inputs to be unlocked. In this way, the three games were varied in their difficulty with Game 1 being easiest to learn, followed by Game 2, and then Game 3.

**Procedure.** The experiment was divided into three phases – the exposure round, the practice round, and the bonus round (Figure 8). In the exposure round, participants were shown the three games sequentially and they were given a small number of guesses to play each game (this number varied by the condition the participants were in). To help participants remember the corresponding key input once they had correctly guessed it, a progress bar was also displayed on each screen such that one portion of the progress bar filled up with the label of the correct key input everytime a correct guess was made. Participants were also informed that this round was only meant to provide them a brief exposure to the three games and they should not be concerned if they



*Figure 8. Design of Experiment 2.* The experiment was divided into three phases. In the exposure round, participants were shown three games sequentially and were given a short amount of guesses to play each game. The three games varied in their difficulty with Game 1 being easiest to learn, followed by Game 2, and then Game 3. In the practice round, all three games were shown in one screen together and participants were given 30 guesses overall to practice the three games to learn and master them. In the bonus round, participants were shown one of the three games and had a chance to earn bonus money. Note that, in all three rounds, the remaining number of guesses, the progress bar, and the active box (highlighted in yellow) were displayed to the participants.

couldn't completely learn all the three games in this round.

In the practice round, all three games were shown in one screen together and participants were given a total of 30 guesses to play the three games. Participants could decide to play any of the three games by pressing keys 1-3. Once they decided to play a game, they could always switch and play another game by simply pressing the spacebar and then again pressing keys 1-3 to make their selection. The participants' goal in this round was to learn the three games as much as possible to maximize their chances of getting a bonus.

In the bonus round, one of the three games was displayed to the participants and they were given a total of 12 guesses to play the game. If a participant successfully solved the game before the guesses ran out (i.e. successfully unlocked the two boxes of the game), they were provided a bonus payment of \$0.25.

At the beginning of the experiment, participants were randomly assigned to three conditions –

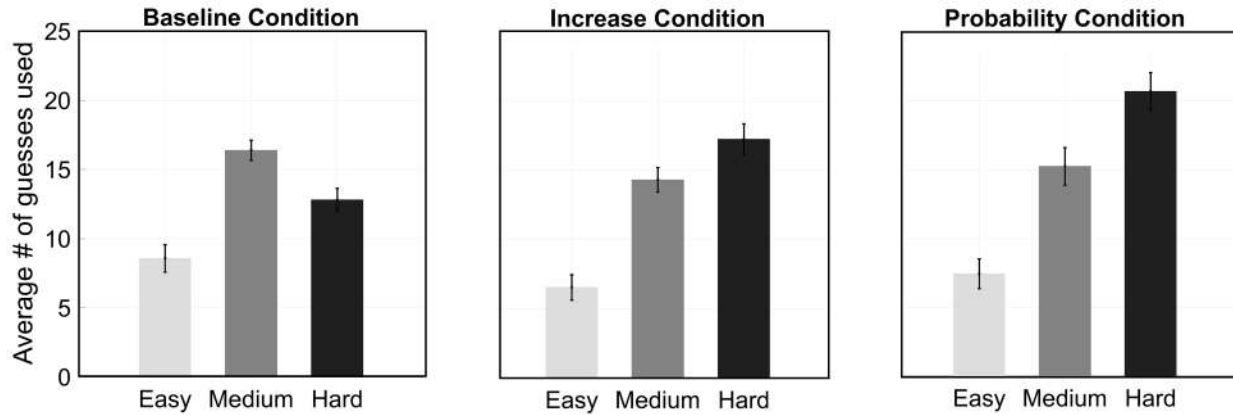
the *baseline*, the *increase*, and the *probability* condition. In the *baseline* condition, participants were given *four* guesses to play each game in the exposure round and the sampling procedure in the bonus round was *random* i.e. any of the three games was equally likely to appear in the bonus round. In the *increase* condition, participants were given *eight* guesses to play each game in the exposure round and the sampling procedure in the bonus round was again *random*. In the *probability* condition, participants were given *four* guesses to play each game in the exposure round. However, the sampling procedure in the bonus round was *biased* towards Game 3 such that it had a 80% chance of appearing in the bonus round. Participants were provided instructions about the bonus round and the sampling procedure before the start of the practice round. To ensure that participants understood the sampling procedure, participants had to answer a multiple-choice question about the sampling procedure after the bonus round instructions. If they gave an incorrect answer, participants were shown the instructions again and had to re-answer the question (this process was repeated until they answered the question correctly).

According to our model, participants' behavior in the practice round should vary significantly across conditions. In the baseline condition, participants should spend a significant amount of time playing the moderately difficult game i.e., Game 2 in the practice round. In the increase condition, participants should spend a greater amount of time playing the hardest game i.e., Game 3 in the practice round. This is simply due to the fact that participants were provided more guesses in the exposure round (compared to baseline condition). This leads to an increase in exposure for the three games and thus curiosity would be more towards the harder stimulus i.e., Game 3. In the probability condition, participants should again spend a greater amount of time playing the hardest game i.e. Game 3 (as the probability of occurrence  $p_k$  of Game 3 is highest thereby leading to higher curiosity).

## Results

To test our predictions, we analyzed participants' behavior in the practice round for the three different conditions. To that end, we report the average number of guesses participants allocated to

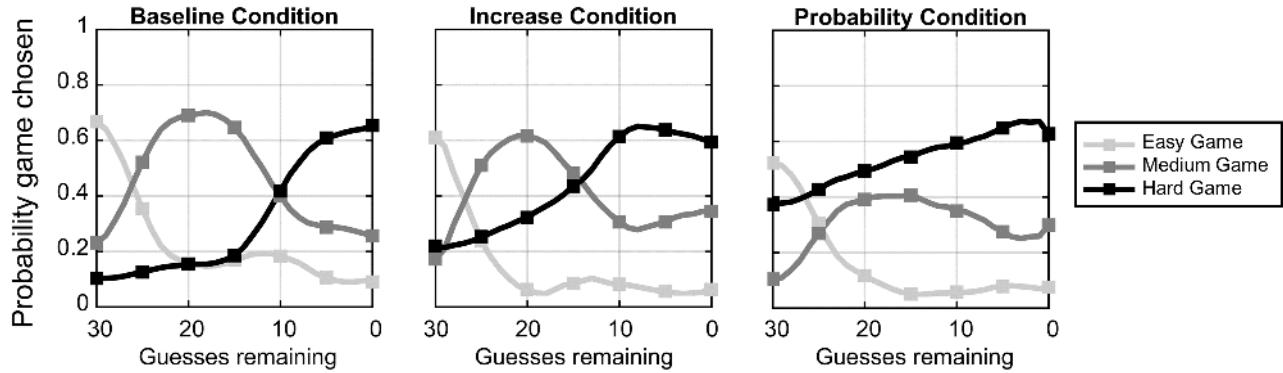




**Figure 9. Empirical results from Experiment 2.** The graph plots the average number of guesses participants spent on playing the three different games for each condition. As per our model's predictions, participants spend the highest amount of guesses on playing Game 2 (i.e. medium difficulty) in the baseline condition and the highest number of guesses on playing Game 3 (i.e. hard difficulty) in both the increase and the probability condition.

each game in the practice round. As shown in Figure 9(a), in the *baseline* condition, participants spent a significantly longer time (in terms of guesses spent) playing the medium game (Game 2) compared to Games 1 and 3 (one way repeated measures ANOVA  $F(2, 188) = 21.39$ ,  $MSE = 46.99$ ,  $p < 0.01$ ). In contrast, in the *increase* condition, participants spent a significantly longer time playing the hard game (Game 3) compared to Games 1 and 2 ( $F(2, 149) = 28.99$ ,  $MSE = 49.23$ ,  $p < 0.01$ ). Furthermore, a paired t-test revealed that participants spent significantly greater time (in terms of guesses) playing Game 3 in the increase condition compared to the baseline condition ( $t(108) = 3.2$ ,  $p < 0.01$ ). Lastly, in line with our model's predictions, participants again spent longer time playing the hard game (Game 3) compared to Games 1 and 2 in the *probability* condition as well ( $F(2, 136) = 26.85$ ,  $MSE = 78.28$ ,  $p < 0.01$ ). A paired t-test revealed that participants spend significantly greater time playing Game 3 in the probability condition compared to the baseline condition ( $t(111) = 5.0$ ,  $p < 0.01$ ).

We next analyzed step-by-step behavior of the participants in the practice round for the three conditions. To do this, we computed the probability of participants choosing the three games as a



*Figure 10. Time course analysis of Experiment 2.* The graph plots participants' probability of choosing the three games (y-axis) as a function of remaining guesses in the practice round (x-axis). As per our model's predictions, participants' curiosity increases progressively. Further, participants switch to playing the hard game much earlier in the increase and probability condition compared to the baseline condition.

function of their remaining guesses. As shown in Figure 10, consistent with our model predictions, participants' curiosity increases progressively. Individual one-sided binomial tests further revealed that for all three conditions, participants probability of choosing Game 1 (easy) was significantly greater than their probability of choosing Game 2 (medium) or Game 3 (hard) at the first time-point i.e. when the total remaining guesses were 30 ( $p < 0.001$  for all tests). In line with our predictions, we also see that participants switch to the hard game much earlier in the *increase* and *probability* condition compared to the *baseline* condition. Separate Chi-square tests revealed that the probability distributions of the Game 2 and Game 3 were significantly different than the probability distribution of Game 1 for all three conditions ( $\chi^2(31) = 23.4$  and  $\chi^2(31) = 26.9$  respectively for baseline,  $\chi^2(31) = 60.7$  and  $\chi^2(31) = 84.9$  respectively for increase, and  $\chi^2(31) = 30.7$  and  $\chi^2(31) = 96.7$  respectively for probability condition,  $p < 0.001$  for all). We next computed the area under the curve (AUC) of the probability distributions for the three games and found that this was consistent with our previous analysis. More specifically, the AUC for Game 2 (medium) was greater than AUC for Game 1 and Game 3 for the *baseline* condition (AUC = 6.66 for easy, 14.04 for medium, and 9.30 for hard respectively). Further, the AUC for Game 3 (hard) was greater than

Game 1 and Game 2 for the *increase* condition (AUC = 4.16 for easy, 12.49 for medium, and 13.34 for hard respectively) as well as for the *probability* condition (AUC = 4.48 for easy, 9.35 for medium, and 16.17 for hard respectively).

## Discussion

Experiment 2 tested and found support for three predictions of our model. First, people are most curious about stimuli that they perceive to be moderate difficulty (baseline condition). Second, as the exposure increases, they eventually become curious about harder to learn stimuli (increase condition). Lastly, people also become curious about harder to learn stimuli if the probability of occurrence of those stimuli is high (probability condition). These results suggest that while people's curiosity is more directed towards stimuli that have moderate difficulty, curiosity can be modulated towards harder to learn stimuli by either manipulating the exposure or probability of occurrence. Whereas results from the *baseline* and *increase* condition are in line with previous findings on curiosity (empirical findings 3 and 4), results from the *probability* condition confirm an important prediction of our model that wasn't captured by previous models – curiosity can be modulated towards harder tasks by manipulation of probability of occurrence.

## General Discussion

Curiosity is one of the hallmarks of human intelligence and is crucial to scientific discovery and invention. Previous theories explained human curiosity by linking it to various psychological factors such as novelty or stimulus complexity. In this article, we presented a rational analysis of curiosity which offers a way to unify previous theories of curiosity into a single framework. Our approach posits that curiosity is the mechanism that approximates the rational behavior of seeking stimuli to maximize an agent's ability to make appropriate responses in the future. In doing so, our approach suggests that curiosity reflects general facts about the environment. For instance, consider an agent that is in an environment where the more something has occurred in the past, the more likely it is to occur in the future. In this case, the behavior of the agent should be better adjusted to frequently recurring events than to rare ones (Shettleworth, 1999). According to our framework,

this will lead curiosity to be driven by stimuli that are of intermediate complexity. On the other hand, if the agent is in an environment where the future is no longer influenced by past events, then the behavior of agent should be better adjusted to events for which it has had less experience with. In this case, our framework predicts that curiosity is driven by novel stimuli. This perspective suggests that rather being separate phenomena requiring different kinds of explanations, novelty-based and complexity-based theories of curiosity can be seen as lying on a continuum influenced by the causal structure of the environment. Along with providing a way to unify previous distinct mechanistic accounts of curiosity, our rational model explains human curiosity in a variety of settings (empirical findings 1-4). To our knowledge, this is the first computational framework that explains all these empirical results.

In the remainder of the paper, we examine how our model captures the subjectivity of human curiosity, discuss implications of our results for teaching, and look at potential shortcomings and future directions of our work.

### **Curiosity and affect: connecting rational model to subjective experiences**

One implication of our work is that it suggests that people ought to be curious about stimuli which maximally increase the usefulness of their knowledge. However, contradictory to these implications, a number of studies have demonstrated that people's curiosity can be sub-optimal in various situations (Hsee & Ruan, 2016; Wiggin et al., 2018). How can our model capture such findings?

Until now, we considered that the various parameters in our rational model such as confidence  $c_k$ , probability of occurrence  $p_k$ , and ease of learning,  $\gamma_k$  were known to the agent. However, these parameters i.e.  $c_k$ ,  $p_k$ , and  $\gamma_k$  are *estimates* computed by the agent and thus the agent needs to infer the value of these parameters. An unbounded agent can compute these values using Bayesian inference, however, since full Bayesian inference is intractable, a bounded agent will rely on approximate Bayesian inference (Griffiths et al., 2015; Sanborn & Chater, 2016). Thus, the estimates computed by the agent are subjective as well as prone to errors (Lieder et al., 2012).

Further, the coupling between curiosity  $\Omega_k$  and the optimal solution i.e. rate of change of value of knowledge with respect to exposure,  $\frac{dV}{dh_k}$  can also be weak (as curiosity is a mechanism that approximates the optimal solution). As a consequence, an agent may not become curious about a stimulus even if it were optimal to do so.

For instance, consider an agent that erroneously estimates the importance of a stimulus i.e.  $p_k$  to be lower than its actual value. Then the agent may not become curious about the stimuli even if its probability of occurrence maybe high. Conversely, an agent that estimates  $p_k$  to be higher than its actual value, may become curious about seemingly irrelevant stimuli. Further, since the agent is performing approximate Bayesian inference, the estimated parameters can also be erroneously influenced by prior experience.<sup>9</sup> For example, if people a priori believe a task to be very hard, then they are less likely to be curious about that task without even trying. Understanding how to influence people's curiosity in such scenarios is an important research question for future work.

While our focus has been on providing a rational analysis of curiosity, our framework can potentially explain phenomena that seem maladaptive. As a case study, consider people's penchant for watching British detective shows. At a first glance, British detective shows seem to be at odds with our framework – knowing the answer to a hypothetical murder mystery is surely not relevant to our future. However, such stimuli offer cues that make them seem very likely to increase the value of our knowledge. In particular, if someone within your social circle was murdered, then it would be extremely valuable to know who committed the murder. Indeed, these shows seem to be hacking into people's curiosity, using great acting to simulate a situation where it is important to identify the true criminal. This is an example of maladaptive curiosity – because it is optimal to have a drive to explore stimuli that increase our value of knowledge, we can get sub-optimally curious about things that may not have any real-world advantage. This behavior is analogous to our drive towards sweet and fatty foods like donuts. Prehistoric humans had to acquire rare nutrients like sweets and fats for survival which resulted in the development of a drive in humans for sweet

---

<sup>9</sup> Here we say erroneously influenced by prior experience because under limited time and computational resources, an agent's posterior estimate can be biased towards the initial value (Lieder et al., 2012).

and fatty foods. However, for humans living in wealthy societies, having a drive towards foods like donuts is not necessarily adaptive. British detective shows are the donuts of curiosity, engineered to hijack drives that would usually have more adaptive consequences.

### **Implications for teaching and education**

Despite theoretical disagreements, curiosity is universally positively regarded, and has been acknowledged as a significant predictor of academic achievement (Von Stumm et al., 2011). In line with this, a number of studies in recent years have explored methods to stimulate curiosity of people in various learning settings (Pluck & Johnson, 2011; Gentry et al., 2014; Law et al., 2016). A common thread among these intervention studies is that they are all based on Loewenstein's information-gap theory (Loewenstein, 1994) i.e. they prompt curiosity in learners by inducing an information-gap. A potentially significant implication of our work is that our framework suggests that in addition to inducing information-gaps, there are various other ways in which curiosity can be readily modulated. First, consistent with complexity-based theories like the information-gap hypothesis, our model suggests that we can influence people's curiosity by inducing an information-gap (also shown in the confidence sampling condition, Experiment 1). Second, our results suggest that if we want to make people curious about tasks or activities that are novel and for which they have little confidence, perhaps subtle changes in the structure of the environment might be a step towards achieving that (i.e. creating an environment where the past is independent of the future - uniform condition, Experiment 1). Third, our work suggests that people's curiosity can be driven towards harder to learn tasks if they perceive them to be useful or important. This is because when people perceive an event or task to be important, then they would estimate those tasks to have a higher probability of occurrence,  $p_k$  thereby leading to higher curiosity for those tasks. For example, consider the case of a student learning how to solve calculus problems. The student is more likely to be curious about solving these problems if he or she wants to pursue mathematics in the future or perceives that they can connect this knowledge to other domains. This is also consistent with empirical studies showing that people's perceived usefulness of information is a

strong predictor of their curiosity (Rossing & Long, 1981; Mehta et al., 2018; Liquin & Lombrozo, 2018). Furthermore, recent empirical work on decision-making has also shown that people are more willing to explore risky prospects if they expect to encounter them again in the future (Rich & Gureckis, 2014). Indeed, as shown in the probability condition of Experiment 2, people's curiosity can be modulated towards harder tasks if they believe that the probability of occurrence of those stimuli to be high. Lastly, our work suggests that curiosity towards harder tasks can be also increased if people's perception about that task's learnability changes. As demonstrated in the baseline condition of Experiment 2, people are maximally curious about tasks that are moderately difficult. Therefore, if we want to make people curious about hard tasks, perhaps an intervention that manipulates the perceived difficulty of those tasks (for e.g. from hard to moderately difficult) might be an effective solution.

In summary, while recent work has explored methods to stimulate curiosity in learners by inducing information-gaps, our approach suggests various other ways in which curiosity can be stimulated. Importantly, while the previous section highlighted that human curiosity can be highly situational and even sub-optimal at times, this section presents an encouraging and optimistic implication of work – that curiosity is not fixed and innate but malleable and can be readily modulated by experience. This emphasis is consistent with the view that curiosity and the habit of exploring the environment can be learned (Simon, 2001). Thus, although human curiosity is prone to errors, our work also suggests how people's curiosity can be driven towards relevant stimuli with strategic interventions. We believe these results can have important implications in the context of education where researchers are concerned with ways to pique curiosity in students.

### **Limitations and future directions**

While our account captures various aspects of curiosity, it also has several limitations. For one, our focus in this article has been on providing a computational-level explanation of curiosity, rather than an account of the representations and algorithms underlying curiosity. In performing our analysis, we suggested that curiosity is a mechanism by which people approximate the rational

behavior derived within our framework. While our analysis remains agnostic about the specific psychological process underlying curiosity, future work should consider how human minds approximate this solution and what those algorithms look like. More specifically, recent computational results hold promise as they provide several interesting approaches to operationalize curiosity within the reinforcement learning framework (Chentanez et al., 2005; Dayan, 2013; Schmidhuber, 2013; Pathak et al., 2017; Colas et al., 2018).

Another limitation of our theory is that it does not explain or capture the epistemic emotion of curiosity. A number of studies have shown that curiosity is often synonymous with a strong epistemic emotion and a feeling of deprivation (Litman & Jimerson, 2004; Litman et al., 2005). Why do we feel a strong emotion when we don't know things? Some proposals have suggested that this could be linked to the fact that people have a strong aversion to uncertainty (Hogg, 2000). Based on this, Bennett et. al., (2016) proposed a model that incorporates this principle of uncertainty aversion to the traditional value of information framework to explain the human tendency to resolve uncertainty as quickly as possible. Future work could look to build a process model of curiosity that also incorporates a form of uncertainty aversion to possibly explain the epistemic emotion of curiosity.

Another interesting future direction could be understanding how and why people avoid information, even when it is free and could improve decision making (Sweeny et al., 2010; Hertwig & Engel, 2016; Golman et al., 2017). A number of studies have shown that in certain settings, people avoid information and chose not to resolve their curiosity even if it might be optimal do so. For example, many people consistently tend to avoid learning about their HIV status even if it comes without any cost or additional effort (Tao et al., 1999; Hightow et al., 2003; Sweeny et al., 2010). Regret aversion theories explain this phenomena by stating that people avoid information if they anticipate that they might experience regret upon obtaining information (Gigerenzer & Garcia-Retamero, 2017). How curiosity interacts with regret aversion and subsequently motivates information search/avoidance will be an interesting future direction to pursue.

We used trivia questions and computer games in our experiments because previous studies



showed the effectiveness of such materials to study curiosity (Kang et al., 2009; Gruber et al., 2014; Baranes et al., 2014; Geana et al., 2016). However, these stimuli can stimulate only a limited aspect of curiosity and curiosity can be different depending on the type of stimulus. Furthermore, both Experiments 1 and 2 manipulated curiosity in an artificial manner which may not necessarily reflect real-world scenarios. In addition, Experiment 2 measured people's curiosity by recording their playing behavior as opposed to directly asking about their curiosity. Future studies should examine the robustness and generalizability of our findings using different types of experimental materials and settings to trigger curiosity.

An alternate explanation of an agent's curiosity being driven by novelty or complexity could be based on the shape of the learning curve as suggested by Son and Sethi (2006). In their work, Son and Sethi provided a formal account of meta-cognitive control of time allocation and suggested that time allocation is influenced by the shape of the learning curve. This perspective could be applied towards our work in that curiosity is driven by the shape of the learning curve. For instance, when the learning curve is concave, then according to Son and Sethi's framework, curiosity would be driven by novel stimuli and when the learning curve is s-shaped, then curiosity would be driven by moderately complex stimuli. We believe that while this viewpoint provides a good alternative explanation of curiosity being driven by novelty or complexity, it doesn't fully capture important aspects of curiosity. For example, as demonstrated in Experiment 1, the learning curve doesn't necessarily have to s-shaped for curiosity to be influenced by complexity. Second, this viewpoint doesn't offer an explanation as to why curiosity is influenced by probability of future occurrence (Experiment 2, probability condition) as well as why curiosity is influenced by value (Liquin & Lombrozo, 2018; Mehta et al., 2018). One likely hypothesis is that both the shape of the learning curve and structure of the environment influences people's curiosity – this is an important question for future. Regardless of these differences, we note that an important common thread between our paper and the work of Son and Sethi is that we both accommodate these two phenomena (i.e. curiosity and time allocation) under a resource-rational framework. In this vein, an important direction for future research would be studying the connection between meta-cognition, time

allocation, and curiosity.

Finally, another important direction of future research will be developing intervention methods to stimulate curiosity in various education and learning settings. Our own current work is building on the framework we have outlined here to develop such intervention methods (Mehta et al., 2018; Dubey et al., 2019).

## **Conclusion**

Philosophers and psychologists have long recognized that without curiosity – inextinguishable curiosity – scientific work cannot sustain the day-to-day effort required to prepare for it and complete it. In this article, we have provided a rational analysis of curiosity which proposes that curiosity is the mechanism by which humans come to act in accordance with the optimal behavior of seeking stimuli that maximally increase the value of knowledge to maximize rewards. This perspective unifies previous distinct theories of curiosity in a common framework by demonstrating that those theories have at their heart the same computational problem. Yet, our approach is not strictly normative and we also explain why people’s curiosity can be highly situational and even sub-optimal at times. Our perspective also emphasizes that curiosity is not fixed but malleable and we suggest various ways to stimulate people’s curiosity. Although we are still a long way from explaining how scientists come to develop and hone their curiosity, our framework provides a step towards making this process less mysterious.

## References

- Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
- Auer, P., Cesa-Bianchi, N., & Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3), 235–256.
- Baldassarre, G., & Mirolli, M. (2013). Intrinsically motivated learning systems: an overview. In *Intrinsically motivated learning in natural and artificial systems* (pp. 1–14). Springer, Berlin, Heidelberg.
- Baranes, A., Oudeyer, P.-Y., & Gottlieb, J. (2014). The effects of task difficulty, novelty and the size of the search space on intrinsically motivated exploration. *Frontiers in neuroscience*, 8, 317.
- Baranes, A., Oudeyer, P.-Y., & Gottlieb, J. (2015). Eye movements reveal epistemic curiosity in human observers. *Vision research*, 117, 81–90.
- Bennett, D., Bode, S., Brydevall, M., Warren, H., & Murawski, C. (2016). Intrinsic valuation of information in decision making under uncertainty. *PLoS computational biology*, 12(7), e1005020.
- Berlyne, D. E. (1950). Novelty and curiosity as determinants of exploratory behaviour. *British Journal of Psychology. General Section*, 41(1-2), 68–80.
- Berlyne, D. E. (1960). *Mcgraw-hill series in psychology. Conflict, arousal, and curiosity*. McGraw-Hill Book Company, New York, NY, US.
- Berlyne, D. E. (1966). Curiosity and exploration. *Science*, 153(3731), 25–33.
- Bubeck, S., Cesa-Bianchi, N., et al. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1), 1–122.
- Chater, N., & Oaksford, M. (1999). Ten years of the rational analysis of cognition. *Trends in cognitive sciences*, 3(2), 57–65.

- Chentanez, N., Barto, A. G., & Singh, S. P. (2005). Intrinsically motivated reinforcement learning. In *Advances in Neural Information Processing Systems* (pp. 1281–1288).
- Coenen, A., Nelson, J. D., & Gureckis, T. M. (2017). Asking the right questions about human inquiry. *PsyArXiv*, 13.
- Cohen, J. D., McClure, S. M., & Angela, J. Y. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 362(1481), 933–942.
- Colas, C., Sigaud, O., & Oudeyer, P.-Y. (2018). GEP-PG: Decoupling Exploration and Exploitation in Deep Reinforcement Learning Algorithms. *arXiv preprint arXiv:1802.05054*.
- Daw, N. D., O’doherly, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, 441(7095), 876.
- Dayan, P. (2013). Exploration from generalization mediated by multiple controllers. In *Intrinsically motivated learning in natural and artificial systems* (pp. 73–91). Springer, Berlin, Heidelberg.
- Dubey, R., Griffiths, T. L., & Lombrozo, T. (2019). If it’s important, then I am curious: A value intervention to induce curiosity. In *Proceedings of the Annual Conference of the Cognitive Science Society*.
- Düzel, E., Bunzeck, N., Guitart-Masip, M., & Düzel, S. (2010). Novelty-related motivation of anticipation and exploration by dopamine (NOMAD): implications for healthy aging. *Neuroscience & Biobehavioral Reviews*, 34(5), 660–669.
- Gabaix, X., Laibson, D., Moloche, G., & Weinberg, S. (2006). Costly information acquisition: Experimental analysis of a boundedly rational model. *American Economic Review*, 96(4), 1043–1068.
- Geana, A., Wilson, R., Daw, N. D., & Cohen, J. (2016). Boredom, information-seeking and exploration. In *Proceedings of the Annual Conference of the Cognitive Science Society*.

- Gentry, J. W., Burns, A. C., Dickinson, J. R., Putrevu, S., Chun, S., Hongyan, Y., . . . Gentry, R. A. (2014). Managing the curiosity gap does matter: What do we need to do about it? *Developments in Business Simulation and Experiential Learning*, 29.
- Gigerenzer, G., & Garcia-Retamero, R. (2017). Cassandra's regret: The psychology of not wanting to know. *Psychological review*, 124(2), 179.
- Gobet, F. (2016). From bounded rationality to expertise. In *Minds, Models and Milieux* (pp. 151–166). Palgrave Macmillan, London.
- Golman, R., Hagmann, D., & Loewenstein, G. (2017). Information avoidance. *Journal of Economic Literature*, 55(1), 96–135.
- Gottlieb, J., & Oudeyer, P.-Y. (2018). Towards a neuroscience of active sampling and curiosity. *Nature Reviews Neuroscience*, 1.
- Gottlieb, J., Oudeyer, P.-Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in cognitive sciences*, 17(11), 585–593.
- Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in cognitive science*, 7(2), 217–229.
- Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive psychology*, 51(4), 334–384.
- Gruber, M. J., Gelman, B. D., & Ranganath, C. (2014). States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit. *Neuron*, 84(2), 486–496.
- Hertwig, R., & Engel, C. (2016). Homo ignorans: Deliberately choosing not to know. *Perspectives on Psychological Science*, 11(3), 359–372.

- Hightow, L. B., Miller, W. C., Leone, P. A., Wohl, D., Smurzynski, M., & Kaplan, A. H. (2003). Failure to return for HIV posttest counseling in an STD clinic population. *AIDS Education and Prevention, 15*(3), 282–290.
- Hogg, M. A. (2000). Subjective uncertainty reduction through self-categorization: A motivational theory of social identity processes. *European review of social psychology, 11*(1), 223–255.
- Howard, R. A. (1966). Information value theory. *IEEE Transactions on systems science and cybernetics, 2*(1), 22–26.
- Hsee, C. K., & Ruan, B. (2016). The Pandora effect: The power and peril of curiosity. *Psychological Science, 27*(5), 659–666.
- Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T. Y., & Camerer, C. F. (2009). The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science, 20*(8), 963–973.
- Kidd, C., & Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron, 88*(3), 449–460.
- Law, E., Yin, M., Goh, J., Chen, K., Terry, M. A., & Gajos, K. Z. (2016). Curiosity killed the cat, but makes crowdwork better. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 4098–4110).
- Lehman, J., & Stanley, K. O. (2011). Abandoning objectives: Evolution through the search for novelty alone. *Evolutionary computation, 19*(2), 189–223.
- Lieder, F., Griffiths, T., & Goodman, N. (2012). Burn-in, bias, and the rationality of anchoring. In *Advances in Neural Information Processing Systems* (pp. 2690–2798).
- Liquin, E. G., & Lombrozo, T. (2018). Determinants and Consequences of the Need for Explanation. In *Proceedings of the Annual Conference of the Cognitive Science Society*.

- Litman, J., Hutchins, T., & Russon, R. (2005). Epistemic curiosity, feeling-of-knowing, and exploratory behaviour. *Cognition & Emotion*, *19*(4), 559–582.
- Litman, J., & Jimerson, T. L. (2004). The measurement of curiosity as a feeling of deprivation. *Journal of Personality Assessment*, *82*(2), 147–157.
- Litman, J., & Spielberger, C. D. (2003). Measuring epistemic curiosity and its diversive and specific components. *Journal of personality assessment*, *80*(1), 75–86.
- Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological bulletin*, *116*(1), 75.
- Marr, D. (1982). *Vision: A computational approach*. San Francisco, CA: Freeman.
- Marvin, C. B., & Shohamy, D. (2016). Curiosity and reward: Valence predicts choice and information prediction errors enhance learning. *Journal of Experimental Psychology: General*, *145*(3), 266.
- Mehta, H., Dubey, R., & Lombrozo, T. (2018). Your liking is my curiosity: a social popularity intervention to induce curiosity. In *Proceedings of the Annual Conference of the Cognitive Science Society*.
- Miščević, N. (2016). Curiosity – The Basic Epistemic Virtue. *Moral and Intellectual Virtues in Western and Chinese Philosophy: The Turn toward Virtue*. London: Routledge, 145–163.
- Murayama, K., FitzGibbon, L., & Sakaki, M. (2018). Process account of curiosity and interest: A reward learning model of knowledge acquisition. *OSF Preprints*.
- Oudeyer, P.-Y., & Kaplan, F. (2006). Discovering communication. *Connection Science*, *18*(2), 189–206.
- Oudeyer, P.-Y., Kaplan, F., & Hafner, V. V. (2007). Intrinsic motivation systems for autonomous mental development. *IEEE transactions on evolutionary computation*, *11*(2), 265–286.

- Pathak, D., Agrawal, P., Efros, A. A., & Darrell, T. (2017). Curiosity-driven exploration by self-supervised prediction. In *International Conference on Machine Learning (ICML)* (Vol. 2017).
- Pierce, J. P., Distefan, J. M., Kaplan, R. M., & Gilpin, E. A. (2005). The role of curiosity in smoking initiation. *Addictive behaviors, 30*(4), 685–696.
- Pierson, E., & Goodman, N. (2014). Uncertainty and denial: a resource-rational model of the value of information. *PloS one, 9*(11), e113342.
- Pluck, G., & Johnson, H. (2011). Stimulating curiosity to enhance learning. *GESJ: Education Sciences and Psychology, 2*.
- Ranganath, C., & Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. *Nature Reviews Neuroscience, 4*(3), 193–202.
- Rich, A., & Gureckis, T. (2014). The value of approaching bad things. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 36).
- Robbins, H., et al. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society, 58*(5), 527–535.
- Rossing, B. E., & Long, H. B. (1981). Contributions of curiosity and relevance to adult learning motivation. *Adult Education, 32*(1), 25–36.
- Sanborn, A. N., & Chater, N. (2016). Bayesian brains without probabilities. *Trends in cognitive sciences, 20*(12), 883–893.
- Schmidhuber, J. (1991). Curious model-building control systems. In *IEEE International Joint Conference on Neural Networks, 1991*. (pp. 1458–1463).
- Schmidhuber, J. (2010). Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development, 2*(3), 230–247.



- Schmidhuber, J. (2013). Powerplay: Training an increasingly general problem solver by continually searching for the simplest still unsolvable problem. *Frontiers in psychology*, 4, 313.
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(4), 623–656.
- Shettleworth, S. J. (1999). Cognition, Evolution, and Behavior. *Trends in Cognitive Sciences*, 3(12).
- Simon, H. A. (1990). Bounded rationality. In *Utility and probability* (pp. 15–18). Palgrave Macmillan, London.
- Simon, H. A. (2001). “Seek and Ye Shall Find” How Curiosity Engenders Discovery. In *Designing for Science* (pp. 19–32). Psychology Press.
- Smock, C. D., & Holt, B. G. (1962). Children’s Reactions to Novelty: An Experimental Study of “Curiosity Motivation”. *Child Development*, 631–642.
- Son, L. K., & Sethi, R. (2006). Metacognitive control and optimal learning. *Cognitive Science*, 30(4), 759–774.
- Sun, Y., Gomez, F., & Schmidhuber, J. (2011). Planning to be surprised: Optimal bayesian exploration in dynamic environments. In *International Conference on Artificial General Intelligence* (pp. 41–51).
- Sweeny, K., Melnyk, D., Miller, W., & Shepperd, J. A. (2010). Information avoidance: Who, what, when, and why. *Review of general psychology*, 14(4), 340.
- Tang, H., Houthoofd, R., Foote, D., Stooke, A., Chen, O. X., Duan, Y., . . . Abbeel, P. (2017). # Exploration: A Study of Count-Based Exploration for Deep Reinforcement Learning. In *Advances in Neural Information Processing Systems* (pp. 2750–2759).

- Tao, G., Branson, B. M., Kassler, W. J., & Cohen, R. A. (1999). Rates of receiving HIV test results: data from the US National Health Interview Survey for 1994 and 1995. *Journal of acquired immune deficiency syndromes*, 22(4), 395–400.
- Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie*, 20(4), 410–433.
- Von Stumm, S., Hell, B., & Chamorro-Premuzic, T. (2011). The hungry mind: Intellectual curiosity is the third pillar of academic performance. *Perspectives on Psychological Science*, 6(6), 574–588.
- Wiggin, K. L., Reimann, M., Jain, S. P., & Herr, P. M. (2018). Curiosity tempts indulgence. *Journal of Consumer Research*, Advanced online publication. doi: 10.1093/jcr/ucy055.
- Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore–exploit dilemma. *Journal of Experimental Psychology: General*, 143(6), 2074.
- Wittmann, B. C., Bunzeck, N., Dolan, R. J., & Düzel, E. (2007). Anticipation of novelty recruits reward system and hippocampus while promoting recollection. *Neuroimage*, 38(1), 194–202.
- Wittmann, B. C., Daw, N. D., Seymour, B., & Dolan, R. J. (2008). Striatal activity underlies novelty-based choice in humans. *Neuron*, 58(6), 967–973.

## Appendix A

### Connections to the multi-armed bandit problem

The information-seeking problem described in our framework is related to a problem that has been studied extensively in decision-making literature in the form of multi-armed bandit problems (Robbins et al., 1952; Daw et al., 2006; Cohen et al., 2007; Bubeck et al., 2012; Wilson et al., 2014). In this section we describe how our information-seeking problem relates to the multi-armed bandit problem and how the optimal solution derived in our problem relates to the optimal strategy for solving the multi-armed bandit problem.

#### Multi-armed bandit

A multi-armed bandit is similar to a traditional slot machine (called the one-armed bandit) but has more than one arm. When pulled, each arm provides a reward drawn from a distribution specific to that arm. The player begins without any knowledge about the reward distribution of the arms and the goal of the player is to maximize the rewards obtained through repeated trials. The optimal strategy to maximize rewards requires choosing between exploration (sampling to obtain information about various options) and exploitation (obtaining the highest known reward by choosing current best option), often referred to as the explore/exploit trade-off.

#### Value of Information

One way of balancing exploration and exploitation is to assign a value to the information gained while pulling a particular arm. The basic idea underlying this is that exploration is useful because it gains information. In the decision-making literature, this notion is called the ‘value of information’ and it measures how much providing an information allows a decision-maker to improve its decision (Howard, 1966). Thus, if the decision-maker knows the value of information, then the decision-maker can trade-off exploration and exploitation optimally. Value of information, *VOI*, can be quantified by measuring the difference between maximum expected payoffs when the

decision-maker is provided with the information and when it is not,

$$VOI = \left( \sum_{e'} P(e'|e) \cdot MEU(e, e') \right) - MEU(e). \quad (11)$$

Here,  $MEU(e)$  is the maximum expected utility i.e. the value of the current best action  $a$  under evidence  $e$  defined as,

$$MEU(e) = \operatorname{argmax}_a \sum_s P(s|a, e) U(s) \quad (12)$$

where  $P(s|a, e)$  is probability of outcome state  $s$  conditional on action  $a$  and evidence  $e$ , and  $U(s)$  is the utility of that state  $s$ . Further,  $MEU(e, e')$  is the value of the new best action after new evidence  $e'$  is obtained,

$$MEU(e, e') = \operatorname{argmax}_a \sum_s P(s|a, e, e') U(s). \quad (13)$$

The explore/exploit tradeoff manifests because an agent has to assess both the expected immediate reward and the VOI associated with each action. If the VOI for an action is large, it may make sense to take that action even if it has low expected reward – exploration. By contrast, if the VOI is small for all actions, the agent should focus on the best action discovered so far – exploitation.

$VOI$  is higher when there is greater uncertainty and therefore it makes sense for a rational agent to explore novel actions more, assuming those actions have similar expected reward. In fact, various algorithms have been proposed to solve multi-armed bandit problems that rely on the principle of exploring arms for which the agent has the highest uncertainty (Auer et al., 2002). However, behavioral results on curiosity present a dilemma – instead of always seeking novel stimuli, people often seek stimuli for which they have moderate confidence (in line with empirical findings 2-4). That is, the behavioral findings on curiosity are in contrast to the predictions made by VOI on the multi-armed bandit problem.

### **Resolving the gap**

Our rational analysis shares similarities to the multi-armed bandit problem but it also reformulates the multi-armed bandit problem by making some crucial adjustments to it. As a

consequence, the resulting optimal solution to our information-seeking problem addresses the above gap between behavioral results on curiosity and VOI.

**Re-considering the multi-armed bandit problem.** The problem considered in our rational analysis is different from the multi-armed bandit problem in two crucial ways. First, in our information-seeking problem, there are no immediate rewards associated with taking any actions – rewards are entirely determined by the stimuli the agent encounters at a later point. Second, those later stimuli are randomly sampled from a probability distribution. In the multi-armed bandit problem, any action undertaken by the agent results in an immediate reward. Furthermore, the agent makes choices about the opportunities it has to receive a reward. This is analogous to an agent *choosing* which stimuli it encounters in the future in our problem.

**Curiosity and Value of Information.** We saw that the optimal solution to our information-seeking problem turned out to be simply computing the change in value of knowledge  $V$  with respect to exposure  $h$ ,

$$\Omega_k \approx p_k \cdot \frac{dc_k}{dh_k}. \quad (14)$$

Expanding the derivative further, we can rewrite Equation 14 as follows,

$$\Omega_k \approx \lim_{x \rightarrow 0} p_k \cdot \frac{c'_k - c_k}{x}. \quad (15)$$

Here,  $c'_k$  is the confidence of the  $k$ th stimulus computed by the agent upon increase in exposure  $h_k$  by  $x$  units. In other words, we can think of curiosity as approximating the difference between value of knowledge  $V$  when the agent explores a stimulus and when it does not,

$$\Omega_k \approx \lim_{x \rightarrow 0} \frac{V' - V}{x}, \quad (16)$$

where  $V'$  is the value of knowledge upon increase of exposure of the  $k$ th stimulus by  $x$  units.

According to our framework, the agent explores the stimulus that maximizes Equation 16, i.e. the stimulus that maximally increases value of knowledge. This is similar to an agent that acts by selecting stimuli that maximize VOI as per Equation 11. Here, utility is comparable to confidence  $c_k$  (since greater confidence implies higher probability of reward) and conditional probability  $P(s|a)$

is similar to probability of occurrence  $p_k$  (as probability of occurrence captures how likely the  $k$ th stimulus is to occur in the future).

While our optimal solution in the form of curiosity shares similarities with VOI, the differences between our information-seeking problem and the multi-armed bandit problem are also reflected in differences in the optimal strategies for solving these two problems (i.e. curiosity and VOI). Recall that, in contrast to the multi-armed bandit problem, in our problem there is no tension between exploration and exploitation. The agent can simply seek to maximize value of knowledge – to explore – without having to worry about the immediate rewards that result from its actions. This results in a major deviation from the optimal strategy in a multi-armed bandit problem (i.e. VOI). Indeed, our analysis shows why exploration should not always be guided by novelty (in contrast to predictions made by VOI). Because the distribution from which future stimuli are drawn is not under control of the agent, and can reflect the frequency with which those stimuli are encountered in the past (Figure 3(a)), it may be rational to focus on gathering information about stimuli that have occurred a moderate number of times already. In contrast, in the multi-armed bandit problem, the agent makes a choice that influences both what it learns and which stimulus it is presented with. The probability of encountering a stimulus in the future in a multi-armed bandit problem is thus by definition independent of the past. When the past and future are independent of each other (Figure 4(a)), then the agent should seek out the most uncertain stimuli i.e. novel stimuli.

In this way, we see that re-considering the multi-armed bandit problem as a problem that is consistent with the problem in our framework, leads to an optimal solution that is closer to people's curiosity and also explains why VOI cannot capture all aspects of people's curiosity. Importantly, while curiosity shares similarities to VOI, unlike VOI, people's curiosity can be driven by moderate amount of uncertainty (and learning progress) as well as probability of occurrence. People presumably encounter situations that are consistent with both of these kinds of problems (i.e. our information-seeking problem and the multi-armed bandit problem). Choosing a restaurant for dinner is a good example of a multi-armed bandit problem, since you get immediate rewards from your meal as well as information about the quality of the restaurant. But choosing what to study to

prepare for next week's test has a structure consistent with the problem we have focused on, since studying is never likely to be immediately rewarding. Our choice of this problem was inspired in part by the tasks used to study curiosity, in which people do not have to exploit rewards in the present but only need to gather information to be used in the future. While the multi-armed bandit problem has offered a plethora of insights on the decision-making behavior of people, we believe that our task can offer important insights into the nature of human curiosity that the multi-armed bandit problem alone cannot capture.

## Appendix B

### Individual-level analysis of Experiment 1 data

In this section, we report individual-level analysis of the Experiment 1 data. We compared the proportion of participants that showed a decreasing relationship vs. an inverted U-shape relationship between curiosity and confidence for the two conditions for both Phase 1 and Phase 2. To do so, we first binned participant's curiosity into three groups – low (confidence rating between 0 and 30), medium (confidence rating between 40 and 70), and high (confidence rating between 80 and 100). We then computed the proportion of participants that showed a decreasing relation i.e. where  $\text{low} > \text{medium}$  and  $\text{medium} > \text{high}$  vs. an inverted U-shape i.e. where  $\text{low} < \text{medium}$  and  $\text{medium} \geq \text{high}$  (low and high could be in any relationship with each other) vs. all other patterns. Note that a total of 13 binning combinations were possible with 1 for decreasing, 4 for inverted U-shape, and 8 for all other patterns (also refer to the horizontal-axis of any plot in Figure 11 for a visual representation of the 13 combinations). After performing the binning, we computed the empirical chance level for each bin by randomly shuffling the data (total of 1000 times) and then calculating the mean proportion of participants in each bin.

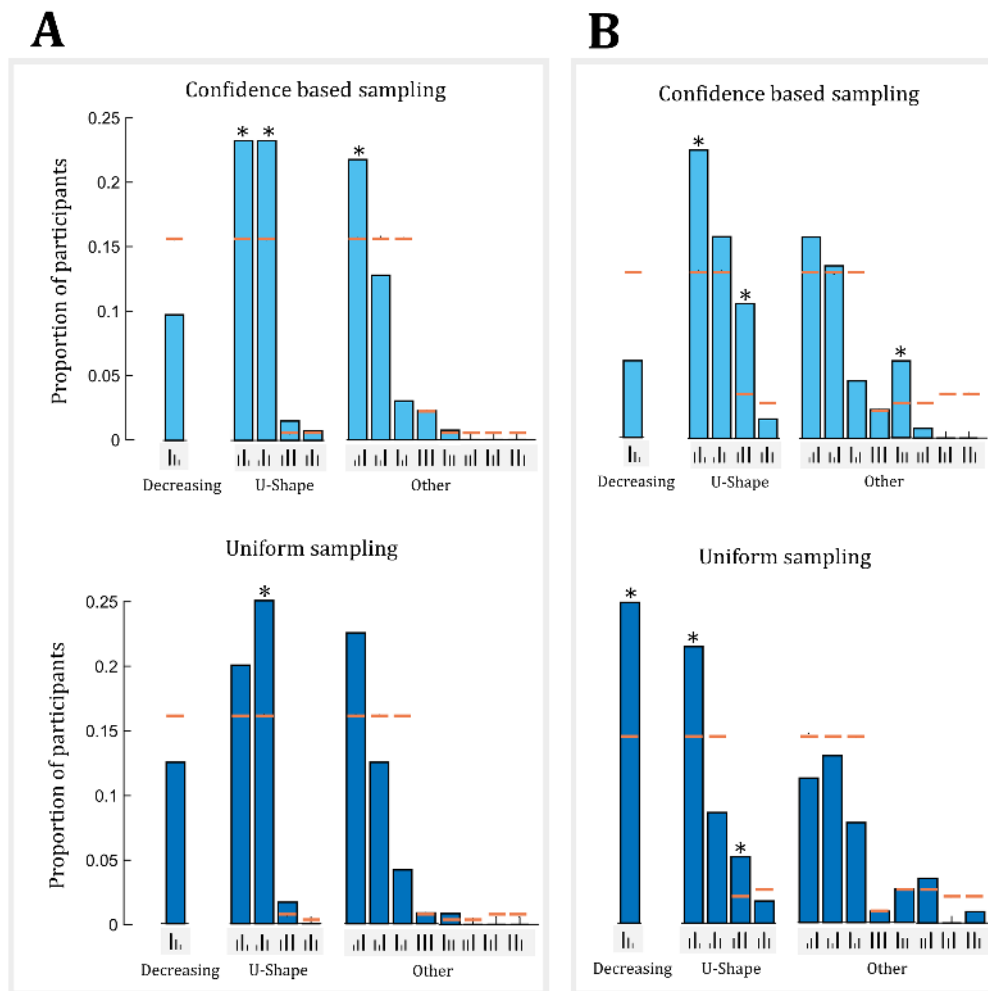
**Phase 1.** Figure 11(a) plots the results of the above analysis for both confidence based and uniform sampling condition. Qualitatively, we observe that these results are consistent with our model predictions in that participants are more likely to show an inverted U-shape pattern (compared to decreasing or other patterns) for both the conditions. We first calculated a Chi-square

test to compare the observed proportion of participants and empirical chance level for both the conditions. For both the confidence based and uniform sampling condition a significant difference was found ( $\chi^2(13) = 35.6, p < 0.01$  and  $\chi^2(13) = 29.1, p < 0.01$  respectively) showing that the observed distributions for both conditions were significantly different than chance. Next, we conducted individual binomial tests to compare all the 13 bins to their respective chance levels for both the conditions. For the confidence based sampling condition, the binomial tests indicated that 2 of the 4 bins within the inverted U-shape pattern were greater than the chance level,  $p = 0.015$  and  $0.014$  (1-sided). Furthermore, only 1 of the 8 bins within the other pattern was greater than the chance level,  $p = 0.04$  (1-sided) (also refer to Figure 11(a)). For the uniform condition, the binomial tests indicated that 1 of the 4 bins within the inverted U-shape pattern was greater than the chance level,  $p = 0.005$  (1-sided) and no bin within the other pattern was greater than the chance level. Importantly, as per our model's prediction, the bin within the decreasing pattern was not significantly greater than the chance level for both the conditions.

**Phase 2.** Figure 11(b) plots the results for the two conditions for Phase 2. Qualitatively, we observe that these results are again consistent with our model predictions. More specifically, for the confidence based condition, participants are again more likely to show an inverted U-shape pattern (compared to decreasing or other patterns). On the other hand, for the uniform condition, a much larger proportion of participants show a decreasing pattern (compared to the confidence based condition). To test this, we first calculated a Chi-square test to compare the observed proportion of participants and empirical chance level for both the conditions. For both the confidence based and uniform sampling condition a significant difference was found ( $\chi^2(13) = 59.4, p < 0.01$  and  $\chi^2(13) = 29.7, p < 0.01$  respectively) showing that the observed distributions for both conditions were significantly different than chance. Next, we again conducted individual binomial tests to compare all the 13 bins to their respective chance levels for both the conditions. For the confidence based sampling condition, similar to Phase 1, the binomial tests indicated that 2 of the 4 bins within the inverted U-shape pattern were greater than the chance level,  $p = 0.002$  and  $0.0002$  (1-sided). Furthermore, only 1 of the 8 bins within the other pattern was greater than the chance level,



$p = 0.03$  (1-sided) (also refer to Figure 11(b)). And again similar to Phase 1, as per our model's prediction, the bin within the decreasing pattern was not significantly greater than the chance level for the confidence based condition. For the uniform condition, the binomial tests indicated that 2 of the 4 bins within the inverted U-shape pattern were greater than the chance level,  $p = 0.02$  and  $0.03$ (1-sided) and no bin within the other pattern was greater than the chance level. However, in contrast to Phase 1, the bin within the decreasing pattern was now significantly greater than the chance level for the uniform sampling condition,  $p = 0.002$ . These results confirm our model predictions and demonstrate that participants show an increased tendency to exhibit a decreasing relationship of curiosity with confidence in the uniform sampling condition.



*Figure 11. Individual level analysis for Experiment 1.* The horizontal-axis shows all of the possible bin patterns and the vertical-axis shows the proportion of participants in each bin. The orange lines indicate the chance level for each bin. Asterisk indicates the bins that were significantly greater than the chance level. (A) Results from Phase 1. Proportion of participants that showed decreasing vs. inverted U-shape vs. other pattern for the confidence sampling condition (top) and the uniform condition (bottom). For both conditions, the proportion of subjects exhibiting the U-shaped patterns is higher than proportion of subjects exhibiting the decreasing patterns. (B) Results from Phase 2. Proportion of participants that showed decreasing vs. inverted U-shape vs. other pattern for the confidence sampling condition (top) and the uniform condition (bottom). Consistent with our model predictions, the proportion of participants that exhibit a decreasing relation rises sharply for the uniform sampling condition (when compared to Phase 1).