Curiosity Killed the Cat, but Makes Crowdwork Better

Edith Law[1], **Ming Yin**[2], **Joslin Goh**[1], **Kevin Chen**[1], **Michael Terry**[1], **Krzysztof Z. Gajos**[2]
[1]University of Waterloo, [2]Harvard University
edith.law@uwaterloo.ca, mingyin@fas.harvard.edu, jtcgoh@uwaterloo.ca, krschen@uwaterloo.ca, mterry@cs.uwaterloo.ca, kgajos@eecs.harvard.edu

ABSTRACT
Crowdsourcing systems are designed to elicit help from humans to accomplish tasks that are still difficult for computers. How to motivate workers to stay longer and/or perform better in crowdsourcing systems is a critical question for designers. Previous work have explored different motivational frameworks, both extrinsic and intrinsic. In this work, we examine the potential for *curiosity* as a new type of intrinsic motivational driver to incentivize crowd workers. We design crowdsourcing task interfaces that explicitly incorporate mechanisms to induce curiosity and conduct a set of experiments on Amazon's Mechanical Turk. Our experiment results show that curiosity interventions improve worker retention without degrading performance, and the magnitude of the effects are influenced by both the personal characteristics of the worker and the nature of the task.

Author Keywords
Motivation; Crowdsourcing; Curiosity.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION
Crowdsourcing systems organize individuals to solve complex problems that are otherwise difficult to solve computationally [30]. A key challenge in creating these systems is to sufficiently incentivize individuals to participate and keep producing high quality work. In support of this goal, a wide range of incentive mechanisms have been designed and studied in crowdsourcing and social computing settings, including monetary payment [21, 57, 62], gamification techniques [59], social comparisons [18], visualizations and facilitation [23, 26, 37, 47], and virtual reward systems [12]. The efficacy of these approaches ranges quite a bit, making this an active, open research area.

In this work, we examine a specific type of intrinsic motivation, namely *curiosity*, to motivate crowd workers. Curiosity is defined as "the desire to know, to see, or to experience that motivates exploratory behavior directed towards the acquisition of information" [31]. Numerous theories have offered explanations for the origin, mechanics, and effects of curiosity [31]. Our work is inspired by *information gap theory* [34], a contemporary model of curiosity which posits that curiosity arises due to a gap between what one knows and what one wants to know. According to this theory, when people are made aware of this gap in their knowledge, they become curious and engage in information-seeking behavior to complete their knowledge and resolve the uncertainty. This innate desire to satisfy one's curiosity suggests a way to design and structure crowdsourcing work: if tasks can be designed to stoke one's curiosity, and completing the task provides the requisite information to satisfy that curiosity, then the system may create a more enriching experience for workers.

In this paper, we report results from a set of studies in which we explicitly incorporate mechanisms to induce curiosity in workers performing an audio transcription task. Importantly, the curiosity stimuli are related to the task itself, creating a synergy between completing the task and satisfying one's curiosity. Our results indicate that curiosity is an effective means of motivating workers. In particular, we

- introduce a framework for studying *information* as a currency for motivating participation in crowdsourcing;
- operationalize the concept of curiosity in the task interface using ideas from information gap theory;
- show that workers are more likely to complete more tasks, while maintaining a high level of accuracy, when presented with curiosity-inducing stimuli;
- present observations showing that individual workers respond differently to curiosity interventions; and
- demonstrate the interaction between curiosity interventions and task characteristics (e.g., its inherent interestingness), showing that the effects of curiosity interventions are larger for tasks that are less interesting.

While this research is centered around curiosity interventions in a paid crowdsourcing environment, our findings have relevance to other crowdsourcing contexts, such as citizen science, where an important goal is to stoke people's curiosity about the broader context (e.g., scientific questions) driving the mundane data processing tasks. The rest of this paper describes related work, then details our studies and findings supporting these contributions, followed by a discussion of the design implications.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CHI'16, May 07–12, 2016, San Jose, CA, USA.
Copyright © 2016 ACM 978-1-4503-3362-7/16/05...$15.00.
DOI: http://dx.doi.org/10.1145/2858036.2858144

BACKGROUND

Curiosity as Motivation

Curiosity is an old, yet critical, concept in the psychology of motivation. Various exploratory or information-seeking behaviors have been defined as "curiosity". For example, animals' orienting response (i.e., their immediate response to changes in their environment, such as change in illumination or unusual sound) is considered "perceptual curiosity", whereas humans' desire for information and knowledge is categorized as "epistemic curiosity" [5]. *Trait* curiosity is a persistent personality attribute, while curiosity aroused by external situations is called *state* curiosity, which is the focus of this study. Curiosity can be triggered by stimuli that are novel (e.g., unexpected changes or violated expectations [25]), conflicting (i.e., arousing two or more incompatible responses), uncertain (i.e., leading to outcomes that one is not sure about), and complex (e.g., presenting variety and diversity) [5]. Although curiosity has been consistently recognized as an important influence on behavior [34], there is no single, agreed-upon model to characterize curiosity's motivational nature [55]. Instead, psychologists have proposed a number of theories [6, 56, 13, 31, 6, 5, 4, 43, 55] to explain curiosity and people's information seeking behavior.

In this work, we focus on one well-established theory of curiosity, Lowenstein's *information gap theory* [34]. This theory posits that curiosity arises when there is an information gap between what one knows (knowledge baseline) and what one wants to know (information goal). The information goal is subjective, meaning that the same stimuli will arouse different levels of curiosity for each person, depending on the individual's perception about what she does or does not know.

More formally, information gap theory represents information as a unidimensional concept quantified by an entropy coefficient, $I = -\sum_{i=1}^{n} p_i log_2 p_i$. Given this, an individual's knowledge gap can be measured by the difference between the entropy of the information goal and knowledge baseline. This quantification of information is approximate and serves as a crude proxy that nonetheless provides a way to make predictions about how curiosity might increase or decrease depending on the availability of information, and an individual's perception of the knowledge gap. Our research adopts a common methodology [34], which only considers ordinal predictions (e.g., "curiosity will increase with information"), and does not attempt to measure the precise magnitude of the information gap.

Using this formulation, curiosity is expected to increase with the accumulation of information, as it creates a sudden shift of attention from focusing on the known (i.e., the existing information) to the unknown (i.e., the missing information). Inspired by approach-gradient theory [42, 27] and Gestalt psychology [41], the theory further predicts that the motivation to seek information becomes most intense as one approaches the answer, creating the urge to "complete" the picture. This is used to explain why curiosity is greater for insight problems (where a single piece of information may resolve the entire problem) than for incremental problems (where a single piece of information only provides small progress towards a solution). For example, in a series of experiments [32], subjects were shown a list of vocabulary words, and asked to evaluate for each word whether they knew the definition, knew the definition only by the tip-of-their-tongue, or did not know the definition. Their results showed that individuals were most curious about those tip-of-their-tongue vocabulary words, for which they had some, but not complete, knowledge. In other words, people are unlikely to be curious about information on a topic that they have zero or complete knowledge of, while curiosity is at its height when the information gap becomes quite small, but is not completely closed.

A second implication is that people are more likely to become curious if they have prior knowledge about a particular domain, since a higher knowledge baseline creates a smaller information gap. Indeed, Jones [24] found that knowledge about a particular domain is correlated with curiosity in that domain, and Berlyne [5] found that questions about more familiar entities evoke greater curiosity.

Finally, curiosity requires *attention* to the information gap: to induce one's curiosity, the information gap must be *salient*, so that the individual can recognize that some information is missing. Also, people will only expose themselves to curiosity-inducing stimuli if there is a non-trivial chance that their curiosity *will be satisfied*, and without long delays. Based on these two implications, Lowenstein suggests that one way to induce curiosity is to ask people to make guesses, which makes the information gap more salient and accurately perceived [34]. The curiosity-inducing designs that we propose and evaluate in this paper are based on these key insights.

Practical Applications and Related Constructs

The idea of withholding information to induce the sense of curiosity have been broadly studied and applied in various settings such as games [35], software engineering [61], interactive designs [16, 58], business [1, 40] and education [36, 45, 51, 66]. For example, the toy company HotWheels designs "mystery cars" for which the identities are unknown until purchase to boost sales, and the social media company LinkedIn binds the upgrade to the premium account with the reveal of hidden profiles of one's followers [1]. Some of the methodologies for stimulating and sustaining curiosity in the classroom include the use of questions, and problem-solving sessions where students each hold partial information, thus requiring them to exchange information in order to accomplish a joint task [45, 51, 66].

In academic research, curiosity is a distinct construct, though closely related to reinforcement schedules [17, 64], as well as goal-setting theories of motivation [33], which have been studied in several social computing [3, 18, 65] contexts. The idea of *suspense* [8, 29] is a close counterpart to curiosity, with the key difference being the level of emotional engagement with the uncertain outcome.

Incentive Mechanisms in Crowdsourcing

The design of incentive mechanisms for engaging workers and motivating them to produce high-quality work has been an extensively studied topic in crowdsourcing. One of the most natural incentives to use in a paid crowdsourcing environment

is monetary rewards. To this end, researchers have conducted a large number of experiments to understand the effects of financial incentives on crowd work quantity and quality [38, 48, 21, 63, 19, 20]. It has been observed that while a larger reward alone incentivizes workers to complete more tasks [38, 48], the effect on the work quality in a particular task depends on a combination of several factors: whether the magnitude of the reward is large enough [20], whether the payment level changes in subsequent tasks [63], and the specific characteristics of the task [20].

Intrinsic motivation has also been examined in paid crowdsourcing contexts. Studies found that when tasks are framed as something meaningful to do, workers are more productive and complete more tasks [2, 9, 48, 53]. Likewise, gamification techniques applied to crowdsourcing task interfaces have been shown to improve both worker engagement and effectiveness [14]. More recently, Dai et al. explored the effects of "micro-diversions" (i.e., breaks during which workers can engage with an enjoyable, task-irrelevant activity) in a long (1+ hours) sequence of tasks [11]. One of the "micro-diversions" examined was an award-winning comic that is shown one page at a time during each break; this curiosity-inducing mechanism was shown to be successful in increasing workers' output. Different from our work, the curiosity-inducing comic is *irrelevant* to the actual tasks that workers are asked to do.

In contrast to paid crowdsourcing systems, volunteer-based crowdsourcing platforms, such as citizen science projects [46, 49, 54], rely mostly on the design of intrinsic motivational drivers to incentivize the crowd. In these contexts, there have been numerous attempts to engineer virtual reward systems [12], add gamification elements such as points, badges and leaderboards [59, 7, 10], and introduce direct communications [52] in order to encourage crowd participation.

OPERATIONALIZING CURIOSITY

The central idea in this research is to embed curiosity-inducing designs in crowdsourcing task interfaces to improve worker engagement and performance. Guiding our designs are the two primary tenets of information gap theory: curiosity can be induced if (1) people are aware of a salient information gap in their knowledge, and (2) people are provided with a means to help them close the gap. Using these constructs as a foundation, we created *curiosity interventions* which consist of three design concepts:

Information Goal To induce curiosity, we create an information gap by posing a *question* that is *relevant* to the current task at hand.

Gap Salience To make workers aware of what they do not know, we prompt them to *guess the answer to the question*.

Incremental Reveal To help workers close the gap, we reveal information as workers progress with their tasks.

We refer to interfaces that employ these three design concepts as *curiosity-inducing stimuli*, and use them to examine the effects that curiosity can have on crowd workers. In this work, we focus exclusively on curiosity-inducing stimuli that support audio transcription tasks. Audio transcription task is a good candidate for this study because it is a common crowdsourcing task that is non-trivial, and at times difficult, for the inexperienced. Hence, effective motivation is especially needed to engage workers. In addition, transcription is a highly *decomposable* task, as an audio clip can be broken down into arbitrarily small units, which enables us to stoke workers' curiosity by using each unit as a piece of the "puzzle." For example, we may reveal one more sentence after each task to complete the story, or one more clue about the identity of the person being discussed in the article. Such techniques are used in the design of our curiosity interventions.

STUDY DESIGN

To understand the effects of curiosity interventions, we conducted a study on Amazon's Mechanical Turk (MTurk). We describe the task that workers were asked to perform, the curiosity interventions that we designed, specific hypotheses driving our study, and various analysis methods we employed.

Task and Procedure

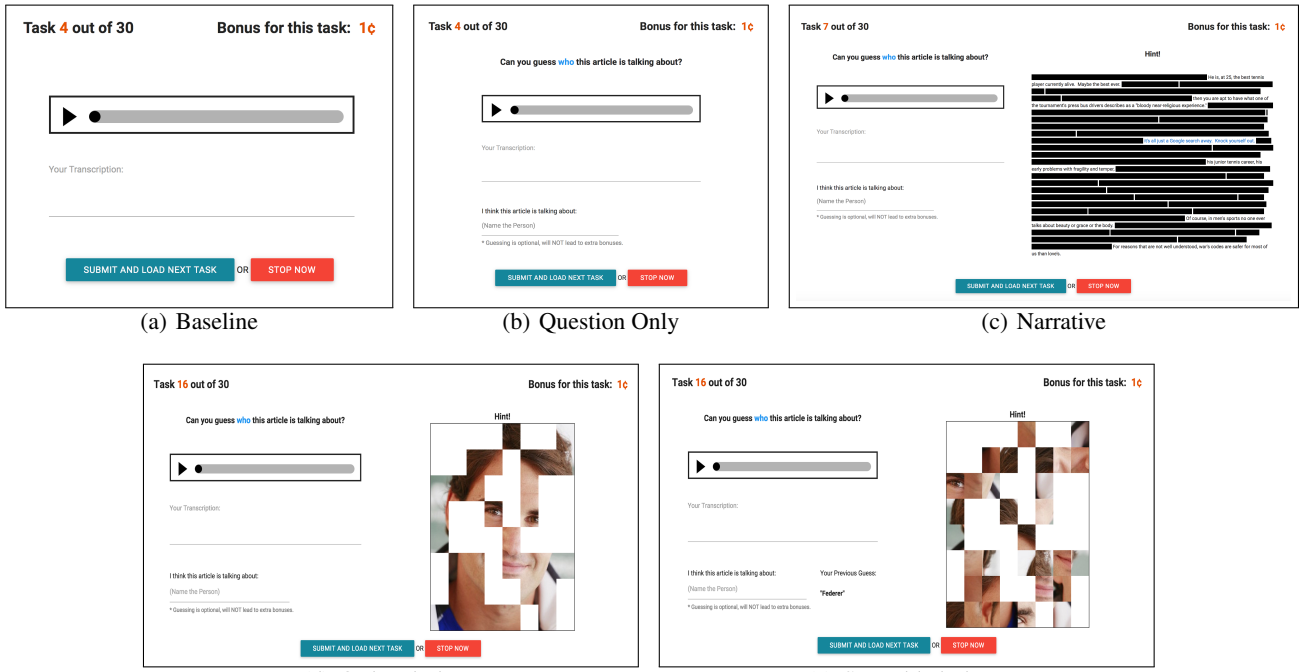
The task used in all of our experiments is audio transcription. The audio files were created by a colleague reading an excerpt from articles drawn from a variety of sources (e.g., novels, editorials, news, textbook). Each audio file is cut into 30 individual audio transcription tasks, which vary in length and difficulty. We then combine the 30 transcription tasks for the same audio file in a fixed, but not sorted, order, and bundle them into one HIT (Human Intelligence Task).

In the HIT, workers are asked to perform *at least* three transcription tasks for a base pay of 45 cents, but can quit at any point thereafter by clicking on the "stop now" button, which brings them to a questionnaire. If workers choose to continue after the 3rd task, they earn a 1-cent bonus for completing each additional transcription task. We indicate that all 30 transcription tasks in the HIT come from the same article. A progress bar present on the interface shows how many tasks are remaining. This design is common for studying intrinsic motivation in paid crowdsourcing environment (e.g., as used in [9])—that is, by implementing a low payment scheme, we can infer that the reasons for workers to continue are intrinsically, rather than extrinsically, motivated. In addition, we avoid variable payments to minimize confounding motivations, e.g., anticipation of surprise bonuses.

Workers are randomly assigned to one of the experimental conditions (described in the next section) as they sign up for the HIT. All our experiments follow a between-subject design, that is, our system ensures that each worker takes our study only once. The HIT is restricted to US workers only. In this work, we are mostly interested in state curiosity and population-level effects; hence, we do not differentiate workers based on personal characteristics or prior experience.

Experimental Conditions: Control and Treatments

As mentioned previously, our curiosity-inducing stimuli consist of three design elements—information goal, gap salience and incremental reveal. By varying *which* of these three design elements are used, we created the following set of control and treatment conditions, as depicted in Figure 1(a) – 1(e).



(a) Baseline (b) Question Only (c) Narrative

(d) Ordered Photo (e) Scrambled Photo

Figure 1. Five experimental conditions with varying curiosity interventions

Baseline (Control) Workers are *not* presented with any curiosity-inducing stimulus.

Question Only Workers are presented with a task-relevant question to induce curiosity, and the ability to guess the answer, but without any incremental reveal of information that would provide a hint to the answer.

Narrative Workers are presented with a task-relevant question, the ability to guess the answer, and a visual representation of the article in which the sentences, except the ones that the worker has transcribed, are obscured (i.e., blacked out).

Ordered Photo Workers are presented with a task-relevant question, the ability to guess the answer, and a partially obscured photo (e.g., the subject of the article) as a hint to the answer. The picture is divided into a 8×8 grid. Two more cells are revealed after each audio segment is transcribed, with the full photo revealed at the last (i.e., 30th) task.

Scrambled Photo Workers are presented with a task-relevant question, the ability to guess the answer, and a partially obscured and *scrambled* photo (e.g., the subject of the article) as a hint to the answer. The grid size and method of reveal is the same as in the ordered photo condition.

The task interface provides an audio player and a textbox to support transcription. For all conditions except the baseline, the task-relevant question is present and serves as an information goal. To make the information gap created by this question more salient, the interface includes a textbox for workers to enter guesses, and continually displays their most recent guess. The system does not provide any accuracy feedback on the guesses, and only reveals the correct answer when the worker reaches the end. The information that is incrementally revealed in the narrative, ordered photo, and scrambled photo conditions serves to reduce the information gap by providing hints to the answer.

Research Questions and Hypotheses

Our study aims to answer three research questions (Q1—Q3).

Q1: Can crowds be motivated by curiosity? We hypothesize that curiosity interventions will affect worker retention and performance in a positive way:

[H1] Workers will complete more tasks if they are presented with a curiosity-inducing stimulus.

[H2] Workers will have a higher probability of completing all 30 tasks if they are presented with a curiosity-inducing stimulus.

[H3] Workers will have similar or better performance (in terms of work quality) if they are presented with a curiosity-inducing stimulus.

Q2: How do individuals respond differently to curiosity interventions? Lowenstein suggests that guessing draws attention to the knowledge gap and may lead to increased curiosity; hence we are particularly interested in understanding how guessing behavior (e.g., whether workers make a correct guess, incorrect guesses, or no guess at all) correlates with the effects of curiosity interventions. We hypothesize that:

[H4] Workers who make a correct guess will complete more tasks, have a higher probability of completion, and have similar or better performance (in terms of work quality) than workers who make an incorrect guess or make no guesses.

Q3: What are the interactions between task characteristics and curiosity interventions? In particular, when the article to transcribe is inherently interesting in and of itself, workers may be eager to know more about it, even *without* curiosity interventions. In contrast, if the article is not interesting, workers are likely to be indifferent, thus requiring explicit interventions to induce their curiosity. Therefore, our final prediction is:

[H5] The effect of a curiosity intervention is larger when the intervention is combined with a task that is less interesting.

Choice of Article

To study Q1—Q3, our first goal is to select a small number of articles that span a wide range in terms of how inherently interesting they are. We considered five candidate articles drawn from diverse sources—an essay about the famous tennis player Roger Federer, a health article on salt and cholesterol, a blogpost about imposter syndrome, a case study from a marketing textbook, and an excerpt from a novel.

We recruited 98 workers from MTurk to perform a series of 10 paired comparisons of the articles (as depicted in Figure 2) as a pilot study. In each task, workers see two articles side-by-side. Each article has all its content blacked out except for 3 randomly chosen sentences. We then ask workers to decide which of the two articles they would like to fully reveal. After 10 rounds of comparisons, we determine the article that each worker wants to reveal the most, and ask her to explain why she finds the article most interesting.

Figure 2. Pilot study: Paired comparison of articles

Given each worker's full ranking of the articles, we assign each article an "interestingness" score, defined as n_{best} - n_{worst}, where n_{best} is the number of workers who voted the article to be the one they most want to reveal, and n_{worst} is the number of workers who voted the article to be the one they least want to reveal. Using this formulation, the interestingness of the imposter syndrome, health and Federer articles are 12, 10 and -22, respectively, indicating that on average, the health and imposter syndrome articles are most inherently interesting, while the Federer article is the least interesting.

An analysis of the textual responses suggests that the reasons for an article being interesting fall under two broad categories: *relevance* and *intrigue*. Some workers found an article to be interesting because it was relevant to them in some way (e.g., "My family has a history of health problems so I would love to read this article," "I used to work in marketing and am interested in online marketing and branding."). In contrast, other workers were intrigued by the revealed sentences (e.g., "It hooked me after I read 'raw energy'. It seemed intriguing," "I'm curious to know what it's all about. Who is the boy? Why is he pretending?", "The sentence started with some mysterious voice the protagonist hears. It kind of made me want to know more about what that voice was about.").

Analysis Methods

Table 1 summarizes the data we use in our analysis, which include worker-initiated actions during the task (e.g., quitting, guessing) as well as responses from the questionnaire. The description column describes the measurement details for the task data and the actual questions / statements associated with the questionnaire data.

In order to capture other factors that may influence retention and performance, we applied a common technique in psychological research to quantify intrinsic motivational factors, and used these factors as covariates in the analysis when appropriate. Specifically, we measure motivational factors using the Intrinsic Motivation Inventory (IMI) [50], a scale that measures factors related to *enjoyment* (how much workers enjoy transcription), *competence* (how competent workers think they are at transcribing) and *effort* (how much effort workers put into the tasks). As shown in Table 1, workers are asked to rate on a 7-point Likert scale about how much they agree with a set of statements related to these three dimensions. For each dimension, we then average the workers' responses (reversing the scores for the negative statements) and use the mean value as the summary statistics for that dimension.

Dependent Variables Quit index (i.e., how many tasks the worker performed before stopping) is used as our dependent variable to understand the effects of curiosity interventions on worker retention, and error rate (i.e., the percentage of errors the worker made in the transcription tasks, as defined in Table 1) is used as our dependent variable to understand the effects of curiosity interventions on worker performance.

Independent Variables The experimental condition the worker was assigned to and the article that the worker was transcribing serve as the independent variables, i.e., factors that are believed to influence worker retention and performance.

Statistical Methods For high-level descriptions of worker retention and performance, we use descriptive statistics (e.g., mean, median) when they are appropriate. We also provide histograms and retention curves to visualize the number of workers who quit or remain after each task.

To examine the effects of curiosity interventions on how many tasks a worker completes and how well a worker performs in the tasks, we conduct one-way analysis of variance (ANOVA) [60] or a Kruskal-Wallis test [28], depending on whether the residuals are normally distributed. Both tests allow us to measure whether there are any statistically significant differences across conditions in terms of the *mean* (or *median*) of the metric that is being examined.

To examine the effects of curiosity interventions on how likely a worker completes all tasks, we first use a proportion test

Task Data	Description
Condition	The experimental condition the worker is assigned to
Quit Index	The number of tasks the worker completes before quitting
Error Rate	The total number of errors the worker makes, divided by the total number of words in the ground-truth transcriptions
Earliest Correct Guess	The number of tasks the worker completes before making a correct guess
Questionnaire Data	**Description**
Enjoyment	Mean value of Likert scale (1-7) responses for the following statements: · *This task did not hold my attention at all. (Negative)* · *While I was doing this task, I was thinking about how much I enjoyed it.* · *This task was fun to do.* · *I thought this was a boring task. (Negative)*
Competence	Mean value of Likert scale (1-7) responses for the following statements: · *I think I did pretty well at this task, compared to other workers.* · *After working at this task for a while, I felt pretty competent.* · *This is a task that I couldn't do very well. (Negative)*
Effort	Mean value of Likert scale (1-7) responses for the following statements: · *I didn't put much energy into this. (Negative)* · *I tried very hard on this task.* · *It was important to me to do well at this task.*
Why Quit	Did you choose to stop before reaching the end (i.e., the 30th task)? If so, why?
Why Persist	Did you stop as soon as you first thought of stopping? If not, why did you persist and continue doing more tasks?

Table 1. Data Summary

[60], which allows us to measure statistically significant differences between conditions in the *proportion* of workers who completed all 30 tasks. We apply Bonferroni correction [15] to account for the bias introduced by multiple comparisons.

By treating whether a worker completes all tasks as a binary variable, we further use a generalized linear model (GLM) [39] with logit link function (also known as the logistic regression model) to model the probability (or odds) of completing all 30 tasks in different conditions. Specifically, to compare the completion probability in conditions with curiosity-inducing stimuli against the baseline condition, we set the baseline condition as the reference. We also control for the influences of other intrinsic motivational factors, such as enjoyment, competence and effort, in our GLM—as an ANOVA on each of these factors suggests that there is no significant difference in them across experimental conditions, we feel comfortable to include them as covariates in our regression model. The fit of each GLM is assessed graphically using the residual plots and other quantitative approaches, including the Hosmer and Lemeshow test [22] and Osius-Rojek test [44], two common goodness-of-fit tests for logistic regression models.[1]

Q1: CAN CROWDS BE MOTIVATED BY CURIOSITY?

Since the Federer article generated the least interest, we selected this article as the worst case scenario to understand whether curiosity interventions affect worker retention and performance. Accordingly, in our experiment, the task-relevant question is "Can you guess who this article is talking about?", and the photo we use in the ordered photo and scrambled photo conditions is the photo of Roger Federer. We recruited 100 workers for each condition, and gathered data from a total of 500 participants. We found 4 workers who were obvious spammers and filtered out these suspicious cases, leaving 496 workers for further analysis.

Effects on Worker Retention

To understand how various curiosity interventions affect worker retention in the transcription tasks, we first plot a histogram showing the number of workers who quit after a certain number of tasks across five experimental conditions (Figure 3). To further illustrate the difference in retention, Figure 4 shows the the retention curves of each condition plotted against the baseline.

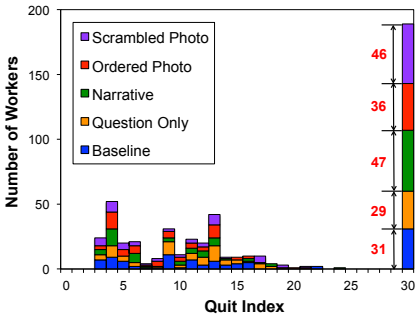
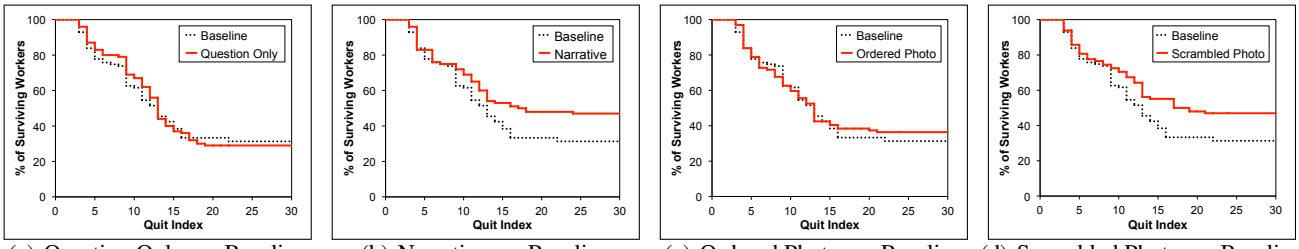


Figure 3. Histogram showing the number of workers who quit after completing X tasks in each experimental condition.

As can be seen in Figure 3 and Figure 4, in general, the majority of the workers either chose to quit when they completed fewer than half of all transcription tasks (i.e. quit before the 15th task), or kept working until they completed all tasks (i.e. quit after completing the 30th task). Table 2 further shows the median, mean, and standard deviation for the number of completed tasks in each experimental condition. The disagreement between means and medians as well as the large standard deviation again indicate the two-sided skew in the distributions of the number of completed tasks, for which measures of center are not the best characterizations. Nevertheless, we still find that workers in experimental conditions with curiosity-inducing stimuli tend to complete a larger number of tasks compared to workers in the baseline condition, which is consistent with our prediction in H1. ANOVA shows that the effects of curiosity interventions on the number of completed tasks is marginally significant, $F(4,491) = 2.17$, $p = 0.07$.

To take a closer look, we first focus on workers who decided to quit before they were halfway through the whole HIT. We find many of them actually quit after they completed the 4th, 9th or 13th task. Interestingly, the sentences that the workers were

[1] Statistically significant results are reported as follows: $p < 0.001$(***), $p < 0.01$(**), $p < 0.05$(*), $p < 0.1$(·).



(a) Question Only vs. Baseline (b) Narrative vs. Baseline (c) Ordered Photo vs. Baseline (d) Scrambled Photo vs. Baseline

Figure 4. Retention curves showing the number of workers who continue working on the tasks (i.e. "survive") after completing X tasks.

asked to transcribe in the 5th, 10th and 14th task are among the longest and most complex sentences in the whole article, implying that workers may have decided to quit because they were deterred by the difficulty of the transcription tasks.

It appears that one of the benefits conferred by presenting curiosity-inducing stimuli on task interfaces is that it nudges workers to persist through difficult spots in the task sequence, to get to the *tipping point* where they feel the urge to complete the whole HIT. Indeed, as shown in Figures 4(b) and 4(d), while 11% of the workers in the baseline condition quit after the 9th task, only 2–3% of workers in the narrative and scrambled photo conditions did so. The same pattern was not observed in the question only or ordered photo conditions.

Conditions	Median	Mean (SD)	Completion Rate
Baseline	13	15.86 (10.38)	31%
Question only	13	15.97 (9.76)	29%
Narrative	17.5	18.85 (11.14)	47%
Ordered Photo	13	16.45 (10.88)	36%
Scrambled Photo	18	19.03 (11.04)	47%

Table 2. Summary of quit index in different conditions.

Next, we shift our focus to workers who completed all tasks to understand the effects of curiosity interventions on completion (H2). Table 2 shows the percentage of workers who completed all 30 tasks in each condition. Compared to the baseline, most other conditions have a larger percentage of workers who completed all 30 tasks, with the question only condition being the only exception. Two-sided proportion test results suggest that the difference in the percentage of completion across conditions is statistically significant, $\chi^2(4, N = 496) = 12.56, p = 0.02$, yet none of the pairwise comparisons against the baseline condition is statistically significant under the multiple comparison test with Bonferroni correction.

We get a better understanding of the effects of different curiosity interventions on completion using a generalized linear model (GLM), by taking into account the influences of other intrinsic motivational factors such as enjoyment, competence and effort. Our GLM is a reasonable fit as we find no obvious pattern when the residuals are plotted against the independent variables and fitted values, and the results for both Hosmer and Lemeshow test and Osius-Rojek test also support our graphical assessment, $\chi^2(8, N = 496) = 5.88, p = 0.66$ and $z = 0.86, p = 0.39$, respectively.

Table 3 reports the results for the GLM. Here, we find that when a curiosity-inducing stimulus is present, workers are more likely to complete all 30 tasks in all cases (i.e., the estimated coefficient $\hat{\beta}$ is positive) compared to workers in the baseline condition. Holding all other explanatory variables constant, workers in the narrative (and scrambled photo) condition have a significantly higher estimated odds to complete all tasks that is $e^{0.98} = 2.65$ (and $e^{0.88} = 2.41$) times as large as that for workers in the baseline condition, while the increase of estimated odds to complete all tasks in the question only and ordered photo condition is not statistically significant. Interestingly, we also find that enjoyment and self-reported competence both have a significant impact on the likelihood of completing all the tasks — the more a worker enjoys the tasks and/or feels competent at the tasks, the more likely she will complete all tasks, which is quite intuitive.

	Model Parameters				
Variable	$\hat{\beta}$	*Std. Error*	*t*	*p-value*	
Question Only	0.10	0.32	0.31	0.75	
Narrative	0.98	0.31	3.10	1.91×10^{-3}	**
Ordered Photo	0.47	0.32	1.49	0.14	
Scrambled Photo	0.88	0.31	2.81	4.91×10^{-3}	**
Enjoyment	0.19	0.10	1.91	0.06	.
Effort	0.02	0.11	0.21	0.83	
Competence	0.52	0.11	4.85	1.24×10^{-6}	***

Table 3. GLM for the probability of completing all 30 tasks in each condition, with the baseline condition being the reference.

Why Quit? Why Persist?

By looking into the questionnaire data, where workers explain their reasons for quitting and persisting despite initial urges to quit, we gain more insights into the diverse factors that may contribute to whether and how our curiosity interventions influence worker retention.

In general, we find a few major categories of reasons for quitting: *low payment* ("The pay is low for the time it was consuming"), *task difficulty* ("I chose to stop because there came a point where I was having difficulty with understanding some of what was being said, and I didn't want to transcribe incorrectly"), *lack of engagement* ("I was bored and did not know who they were talking about"), and *external factors* ("I was interrupted by someone at my door.")

As for persisting to work, the major reasons cited are *payment* ("I wanted to make more money from the bonuses"), *learning* ("It was good practice, I felt I was getting better as I went on"), and a *completionist attitude* ("I don't like leaving things half-finished").

Importantly, we notice that when curiosity-inducing stimuli are present, many workers actually cite *curiosity* as their primary reason for continuing to work on the tasks. For example:

- "I wanted to see if I could figure out the player's name."
- "I persisted because I was curious about how the article would unfold."
- "I was addicted to transcribing the next sentence to reveal the article's subject, and once I knew who the article's subject was, I just wanted to complete the article."
- "I wanted to know who the article is about. It was like getting a puzzle piece and putting it all together."
- "I thought of stopping several times, but my desire to do a good job, earn the maximum bonus, and to be frank, my curiosity kept me going."

In other words, our curiosity interventions are shown to be effective on many workers as they become eager to find out the answer to the question, and they also take the satisfaction of their curiosity into the consideration when making the cost-benefit analysis on whether to continue or not.

Furthermore, the lessening or lack of curiosity plays a role in why workers *quit*. Some workers cited their inability to guess the answer to the question ("I had no idea who it could be and stopped caring") or their certainty about the answer ("I thought I had already figured out who it was talking about and didn't want to transcribe any more") as reasons for quitting. In other words, when curiosity cannot be satisfied or if curiosity dies, workers would end up quitting. This is in line with Loewenstein's observations [34]—the accumulation of new information may cause the information goal (what one wants to know) to change, or the objective value of the missing information to decrease (because one can infer the answer), thus diminishing curiosity. In other words, curiosity increases with information, but curiosity may also die as the gap is dynamically re-adjusted in light of new information.

Effects on Worker Performance

The median error rates in baseline, question only, narrative, ordered photo and scrambled photo conditions are 4.29%, 3.91%, 3.55%, 3.96% and 3.54% respectively. While workers seem to perform better in experimental conditions with curiosity-inducing stimuli, the Kruskal-Wallis test results suggest that the difference in error rate across conditions is actually not statistically significant, $\chi^2(4, N = 496) = 6.85, p = 0.14$[2]. Our findings on worker performance supports hypothesis H3, that is, workers are able to maintain a high level of accuracy when presented with curiosity-inducing stimuli. Meanwhile, we also notice that the error rate in all experimental conditions is already *very low*, which implies that the space for performance improvement can be very limited. Further examination on task duration shows that each transcription task takes around 1 minute to complete on average, and there are no statistically significant difference across conditions, $F(4, 491) = 0.52$, $p = 0.72$.

[2]An one-way ANOVA is not suitable here due to the non-normally distributed residuals.

Q2: HOW DO INDIVIDUALS RESPOND DIFFERENTLY TO CURIOSITY INTERVENTIONS?

The information gap theory predicts that making guesses leads to increased curiosity. Hence, in this study, we are particularly interested in understanding guessing behavior and its connection to retention and task performance. Specifically, we make comparisons across three groups of workers in experimental conditions with curiosity-inducing stimuli: workers who made *correct* guesses to the question, workers who made *incorrect* guesses, and workers who made no guess. Since the correct answer to the question (i.e. the word "Federer") is revealed at the 25th sentence for the first time, we restrict our attention to guessing behavior before that task.

Metrics		Correct Guess	Incorrect Guess	No Guess
Quit Index	Median	30	11	13
	Mean (σ)	22.63 (9.84)	13.60 (10.03)	15.37 (10.23)
Completion	%	62	24	29
Error Rate	Median	0.03	0.05	0.04
	Mean (σ)	0.04 (0.04)	0.06 (0.04)	0.05(0.03)

Table 4. Retention and performance by guessing behavior.

Among 397 workers, 149 workers made a correct guess, 117 made incorrect guesses and 131 did not make a guess. Table 4 and Figure 6 report the comparison across these three groups of workers. Results show that curiosity interventions affect workers who made correct guesses the most, as they complete significantly more tasks ($t(247) = 7.35, p = 2.82 \times 10^{-12}$ and $t(270) = 6.04, p = 5.16 \times 10^{-9}$ with Bonferroni correction), are significantly more likely to complete *all* tasks ($\chi^2(1, N = 266) = 36.33, p = 3.32 \times 10^{-9}$ and $\chi^2(1, N = 266) = 28.74, p = 8.30 \times 10^{-8}$ with Bonferroni correction), and are more accurate in their transcriptions ($t(116) = -14.63, p = 4.4 \times 10^{-16}$ and $t(130) = -17.15, p = 4.4 \times 10^{-16}$ with Bonferroni correction) than workers who made incorrect guesses or no guess.

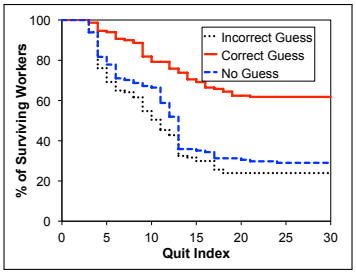
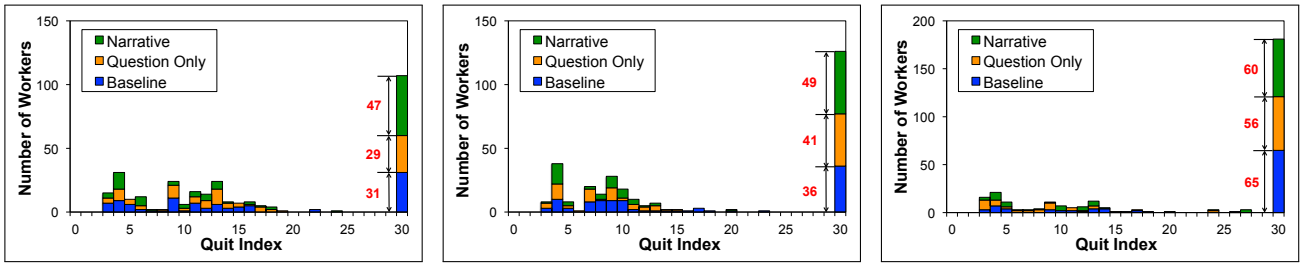


Figure 6. Retention curves: correct, incorrect and no guesses.

A variety of reasons can account for these differences. First, we observe that correct guessers made their first guesses (median at task 1) as well as their first correct guesses (median at task 5) quite early, suggesting that they might have certain prior knowledge about the tennis player Federer and hence were more curious. Second, correct guessers self-reported higher levels of competence in the questionnaire than incorrect guessers ($t(236) = 3.15, p = 0.002$), implying that they had more confidence in their performance than incorrect guessers. Our conjecture is that the effects of curiosity interventions can vary depending on individual worker's level of prior knowledge as well as actual and perceived competence.



(a) Federer article (b) health article (c) imposter article

Figure 5. Histogram showing the number of workers who quit after completing X tasks: variation between different tasks (i.e. different articles).

Q3: WHAT ARE THE INTERACTIONS BETWEEN TASK CHARACTERISTICS AND CURIOSITY INTERVENTIONS?

Our final study examines how the effects of curiosity interventions are influenced by task characteristics, such as the inherent interestingness of the tasks. Our previous experiment shows the effects of curiosity interventions when the task itself is *not* interesting (i.e., the Federer article is the least interesting article according to our pilot study). We now repeat our experiment for the health and imposter syndrome articles—the two articles that generated most interests among workers in our pilot study—to understand how the effects of our curiosity interventions may differ. In particular, we include three experimental conditions for each of these two articles: baseline, question only and narrative. The task-relevant questions we pose for the health and imposter syndrome articles are "Is salt good or bad for you?" and "What psychological condition is the article talking about?", respectively.

We recruited 100 workers per condition per article and randomly assigned them to one of the three experimental conditions. As in the previous experiment, workers are asked to complete at least 3 tasks to get the base payment of 45 cents. After the 3rd task, workers may choose to stop at any time or complete more tasks in exchange for an extra 1-cent bonus per task. Our system explicitly prevents workers in our previous experiment (who worked on transcribing the Federer article) from participating in this experiment, and each worker is only allowed to take this experiment once.

Figure 5 illustrates the number of workers who quit after a certain number of tasks for the three articles. Visually, we can see that the proportion of workers who completed all the tasks is much higher for the health and imposter articles. In addition, for these two inherently interesting articles, the completion rates across experimental conditions appear to be similar.

To validate our visual intuition, we combine our data in all three tasks and again create a GLM with logit link function to model the probability of completing all tasks, with the baseline condition and the Federer article as the reference. In particular, we use task (i.e., the article used for the transcription) as an independent variable, and the interaction terms between tasks and curiosity interventions are also included in the model. Significant interaction terms then imply that the effects of curiosity interventions depend on the tasks. Both visual examination on the residual plots and the two goodness-of-fit test results suggest that our model is a reasonable fit. Results for this GLM is shown in Table 5.

	Model Parameters				
Variables	$\hat{\beta}$	*Std. Error*	*t*	*p value*	
Question Only	0.09	0.32	0.27	0.79	
Narrative	0.95	0.31	3.07	2.16×10^{-3}	**
Health Article	0.12	0.32	0.38	0.71	
Imposter Article	1.33	0.32	4.27	1.93×10^{-5}	***
Question Only × Health Article	0.35	0.44	-0.79	0.43	
Narrative × Health Article	-0.29	0.43	-0.68	0.50	
Question Only × Imposter Article	-0.18	0.44	-0.42	0.68	
Narrative × Imposter Article	-1.10	0.43	-2.53	0.01	*
Enjoyment	0.19	0.07	2.50	0.01	*
Effort	0.01	0.08	0.17	0.87	
Competence	0.47	0.08	6.02	1.76×10^{-9}	***

Table 5. Generalized linear model: 3 articles

As expected, compared to workers performing the less interesting task (i.e., transcribing the Federer article), workers in more interesting tasks are more likely to complete all tasks, even *without* explicit curiosity interventions (i.e., the estimated coefficient $\hat{\beta}$ is positive for both the imposter syndrome article and the health article), and this difference is statistically significant for the task of transcribing the imposter syndrome article, $t(881) = 4.27, p = 1.93 \times 10^{-5}$. Furthermore, while introducing curiosity interventions improves the probability of completion in the question only and narrative conditions for the Federer article (i.e., in both cases, the estimated coefficient $\hat{\beta}$ is positive), almost all the estimated coefficients for interaction terms between tasks and curiosity interventions are *negative*, showing that the curiosity interventions are more effective for inherently uninteresting tasks, which is consistent with our hypothesis H5.

DISCUSSION

Our study shows that stimulating curiosity can be an effective way to incentivize crowd workers. Specifically, we find a close relationship between curiosity and worker retention and performance—given curiosity interventions, workers completed more tasks, while maintaining a high level of performance. Furthermore, we found that workers who made a correct guess complete a significantly larger number of tasks with significantly higher quality than those who made incorrect guesses or no guess. Finally, the effects of our curiosity interventions also depend on the characteristics of the tasks; namely, the effects are larger when the interventions are introduced to tasks that are less interesting.

Why Are Some Interventions More Effective?

Some interventions (e.g., narrative and scrambled photo) seem to be more effective than other interventions (e.g., question

only and ordered photo) in incentivizing crowd workers. In the question only condition, while we create an information goal and make the information gap salient by allowing workers to make guesses, we do *not* explicitly provide any information to help workers answer the question. Thus, the fact that question only is not as effective as other interventions indicates that simply setting an information goal, without providing the means for workers to satisfy their curiosity in the interim, is not an effective approach.

A related question is why the ordered photo condition is not that effective. We found that workers in the ordered photo condition attempted to make their first guesses *and* figured out the answer to the question much earlier than workers in other conditions – the medians are 2, 5 and 4.5 for the number of tasks completed before the first guess, and 5, 21 and 24 for the number of tasks completed before the first correct guess for the ordered photo, narrative and scrambled photo conditions respectively. There are also more workers who made a correct guess (68%) in the ordered photo condition than any other conditions (i.e., 43%, 58% and 54% for the question-only, narrative and scrambled photo condition). That is to say, compared to other conditions, workers in the ordered photo condition are given too much information that enables them to satisfy their curiosity early on, making the effects of the curiosity interventions there virtually not any different from the question only condition, where workers are not given any information at all. This reflects the subtlety in designing curiosity interventions as information can be a double-edged sword: Too little information is not enough for inducing curiosity, yet too much information could satisfy one's curiosity too soon and diminish curiosity. In general, both personal characteristics (e.g., prior knowledge) and beliefs (e.g., how certain one is about the answer) can influence the effectiveness of curiosity interventions.

Design Space and Generalizability

Our curiosity interventions consist of three design elements—a question that is answerable by the information contained in a series of tasks (which serves as the information goal), a mechanism for eliciting guesses (which increases the salience of the information gap), and an incremental reveal of information (which closes the information gap as workers do more tasks). This conceptual approach is quite general, and is applicable to any setting in which the requester has some basic knowledge about the dataset. For example, the question can be about (i) a feature/property shared by some or all tasks, e.g., the neighborhood that a set of images depict, (ii) the global picture of how the individual tasks fit together, e.g., a design task where the identity of the larger system is not revealed until all the submodules are completed, (iii) some global statistics computed from individual data points gathered from the tasks, e.g., a counting task where participants assess the number of flowers on a herbarium specimen, each of which contributes a data point towards testing the hypothesis"flowering time is becoming earlier over time due to effects of climate change." The idea is to obscure by hiding or scrambling certain information that is incrementally revealed as tasks are being completed, and by making the obfuscation salient such that people notice and become curious about the missing information.

Our work demonstrates that information gap theory can be operationalized to affect the behavior of crowd workers. However, the generalizability of the specific designs that we explored in this study, i.e., the use of scrambled photo and obscured article, is limited to transcription tasks and situations where the answer to the question has a visual representation (e.g., the article is about Roger Federer).

In practice, the design elements are *knobs* that can be tuned—the particular choice of questions, feedback mechanisms for responding to guesses, or the frequency of information reveal all have subtle impacts on the extent to which workers feel curious. Additional *knobs* include: (1) questions that reveal a different *type* of information, such as social comparison statistics (e.g., "how much do your transcriptions agree with other workers"); (2) the number of curiosity stimuli to present, that is, whether we create one primary information gap or bite-sized information gaps that are routinely revealed and satisfied (e.g., giving workers a new puzzle to solve when they guessed the answer correctly); (3) the amount and complexity of information to present (e.g., varying the number of cells to reveal and the extent to which the photo is scrambled); and (4) whether or not to provide regular accuracy feedback about the guesses to further increase the attention to the knowledge gap.

Ethical Implications

Prior work has explored non-monetary mechanisms to motivate workers on paid crowdsourcing platforms, e.g., by offering micro-diversions [11] or providing an altruistic purpose for the task [9, 53, 2, 48]. On the one hand, one can argue that these intrinsic motivators (e.g., enjoyable activities during breaks, a meaningful purpose, the desire to read the entire article) serve as *extra payment*, i.e., workers are rewarded with a valuable experience. On the other hand, these mechanisms for increasing the intrinsic motivation of *extrinsically motivated workers* raise ethical concerns—workers may, unknowingly, be doing more work for less pay. In this work, we chose task-relevant questions as curiosity stimuli because we want workers to feel a sense of engagement with the task at hand. Nevertheless, curiosity interventions would find a more natural home in volunteer-based crowdsourcing settings (e.g., citizen science), where the goal is to stoke people's curiosity about science in addition to collecting data to facilitate discoveries.

CONCLUSION

Studying information as a driver for motivation has the potential of unifying our way of looking at the myriad of incentive mechanisms (e.g., gamification, goal-setting, social comparisons) as different ways to utilize information to draw participants in. Inspired by information gap theory, we introduced a set of curiosity interventions and studied their effects on worker retention and performance in a crowdsourcing environment. Our results highlight the potential for information to be used as a *currency* for incentivizing workers. Future work includes developing techniques that automatically create curiosity-inducing stimuli by tuning design parameters, adapting curiosity interventions to account for individual differences in knowledge and interest, investigating ways to not only induce but sustain curiosity, and applying curiosity interventions to other crowdsourcing settings (e.g., citizen science).

REFERENCES

1. Anderson, S. Applying curiosity to interaction design: Tell me something I don't know, 2009. Last Retrieved January 7, 2016 from http://johnnyholland.org/2009/08/curiosity-and-interaction-design/.
2. Ariely, D., Kamenica, E., and Prelec, D. Man's search for meaning: The case of legos. *Journal of Economic Behaviour and Organization 67*, 3 (2008), 671–677.
3. Beenen, G., Ling, K., Wang, X., Chang, K., Frankowski, D., Resnick, P., and Kraut, R. E. Using social psychology to motivate contributions to online communities. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*, CSCW '04, ACM (New York, NY, USA, 2004), 212–221.
4. Berlyne, D. An experimental study of human curiosity. *British Journal of Psychology 45* (1954), 256–265.
5. Berlyne, D. A theory of human curiosity. *British Journal of Psychology 45* (1954), 180–191.
6. Berlyne, D. *Conflict, arousal and curiosity*. McGraw-Hill, London, 1960.
7. Bowser, A., Hansen, D., He, Y., Boston, C., Reid, M., Gunnell, L., and Preece, J. Using gamification to inspire new citizen science volunteers. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications*, Gamification '13, ACM (New York, NY, USA, 2013), 18–25.
8. Caplin, A., and Leahy, J. Psychological expected utility theory and anticipatory feelings. *Quarterly Journal of Economics* (2001), 55–79.
9. Chandler, D., and Kapelner, A. Breaking monotony with meaning: Motivation in crowdsourcing markets. *Journal of Economic Behaviour and Organization 90* (2013), 123–133.
10. Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., Leaver-Fay, A., Baker, D., Popovic, Z., and Players, F. Predicting protein structures with a multiplayer online game. *Nature 466* (August 2010), 756–760.
11. Dai, P., Rzeszotarski, J. M., Paritosh, P., and Chi, E. H. And now for something completely different: Improving crowdsourcing workflows with micro-diversions. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work*, CSCW '15, ACM (New York, NY, USA, 2015), 628–638.
12. Easley, D., and Ghosh, A. Incentives, gamification, and game theory: An economic approach to badge design. In *Proceedings of the Fourteenth ACM Conference on Electronic Commerce*, EC '13, ACM (New York, NY, USA, 2013), 359–376.
13. Festinger, L. A theory of social comparison processes. *Human Relations 7* (1954), 117–140.
14. Feyisetan, O., Simperl, E., Van Kleek, M., and Shadbolt, N. Improving paid microtasks through gamification and adaptive furtherance incentives. In *Proceedings of the 24th International Conference on World Wide Web*, International World Wide Web Conferences Steering Committee (2015), 333–343.
15. Frank Bretz, T. H., and Westfall, P. *Multiple Comparisons Using R*. Chapman & Hall/CRC, 2011.
16. Gaver, W. W., Beaver, J., and Benford, S. Ambiguity as a resource for design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, ACM (New York, NY, USA, 2003), 233–240.
17. Gollub, L. Information on conditioned reinforcement. *Journal of the Experimental Analysis of Behaviour* (2001), 361–372.
18. Grevet, C., Mankoff, J., and Anderson, S. D. Design and evaluation of a social visualization aimed at encouraging sustainable behavior. In *Proceedings of the 2010 43rd Hawaii International Conference on System Sciences*, HICSS '10, IEEE Computer Society (Washington, DC, USA, 2010), 1–8.
19. Harris, C. You're hired! an examination of crowdsourcing incentive models in human resource tasks. In *ACM International Conference on Web Search and Data Mining*, WSDM '11 (2011), 15–18.
20. Ho, C.-J., Slivkins, A., Suri, S., and Vaughan, J. W. Incentivizing high quality crowdwork. In *Proceedings of the 24th International Conference on World Wide Web*, International World Wide Web Conferences Steering Committee (2015), 419–429.
21. Horton, J. J., and Chilton, L. B. The labor economics of paid crowdsourcing. In *Proceedings of the 11th ACM Conference on Electronic Commerce*, EC '10, ACM (New York, NY, USA, 2010), 209–218.
22. Hosmer, D. W., and Lemeshow, S. Goodness of fit tests for the multiple logistic regression model. *Communications in Statistics: Theory and Methods 9*, 10 (1980), 1043 – 1069.
23. Huang, S.-W., and Fu, W.-T. Don't hide in the crowd!: Increasing social transparency between peer workers improves crowdsourcing outcomes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, ACM (New York, NY, USA, 2013), 621–630.
24. Jones, S. Curiosity and knowledge. *Psychological Reports 45* (1979), 639–642.
25. Kang, M., Hsu, M., Krajbich, I., Loewenstein, G., McClure, S., Wang, J., and Camerer, C. The wick in the candle of learning: epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science 20*, 8 (2009), 963–973.
26. Kinnaird, P., Dabbish, L., Kiesler, S., and Faste, H. Co-worker transparency in a microtask marketplace. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, CSCW '13, ACM (New York, NY, USA, 2013), 1285–1290.

27. Koffka, K. *Principles of Gestalt Psychology*. Harcourt-Brace, New York, 1935.

28. Kruskal, W., and Wallis, W. A. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association 47*, 260 (1952), 583–621.

29. Langer, K. Suspenseful Design: Engaging Emotionally with Complex Applications through Compelling Narratives. Master's thesis, University of Waterloo, Waterloo, Ontario, 2014.

30. Law, E., and von Ahn, L. *Human Computation*. Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan and Claypool, June 2011.

31. Litman, J. Curiosity and the pleasures of learning: Wanting and liking new information. *Cognition and Emotion 19*, 6 (2005), 793–814.

32. Litman, J., Hutchins, T., and Russon, R. Epistemic curiosity, feeling-of-knowing, and exploratory behavior. *Cognition and Emotion 19*, 4 (2005), 559–582.

33. Locke, E., and Latham, G. Building a practically useful theory of goal setting and task motivation: A 35 year odyssey. *American Psychologist 57*, 9 (2002), 705–717.

34. Loewenstein, G. The psychology of curiosity: A review and reinterpretation. *Cognition and Emotion 19*, 6 (2005), 793–814.

35. Malone, T. Toward a theory of intrinsically motivating instruction. *Cognitive Science 4* (1981), 333–369.

36. Markey, A., and Lowenstein, G. Curiosity. In *International Handbook of Emotions in Eduction*, R. Pekrun and L. Linnenbrink-Garcia, Eds. Rutledge, New York, 2014.

37. Marlow, J., and Dabbish, L. A. The effects of visualizing activity history on attitudes and behaviors in a peer production context. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work*, CSCW '15, ACM (New York, NY, USA, 2015), 757–764.

38. Mason, W., and Watts, D. Financial incentives and the "performance of crowd". In *ACM SIGKDD Workshop on Human Computation (HCOMP)* (2009), 77—85.

39. McCullagh, P., and Nelder, J. *Generalized Linear Modesl*, 2 ed. Chapman & Hall/CRC, 1989.

40. Menon, S., and Soman, D. Managing the power of curiosity for effective web advertising strategies. *Journal of Advertising 31*, 3 (2002), 1–14.

41. Metzger, W. *Laws of Seeing*. MIT Press, Cambridge, MA, 2006.

42. Miller, N. Liberalization of basic s-r concepts: Extensions to conflict behavior, motivation and social learning. *Psychology: the study of a science* (1959), 196–292.

43. Naylor, F. A state-trait curiosity inventory. *Australian Psychology 16* (1981), 172–183.

44. Osius, G., and Rojek, D. Normal goodness-of-fit tests for multinomial models with large degrees of freedom. *Journal of the American Statistical Association 87*, 140 (1992), 1145 – 1152.

45. Pluck, G., and Johnson, H. Stimulating curiosity to enhance learning. *GESJ: Education Sciences and Psychology 2*, 19 (2011), 1512–1801.

46. Raddick, M., Bracey, G., Gay, P., Lintott, C., Cardamone, C., Murray, P., Schawinski, K., Szalay, A., and Vandenberg, J. Galaxy zoo: Motivations of citizen scientists. *Astronomy Education Review 12*, 1 (2013).

47. Rashid, A. M., Ling, K., Tassone, R. D., Resnick, P., Kraut, R., and Riedl, J. Motivating participation by displaying the value of contribution. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, ACM (New York, NY, USA, 2006), 955–958.

48. Rogstadius, J., Kostakos, V., Kittur, A., Smus, B., Laredo, J., and Vukovic, M. An assessment of intrinsic and extrinsic motivation on task performance in crowdsourcing markets. In *AAAI Conference on Weblogs and Social Media* (2011).

49. Rotman, D., Preece, J., Hammock, J., Procita, K., Hansen, D., Parr, C., Lewis, D., and Jacobs, D. Dynamic changes in motivation in collaborative citizen-science projects. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW) '12 (2012), 217–226.

50. Ryan, R. Control and information in the interpersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology 42* (1982), 450–461.

51. Schmitt, F., and Lahroodi, R. The epistemic value of curiosity. *Education Theory 8* (2008), 125–148.

52. Segal, A., Simpson, R., Gal, Y., Homsy, V., Heartwood, M., Page, K., and Jirotka, M. Improving productivity in citizen science through controlled intervention. In *WWW* (2015), 331–337.

53. Shaw, A. D., Horton, J. J., and Chen, D. L. Designing incentives for inexpert human raters. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, CSCW '11, ACM (New York, NY, USA, 2011), 275–284.

54. Shirk, J., Ballard, H., Wilderman, C., Phillips, T., Wiggins, A., Jordan, R., and Bonney, R. Public participation in scientific research: A framework for intentional design. *Ecology and Society 17*, 2 (2012), 29–48.

55. Silvia, P. J. Curiosity and motivation. In *Oxford Handbook of Motivation*, R. M. Ryan, Ed. Oxford University Press, New York, 2014.

56. Speilberger, C., and Starr, L. Curiosity and exploratory behavior. *Motivation, Theory and Research* (1994), 221–243.

57. Teodoro, R., Ozturk, P., Naaman, M., Mason, W., and Lindqvist, J. The motivations and experiences of the on-demand mobile workforce. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work*, CSCW '14, ACM (New York, NY, USA, 2014), 236–247.

58. Tieben, R., Bekker, T., and Schouten, B. Curiosity and interaction: Making people curious through interactive systems. In *Proceedings of the 25th BCS Conference on Human-Computer Interaction*, BCS-HCI '11, British Computer Society (Swinton, UK, UK, 2011), 361–370.

59. von Ahn, L., and Dabbish, L. Designing games with a purpose. *Communications of the ACM 51*, 8 (2008), 58–67.

60. Wasserman, L. *All of Statistics: A Concise Course in Statistical Inference*. Springer-Verlag, Berlin, 2003.

61. Wilson, A., Burnett, M., Beckwith, L., Granatir, O., Casburn, L., Cook, C., Durham, M., and Rothermel, G. Harnessing curiosity to increase correctness in end-user programming. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, ACM (New York, NY, USA, 2003), 305–312.

62. Yin, M., and Chen, Y. Bonus or not? learn to reward in crowdsourcing. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence*, IJCAI '15 (2015), 201–207.

63. Yin, M., Chen, Y., and Sun, Y. The effects of performance-contingent financial incentives in online labor markets. In *Conference on Artificial Intelligence*, AAAI '13 (2013), 1191–1197.

64. Zeiler, M. Schedules of reinforcement. In *Handbook of Operant Behaviour*, W. Hong and J. Staddon, Eds. Prentice Hall, Englewood Cliffs, NJ, 1977.

65. Zhu, H., Kraut, R., and Kittur, A. Organizing without formal organization: Group identification, goal setting and social modeling in directing online production. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW '12, ACM (New York, NY, USA, 2012), 935–944.

66. Zion, M., and Sadeh, I. Curiosity and open inquiry learning. *Journal of Biological Education 41* (2007), 162–8.