Towards Fair and Equitable Incentives to Motivate Paid and Unpaid Crowd Contributions

Shaun Wallace*
Computer Science and Statistics
University of Rhode Island
Kingston, Rhode Island, USA
shaun.wallace@uri.edu

Talie Massachi
Department of Computer Science
Brown University
Providence, Rhode Island, USA
talie.massachi@gmail.com

Jiaqi Su
Department of Computer Science
Brown University
Providence, Rhode Island, USA
jiaqi_su@alumni.brown.edu

Dave B Miller
Department of Mechanical Engineering
Tufts University
Medford, Massachusetts, USA
dave.miller@tufts.edu

Jeff Huang
Department of Computer Science
Brown University
Providence, Rhode Island, USA
jeff_huang@brown.edu

Abstract

Researchers commonly rely on contributions from either unpaid contributors or work done by paid crowdworkers. Rarely are the motivations of these workers and the accuracy of their contributions studied simultaneously in the wild over time. We maintain a public system where anyone can edit an evolving tabular dataset of Computer Science faculty profiles useful for the field of CS, and in this work, we analyze both the accuracy of contributions and the motivations of paid crowdworkers and unpaid contributors, combining data from real-world edit histories and a discrete choice experiment. The accuracy of edits made by unpaid contributors was 1.9 times higher than that of paid crowdworkers for difficult-to-find data and 1.5 times greater for data requiring domain-specific expertise. Our discrete choice experiment reveals that while both groups are motivated by common attributes describing a contribution task: pay level, estimated completion time, interest, and the ability to help others, they make different trade-offs between these attributes when choosing crowd contribution tasks. We provide recommendations to build hybrid data systems that mix extrinsic and intrinsic motivators to motivate highly accurate contributors, whether paid or unpaid.

CCS Concepts

• Human-centered computing → Human computer interaction (HCI); **Computer supported cooperative work**; **• Information systems → Crowdsourcing**; *Data cleaning*; *Asynchronous editors*; Incomplete data.

Keywords

Tabular Data; Data Maintenance; Discrete Choice Experiment; Unpaid Contributions; Paid Crowdworkers; Crowdsourcing; Peer Production, Intrinsic Motivation, Extrinsic Motivation

ACM Reference Format:
Shaun Wallace, Talie Massachi, Jiaqi Su, Dave B Miller, and Jeff Huang. 2025. Towards Fair and Equitable Incentives to Motivate Paid and Unpaid Crowd Contributions. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 32 pages. https://doi.org/10.1145/3706598.3714195

1 Introduction

Data is rarely static. And the user's motivation to maintain it can be fickle. However, understanding what motivates people to contribute to evolving data is necessary. Ensuring that evolving datasets and information remain current requires ongoing human effort, and motivating people to maintain and update this data is a persistent challenge. You can pay crowdworkers or rely on peer production systems to attract and motivate users to contribute and edit the data. Wikipedia and citizen science projects on Zooniverse thrive on the free contributions of everyday people. Many visitors to public systems (often called "lurkers") consume data without contributing, leaving platforms reliant on a relatively small group of active contributors [5]. Even Wikipedia has a difficult time maintaining enough editors for specific articles [22, 65]. Maintaining user engagement, accuracy, and fairness becomes more complex as these systems scale. How can systems motivate users—both unpaid contributors and paid crowdworkers—to provide accurate, high-quality contributions? Both are viable options, but we rarely design systems to cater to the motivations of both types of contributors. Understanding why people choose to contribute and what motivates highly accurate contributions is essential for designing equitable public systems that balance incentives, motivations, and practical trade-offs.

This challenge is particularly pressing in systems like ours, Drafty, a system housing a publicly editable spreadsheet of Computer Science faculty profiles that we have developed and maintained for over 9 years. Our system has attracted over half a million visitors. How do we develop features to attract visitors, increase

*Work began while the author was at Brown University

CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3714195

their motivation to contribute, and engage highly accurate contributors? Our tabular dataset provides utility to the Computer Science community. Students have used our systems' data to find advisors, departments to justify hiring new professors, and the NSF for studying the diversity gaps among Computer Science departments [75]. It is a constant struggle to understand what motivates our everyday unpaid contributors and how we can design both the system and tasks to attract highly accurate contributors (both paid and unpaid) to contribute their time, effort, and knowledge to maintain its evolving tabular data.

Converting visitors who do not contribute to a public system (i.e., lurkers) into contributors is a complex, multi-faceted challenge [5]. The choice someone makes for why they visit a public data system can differ from their choice to contribute. People contribute to research efforts and datasets for various extrinsic and intrinsic reasons [69, 70]. Researchers and practitioners constantly balance task design, motivational factors, and incentive mechanisms to elicit accurate contributions from the crowd [52]. When building systems for crowdsourcing or peer production tasks, assessing how these systems provide fair and equitable incentives to motivate crowd contributions can be challenging. Multiple factors are at play, from recruiting a large enough active user base of everyday visitors to maintaining the software and then accounting for all real-world events that cause your systems' data and user preferences to evolve over time. Calling on recent research, we want to avoid increasing the invisible cost of labor [99] or people contributing to public systems. We want to use quantitative methods to tease apart the motivations for different users (paid and unpaid) to choose different tasks to contribute their time and effort to edit public data.

Borrowing from Healthcare and Economics research, in this paper, we design and deploy a discrete choice experiment within our real-world system to study user motivation. A discrete choice experiment is a quantitative method used to analyze user preferences [53]. We employ this methodology to investigate the tradeoffs in preferences and motivations behind crowd labor. In a discrete choice experiment, users choose between two hypothetical alternative scenarios. For example, as shown in Table 1, which of these two tasks do respondents choose over the other? These tasks are defined with the following attributes [and levels]: pay level per hour [$0.00 vs. $12.00], time to complete [5 vs. 15 minutes], task difficulty [hard vs. easy], and level of interest [high vs. low].

	Task A	**Task B**
Pay Level per Hour	$0.00	$12.00
Time to Complete	5 minutes	15 minutes
Task Difficulty	hard	easy

Table 1: ***Which of these two tasks would you rather complete? A simple example of a choice set featuring two hypothetical tasks a crowd contributor could select. Choice sets are used in a discrete choice experiment to elicit user preferences. Users would select Task A or Task B to indicate their stated preference, weighing the attributes of the individual choices. See Figure 1 for an actual choice set presented to participants in our discrete choice experiment survey study within Qualtrics.***

The attributes per task help us see what specific features matter to users (e.g., pay level per hour, time to complete, task difficulty). When participants choose one of these two hypothetical alternative scenarios, we can compute utility scores per level (time to complete [5 vs. 15 minutes]) to better understand user preference and motivation. For example, how much more does someone prefer a 5-minute task to a 15-minute one? By exploring enough attributes and levels that describe common crowd contribution tasks, we can explore the trade-offs people make when selecting between different tasks and even recommend optimized tasks for different groups of people. Imagine knowing how to change a task to motivate a paid crowdworker versus one that will attract an everyday visitor to become a contributor to a public system.

Researchers, practitioners, and system builders can conduct a discrete choice experiment with their system's users. Then, use these results and observations to further develop fair and equitable rewards and incentive mechanisms within their related real-world crowd-powered and peer production systems. Our study uses discrete choice experiments to provide quantitative evidence of contributor motivations. We also study these users' real-world editing behaviors as they interact with Drafty.

In spring 2023, everyday visitors (unpaid contributors) and paid crowdworkers visited, made contributions to the dataset, and could freely choose to complete a discrete choice experiment using our real-world public data system, Drafty. Thus, we are able to directly compare the accuracy of edits between unpaid contributors and paid crowdworkers using the same system at the same time. Our study's discrete choice experiment also allows us to determine what motivated those crowd contributors to contribute, comparing users across different pay levels and accuracy rates. By combining real-world editing behaviors and this discrete choice experiment, we can answer the following research questions:

- **RQ1** Who makes more accurate contributions: paid crowdworkers or unpaid contributors?
- **RQ2** How do the motivations of highly accurate vs. inaccurate contributors differ?
- **RQ3** What attributes and levels for crowd contribution tasks universally motivate a public data system's paid crowdworkers and unpaid contributors?
- **RQ4** What attributes and levels for crowd contribution tasks should be individuated per group of users (paid crowdworkers vs. unpaid contributors)?

Our results study the simultaneous editing behaviors of Drafty contributors in the wild. Our results show that unpaid contributors are more accurate than paid crowdworkers in a direct simultaneous comparison. Our discrete choice experiment reveals that paid crowdworkers and unpaid contributors share some similar motivations for selecting crowd contribution tasks. Our discrete choice experiment results show universal motivators among paid crowdworkers and unpaid contributors: the pay level, estimated completion time, and a user's perception of a crowd contribution task. However, while pay level is most important to paid crowdworkers, the perception of the task is most important to unpaid contributors. For example, unpaid contributors prefer tasks they perceive as interesting, ethical, and that will help others. Regardless of how much someone might be paid for a task, we find that highly accurate contributors (both paid and unpaid) are motivated by intrinsic factors such as task interest and helping others. Overall, highly accurate users, as well as paid and unpaid users, make different trade-offs when selecting tasks to contribute. Paid crowdworkers are motivated by tasks where they can collaborate with AI, while unpaid contributors prefer tasks where they can collaborate with others. Highly accurate contributors tend to favor completing tasks where they own the data they contributed, mirroring our system design and that of Wikipedia. Overall, our results and discussion illustrate universal factors to motivate contributors while providing recommendations for building different system features and tasks for paid or unpaid contributors. Thus bridging the gap between paid crowdsourcing and peer production system designs.

2 Related Work

2.1 Eliciting Crowd Contributions

Dataset curators, also called requesters, often recruit paid crowdworkers by posting short, repeatable micro-tasks to quickly collect information to improve the quality of datasets [27, 52]. Historically, decomposing complex tasks into simpler ones aligns with the idea of piecework [2]. This idea has evolved to account for distributed "crowd" workers, both paid and unpaid, who make contributions. Thus, as part of the crowd, people contribute their time and knowledge to research efforts and maintaining public datasets for a number of reasons, such as because their contributions match their interests [19], they can learn something new [95], or their contributions will help others [70].

These **crowd contributions** can take various forms, from completing surveys to editing data on public Google Sheets or Wikipedia. Or even people editing data on custom systems like ours. The reason a user chooses to visit a public data system can differ from their choice to contribute their labor to maintain it. People contribute to research efforts and datasets for various extrinsic and intrinsic reasons [69, 70]. This section presents an overview of recent research on crowd-powered systems and methods to motivate and improve the quality of crowd contributions among different populations. We view any user-made contribution to a peer production or paid crowdsourcing effort as a crowd contribution.

2.2 Approaches to Improve Data Quality: Paid Crowdsourcing and Peer Production

Paying crowdworkers to edit and hopefully improve the accuracy of crowdsourced datasets is a common method among researchers [84]. Prior research includes examples of maintaining evolving data, including crowdsourcing and peer production [1, 48, 96]. Crowdsourcing researchers often find that paid crowdworkers balance pay level per hour with the estimated time to complete tasks [58, 63]. In contrast, peer production researchers often attribute various intrinsic motivators to why people spend their free time contributing to sites like Wikipedia and other public resources [4, 51]. While Wikipedia and WikiData are success stories of popular peer production platforms, they rely on a small number of contributors to maintain each topic [54, 85]. Attracting and retaining new contributors is essential to their continued success. Still, insights drawn from the experience of Wikipedia and WikiData show that it is challenging to maintain contributors within a large popular system [3, 81], for example:

(1) A lack of structured onboarding and low collaboration within the established community repels new contributors.
(2) It is technically difficult to contribute (i.e., editors must learn markup language).
(3) There are too many guidelines and confusing policies.
(4) Deletionists, bad actors, or bots vandalize data or quickly revert new contributions from the community.

There are many short-term efforts to study how extrinsic and intrinsic motivators can increase the quantity and quality of crowd contributions.

Paid crowdsourcing is an alternative to uncompensated peer production, where people accept and complete small micro-tasks for money [50]. However, in peer production, poor quality contributions often result from mixed motivation scenarios and decentralized task creation [4]. Research in this area often seeks to optimize pay levels or improve task instructions to improve the participation and quality of paid crowdworker contributions [14, 41, 113]. However, paid crowdworkers often lack the domain-specific knowledge required to make accurate contributions to data that are difficult to interpret [103]. Paying crowdworkers for repeated tasks over time is not financially feasible for many public datasets [23, 63, 92]. Other issues commonly faced in paid crowdsourcing situations include varying motivation levels and effort of crowdworkers [7, 71]. Recent research has covered the downward trend in the quality of work from paid crowdworkers in Amazon Mechanical Turk and their shifting motivations [31, 42, 62].

Unpaid peer production can be compared with paid crowdsourcing to study the long-term benefits of reciprocity [32]. Prior work shows that reciprocity increases when persistent and transparent reputation mechanisms are created from user behaviors and data [4]. This incentive structure rewards positive behaviors and applies penalties for negative ones–often providing good actors with improved reputations and leaving bad actors with sullied ones. These mechanisms move beyond using badges and credits used in gamification [64]; instead, this approach leverages the desire to get credit for doing something of value to the community. Instead of getting a badge, you gain (or lose) credibility. An adjacent area to peer production is Learnersourcing, where a specialized crowd of learners is naturally motivated as part of their learning process and make contributions in their area of knowledge [49].

Across various paid and unpaid research efforts, researchers are trying to fine-tune task design, motivational factors, and incentive mechanisms to generate accurate contributions from the crowd [52]. When building systems to run crowdsourcing or peer production tasks to elicit crowd contributions, assessing how these systems provide fair and equitable incentives to motivate crowd contributions can be challenging. As systems, tasks, and user motivations evolve, this will always be an ongoing open research question [24, 42, 43]. Our paper investigates a range of factors that affect people's motivation to complete crowd contribution tasks by combining a discrete choice experiment with our system to study anonymous everyday users and paid crowdworkers editing a public tabular dataset.

3 Method

We discuss our sequential steps to design and run a discrete choice experiment with paid crowdworkers from Prolific and our normal everyday unpaid contributors within our system Drafty in the wild. First, we reviewed the literature on what motivates paid crowdworkers and unpaid contributors to choose crowd contribution tasks. We created an initial set of attributes and levels to describe crowd contribution tasks. Then, we conducted a pilot study in the wild with paid crowdworkers and unpaid contributors from Drafty to evaluate these attributes and levels. Then, we combined these results by consulting the literature again to arrive at the final set of attributes and their associated levels. We try to balance previous literature with upcoming trends for what might motivate users. Finally, we conducted the main study in the wild. Paid crowdworkers and unpaid contributors within Drafty freely chose to complete the discrete choice experiment. By recruiting our system's actual users, we can compare what motivates them to complete crowd contribution tasks with their actual editing behaviors and accuracy from Drafty.

Our Institutional Review Board reviewed and approved our methods, procedures, and proposed analysis.

3.1 Study Design: Terminology for a Discrete Choice Experiment

Discrete choice experiments are a preference elicitation technique where participants choose between two or more hypothetical alternatives that vary systematically on multiple dimensions (see Figure 1). Originally developed by Louviere and Woodworth for health economics, discrete choice experiments are commonly used in economics, health, and market research [57]. This section covers the basic terminology and methods used in a discrete choice experiment. A discrete choice experiment better resembles people's real-world decisions compared with other stated preference methods such as ranking or questionnaires using Likert-type scales [59]. Since discrete choice experiments are seldom used in HCI, we have provided extensive definitions to help familiarize readers.

Choice Set is at least two hypothetical alternatives where each alternative has the same attributes but different associated levels per attribute. Each alternative, or choice set, has at least two attributes, and each attribute has at least two levels. By presenting participants with a series of choice sets, they reveal their preferences, having compared the attributes and levels–for example, preferring higher pay for more work over less pay for completing a shorter task. Our study features only two hypothetical alternatives per choice set due to our higher number of attributes.

Attributes are the independent variables that are being tested in the discrete choice experiment [59]. For example, a crowd contribution task's attributes could be pay level, estimated time to complete the task, and perceived difficulty. Another example could be comparing breakfast cereals: attributes could be price, amount of sugar, and brand name vs. store brand. Attributes are often identified by reviewing related research and real-world observations [20].

Levels are an attribute's options, increments, or possible values. They can be continuous, ordinal, or binary. Generally, levels are selected to reflect the values people encounter in the real world, such as on the back of the cereal boxes [20]. In our experiment, we look for real-world values from crowdwork. For example, Prolific enforces a minimum pay level per hour of $8 per hour, while their recommended pay level is $12 per hour. Meanwhile, Visitors to Drafty (i.e., unpaid contributors) make contributions for a pay level of $0 per hour. A discrete choice experiment uses comparisons between attributes and levels to determine utility scores to understand what levels negatively and positively affect someone's preference for a given scenario.

Crowd contribution task is a task where a person interacts with a computer or device to contribute information or data. This person can be paid or unpaid for the completion of this task. For example, people contributing to Wikipedia would be an unpaid crowd contribution task. A crowd contribution task could also involve a paid crowdworker recruited through Prolific who completes a survey about technology use. Whether paid or unpaid, all visitors to Drafty are completing crowd contribution tasks when editing Drafty's data.

3.2 Designing our Discrete Choice Experiment: Pilot Study and Related Work

3.2.1 Initial Design and Pilot Study Overview. In designing a discrete choice experiment, we must select the set of relevant attributes and their levels [59]. An attribute could be the pay level per hour for a task. Its levels could be $8, $12, or $16 per hour. We first conducted a pilot study to design our discrete choice experiment to consider multiple attributes and levels using a Maximum Differential technique following prior work [59]. In Maximum Differential (MaxDiff), participants select their most preferred and least preferred level per attribute. We conducted the pilot study with paid crowdworkers and unpaid contributors within Drafty using the same recruitment methods as our main study.

We selected these final sets of attributes and levels after reviewing related research and reviewing the results from our pilot study. In this pilot study, paid crowdworkers and unpaid contributors (i.e., everyday visitors) using Drafty voluntarily chose to take an initial survey. Our pilot study consists of 34 participants: 18 paid crowdworkers recruited through Proflic and 16 unpaid contributors who freely visited Drafty. Our pilot survey features MaxDiff (Maximum Differential) questions where participants select what motivates them the most and the least to contribute. These participants voluntarily chose to take the pilot survey by selecting a blue button in a banner within Drafty; see Figure 3. Participants who completed the survey could submit their email for a one in four chance to receive a $25 Amazon gift card as compensation.

Two attributes (Reward and Time Already Spent on a Task) we considered for the pilot study were not included in the main study. Our idea to include "Time Already Spent on a Task" was directly motivated by recent research on the sunk costs of time already invested by crowdworkers [99]. Pilot study results showed that participants found the attribute difficult to understand. This was paired with pilot results showing that this attribute did not strongly affect user motivation. Thus, we chose not to include it as an attribute in our main study.

We similarly did not include the "Reward" attribute in the main study, as pilot participants found it confusing when compared to pay level per hour (a separate attribute). However, we did keep some of the "Reward" levels that showed promising trends in the pilot as levels in other attributes. Below are the levels we chose to keep *(* indicates what was kept or moved to different attributes)*:

Which task would you be more likely to complete? (choose your most preferred)

	Task 1	Task 2
Pay Level	$8.00 per hour	$12.00 per hour
Estimated Time to Complete	15 minutes	5 minutes
Task Difficulty	Not difficult (easy) to complete	Very difficult to complete
Your Reason to Complete a Task	You will learn a new or special skill.	The task is part of a hobby
Task Requirement	To complete the task by yourself	Complete the task with Artificial Intelligence
Who Asks you to Complete the Task	A friend	A for-profit company
What Happens with your Contribution	Your contribution could be rejected	Your contribution is automatically accepted
Your Perception of the Task	The task looks interesting	The task looks boring
	○	○

Figure 1: A screenshot of choice set from Qualitrics used in the main study. It shows two hypothetical alternative crowd contribution tasks presented to survey participants. The attributes are on the leftmost column, and their associated levels are under the columns labeled "Task A" and "Task B."

	Task 1	Task 2
Pay Level	$8.00 per hour	$12.00 per hour
Estimated Time to Complete	15 minutes	5 minutes
Task Difficulty	Not difficult (easy) to complete	Very difficult to complete
Your Reason to Complete a Task	You will learn a new or special skill.	The task is part of a hobby
Task Requirement	To complete the task by yourself	Complete the task with Artificial Intelligence
Who Asks you to Complete the Task	A friend	A for-profit company
What Happens with your Contribution	Your contribution could be rejected	Your contribution is automatically accepted
Your Perception of the Task	The task looks interesting	The task looks boring

Attributes and their associated Levels for one choice set.

Figure 2: We color-coded the attributes (purple box) and their associated levels (teal box) in a screenshot of a choice set from our discrete choice experiment run in Qualtrics.

(1) You learn a new skill *(* kept for Reason to Complete a Task)*
(2) You will be paid for doing exceptional work *(* modified for Reason to Complete a Task)*
(3) You will be paid fairly *(* removed)*
(4) Your contribution helps others *(* modified for Perception of a Task)*
(5) Your contribution benefits you *(* kept for Reason to Complete a Task)*
(6) You get a personalized recommendation *(* modified for Reason to Complete a Task)*
(7) Nothing *(* removed)*

We also considered levels across a few different attributes regarding either a task requiring completion multiple times or a task being related to a previous task. These levels were chosen based on previous research and common practices in paid crowdsourcing. However, none of these attribute levels were found to have a strong effect in the pilot study and thus were removed from the main study.

Specifically, the level "You have completed similar tasks before" was initially included as a reference to the idea that paid crowdworkers tend to do work for the same requester multiple times [8]; the level "You complete the task once" was included as a common scenario when a crowdworker completes an activity or a survey one time [50]; and "You complete the task multiple times" was intended to mirror the scenario where a user edits multiple pieces of data on Wikipedia [3] or multiple cells in a spreadsheet [78].

Finally, some levels included in the pilot study were removed from the main study because they had a low effect on motivation, such as: "You will be paid fairly (at or above minimum wage where you live)" and "You will be underpaid (below minimum wage where you live)" for the Pay Level attribute. These two attribute levels are difficult to interpret and would be challenging to translate into effective task designs since requesters do not know the locality of every individual crowdworker. Instead of the more ethereal "underpaid" statement, our attribute Pay Level instead includes the $4.00 per hour level, which Prolific's guidelines recommend as "unfair". We also removed two levels from the attribute of Task Perception about if a task is fun or not fun. In the pilot, these levels showed a much lower effect on user motivation than on task interest.

3.2.2 Selecting the Main Study's Attributes and Levels. After the pilot study, we selected eight attributes to describe crowd contribution tasks from a user's perspective: pay level per hour, estimated time to complete the task, task difficulty, their reason to complete the task, what the task requires them to do, who asked them to complete the task, what happens to their contribution, and their perception of the task. The associated levels per attribute are described in the subsections below.

3.2.3 Attribute: Pay Level per Hour.

(1) $0.00 per hour
(2) $4.00 per hour
(3) $8.00 per hour
(4) $12.00 per hour
(5) $16.00 per hour

Pay level per task is among the most frequently researched topics in crowdsourcing [50, 63]. Specifically, requesters on paid crowdsourcing platforms find it challenging to assess what is a fair payment [89]. Paid crowdworkers often use resources like TurkerView to view the hourly pay rate a requester offers and if that rate is fair [90]. While research has suggested alternative payment schemes such as payments in bulk, payment per task is still the most common because it is easily understood [44]. Dynamo [89], like Prolific, recommends a minimum hourly wage. However, there is a diminishing return in increasing payment to elicit higher quality contributions [38]. Linearly increasing pay levels can help study what other task attributes and levels must be present to motivate people fairly when the pay is not high enough. Most importantly, regarding unpaid contributors using Drafty, what motivates them when monetary compensation is absent (i.e., $0.00 per hour)?

Prior research on pay level in crowd contributions has focused on perceived social good as a motivator for unpaid contributions [5] or on the effect of paying a system's existing already intrinsically motivated users [47, 105]. However, the choice is driven by more than intrinsic motivations and monetary reimbursement for labor. While some choice paradigms have been studied, there is still a lack of understanding about the trade-offs people make when choosing tasks in crowdsourcing and peer production scenarios.

Selecting Levels and Reflections from Pilot Study: In this study we use pay level per hour instead of a single dollar amount for many reasons. First, our levels mirror Prolific's pay level per hour recommendations. Prolific displays a recommended pay level per hour between $8.00 and $16.00 per hour of estimated time for task completion, with the default recommendation at $12.00 per hour. While a requester may pay over $16.00 per hour, Prolific enforces fair compensation practices by requiring the requester to pay crowdworkers at least $8.00 per hour if the estimated task time is incorrect. Second, most crowdsourcing research papers measure crowdworker pay in payment per hour. Lastly, we recruit paid crowdworkers from Prolific, paying them $8.00, $12.00, or $16.00 per hour. This enables us to study how the stated preferences shift among crowdworkers who accept tasks at different payment levels. $4.00 per hour was additionally chosen to represent a value that should be perceived as underpaid by a crowdworker on Prolific. We also considered intervals of $1, $5, and $10 per hour in the pilot. However, we wanted our levels to mirror what actual requesters and crowdworkers see in real-world platforms.

3.2.4 Attribute: Estimated Time to Complete.

(1) 1 minute
(2) 5 minutes
(3) 15 minutes
(4) 30 minutes
(5) 60 minutes

The time to complete a task is often cited as an essential motivator in choosing to complete paid [29] and unpaid [85] tasks. Estimated time to complete the task is also a requirement for requesters to post to platforms like Prolific and Amazon Mechanical Turk. However, predicting the time required to complete a task is difficult for both paid crowdworkers and requesters alike [99]. Prior research also shows unpaid users are more likely to spend more time on a task than paid crowdworkers due to intrinsic motivators such as wanting to help others or their community [46].

Selecting Levels and Reflections from Pilot Study: We chose five gradually increasing intervals between each "estimated time to complete" time unit. While the pilot study featured eight levels, we removed the levels of 10, 20, and 45 minutes for the main study because these did not make a difference in user preference, per recommendations for selecting levels [59]. We settled on five different levels to ensure we represented both very long tasks (60 minutes) and variations of shorter tasks (1 and 5 minutes). In 1 minute, it is reasonable to assume someone can visit a public Google Spreadsheet and add one piece of data they already know in a cell. In contrast, in 60 minutes, someone could visit Wikipedia, read one article, search the various facts and sources to verify each is correct, and correct any inaccuracies within the article. The intention is for the "estimated time to complete" attribute to cover this range of times to help future requesters and system designers understand the impact task completion times can have on ones willingness to contribute.

3.2.5 Attribute: Task Difficulty.

(1) Not difficult (easy) to complete
(2) Moderately difficult to complete

(3) Very difficult to complete

Perceived task difficulty is one of the most common factors requesters and systems designers try to optimize [39]. Prior research shows that task difficulty relates to the effort required to complete a task [18, 72]. For example, a paid crowdworker could be required to label ambiguous images or to interpret subjective data, such as a professor's research area. When deciding if they should complete a crowd contribution task, crowdworkers might favor more straightforward tasks than higher-paid but more difficult ones. Liu et al. showed that increased task difficulty could adversely increase the time to complete a task [56]. This is notable because many paid crowdworkers are attempting to maximize the amount they are paid per unit of invested time. By completing easy tasks, they can better optimize their time and compensation. By understanding this attribute, requesters and system builders can better decide whether they should spend more time improving the difficulty of their crowd contribution tasks to encourage task adoption.

Selecting Levels and Reflections from Pilot Study: We chose three mutually exclusive categorical levels for task difficulty to keep participants' answers easier to interpret for future researchers. These levels represent the options a participant might see on a three-point Likert-type scale representing a task's difficulty level. We limit to only three points to reduce participants' cognitive load and provide more statistical power for analyzing our discrete choice experiment results [59]. These levels were kept verbatim from our pilot study as our results showed a linear relationship between easy, moderate, and difficult tasks.

3.2.6 Attribute: Reason to Complete a Task.

(1) You might be paid for doing exceptional work
(2) Your contribution benefits you personally
(3) You will learn a new or special skill
(4) You will get a personal recommendation or learn something new about yourself
(5) You get reputation points in a system (special badge, points, credit, etc.)
(6) The task is part of your job
(7) The task is part of a hobby

While all other attributes address the degree to which different motivators affect motivation, this attribute instead questions which additional reasons to contribute are most impactful for crowdworker motivation. Each level chosen in this attribute is based on previous research into crowdworker motivation. We included a level about potential pay for doing exceptional work as it mirrors the idea of providing possible bonuses to paid crowdworkers to increase motivation [78, 110]. The option "Your contribution benefits you personally" builds on the findings from prior research showing that people can be intrinsically motivated to contribute because it benefits themselves [26]. Similarly, there are peer production systems [87] and crowdsourcing efforts that motivate participation by promising workers they will get a personal recommendation or learn something new about themselves. Reputation points within a system are one representation of gamified systems, a common practice used to elicit crowd contributions [64, 66]. Results showing the effectiveness of gamification from prior research are mixed [58]. Additionally, prior paid verification strategies offer the promise of higher-paying tasks if paid crowdworkers complete an initial set of tasks [61, 103], a system which also mirrors the idea of building "reputation" or points. Finally, the idea that a task might mirror a hobby aligns with some of the motivation from learnersourcing [49], where users voluntarily contribute their labor while learning in their free time.

Selecting Levels and Reflections from Pilot Study: Several levels from various attributes were included here after the pilot study, though for many, wording was changed based on feedback from the pilot. For example, we changed the level "You will be paid for doing exceptional work" to "You might be paid for doing exceptional work", since bonuses for paid crowdworkers are not always guaranteed.

3.2.7 Attribute: The Task Requires you to.

(1) Collaborate with other people to complete the task
(2) Complete the task with Artificial Intelligence
(3) Complete the task by yourself
(4) Learn something new
(5) Contribute or use specialized knowledge you already know
(6) Provide your personal information

Many different types of tasks exist in crowdworking platforms. As each type of task has different requirements, there may be preferences among paid crowdworkers for certain types of tasks to complete or requirements to avoid. For example, many surveys or systems require participants to provide part of their personal information (age, email, username, IP address) for payment, attribution purposes, or to improve the validity and generalization of results. Prior research shows that paired crowdworkers sharing their personal demographic information produce higher quality work [40]. However, specific paid crowdsourcing platforms like Prolific have begun providing identifying information that is not personally identifiable data to help protect worker privacy and reduce the effort required of users. This strive for privacy may drive workers towards more privacy-preserving tasks.

With regards to contributing using specialized knowledge, prior research on knowledge-intensive crowd contribution tasks [23] supports the idea that crowds [34, 76] can provide more accurate contributions to datasets requiring domain-specific knowledge when the crowdworkers have the required knowledge. Prior research using discrete choice experiments also found a relationship between expertise and stated preference when selecting healthcare treatments [20]. It is also common for requesters to require paid crowdworkers to learn new skills or knowledge to make accurate contributions for a paid micro-task [23, 111].

While prior research explores collaborative editing behaviors among paid crowdworkers [103], recent research is studying where people and AI collaborate on tasks [109]. Often, peer production tasks produce a sense of collaboration among contributors [4]. Prior efforts show it is common for paid crowdworkers to learn something new to contribute [52]. Even unpaid citizen science efforts often require users to gain new knowledge to contribute [15].

Selecting Levels and Reflections from Pilot Study: These levels are primarily based on current research. They are categorical variables and not all are mutually exclusive. Our pilot study showed each level was important, but contributing or using specialized knowledge you already know had the greatest affect on preference.

3.2.8 Attribute: Who Asks you to Complete the Task.

(1) A friend
(2) A family member
(3) Someone you do not know
(4) People on social media
(5) A bot (not a person) on social media
(6) A for-profit company
(7) A non-profit company
(8) A system (e.g., Wikipedia or Drafty)
(9) Paid Crowdsourcing system (i.e., Prolific)

There are a plethora of systems and methods to ask someone to make a crowd contribution task, ranging from friends, social media, and popular paid crowdsourcing platforms such as Prolific or Mechanical Turk. Direct human recruiting can be highly effective; prior research into "friendsourcing" has found that asking your friends to contribute increases the quality and likelihood of contributions [10, 11]. Even public systems, like Wikipedia, have had research efforts asking people to contribute [25]. Bots on Twitter have also been used to recruit crowd contributors, replacing human connections [91]. Beyond friendsourcing or using virtual agents, Rogstadius et al. studied the quality and likelihood of crowd contributions of paid crowdworkers when posting the tasks as either for-profit or non-profit companies [88]. Understanding the impact of the source of the task completion request will help define how vital an individual's network is when attracting people to make crowd contributions.

Selecting Levels and Reflections from Pilot Study: These levels are categorical variables, and not all are mutually exclusive. We selected each level based on previous research papers comparing recruitment methods. We included if a company was for-profit or not due to the increasing number of companies who engage with paid crowdworkers to evaluate systems for label data [79].

3.2.9 *Attribute: What Happens with your Contribution?*

(1) You own the data you contributed (you can see and edit it)
(2) You do not own the data you contribute (you cannot see or edit it)
(3) A public community owns the data you contributed (anyone can see and edit it)
(4) You receive no credit for your contribution (it is anonymous)
(5) Your name or username is attached to your contribution (not anonymous)
(6) Your contribution is automatically accepted
(7) Your contribution could be rejected

Work around ownership, credit, and outcomes in paid crowdsourcing systems has shown that these can have a strong impact on user motivation to contribute. It is common in paid crowdsourcing scenarios for someone to complete a task and not have access to see or edit their contribution once made [50]. However, if we imagine a paid crowdworker labeling hundreds of images but unable to correct a mistake, we can see why these types of tasks may be demotivating. An alternative model used in individual creativity support tools, like Sketchy [102], allows only the original contributor to edit their contributions. Surveys often allow users to see and edit their submissions; in this scenario, people control or own the data they contributed. In contrast, systems like Drafty, Wikipedia, and WikiData allow users and the entire community of interested users to see or edit any contribution. This idea of community ownership of the data is a hallmark of peer production [4, 108].

Anonymity may also have a strong impact on contributions. Many paid crowdsourcing systems like Prolific and Amazon Mechanical Turk provide paid crowdworkers anonymous IDs to safeguard their personal information. Websites similar to StackOverflow require users to create accounts to possibly identify themselves, or remain anonymous with an ID based on individual user preference. Similarly, systems such as Wikipedia and WikiData require contributors to make an account in order to contribute [37]. In Drafty, users make contributions anonymously by default. This additionally is beneficial as it reduces the effort required to contribute.

Rejection or acceptance of edits has been shown to act as a strong source of motivation for crowd contributors. Prior research shows that a contribution being rejected can be de-motivating for contributors [60]. New Wikipedia users have also be dissuaded by the possibility of another user quickly reverting or rejecting their edit [85]. In contrast, Drafty employs the model that someone's contribution is automatically accepted.

Selecting Levels and Reflections from Pilot Study: These levels are categorical variables where some are mutually exclusive (a task automatically being accepted or rejected). We chose these levels to match what happens with contributions across paid crowdsourcing and peer production platforms. Thus, our results should translate to system designs. The pilot study showed each had some effect on user preference.

3.2.10 *Attribute: Your Perception of the Task.*

(1) The task looks interesting
(2) The task looks boring
(3) The task might be unethical
(4) The task is likely ethical
(5) Your contribution might help people you do not know
(6) Your contribution might help your peers or community

A contributor's perception of a crowd contribution task covers multiple possible levels, many of which are intrinsic motivators. In this attribute we focus on ethical implications, interest, and whether the task might help others as potential intrinsic motivators. While many researchers use paid crowdworkers to complete tasks, there are lingering questions about how the ethical use of data collected by paid crowdworkers could affect their motivation to contribute [33].

Prior research on Drafty shows user interest is correlated with higher quality and likelihood of someone contributing [100, 103, 104]. Research has further shown that perceiving a task as boring is one of the main reasons paid and unpaid workers quit a task [60], and that paid crowdworkers [19] and older adults [12] are more likely to contribute if the task is relevant to their interests. Beyond interest, one of the best motivators to appeal to a user's intrinsic motivation is showing how their contribution will help others [70, 91]. Among paid crowdworkers, prior work shows helping others can intrinsically motivate them to contribute their time even when paying less money [88].

Selecting Levels per Attribute & Reflections from Pilot Study: We selected levels to try and balance positive and negative perceptions. In the pilot study, one level for reward had a very high effect on user motivation: "Your contribution helps others". However, who are these others? We made two more specific options: Your contribution "might help people you do not know" or "might help your peers or community" in the hope the main study can further tease apart user motivation. We moved these within the "Perception of a Task" attribute because contributors can only perceive and not directly control if their contribution helps others [3].

3.2.11 Creating the Choice Sets from the Attributes & Levels. We use Qualtrics to create and conduct the discrete choice experiment survey. Qualtrics automatically builds the choice sets from the provided attributes and levels. It uses a randomized, balanced design approach to ensure the choice sets are varied and present all levels to each participant. Per Qualtrics recommendations, participants completed twenty choice sets (i.e., selected their preferred crowd contribution task). Each choice set is binary; it contains two hypothetical scenarios. Each choice set contains the same number of attributes and one level per attribute (see Figure 1). Qualtrics automatically balances the levels per attribute to ensure valid results; see Appendix section A for more details.

3.3 Modifying our Public System Drafty to Recruit Participants

We aim to recruit normal users of Drafty to take the survey. This includes both paid crowdworkers and unpaid contributors to Drafty. To accomplish this, Drafty shows users a banner; see Figure 3.

3.3.1 Drafty: System Description and History. We have engaged in a longitudinal effort from 2014 to the present to build and maintain a public system to engage anonymous visitors to maintain an evolving tabular dataset of Computer Science professor profiles. This paper runs the third version of our interactive spreadsheet web application, "Drafty". Drafty is a public data system that supports free access and open edits to large tabular datasets (thousands of rows) using a custom spreadsheet interface. It does not record any personally identifiable information. Visits per user are recorded per browser using cookies. During this paper's study, Drafty hosts a publicly editable tabular dataset the CS community has had access to since 2014. The data consists of academic profiles of tenure-track Computer Science from top Universities in the US and Canada. Each profile (row in the tabular dataset) consists of their full name, the university they are employed at, the year they joined that university as a tenure-track professor, their research area of expertise (i.e., subfield), and the name of the institutions that awarded their Bachelor's and Doctorate degrees.

3.4 Study Procedures for our Discrete Choice Experiment within Drafty

3.4.1 Crowd Contribution Tasks within Drafty. This study features paid crowdworkers and unpaid contributors editing Computer Science professor data within Drafty. Within the system, they can add, update, or delete individual cells and rows of data in Drafty in one of six possible ways:

(1) Fill in an empty cell within Drafty (fix one empty cell)
(2) Add a new row of data (add a new Professor)
(3) Delete a row of existing data (remove a Professor)
(4) Review a row of existing data (review and fix a Professor's data)
(5) Add a note to a row (add a note about a Professor)
(6) Contributed using the "Help Us!" feature within Drafty

In Prolific, paid crowdworkers are instructed to "Add a new row of data." Thus, each paid crowdworker can find and add six pieces of information for a Computer Science professor not currently listed in Drafty. We made this choice for two reasons. First, at the time, our tabular dataset had very few empty cells for paid crowdworkers to find and edit. Second, in adding a new row of data crowdworkers make six individual edits, which provides more edits per paid crowdworker for us to evaluate for accuracy by hand.

3.4.2 Recruiting Drafty's Visitors: Unpaid Contributors and Paid Crowdworkers. While discrete choice experiments have good internal validity, we need observable data to improve our result's ecological validity. To enhance ecological validity, we used a convergent design. Visitors to Drafty can freely choose to take the discrete choice experiment survey. These visitors include Drafty's regular unpaid contributors and paid crowdworkers recruited through Prolific. Unpaid contributors are the normal everyday visitors who arrive at Drafty. During the recruitment period, Drafty contains a banner to recruit participants (see Figure 3). Anyone who completes the survey has a one-in-four chance to win a $25 Amazon Gift Card, providing a monetary incentive to participate in our survey study.

Paid crowdworkers recruited through Prolific were asked only to review and edit data on Drafty. Completion of the discrete choice experiment survey was fully optional and not mentioned in the Prolific task. We used Prolific's built-in pre-screening to ensure paid crowdworkers met the following criteria:

(1) Minimum 95% approval rate.
(2) Minimum 100 tasks completed.
(3) Minimum age of 18.
(4) They could not have completed any of our prior tasks, to ensure new unique participants each time.
(5) Are from the USA. (Because Drafty mainly features universities from the US and Canada.)
(6) Use a Desktop device because the user interface design of Drafty and the survey are optimized for desktop screen sizes.

For all tasks, we advertised a 15-minute estimated completion time based on timings from our pilot study. When creating the paid tasks, participants were compensated within Prolific's payment guidelines at $8.00, $12.00, or $16.00 per hour. At the time of this study, Prolific's interface recommended these pay levels per hour to requesters: minimum $8.00, default $12.00, and maximum $16.00. Our compensation scheme aligns with real-world examples of recruiting paid crowdworkers across Prolific and the levels we used for the Pay Level attribute in our discrete choice experiment.

Each paid crowdsourcing task asked paid crowdworkers to review Drafty's data and add a missing professor from a university. Table 6 in Appendix section B.1 lists the university name and payment per task. The universities for the paid crowdsourcing task were selected because they had either not had a new professor added recently or they were new universities not currently included in Drafty (i.e., Drafty has no data for this university). Before releasing the tasks, we additionally hand-checked each university for missing professors and new assistant professors to ensure there were

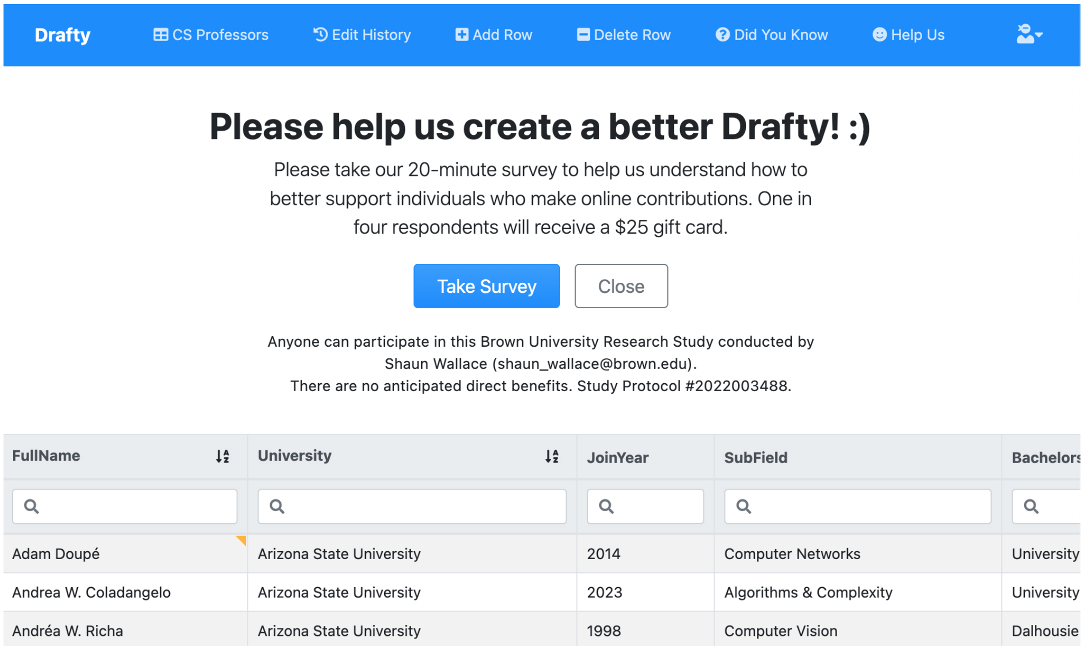

Figure 3: Anyone can voluntarily choose to take the survey and will have a one in four chance to receive a $25 Amazon gift card as compensation. To be eligible, participants must enter their email in a separate survey that is provided after and not connected to the main discrete choice experiment survey.

enough professors to add following a similar study and recommendations from Papoutsaki et al. [78].

The exact instructions per task posted on Profilic are included in Appendix B.

4 Results

Our results combine participants' answers to our survey-based study with their real-world edit histories from Drafty for paid and unpaid tasks. Our analyses of the discrete choice experiment follow the recommendations of prior research [86] to help us uncover what motivates different contributors within a real-world system.

4.1 Discrete Choice Experiment: Data Overview

We conducted the study and gathered data from March 11th to April 25th, 2023. We recruited 115 paid crowdworkers from Prolific to add new rows of data to Drafty. Also, 2,723 unpaid contributors freely visited Drafty to view or edit data within the system. These everyday active visitors freely chose to visit Drafty and made at least one interaction within Drafty (i.e., click, search, edit, etc.). All visitors (paid crowdworkers and unpaid contributors) to Drafty made 33,283 total interactions and 1,048 edits during this study period. We manually hand-checked edits made within Drafty from paid crowdworkers and unpaid contributors.

During this study period, anyone who arrived at our public web-based spreadsheet system Drafty could freely choose to complete the discrete choice experiment survey. For the discrete choice experiment, our survey participants come from two distinct groups: 1) paid crowdworkers we recruited from Prolific to add new rows of data in Drafty and 2) unpaid contributors (i.e., Drafty's normal everyday visitors).

Survey Data Pre-Processing. The following section describes the survey study data's review process and removal criteria. A total of 149 people freely chose to take the survey within Drafty. Taking the survey was not part of the paid crowdworkers' task posted on Prolific. Based on the session and profile IDs tracked from Drafty within the survey, the same participant never took the survey multiple times. Among the initial 149 participants, 16 were removed for not completing the survey. After that, 10 additional participants were removed for entering the same response for ten Likert-type scale questions. Finally, 2 additional participants were removed for providing nonsensical answers to qualitative questions, indicating they might be bots or speeding through the survey. For example, their answers did not answer the survey questions and featured significantly longer responses than other participants. See Appendix section C for more details. After answering a sample question, the remaining 121 participants reported understanding the instructions for the discrete choice experiment. The following metrics to evaluate our discrete choice experiment outlined in Section 4.2 are computed using the remaining 121 participants.

Survey Participants. The following summary demographics are for the 121 participants who completed the survey remaining after data cleaning. The average age per participant was 36 years, ranging from 19 to 78. The average completion time for the survey is 19 minutes and 6 seconds. Among several options to indicate gender identity, 58 participants identified as Male and 43 as Female, while 20 preferred not to say or chose another option. Among all participants, 67 indicated payment for a task is for extra spending money, while 26 indicated it helps pay some of their bills.

4.2 Discrete Choice Experiment: Evaluation Metrics used in Results

Below are common metrics and analyses used to evaluate user choice data from discrete choice experiments [77]. The terms utility and preference relate to someone's motivation to complete a crowd contribution task. We have selected metrics that are automatically computed within Qualtrics [106] and we provide how we computed trade-offs using methods recommended by prior research [86].

Average Utility Scores the average utility score of each level across all participants. A utility score is expressed as a linear combination of an attribute's levels. They are reported as xx.X (i.e, 11.6). Higher utility scores indicate more preferred levels within an attribute. The relative differences between utility scores help explain the trade-offs participants make between attributes. The utility scores show the relative preference between levels within an attribute. A level with a positive utility score will increase someone's motivation to complete a crowd contribution task, while a negative score will decrease it. Qualtrics uses a Hierarchical Bayes estimation method to calculate individual respondent utility scores, which are then averaged to get the overall average utility.

Attribute Importance represents the relative weight that participants assign to each attribute when choosing between two hypothetical scenarios. This metric helps determine which attributes most influence their preference. They are reported as xx.X% (i.e, 22.1%). Attribute importance is the proportion of its utility range compared to the total utility ranges across all attributes. Attribute importance is computed by taking the distance of the highest and lowest average utility score per level in an attribute. Imagine if $4 per hour was the lowest at -5.0, and $16 per hour was the highest at 10.0. Then, the attribute importance for pay level would be 15.0%, using the distance between -5.0 and 10.0. If an attribute's importance were 0.0%, it would not affect someone's preference. The greater the attribute importance, the more influence its levels have over someone's preference and motivation. For example, if the pay level per hour had a higher attribute level than the estimated completion time, then pay level would have a bigger influence on user preference and motivation. Our results feature four Figures (5, 6, 7, and 8) showing the attribute importance per attribute across different participant groups and their behaviors.

Optimal Task is the combination of attributes and levels that, based on the data collected from participants, is predicted to be the most preferred choice by the target population. Qualitrics computes this automatically. Normally, this is the level per attribute with the highest average utility score. This metric, optimal task, can help you understand how to design contribution tasks for specific groups of users. For example, maybe unpaid contributors prefer ethical tasks that help their community.

Preference Share measures the probability that a level would be chosen over another when all other attribute levels are the same. It is computed using a Multinomial Logistic Regression model and the utility scores per level within an attribute. In section 4.5, the analysis uses preference share to simulate the total utility score (i.e., level of motivation) between two different crowd contribution tasks. We use preference share to compute trade-offs people make when choosing different tasks. We include a brief explanation of computing trade-offs below.

Trade-offs between two attributes and their levels can be computed using partworth functions [86]. These results can help inform if one level is changed within attribute "A" and how the level of attribute "B" needs to increase or decrease to produce the same utility score (i.e., level of motivation). For example, if the estimated time to complete a task is increased, how much should the pay level increase to maintain the same level of motivation? The values of the partworth functions can be normalized (assigning 0.0 to the smallest and 1.0 to the largest value) to make results comparable. Equation 1 below represents the partworth utility equation for a given individual i choosing option j, where β_1 and β_2 are the partworth utilities for pay and time, respectively.

$$U_{ij} = \beta_1 \cdot \mathrm{PayLevel}_i + \beta_2 \cdot \mathrm{Time}_j + \epsilon_{ij} \tag{1}$$

4.3 Universal Incentives are Pay Level, Time to Complete, and Task Perception

Regardless of whether someone is a paid crowdworker or unpaid contributor, our results show three attributes that should increase someone's motivation to contribute: focus on improving the pay level, reducing the estimated time to complete, and improving someone's perception of the task. Task perception can be if someone finds a task interesting or if they believe their contribution will help others. Figure 4 shows the most important attributes and levels that are universal across users who completed the survey. However, when scrutinizing the results, unpaid contributors and paid crowdworkers make different trade-offs every day when selecting between crowd contribution tasks.

Task perception was most important to unpaid contributors, and pay level was most important to paid crowdworkers. For paid crowdworkers, there was a 74% increase in the attribute importance for pay level compared to the estimated time to complete (See Figure 5; Attribute Importance of Pay Level for Paid Crowdworkers is 174% of Attribute Importance for Estimated Time to Complete). At the same time, there was only a 10% increase among unpaid contributors. This finding that paid crowdworkers value pay level the most mirrors prior research [63]. Also, when comparing attribute importance for pay level and task perception, there is a 70% increase among paid crowdworkers for pay level. At the same time, there is a 40% decrease among unpaid contributors for pay level when comparing to task perception (Figure 5). While the pay level motivates unpaid contributors to complete crowd contribution tasks, task perception plays a larger role. The levels for task perception mainly focus on intrinsic motivators, thus showing our unpaid contributors are influenced by intrinsic motivators [69, 103].

When comparing feature importance for those who only made 100% accurate edits on Drafty, we found that unpaid contributors and paid crowdworkers were less influenced by the pay level per task compared to the same populations who were not accurate (See Figure 4, 100% Accurate Editors columns). Both unpaid contributors and paid crowdworkers were more influenced by task perception. Indicating someone's level of interest and intrinsic motivators can

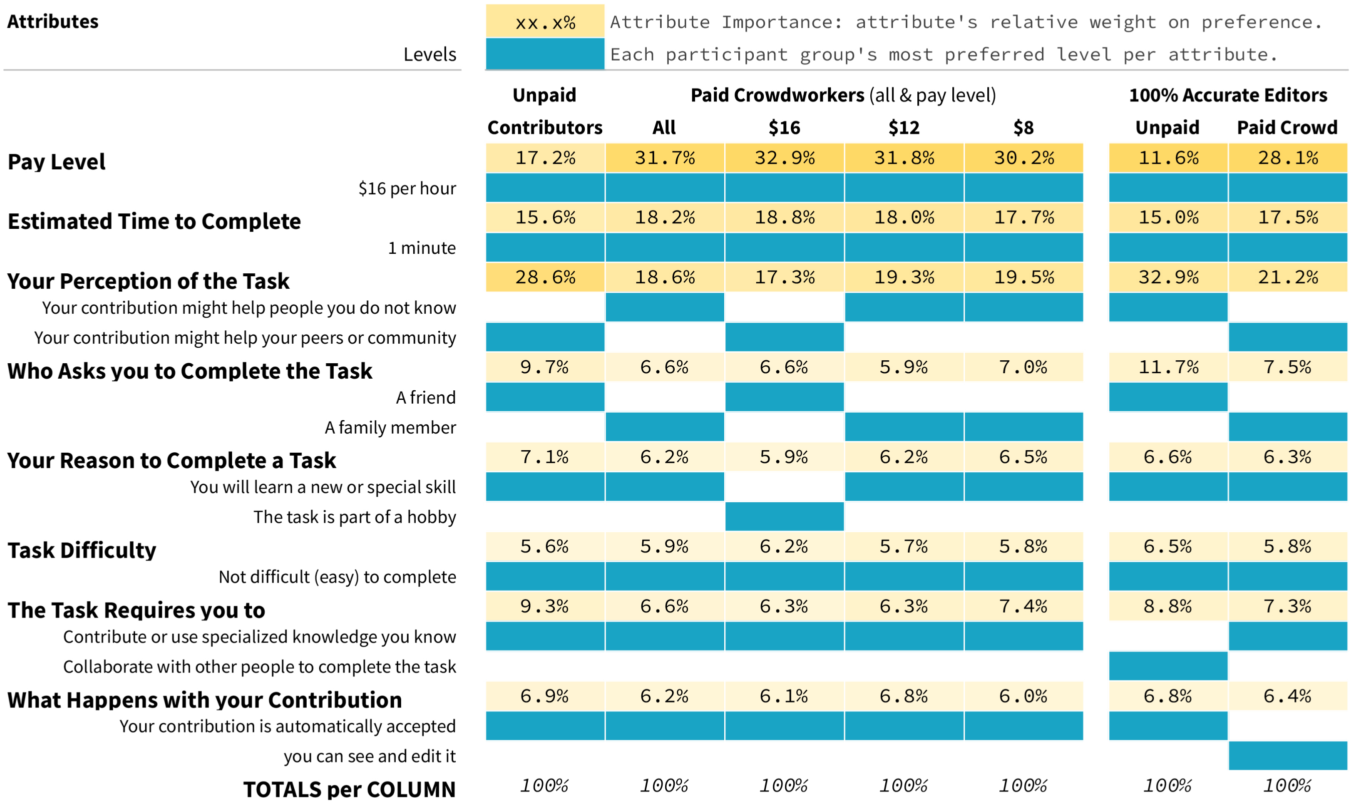

Figure 4: Optimal Crowd Contribution Tasks for Different Populations: This figure shows the optimal crowd contribution task per population. The gold boxes show each attribute's relative importance measure [86], or how much each attribute influenced someone's preference. For example, "Pay Level" is 33.6% for paid crowdworkers whom we compensated $16 per hour. The teal box is the optimal level per attribute. For example, when comparing those who only made 100% accurate edits in Drafty, unpaid contributors preferred tasks where they could collaborate, compared to paid crowdworkers who preferred tasks where they could contribute their specialized knowledge. Overall and unsurprisingly, all participants preferred the highest pay level per hour, the lowest estimated time to complete, and the easiest tasks. While task perception was the most important attribute for unpaid contributors, paid crowdworkers were mainly influenced by the pay level per hour. The teal square indicates the level per attribute with the highest average utility. Each column then constructs the optimal or ideal crowd contribution task per group of people. NOTE: Each column's attribute importance adds up to 100%.

influence their motivation more than money. This result mirrors prior research on paid crowdworkers [19] and demonstrates a similar result among unpaid contributors. Notably, the feature importance between pay level and who is asking for a contribution is similar among unpaid contributors who made 100% accurate edits on Drafty (See Figure 4. This result mirrors prior research by Brady et al. [10] showing that asking friends to contribute can motivate accurate contributions.

The analysis also splits participants (paid crowdworkers and unpaid contributors) by their self-reported knowledge of Computer Science. For participants who self-reported High or Very High, their optimal task design mirrored the attributes and levels of our unpaid contributors. Unpaid contributors with high levels of knowledge or interest in Computer Science were more influenced by tasks that required them to collaborate with people. This observation mirrors our long-term observations of Drafty and other peer production-inspired systems, where collaboration can motivate contributions [4]. Table 2 shows the average responses from our participants, showing a trend that unpaid contributors and paid crowdworkers ($8.00 per hour) self-reported more Knowledge and Interest in various Computer Science related topics and areas.

4.4 Paid Crowdworkers & Unpaid Contributors Share Similar Motivations but Make Different Trade-offs When Selecting Tasks

We compare the trends in average utility scores between unpaid contributors and paid crowdworkers across both groups and among the subset of users who only made 100% accurate edits. The levels of each attribute show patterns about how each group differs. We created figures showing heatmaps of the average utility scores per group; see Figures 5, 6, 7, and 8.

This analysis compares groups where one group (the ideal group) is likely to benefit Drafty more than the other (less ideal). The ideal

	Knowledge about Computer Science					Interest in Computer Science				
Participant Type	**General**	**Profs.**	**Univ.**	**Res.**	**Jobs**	**General**	**Profs.**	**Univ.**	**Res.**	**Jobs**
Unpaid Contributor	3.8	3.3	3.4	3.5	3.1	4.3	3.8	3.9	4.0	3.6
Paid Crowdworkers ($8)	2.9	2.1	2.5	2.7	2.7	3.0	2.3	2.3	3.0	2.7
Paid Crowdworkers ($12)	2.4	2.2	2.4	2.3	2.4	2.8	2.3	2.4	2.7	2.9
Paid Crowdworkers ($16)	2.4	1.8	2.1	2.1	2.4	2.9	2.2	2.2	2.6	2.4

Table 2: Mirroring our quantitative results, unpaid contributors and paid crowdworkers ($8 per hour) had more interest and knowledge in Computer Science than the more highly paid crowdworkers. This table shows participants' average answer to 5-point Likert scale questions (Very Low to Very High). The questions are: "Your level of knowledge about Computer Science" and "Your level of interest in Computer Science" across five areas: overall (in general), professors (faculty members), universities and departments, research, and jobs.

group made only 100% accurate edits. Figures 6 and 8 show pairs of these groups separated by a column of white space, where the group on the left is ideal while the group on the right is less ideal.

Pay Level Per Hour. Prolific's fair payment guidelines match the motivations of unpaid contributors. Across all groups in Figures 5 and 6, a minimum pay level ($8) generated a positive average utility score. The average utility score for the highest pay level per hour ($16) was 2 times higher for paid crowdworkers compared to unpaid contributors (See Figure 5; utility score for All Paid Crowdworkers evaluating Pay Level $16 [i.e., 13.2] is 2x that of Unpaid Contributors [i.e., 6.5]). This trend continues when splitting users based on their real-world usage of Drafty (i.e., did they make an edit, were they accurate). The average utility scores for pay level ($16) for unpaid contributors who did not make edits were 2.2 times higher than those who only made accurate edits. Likewise, The average utility scores for pay level ($0) for unpaid contributors who did not make edits were 1.7 times lower than those who only made accurate edits. The ideal unpaid contributors (i.e., those who only make accurate edits) were more motivated by attributes other than pay level. These exact trends continue for $16 and $0 per hour when comparing the paid crowdworkers who made 100% accurate edits on Drafty. This indicates that accurate users are motivated by more than money [70]. To elicit accurate paid contributions, fairly priced tasks that benefit others are more important than highly paid tasks.

Estimated Time to Complete. For estimated time to complete all groups preferred tasks under 15 minutes. While Figure 5 shows a trend where less time yields more preference among paid crowdworkers, this trend does not exist among unpaid contributors. For unpaid contributors, the average utility score for a 15 minute task is .5 points higher than for 5 minute task. This result replicates prior research showing unpaid contributors will spend more time on a task than paid crowdworkers [46]. Paid crowdworkers who submitted edits to Drafty had the same average utility score for 15 minutes tasks and 5 minute tasks. The paid crowdworkers who did not submit edits were the ones who chose to complete the survey for a possible gift card but abandoned the initial editing task. Viewing their survey responses they felt it was difficult to contribute and find the professor's information. Since their preferences indicate they prefer shorter tasks, it is not a surprise they abandoned the task of editing data on Drafty. There was no difference in trends among paid crowdworkers who submitted 100% accurate edits and those who did not.

Your Perception of the Task. There are multiple trends among the levels across groups for task perception. In every comparison, the ideal group prefers tasks that are ethical, look interesting, and will help others compared to another group. Hence, this is why the attribute importance for task perception is higher among the ideal groups. While we never condone creating tasks or gaining contributions for unethical reasons, unpaid contributors who only made accurate edits valued ethical tasks that help others more than any other group. Also, building on prior research, their interest in the task is more motivating than any other group [19, 104]. Peer production environments rely on people freely engaging and contributing data [4]. These results show how systems like Drafty motivate accurate contributions.

Who Asks you to Complete the Task. Across all groups, people preferred tasks where they volunteered or a friend, family member, or a system (i.e., Wikipedia or Drafty) asked them to contribute. While unpaid contributors on Drafty were the most motivated by freely choosing to contribute, being asked by family and friends was the most motivational look at all users. This is not surprising, as Brady et al. showed that when a friend asks, this can elicit accurate contributions [11]. However, our results also show this trend across paid crowdworkers.

The average utility score for when a friend asks is 2.2 times greater for unpaid contributors compared to paid crowdworkers and 2 times greater for unpaid contributors who only submitted accurate edits compared to those who did not (See Figure 6). Mirroring prior research, all groups showed a strong dislike for tasks where a bot (not a person) on social media [91] or a profit for-profit company [88] asks them to contribute. While Rogstadius et al. showed that people prefer non-profit companies compared to for-profit companies, our results indicate this alone is not enough to motivate someone to complete a task and is relatively similar to people on social media asking for contributions.

Your Reason to Complete a Task. All groups preferred tasks where they would learn a new or special skill, or it is a hobby. This is similar to learnersourcing, where people simultaneously learn new skills about their hobby or an area of interest and then make contributions [49].

[Attribute Importance xx.X%] distance between highest & lowest level per attribute

Red decreases motivation | Blue increases motivation

Change in Preference / Motivation to accept a task

Attributes / Levels	Unpaid Contributors	Paid Crowdworkers (all & pay level) All	$16	$12	$8
Pay Level	[17.2%]	[31.7%]	[33.0%]	[31.8%]	[30.2%]
$0 per hour	-10.7	-18.5	-19.1	-18.4	-18.0
$4 per hour	-2.1	-4.8	-5.4	-4.9	-3.9
$8 per hour	1.2	1.7	1.6	1.7	1.7
$12 per hour	5.0	8.5	9.1	8.3	8.0
$16 per hour	6.5	13.2	13.9	13.4	12.2
Estimated Time to Complete	[15.6%]	[18.2%]	[18.8%]	[18.0%]	[17.7%]
1 minute	6.5	7.4	7.6	7.4	7.1
5 minutes	2.0	3.8	4.0	3.4	3.9
15 minutes	2.5	2.7	2.6	2.7	2.7
30 minutes	-2.0	-3.0	-3.1	-2.9	-3.1
60 minutes	-9.1	-10.8	-11.2	-10.6	-10.6
Your Perception of the Task	[28.6%]	[18.6%]	[17.3%]	[19.3%]	[19.5%]
The task looks boring	-1.5	-0.8	-1.1	-1.2	0.1
The task looks interesting	4.5	3.0	3.7	2.9	2.1
The task might be unethical	-20.6	-13.6	-12.6	-13.8	-13.8
The task is likely ethical	2.1	1.4	1.6	1.7	0.9
Your contribution might help people you do not know	7.4	5.0	3.8	5.5	5.7
Your contribution might help your peers or community	8.0	5.0	4.7	4.9	5.0
Who Asks you to Complete the Task	[9.7%]	[6.6%]	[6.6%]	[5.9%]	[7.0%]
A friend	4.7	2.1	2.7	1.0	2.6
A family member	4.5	2.8	2.7	2.6	2.9
Someone you do not know	-1.1	-0.7	-0.7	-0.6	-0.8
Volunteer	1.9	1.4	1.5	1.3	1.5
People on social media	-0.7	-0.2	-0.2	-0.1	-0.3
A bot (not a person) on social media	-4.9	-3.8	-3.9	-3.3	-4.1
A for-profit company	-5.0	-2.5	-3.1	-1.7	-2.6
A non-profit company	-0.2	0.2	-0.2	-0.3	-0.2
A system (i.e., Wikipedia or Drafty)	1.3	1.3	1.3	1.2	1.3
Paid Crowdsourcing system (i.e., Prolific)	-0.5	-0.2	-0.2	-0.2	-0.2

Figure 5: Preference heatmap Unpaid Contributors and Paid Crowdworkers Part 1 (four attributes with the highest attribute importance). This Figure show's that paid crowdworkers are more extrinsically motivated (pay level per hour) compared to the primarily intrinsic motivations of unpaid contributions (task perception). Also, across all contributors, the estimated time to complete a task was more important than the task's difficulty. This Figure is a heatmap for the Attribute Importance and Utility Score per Level comparing unpaid contributors and paid crowdworkers. The four attributes with the highest Attribute Importance had the highest effect on preference (user motivation). See Figure 7 for the other four attributes. The attribute name is in the leftmost column. Then, the heatmap shows the levels per attribute. In the same row as the attribute is the attribute importance formatted as [xx.X%]. Each level's row contains its utility score.

Attributes — **[Attribute Importance xx.X%] distance between highest & lowest level per attribute**

Red decreases motivation / Blue increases motivation

Change in Preference / Motivation to accept a task

Attributes / Levels	Unpaid Contributors		Paid Crowdworkers		Paid Crowdworkers	
	Made Edits*	**No Edits**	**Made Edits**	**No Edits**	**100% Acc.**	**<100% Acc.**
Pay Level	[11.6%]	[19.6%]	[27.9%]	[33.7%]	[28.1%]	[31.9%]
$0 per hour	-7.0	-12.0	-15.5	-20.2	-16.6	-18.7
$4 per hour	-2.6	-1.9	-5.5	-4.3	-4.0	-4.7
$8 per hour	1.5	1.1	1.4	1.8	1.5	1.7
$12 per hour	4.6	5.2	7.2	9.2	7.6	8.4
$16 per hour	3.5	7.6	12.4	13.5	11.5	13.2
Estimated Time to Complete	[15%]	[15.9%]	[17.1%]	[18.7%]	[17.5%]	[18.2%]
1 minute	5.9	6.9	6.5	7.8	7.2	7.4
5 minutes	2.5	1.7	3.3	4.0	3.2	3.8
15 minutes	2.3	2.6	3.3	2.4	2.6	2.7
30 minutes	-1.6	-2.2	-2.6	-3.3	-2.8	-3.1
60 minutes	-9.1	-9.0	-10.6	-10.9	-10.3	-10.8
Your Perception of the Task	[32.9%]	[26.7%]	[21.3%]	[17.3%]	[21.2%]	[18.5%]
The task looks boring	-2.2	-1.1	-0.3	-1.0	-0.9	-0.7
The task looks interesting	5.2	4.3	2.9	3.0	3.5	3.0
The task might be unethical	-23.6	-19.3	-15.5	-12.6	-15.5	-13.5
The task is likely ethical	2.9	1.8	1.7	1.3	1.6	1.4
Your contribution might help people you do not know	8.4	6.9	5.8	4.7	5.6	4.8
Your contribution might help your peers or community	9.3	7.4	5.5	4.6	5.7	5.0
Who Asks you to Complete the Task	[11.7%]	[8.9%]	[7.3%]	[6.1%]	[7.5%]	[6.5%]
A friend	6.3	3.1	2.8	1.8	2.8	2.4
A family member	3.5	4.3	3.3	2.5	3.3	2.7
Someone you do not know	-0.8	-1.1	-0.6	-0.7	-0.8	-0.7
Volunteer	1.5	1.8	1.6	1.3	1.6	1.4
People on social media	-1.2	-0.4	-0.5	0.0	-0.3	-0.2
A bot (not a person) on social media	-4.1	-4.6	-4.0	-3.6	-4.2	-3.8
A for-profit company	-5.4	-4.0	-3.0	-2.2	-3.2	-2.5
A non-profit company	-0.2	-0.1	-0.6	0.0	-0.2	-0.3
A system (i.e., Wikipedia or Drafty)	0.9	1.3	1.4	1.2	1.3	1.2
Paid Crowdsourcing system (i.e., Prolific)	-0.5	-0.4	-0.3	-0.2	-0.3	-0.2

Figure 6: Preference heatmap Edits and Accuracy Part 1 (four attributes with the highest attribute importance). Among all contributors, highly accurate contributors were less motivated by pay level and more motivated by task perception. Task perception feature intrinsic motivators such as a task looks interesting, ethical, and the contribution might help others. This Figure is a heatmap for the Attribute Importance and Utility Score per Level comparing contributors based on the accuracy of their edits, or if they edited data on Drafty. The four attributes with the highest Attribute Importance had the highest effect on preference (user motivation). See Figure 8 for the other four attributes. The attribute name is in the leftmost column. Then, the heatmap shows the levels per attribute. In the same row as the attribute is the attribute importance formatted as [xx.X%]. Each level's row contains its utility score. *Make Edits includes unpaid contributors who only made 100% accurate edits in Drafty.

While learning communities have successfully used gamification to increase engagement by awarding reputation points (badges, points, credit, etc.) [66], Drafty's users (paid and unpaid) did not prefer this motivator (See Figure 7). This result mirrors prior research showing how gamification mechanisms can decrease the quantity and quality of crowd contributions over time [58]. Considering how strong of an intrinsic motivator helping others is, the extrinsic motivator of earning badges does not matter in Drafty's context.

Unpaid contributors who only submitted accurate edits were almost twice as motivated by tasks where they might be paid for doing exceptional work (See Figure 8). This idea of possibly paying others for doing exceptional work produced a average utility score that is 181% greater among unpaid contributors who only submitted accurate edits compared to those who did not. This is additional evidence that mixing extrinsic rewards in Drafty could further motivate its existing highly accurate visitors [13].

Task Difficulty. Task difficulty is one of the most consistent attributes across this study in terms of affecting the preferences and motivation of all users. While task difficulty is often optimized by system designers [39], it is also cited as affecting people's perceptions of task time [56]. It is not a surprise, crowdworkers paid $16 per hour valued easy tasks more than others. They are trying to optimize their actual earnings per hour. Among unpaid contributors, those who did not submit edits were the only group that showed a positive average utility score for moderately difficult tasks (See Figure 8; Unpaid Contributors, No Edits column). Maybe this group did not edit data on Drafty, because the tasks to contribute were too simple and straightforward. Future research could assess how to appeal to users who only become contributors over time.

The Task Requires you to. Drafty often requires people to contribute or use specialized knowledge you know for data types requiring domain-specific expertise. While this positively motivates every group, unpaid contributors, who only made accurate edits preferred to collaborate with others the most (See Figure 8). Perhaps collaboration can lead to a greater sense of community and contributions to helping others, which is why this group preferred collaboration the most [3, 55]. The average utility score for collaborating reported by unpaid contributors was 10 times higher than paid crowdworkers. It is possible, paid crowdworkers might view that collaborating with others could also increase the time to complete. Looking at possible future AI-centric trends in crowdsourcing, while Drafty does not integrate an AI to help people complete tasks, there were trends in participant responses. Unpaid contributors who only made accurate edits had the strongest negative preference for tasks requiring collaborating with an AI, while they had the strongest positive preference for collaborating with other people. Paid crowdworkers were open to the idea of collaborating with AI to complete a task (Figure 8). This mirrors a growing trend where AI is being used to recommend tasks [16].

What Happens with your Contribution. When people complete crowd contribution tasks, what happens to the data they contributed? Is it publicly available in a system for them to go back and see or edit? Or is it stored in a private database or repository and never seen again? It can be common in paid crowdsourcing, where people make contributions and never access the data they contributed again. Bernstein raised this concern when discussing his crowd-powered system Soylent [8]. In our results, all groups had a negative preference for tasks where they could not see or edit their contribution again. The average utility score for crowdworkers paid $12 an hour was 25% lower than those paid $16 per hour (See Figure 7). The more crowdworkers were paid, the less ownership of their contribution mattered. However, the average utility scores were similar when comparing paid crowdworkers who only made accurate edits versus those who did not.

Ensuring contributors (paid and unpaid) can see and edit their contributions will motivate accurate contributions. The average utility score for contributions where people can see and edit was 5 times greater for paid crowdworkers who submitted edits versus those who did not (Figure 8). This trend continued among unpaid contributors, where the average utility score was 1.6 for those who made accurate edits compared to 0.0 for those who did not.

Overall, everyone preferred tasks where they could remain anonymous and their contribution is automatically accepted (see Figure 7). The main spreadsheet of Drafty enforces these interactions.

4.5 Computing Trade-offs to Optimize Incentives for Paid Crowdworkers and Unpaid Contributors

The previous section identifies and discusses how groups make different trade-offs when selecting crowd contribution tasks. This section computes the trade-offs for pay level and estimated time to complete across various attributes and levels. Figure 9 shows how preferences and motivations shift as pay levels and task time are manipulated. See Appendix section D for more details.

How to compute trade-offs? A simple method to compute trade-offs is comparing the preference share (i.e., total preference or utility) for a given crowd contribution task compared to a group's optimal crowd contribution task. We use partworth functions to compute trade-offs from the results of a discrete choice experiment [86]. We can measure the trade-offs paid crowdworkers and unpaid contributors make between two attributes by using their levels' average utility scores to understand how to change pay levels or task times [86].

In this section, we will walk through an initial example of using pay level per hour and estimated time to complete a task. These two highly motivating attributes are easy to manipulate when designing new crowd contribution tasks.

How does reducing estimated task completion time affect the pay level? We have a task that pays $12 per hour and takes an estimated 30 minutes to complete. If we can reduce the estimated time to complete from 30 to 15 minutes, what is the maximum we can decrease the pay level per hour to maintain the same level of motivation for this new task that takes less time to complete?

The average utility scores can be used to compute this trade-off. Figure 5 shows the average utility scores for paid crowdworkers and unpaid contributors.

For unpaid contributors, using the average utility scores for the estimated time to complete, we note that reducing the estimated time will increase the utility (user motivation) to complete the task by 5.5. How much should we modify the pay level per task to maintain the same level of utility? Since the pay level's average

[Attribute Importance xx.X%] distance between highest & lowest level per attribute

Red decreases motivation — Blue increases motivation

Change in Preference / Motivation to accept a task

Attributes / Levels	Unpaid Contributors	Paid Crowdworkers (all & pay level) All	$16	$12	$8
Your Reason to Complete a Task	[7.1%]	[6.2%]	[5.9%]	[6.2%]	[6.5%]
You might be paid for doing exceptional work	-0.7	-1.5	-1.1	-1.0	-2.4
Your contribution benefits you personally	0.7	0.6	0.2	0.8	0.6
You will learn a new or special skill	3.7	3.3	2.6	3.0	4.1
Personal rec. or learn something new about you	-0.2	0.0	0.4	-0.2	-0.1
Reputation points in a system (badges, points, credit)	-3.4	-2.9	-2.9	-3.2	-2.4
The task is part of your job	-1.7	-1.9	-2.2	-1.5	-1.8
The task is part of a hobby	1.6	2.5	3.0	2.2	2.1
Task Difficulty	[5.6%]	[5.9%]	[6.2%]	[5.7%]	[5.8%]
Not difficult (easy) to complete	2.5	2.9	3.2	2.7	2.9
Moderately difficult to complete	0.6	0.0	-0.2	0.3	0.0
Very difficult to complete	-3.1	-3.0	-3.0	-3.0	-2.9
The Task Requires you to	[9.3%]	[6.6%]	[6.3%]	[6.3%]	[7.4%]
Collaborate with other people to complete the task	3.2	-0.4	-0.4	-1.1	0.2
Complete the task with Artificial Intelligence	-1.1	0.2	0.1	0.6	-0.2
Complete the task by yourself	-0.5	-0.3	-0.1	-0.4	-0.4
Learn something new	1.1	1.5	1.0	2.2	1.5
Contribute or use specialized knowledge you know	3.3	2.8	2.9	2.5	3.1
Provide your personal information	-6.0	-3.8	-3.4	-3.8	-4.3
What Happens with your Contribution	[6.9%]	[6.2%]	[6.1%]	[6.8%]	[6.0%]
You do not own it (you cannot see or edit it)	-4.1	-3.4	-3.0	-4.0	-3.5
You own it (you can see and edit it)	0.5	0.3	-0.9	1.3	0.9
A public community owns it (anyone can see and edit it)	0.2	0.2	0.3	0.3	0.2
Your name / username is attached to it (not anonymous)	-0.2	-0.9	-0.5	-1.5	-0.7
You receive no credit for it (it is anonymous)	0.9	1.1	1.2	1.3	0.7
Your contribution could be rejected	-0.2	-0.1	-0.1	-0.2	0.1
Your contribution is automatically accepted	2.8	2.8	3.1	2.8	2.5

Figure 7: Preference heatmap Unpaid Contributors and Paid Crowdworkers Part 2 (four attributes with the lowest attribute importance). This Figure is a heatmap for the Attribute Importance and Utility Score per Level comparing unpaid contributors and paid crowdworkers. The four attributes with the lowest Attribute Importance had the lowest effect on preference (user motivation). See Figure 5 for the top four attributes. The attribute name is in the leftmost column. Then, the heatmap shows the levels per attribute. In the same row as the attribute is the attribute importance formatted as [xx.X%]. Each level's row contains its utility score. *Make Edits includes unpaid contributors who only made 100% accurate edits in Drafty.

utility scores decrease as the pay level decreases, we can look at the difference in relative utility between $12 per hour and $8 per hour (the next lowest pay level per hour). Going from $12 to $8 per hour will decrease utility by 3.8.

Equation 2 shows the trade-off for pay level per hour when reducing task completion time from 30 to 15 minutes for unpaid contributors to Drafty. The numerator is the difference in the average utility scores for the two levels we will modify (i.e., estimated time to complete, 30 minutes to 15 minutes). The denominator is the difference in the average utility scores for the two levels we will modify (i.e., estimated time to complete, 30 minutes to 15 minutes). The result of this fraction will be multiplied by the difference

Attributes — **[Attribute Importance xx.X%] distance between highest & lowest level per attribute**

Red decreases motivation — Blue increases motivation

Change in Preference / Motivation to accept a task

Attributes / Levels	Unpaid Contributors Made Edits*	Unpaid Contributors No Edits	Paid Crowdworkers Made Edits	Paid Crowdworkers No Edits	Paid Crowdworkers 100% Acc.	Paid Crowdworkers <100% Acc.
Your Reason to Complete a Task	[6.6%]	[7.4%]	[6.9%]	[5.8%]	[6.3%]	[6.1%]
You might be paid for doing exceptional work	1.8	-2.2	-1.5	-1.5	-1.2	-1.8
Your contribution benefits you personally	0.4	0.9	0.8	0.4	0.6	0.6
You will learn a new or special skill	2.9	4.2	3.6	3.1	3.4	3.3
Personal rec. or learn something new about you	-0.8	0.1	0.0	0.0	-0.1	0.1
Reputation points in a system (badges, points, credit)	-3.7	-3.2	-3.3	-2.7	-2.9	-2.8
The task is part of your job	-1.5	-1.7	-2.0	-1.8	-1.8	-1.9
The task is part of a hobby	1.0	2.0	2.5	2.4	2.1	2.5
Task Difficulty	[6.5%]	[5.1%]	[6.1%]	[5.8%]	[5.8%]	[6.0%]
Not difficult (easy) to complete	3.3	2.0	3.1	2.9	2.8	3.0
Moderately difficult to complete	-0.1	1.1	-0.1	0.0	0.2	0.0
Very difficult to complete	-3.2	-3.1	-3.0	-2.9	-3.0	-3.0
The Task Requires you to	[8.8%]	[9.5%]	[7.1%]	[6.4%]	[7.3%]	[6.6%]
Collaborate with other people to complete the task	4.0	2.4	-0.8	-0.2	0.2	-0.5
Complete the task with Artificial Intelligence	-2.4	-0.1	0.5	-0.1	-0.2	0.1
Complete the task by yourself	-0.7	-0.4	-0.3	-0.3	-0.3	-0.3
Learn something new	1.0	1.1	1.7	1.5	1.5	1.6
Contribute or use specialized knowledge you know	2.8	3.3	3.0	2.8	3.0	2.8
Provide your personal information	-4.8	-6.2	-4.1	-3.6	-4.3	-3.8
What Happens with your Contribution	[6.8%]	[7.0%]	[6.1%]	[6.3%]	[6.4%]	[6.3%]
You do not own it (you cannot see or edit it)	-3.9	-4.2	-3.8	-3.2	-3.5	-3.4
You own it (you can see and edit it)	1.6	0.0	2.0	-0.5	0.2	0.1
A public community owns it (anyone can see and edit it)	0.4	0.2	0.2	0.3	0.2	0.2
Your name / username is attached to it (not anonymous)	-1.8	0.5	-1.3	-0.7	-0.6	-0.8
You receive no credit for it (it is anonymous)	0.9	0.9	0.8	1.2	1.0	1.1
Your contribution could be rejected	-0.1	-0.2	-0.2	-0.1	-0.2	-0.1
Your contribution is automatically accepted	2.9	2.8	2.3	3.1	2.9	2.9

Figure 8: Preference heatmap Edits and Accuracy Part 2 (four attributes with the lowest attribute importance). All survey participants preferred tasks that primarily focus on intrinsic motivators: contributing their knowledge, completing tasks for friends and family, or on public/non-profit systems like Wikipedia or Drafty, and making anonymous contributions that are automatically accepted. Although, paid crowdworkers preferred collaborating with AI, while unpaid contributors preferred collaborating with other people. Both paid crowdworkers and unpaid contributors who made edits preferred tasks where they can see and edit their contributions compared to those who did not edit data on Drafty. This Figure is a heatmap for the Attribute Importance and Utility Score per Level comparing contributors based on the accuracy of their edits or if they edited data on Drafty. The four attributes with the lowest Attribute Importance had the lowest effect on preference (user motivation). See Figure 6 for the top four attributes. The attribute name is in the leftmost column. Then, the heatmap shows the levels per attribute. In the same row as the attribute is the attribute importance formatted as [xx.X%]. Each level's row contains its utility score. *Make Edits includes unpaid contributors who only made 100% accurate edits in Drafty.

between the level for the attribute we want to know the tradeoff for (i.e., pay level of \$4 per hour).

$$\text{tradeoff} = \frac{|-2.0 - 2.5|}{|5.0 - 1.2|} \cdot \$4 = \$4.74 \tag{2}$$

Unpaid Contributors

Time & Pay	$0	$4	$8	$12	$16
1	14%	30%	29%	45%	**50%**
5	11%	20%	18%	29%	32%
15	12%	22%	22%	34%	36%
30	11%	13%	9%	12%	16%
60	6%	10%	10%	8%	10%

Unpaid Contributors 100% Accurate

Time & Pay	$0	$4	$8	$12	$16
1	20%	23%	32%	**50%**	44%
5	16%	23%	32%	41%	45%
15	16%	19%	36%	43%	43%
30	15%	7%	9%	16%	22%
60	0%	0%	0%	5%	7%

Paid Crowdworkers

Time & Pay	$0	$4	$8	$12	$16
1	1%	7%	14%	34%	**50%**
5	4%	4%	8%	26%	46%
15	4%	4%	7%	23%	42%
30	2%	3%	4%	10%	20%
60	3%	3%	4%	6%	10%

Paid Crowdworkers 100% Accurate

Time & Pay	$0	$4	$8	$12	$16
1	1%	7%	13%	42%	**50%**
5	0%	1%	16%	20%	44%
15	0%	4%	7%	22%	33%
30	0%	1%	0%	19%	7%
60	0%	0%	0%	1%	0%

Figure 9: This Figure shows the Preference Share when comparing the optimal crowd contribution task against different combinations of pay level and estimated time to complete across unpaid contributors and paid crowdworkers (overall and those who only submitted accurate edits). The teal cell with 50% is the optimal task per group. It is 50% because comparing Preference Share across two identical tasks would result in a perfect split of preference. People with 100% accuracy prefer shorter task times but are more willing to do a 5 to 15 minute task for a lower pay level.

Suppose we can reduce the estimated time to complete by 15 minutes for a group of Drafty's unpaid contributors. In that case, we can also reduce the pay level per hour by a maximum of $4.74 per hour so the new task maintains the same level of utility or motivation to the user. For paid crowdworkers following the same method, reducing estimated task completion from 30 to 15 minutes would allow us to only reduce by a maximum of $3.35 per hour. That is because the pay level per hour has a greater effect on paid crowdworkers' motivation compared to unpaid contributors; see the average utility scores for $8 and $12 per hour in Figure 5. For paid crowdworkers, a reduction of $3.35 per hour from a base level of $12 would still fit within Prolific's fair payment guidelines.

How does making a task more interesting affect the estimated pay level per hour? Maybe a requester has had difficulty attracting paid crowdworkers for a task that appears boring even though it pays $16 per hour. If they spent time to make the task interesting, they could decrease the pay level by a maximum of $3.23 per hour for paid crowdworkers. Prior research shows that unpaid contributors will make contributions matching their interests [19, 103]. If a task paying $16 per hour was redesigned to be more interesting to unpaid contributors, the pay level could be reduced by a maximum of $16 per hour. For unpaid contributors who contribute their time and knowledge, this affects task perception and aligns contributions with their interests.

How does making a task more interesting affect the estimated completion time? What if you are designing a public data system with lengthy tasks that the average contributor finds boring? If a 15-minute task was made more interesting, how much longer would someone spend on the task? This assumes they maintain the same level of motivation. For unpaid contributors, the task could be 20 minutes longer, while for a paid crowdworker, the task could be 10 minutes longer. Unpaid contributors are willing to spend twice as long on an interesting task because they are more intrinsically motivated. Regardless, a requester could save time and money by increasing paid crowdworkers' interest in a task.

4.6 Comparing Paid Crowdworker's Motivations Across Different Pay-Levels

This section compares the trade-offs paid crowdworkers make based by splitting them into three groups based on the pay level per hour they received on Prolific ($8, $12, or $16 per hour) to complete editing tasks on Drafty.

Across all groups of paid crowdworkers, the attributes with the highest feature importance are pay level per hour, the estimated time to complete a task, and then the perception per task. Overall, paid crowdworkers' motivations are similar when paying them fairly based on Prolific's guidelines of a pay level of $8, $12, or $16 per hour. Across all paid crowdworkers, there was a negative effect on their motivation when the pay level per hour was below Prolific's minimum of $8 per hour. Thus, Prolific's pay level per hour recommendations mirror the preferences among the paid crowdworkers who completed our survey.

When looking closer, there are small trends in motivations per attribute and level when increasing or decreasing the pay level per hour. The higher we paid a crowdworker to complete a task, the more they valued a higher pay level when selecting between hypothetical tasks (see Appendix section D.2). However, this did not increase their accuracy.

Results show when recruiting paid crowdworkers, the higher the pay level per hour ($8, $12, and $16 per hour) they were paid, they also preferred a slightly higher pay level. When increasing the pay level per hour they also preferred shorter tasks. There is a 6.2% increase in the attribute importance for estimated time to complete a task between paid crowdworkers paid $8 per hour versus those paid $16 per hour. If you reduced the task time from 30 to 15 minutes for a task that pays $12 per hour, what is the maximum reduction in pay level per hour per paid crowdworker that would not negatively affect their preferences? The pay level per hour could be reduced by a maximum of $3.04 per hour for crowdworkers paid $16 per hour, $3.39 per hour for crowdworkers paid $12 per hour, and $3.68 per hour for crowdworkers paid $8 per hour. The higher-paid crowdworkers are likely trying to maximize their earnings per hour [39, 44]. Table 3 shows the lower the pay level per hour, the more likely paid crowdworkers used the money earned from completing paid crowdsourcing tasks on Prolific as extra spending money. Overall, our results mirror prior research showing the pay level per hour is one of the most important attributes for motivating paid crowdworkers to select a task[90].

4.7 When Pay Level is Lower, Crowdworkers are more Motivated by Altruistic Tasks than Interesting Tasks

Figure 5 outlines paid crowdworkers' perception of the task using the attribute importance per attribute; see the "Your Perception of a Task" part of the Figure. For example, how motivating is a task where "your contribution might help you do not know"? This is an example of altruistic motivation [69]. This also includes how "interesting" or "boring" tasks affect motivation.

The more we paid crowdworkers per hour, the greater their interest in a task influenced their preference. The more we paid a crowdworker, the greater the difference in their preference between boring and interesting tasks (increases of 2.0, 4.1, and 4.8 for motivation among crowdworkers we paid $8, $12, and $16 per hour, respectively). For crowdworkers paid $16 per hour, the interest per task had a 2.4 times greater effect on preference and motivation than it did for crowdworkers paid $8 per hour. This does not mean that crowdworkers paid $8 per hour prefer boring tasks but that higher paid crowdworkers are seeking more interesting tasks. If we look at the trade-offs for task interest and pay level, if a task paying $12 per hour was modified to change it from boring to interesting, the pay level per hour could be reduced by a maximum of $4.08 per hour for paid crowdworkers paid $16 per hour, and by a maximum of $1.9 per hour for paid crowdworkers paid $8 per hour to maintain the same level of preference for the task. Paid crowdworkers, compensated at $8 per hour, were motivated the least by "boring" tasks. They were primarily motivated by tasks where their contribution "might help people they do not know".

One noticeable difference for crowdworkers (paid $8 per hour vs. $16) is the preference for tasks where their contribution might help people they do not know (3.8 vs. 5.7). This is a 50% increase for those paid $8 per hour. Prior work shows that paid crowdworkers are motivated by altruistic tasks where they are helping others and not themselves [69]. However, our results tease apart this observation further. Crowdworkers, paid fairly but lower, are more motivated by tasks with altruistic contributions than tasks they find interesting. A limitation of this observation is that maybe the crowdworkers who self-select lower-paying tasks are the ones who gravitate towards tasks they feel are altruistic. At the minimum, as a requester, paying lower but fair wages can help you attract more altruistically motivated crowdworkers.

4.8 Paid Crowdworkers Higher Pay does not Increase Accuracy

Prior research shows there is a law of diminishing return when it applies to pay level and the quality of work submitted by paid crowdworkers [63]. We compensated paid crowdworkers based on the pay levels per hour ($8, $12, and $16 per hour) required by Prolific. Among paid crowdworkers, the overall accuracy per pay level is 73% for $16 per hour, 76% for $12 per hour, and 79% for $8 per hour. When comparing the number of correct and incorrect edits, a Shapiro-Wilk test for normality reveals the data is not normally distributed. A Kruskal-Wallis H-test reveals no difference between the pay level awarded per paid crowdworker and the accuracy of their edits. There are also no noticeable differences across accuracy per data type among the different pay levels. These results show that while paid crowdworkers indicated increased pay levels are highly motivating, this did not result in more accurate contributions.

4.9 Unpaid Contributors are More Accurate than Paid Crowdworkers

Table 4 shows that everyday unpaid visitors were more accurate across contributions to every data type (i.e., column) within Drafty. Notably, everyday unpaid contributors were 1.9 times more accurate than paid crowdworkers when submitting edits for a professor's join year. Also, everyday unpaid contributors were 1.5 times more accurate than paid crowdworkers when submitting edits for a professor's subfield area of expertise. These results mirror prior research on a similar dataset [103]. However, our paper's results provide more internal validity because all edits were submitted to the same system during the same period using the same system, answering a call for future research [103].

Among unpaid contributors and paid crowdworkers, one common error was submitting nothing for Bachelor's degree when adding new rows of data. Hand-checking edits for Bachelor's or join year often required visiting and reviewing multiple sources, such as a professor's webpage or CV. This increased effort to find information likely demotivated people to take the extra time to verify the correctness of their proposed edits. Common errors among paid crowdworkers included:

(1) Submitting another year from a professor's webpage or materials as their join year.
(2) Submitting blank values for where a professor received their Bachelor's degree.
(3) Confusing Artificial Intelligence and Machine Learning.
(4) Cannot identify a primary research area among multiple reported research areas on a professor's materials.
(5) Adding non-tenure track or emeritus faculty as new rows.

Some of these errors indicate a lack of effort to keep looking for the correct information. While other errors, especially for subfields (i.e., primary research area), demonstrate a lack of domain-specific knowledge among paid crowdworkers. While paying more money might provide the incentive to keep looking for difficult-to-find information, directly integrating data sources into Drafty using services like WikiData or ChatGPT might prove to be more beneficial in reducing the time and effort required to verify and submit accurate edits.

Paid crowdsourcing income used for...	$16 per hour	$12 per hour	$8 per hour
Extra spending money	50.0%	74.1%	82.1%
Helps pay some bills and expenses	40.0%	22.2%	17.9%
Helps pay the majority of bills and expenses	3.3%	0.0%	0.0%
Only source of income	6.7%	3.7%	0.0%

Table 3: Among the paid crowdworkers who completed the survey, the higher they were paid per hour, the more likely they relied on the money from completing tasks to pay at least part of their expenses. Notably, half of the highest paid crowdworkers ($16 per hour) relied on money earned from paid crowdsourcing to pay at least part of their expenses.

	All Edits	Add Row	Delete Row	Full Name	University	Join Year	Subfield	Bachelors	PhD
Unpaid Contributors									
Accuracy	96%	100%	100%	98%	98%	94%	96%	87%	98%
Edits Checked	341	46	7	41	48	50	52	54	43
Paid Crowdworkers									
Accuracy	76%	73%		94%	98%	50%	63%	78%	79%
Edits Checked	340	63		46	46	46	46	46	47
Chi-Squared Test Comparing Unpaid Contributors and Paid Crowdworkers									
p-value	<0.001	<0.001		1.0	<0.001	<0.001	<0.001	0.370	<0.001
χ^2	51.5	11.3		0	0	21.3	15.1	0.8	5.9

Table 4: Unpaid contributors (i.e., everyday visitors) to Drafty made accurate edits at a higher rate than paid crowdworkers. Paid crowdworkers often misinterpreted a professor's subfield area of expertise and submitted nothing or the incorrect year a professor joined a university. The most common error among all users was leaving the Bachelor's blank. After reviewing individual professors' web pages, finding what university granted someone's Bachelor's degree often required viewing their CV or LinkedIn profile page. There were no tasks that required paid crowdworkers to delete rows of data.

4.10 Qualitative Results: The Lower the Pay the more Intrinsic Motivation

During the survey, participants could freely answer two open-ended questions: Q1) "What motivates you to voluntarily contribute to public data systems like Wikipedia, WikiData, or Drafty?" and Q2) "What motivates you to complete paid crowdsourcing tasks on platforms such as Prolific, Amazon Mechanical Turk, or UserTesting.com?". Out of the 121 participants, 100 provided an answer to at least one of these questions. We performed a content analysis on Q1 to determine if participant motivation to contribute to public data systems was extrinsic or intrinsic. Table 5 summarizes the results, showing that while all types of our participants shared some intrinsic motivators, the more someone was paid, the more they were likely to be extrinsically motivated. These results mirror our other results, showing how even among paid crowdworkers, those who were paid more were also more extrinsically motivated to contribute.

Participant responses aligned overall with the attributes and levels of the discrete choice experiment. Some participants mentioned contributing because of monetary compensation, as P11 (paid crowdworker) explained, "[I contribute for] rewards systems, money, if those two things are not available if I know the information without having to spend time researching it [then I will contribute]." Other participants mentioned that contributing might benefit others or their community. P65 (paid crowdworker) highlighted the public benefit of their contribution, "[I contribute because] it's for the public good. Someone, somewhere, will eventually benefit from the information I'm adding." Some participants highlighted their own interest, either in the platform or in how the data might be helpful to them. P58 explained that they felt motivated to contribute "if the task is at least somewhat interesting and if I feel that my contribution is helpful/useful in any way".

5 Key Takeaways

In this study, we identify the factors and trade-offs that impact the motivation of unpaid contributors and paid crowdworkers when selecting tasks to contribute their time, effort, and knowledge. In doing so, we better understand how crowd contribution tasks should be developed to motivate accurate contributions.

We find that tasks related to crowdworker interests or expertise are more motivating. Subsequently, when tasks are more intrinsically motivating to paid crowdworkers, they are more willing to complete longer tasks or have a lower pay rate. Our results also show that highly accurate paid crowdworkers were more intrinsically motivated than inaccurate paid crowdworkers. Their primary motivation was to help others, which was equally important to the estimated task length. Furthermore, we found that among our paid crowdworkers, what we paid them per hour did not impact accuracy. For a **Systems Recommendation**, we suggest that paid crowdsourcing platforms should develop filtering criteria to match

Participant Type	Portion of responses with an intrinsic motivation to contribute	
	to public data systems	to paid crowdsourcing platforms
Unpaid Contributor	100%	50%
Paid Crowdworker ($8)	89%	31%
Paid Crowdworker ($12)	70%	22%
Paid Crowdworker ($16)	69%	7%

Table 5: We labeled survey responses to the optional questions "What motivates you to voluntarily contribute to public data systems like Wikipedia, WikiData, or Drafty?" and "What motivates you to complete paid crowdsourcing tasks on platforms such as Prolific, Amazon Mechanical Turk, or UserTesting.com?". Our qualitative results show that while everyone shares some intrinsic motivators, the more someone was paid, the more they were intrinsically motivated. This mirrors prior results from our discrete choice experiment. We note that only 100 total participants across all survey respondents answered at least one question. Therefore, our reported percentages only reflect a subset of participants represented in our experiment.

the content and requirements of tasks to the interests of paid crowdworkers. This way, requesters can find a pool of workers whose intrinsic motivations match their task, leading to more motivated and accurate workers.

In this study, we compared paid crowdworkers and unpaid contributors in a controlled environment and over the same time period. We found that in our study, unpaid contributors are more accurate overall than paid crowdworkers, regardless of pay rate or interest in the task. Interestingly, in line with our finding that more accurate paid crowdworkers were more intrinsically motivated, we also find that unpaid contributors, in general, are more motivated by intrinsic factors, such as helping their community or their perceived usefulness of the data. The motivation of unpaid contributors was also less impacted by the extrinsic factors of pay rate or time to complete the task than paid crowdworkers. For a **Systems Recommendation**, we recommend the long-term maintenance of data and information; creating a community of intrinsically motivated unpaid contributors is more beneficial than continuously relying on paid crowdworkers. This mirrors prior research from Wikipedia, WikiData, and MovieLens [3, 112].

Across all contributors, we find a preference for control over one's contribution and over one's own self-presentation on the platform. Contributors preferred to have credit for their contributions, the ability to edit their contributions after the fact, and the ability to choose to have their edits be anonymous. Among paid crowdworkers, the higher their current pay, the less these preferences impacted their motivation. We also find that the gamification mechanism of points or credits built over time by completing tasks had the strongest negative effect on user motivation across the whole study. This finding suggests that in the context of our niche system and dataset, gamification may not be an effective method to encourage participation. Finally, as mentioned previously, pre-existing knowledge or interest in a topic was a strong motivator to contribute and an indicator of accuracy across all contributors. This suggests that researchers should strive to recruit people with pre-existing knowledge or interest in the task. Contributors were also motivated to a lesser extent by the opportunity to learn new skills as part of their contribution, presenting another option for increasing motivation when pre-existing knowledge or special skills are not useful. For a **Systems Recommendations**, we encourage designing systems and mechanisms to create intrinsic motivators from the system's data. For example, creating publicly shareable insights from the information or data. This should provide a perpetual mechanism to recruit a community of like-minded and accurate contributors.

Overall, our findings suggest that crowd-powered systems can increase user motivation and appeal to highly accurate contributors by focusing on improving contributor perceptions of the task. Contributors find tasks that are interesting, anonymous, educational, or beneficial for their community to be more motivating. Meaning that they are more willing to take on longer tasks for less pay when these motivating factors exist. By that same note, tasks with less intrinsic motivation can be compensated for by improving the pay level or reducing the time to complete them. From the opposite perspective, our findings suggest that tasks should avoid asking for personal information, requesting actions that contributors find unethical, and requesting contributions through online bots or for-profit companies. In Section 6 we provide additional **Systems Recommendations** for how systems, platforms, and researchers can design hybrid systems appealing to both paid crowdworkers and unpaid contributors.

6 Recommendations for Hybrid Paid and Unpaid Public Data Systems

Our study aims to better understand the motivators behind crowd contributions in public data systems. This study was possible because of our longitudinal effort to create an engaged community of unpaid contributors in Drafty, overcoming a challenge posed by previous researchers [30, 103].

To design our discrete choice experiment, we combined a long history of prior research on intrinsic and extrinsic motivators in crowd power work (see Section3.2.2). Our work, to the best of our knowledge, is the first within human-computer interaction to simultaneously study these motivational factors using a discrete choice experiment with intrinsically and extrinsically motivated contributors in a real-world system. Compared to prior work, our discrete choice experiment's results provide quantitative evidence of how each motivational factor affects user preferences for crowd contribution tasks. Future researchers, requesters, and system designers can use out attribute importance and utility scores to develop and improve fair and equitable crowd contribution systems.

To this end, we specifically asked:

RQ1 Who makes more accurate contributions: paid crowdworkers or unpaid contributors?
RQ2 How do the motivations of highly accurate vs. inaccurate contributors differ?
RQ3 What attributes and levels for crowd contribution tasks universally motivate a public data system's paid crowdworkers and unpaid contributors?
RQ4 What attributes and levels for crowd contribution tasks should be individuated per group of users (paid crowdworkers vs. unpaid contributors)?

Here, we address all of these questions and how these findings may inform future research and development.

6.1 RQ1) Unpaid Contributors are More Accurate than Paid Crowdworkers

As mentioned previously, within our niche system focusing on people editing a tabular dataset of Computer Science faculty profiles, we find that unpaid contributors are more accurate than paid crowdworkers. This result holds regardless of crowdworker pay rate or interest in the task. It is also true across all column types, even subfield areas of research, which require domain-specific knowledge to understand. Past research shows a similar relationship when studying paid crowdworkers [34, 92]. Our results provide more internal validity for other findings that compared the editing behaviors of paid crowdworkers and unpaid contributors using different systems, different versions of a dataset, and at different times [103]. Our study controls for all of these validity issues by studying all contributors (paid and unpaid) using the same system and dataset simultaneously.

Our research focuses on a singular tabular dataset of Computer Science profiles with hundreds of thousands of visitors, much like another singular system, Movielens [112]. Within our narrow community of Computer Science, we believe the unpaid contributors are more accurate because our system, Drafty, has been developed over 9 years to cater to the community's needs and requests. People often request the raw dataset to run analyses, many of which are publicly available through various websites and sources[1]:

- Data analysis about professors, rankings, best papers, and stipends[2].
- CS Faculty Composition and Hiring Trends [source data: Drafty CS Professors [104]].
- Bias in Computer Science Rankings [source data: CS Open Rankings [101]].

Our system continually benefits from unpaid contributors with a strong interest in the dataset because its accurate data benefits our community. We do not require user logins, so anyone can arrive at the system and immediately add or edit data. Our system mirrors many of the positive attributes and levels discussed in this study.

6.2 RQ2) Accurate Contributors are More Motivated by Intrinsic Factors, Inaccurate Contributors are More Motivated by Extrinsic Factors

We find that unpaid contributors are more intrinsically motivated than paid crowdworkers. This may be due to selection bias, as paid crowdworkers could only find the task based on the implied extrinsic motivation of pay to complete the task. Thus, paid crowdworkers are more likely to be motivated by extrinsic factors. However, even within paid crowdworkers, we find that more accurate contributors are intrinsically motivated. This replicates prior research [69, 88] while adding new insights for requesters since our crowdworkers were paid according to the guidelines from Prolific. This may indicate that intrinsic motivation leads to motivation to complete the task well. It may also imply that people more knowledgeable about this kind of task are more likely to complete it correctly and are more intrinsically motivated; thus, they self-selected the task. Regardless, this points to the idea that one way to increase accuracy for paid crowdworkers is to directly appeal to their intrinsic motivators.

6.3 RQ3) Universal Motivators are Pre-Existing Knowledge, Interest in the Topic, and Contributions that Will Benefit Others

Our results provide evidence of attributes that universally motivate extrinsically motivated users (paid crowdworkers) and intrinsically motivated users (unpaid everyday visitors). We find universal attributes describing a crowd contribution that motivate everyone (helping others, contributing your specialized knowledge, the task is interesting, tasks are quick to complete, completing tasks for friends, automatically accepting contributions). Our results build on prior research showing how important intrinsic motivators and altruistic causes are to elicit high-quality contributions [70]. While some results mirror previous research, like friendsourcing, by Brady et al. [10], our quantitative results allow someone to know exactly how much friendsourcing can motivate contributors compared to other methods like Bots on social media advertising tasks to contribute [91]. Compared to prior work, we study all of these motivations simultaneously using a discrete choice experiment [19, 88] and ground these results by observing the real-world editing behaviors of our participants in our system Drafty. We build on prior work and answer calls for future research to mix extrinsically and intrinsically motivated contributors in our real-world system [13, 30, 88, 103]. Our study provides quantitative evidence of what universally motivates contributors.

Building on our findings on what universally motivates contributors regardless of pay level, we developed the following recommendations to help future system designers and requesters develop fair and equitable crowd contribution tasks for paid crowdworkers and unpaid contributors:

(1) Task Perception: Communicate clearly how the task is ethical and how it will be used to help others.
(2) Task Perception and Reason to Complete: Engage potential contributors in online communities with related interests, expertise, or hobbies related to your dataset. Gain feedback

[1]https://github.com/brownhci/drafty
[1]https://jeffhuang.com/computer-science-open-data/

to develop features to help contributors learn new or special skills while contributing.

(3) Pay Level and Task Perception: If payment is not an option, you can rely on people voluntarily contributing or relying on a public system, for example, Wikipedia, WikiData, Drafty, public Google Sheets, or MovieLens.
(4) Pay Level: If payment is an option to recruit contributors, pay at least $8 an hour, although $12 is recommended.
(5) Task Time: Develop features to allow contributors to contribute in 1 minute or less. Tasks should take no longer than around 15 minutes to complete.
(6) Task Time and Effort: When developing features for people to contribute, reducing the time to complete is better than reducing the effort required.
(7) Task Requirements: Recruit users with the interest and expertise who can contribute or use their specialized knowledge to increase the quality of your dataset.
(8) Task Requirements: Do not require people to create user accounts or provide personal information.
(9) What Happens with a Contribution: If possible, contributions should remain anonymous, be automatically accepted, and be stored in a public place where others can see and edit the contributions.
(10) Who is Asking: When asking people to contribute, ensure there is a human representing the request, not a bot or a company. If possible, create an engaged community of contributors and employ social features to create friendships and a sense of community.

Our public data system we developed over 9 years mirrors these recommendations. It has appealed to a community of users interested in Computer Science faculty profiles. Students go to Drafty to find advisors; others use its data to analyze hiring trends in different research areas. Drafty allows anyone to edit anything freely and anonymously, creating an ongoing sense of community and collaboration [3]. We slowly built a public data system over multiple years, creating a community that wants a quality dataset because it can benefit them.

6.3.1 Recommendations for Popular Platforms and Systems to Universally Motivate Contributors. Based on our findings we provide several system recommendations for popular platforms and systems. We preface this by emphasizing to the readers that these are all amazing systems with strong contributions to the scientific community and society.

We propose that Wikipedia and WikiData could support truly anonymous contributions. They should eliminate the need to create an account or publicly post contributors' IP addresses if that contributor does not have an account [3, 37]. In Drafty, we implemented anonymous browser cookies where information was stored securely on the server side, so contributors are only linked to the browser they used. We also support a non-email email user login system, where individuals can create accounts using only a username they choose. We additionally propose that WikiData could simplify its interface to reduce the time and effort to contribute to its open knowledge graph. Currently, its interface hinders users' understanding of the various links and references between data [68]. Recent research has developed interfaces to reduce the effort of Bi-lingual WikiData editors [45]. We recommend developing interfaces to automatically and bi-directionally translate tabular data to and from Wikidata to ease the initial burden of editing while educating contributors on the complex graph references in WikiData [74, 80, 98].

MovieLens is a public data system with a long history of amazing research contributions to personalized recommendations, and the homepage shows how its data is used for others [17]. However, like Wikipedia and WikiData, users must provide personal information to access the data and contribute. Movielens does an excellent job of providing quick, low-effort contribution tasks, like tagging and rating movies. However, Movies can often be a social activity. While Movielens provides a sense of collaboration through public lists and tags, it lacks social features. Based on our results, the developers could add collaboration features that allow friends and family members to contribute. Additionally, Movielens has a cold start problem for providing personalized recommendations [73]. Providing social aspects to motivate contributions might provide a human-centric method to help solve this issue.

The Zooniverse platform has attracted numerous volunteers to contribute to citizen science projects focusing on challenging scientific labeling tasks [94]. It does not require people to create accounts to contribute, thus allowing anonymous contributions to aid datasets for the common good and scientific communities. However, newcomers often lack specialized knowledge and face technical hurdles that increase the time and effort to overcome this knowledge barrier [97]. We recommend presenting simpler tasks with previous answers to give the contributors an initial sense of accomplishment. Thus helping them learn something new with as little effort as possible.

We recommend that paid crowdsourcing platforms like Amazon Mechanical Turk could adopt Prolific's fair payment guidelines. Or, at the minimum, requesters can ensure their tasks, regardless of the invisible labor in crowd work [99], are fairly paid. While these platforms feature messaging systems, they could support more social features to foster long-term relationships between requesters and paid crowdworkers. Previous research shows improved relationships and trusted crowdworkers are highly accurate [8, 78, 103]. Also, these platforms could provide improved filters, focusing on intrinsic motivators or showing the task provides AI assistance to match requesters and paid crowdworkers on tasks. The platforms could allow paid crowdworkers to see a task's estimated time to complete, effort, and pay level per hour update in real-time as others complete the task.

6.4 RQ4) Individuated Suggestions for Paid Crowdworkers and Unpaid Contributors

Every crowd-powered system has different constraints. Some must rely on paid crowdworkers to complete lengthy tasks, and others rely on unpaid contributors. We recommend identifying your system's constraints and building features to increase user motivation to overcome them. However, these constraints make building systems complex [6]. This is because different attributes affect contributor motivation differently (see Figure 10), and tasks must be individuated when possible to maximize user motivation.

Teal is More motivation / preference than paid crowdworkers.

Gold is Less motivation / preference than paid crowdworkers.

Attributes	Unpaid Contributors	Paid Crowdworkers (all & pay level) All	16.0	12.0	8.0
Pay Level	less(↓)	84%	92%	85%	76%
Estimated Time to Complete	less(↓)	17%	21%	15%	13%
Your Perception of the Task	more(↑)	54%	65%	48%	47%
Task Difficulty	less(↓)	5%	11%	2%	4%
Your Reason to Complete a Task	more(↑)	15%	20%	15%	9%
The Task Requires you to	more(↑)	41%	48%	48%	26%
Who Asks you to Complete the Task	more(↑)	47%	47%	64%	39%
What Happens with your Contribution	less(↓)	23%	33%	43%	7%

Unpaid contributors are xx% (more or less) motivated by an attribute than paid crowdworkers?

Figure 10: Preference Differences for Attributes: Unpaid Contributors vs. Paid Crowdworkers. Compared to paid crowdworkers, unpaid contributors are more motivated by their perception of the task, what the task requires them to do, and who asks them to complete the task. Developing systems for unpaid contributors could only focus on features that align with these attributes. This figure shows the percentage difference in attribute importance comparing unpaid contributors to paid crowdworkers.

You can design systems to overcome these constraints by understanding how different attributes and levels for crowd contribution tasks should be individuated for your contributor population (paid crowdworkers or unpaid contributors). Using our utility scores per level (see Figures 5 and 7) to understand what motivates your targeted contributor population, you can optimize your tasks to increase user motivation. Below, we describe two hypothetical task design scenarios for paid crowdworkers and unpaid contributors.

We will develop features and change our task design using the utility scores from our discrete choice experiment as a guideline. Thus, we show a practical approach for using our results that others can replicate when designing their systems and tasks. The lists of attributes, levels, and utility scores below are formatted as **attribute**, level (utility score) from the column "Unpaid Contributors" or "Paid Crowdworkers" in Figure 5 or Figure 7). We identify levels relating to a task's *[constraint]*'s and levels that should maximize motivation for each group of contributors.

6.4.1 Paid Crowdworkers: Limited Funds and Lengthy Tasks. Imagine you want to create a new tabular dataset of STEM faculty profiles from the top 500 universities in the world. You lack the time and expertise to create a full system, so you use Google Sheets to recruit paid crowdworkers to collect the dataset [78]. You remove access to Google Sheets after data collection to protect your data. Prior research shows that paid crowdworkers editing tabular and structured data can be innacurate [103, 112]. So, to motivate the crowdworkers, you offer small bonuses if they do exceptional work [110]. However, because of this decision and your limited budget, you can only pay the crowdworkers $8 per hour. When you pilot the task, it takes around 30 minutes to complete because you have lengthy instructions and a survey collecting demographics, and it can be complex to understand STEM faculty's research areas across multiple disciplines [103]. Worst of all, you realize it may be mundane to find and read faculty websites repeatedly. Your task for paid crowdworkers paid $8 per hour could be represented as:

(1) **Pay Level:** $8 per hour (1.7) *[constraint]*
(2) **Estimated Time to Complete:** 30 minutes (-3.1) *[constraint]*
(3) **Your Perception of the Task:** The task looks boring (0.1) *[constraint]*
(4) **Who Asks you to Complete the Task:** Paid Crowdsourcing system (i.e., Prolific) (-0.2) *[constraint]*
(5) **Your Reason to Complete a Task:** You might be paid for doing exceptional work (-2.4)
(6) **What Happens with your Contribution:** You do not own it (you cannot see or edit it) (-3.5) *[constraint]*
(7) **Task Difficulty:** Moderately difficult to complete (0.0)
(8) **The Task Requires you to:** Provide your personal information (-4.3)

The sum of the utility scores from above is -11.7. You fear it might be challenging to motivate and recruit accurate paid crowdworkers. To improve your task design in Prolific we can modify parts of the and develop new systems to improve our constraints. 1) First, we remove the survey so they do not have to provide personal information. 2) We remove the bonus structure and use the extra money to increase our pay level to $12 per hour. 3) We use Google Apps Script to add a feature to automatically search the web and use AI to look up potential professors from universities preemptively. This makes the task less boring, reduces instructions, and helps people complete it in 15 minutes [99]. 4) We designed a new feature using Google Apps Script when someone finishes adding data, the feature automatically computes insights using the dataset and the paid crowd workers contributions so they can learn something new [21, 82, 93]. 5) The last change is we decide to inform the paid crowdworkers we are leaving the Google Sheets open and they get a copy of their responses. How would our task now look to a paid crowdworker paid $12 an hour instead of $8?

(1) **Pay Level:** $12 per hour (8.3) *[constraint]*
(2) **Estimated Time to Complete:** 15 minutes (2.7) *[constraint]*
(3) **Your Perception of the Task:** Your contribution might help people you do not know (5.5)
(4) **Who Asks you to Complete the Task:** Paid Crowdsourcing system (i.e., Prolific) (-0.2) *[constraint]*
(5) **Your Reason to Complete a Task:** Your contribution benefits you personally (0.8)
(6) **What Happens with your Contribution:** You own it (you can see and edit it) (1.3) *[constraint]*
(7) **Task Difficulty:** Moderately difficult to complete (0.3)
(8) **The Task Requires you to:** Complete the task with Artificial Intelligence (0.6)

We have improved the sum of the utility scores to 19.3 by creating two new features with Google Apps Script to piggyback off Google Sheets [28] and making other simple changes that have a positive effect specifically for crowdworkers paid $12 per hour. We primarily modified the attributes that mattered the most to paid crowdworkers paid $12 per hour, improving our task's extrinsic and intrinsic motivations based on the attributes from Figure 10.

6.4.2 Unpaid Contributors: Challenging Tasks for Unfamiliar People. Imagine you are leading a new citizen science effort to attract unpaid contributors on Zooniverse to label challenging scientific data [94]. For example, unpaid contributors must learn to identify exoplanets (to find potentially habitable planets) from satellite images [15]. It takes 60 minutes to learn a new skill and label the images. And once they have submitted a contribution, it typically cannot be directly edited. However, their efforts will contribute to data that could benefit their friends, family, and humanity one day. Your task for unpaid contributors could be represented as:

(1) **Pay Level:** $0 per hour (-10.7) *[constraint]*
(2) **Estimated Time to Complete:** 60 minutes (-9.1) *[constraint]*
(3) **Your Perception of the Task:** Your contribution might help your peers or community (8.0)
(4) **Who Asks you to Complete the Task:** Volunteer (1.9)
(5) **Your Reason to Complete a Task:** You will learn a new or special skill (3.7)
(6) **What Happens with your Contribution:** You do not own it (you cannot see or edit it) (-4.1) *[constraint]*
(7) **Task Difficulty:** Very difficult to complete (-3.1) *[constraint]*
(8) **The Task Requires you to:** Learn something new (1.1)

The sum of the utility scores from above is -12.3. Thus, it is unlikely that unpaid contributors will be motivated to complete this task. To improve your task design in Zooniverse we can develop two features to increase user motivation for our task. 1) We reduce task time to 30 minutes by creating a new personalized learning sequence to help train new users by comparing their answers with expert annotators building on ideas from LabInTheWild [87]. 2) We create a subreddit where unpaid contributors can make friends with other contributors and collaborate with them [107]. Unlike prior research on Reddit [9], our results show that unpaid contributors are highly motivated by collaborating with others and helping friends and family. What would be some new attributes we could emphasize based on these features?

(1) **Estimated Time to Complete:** 30 minutes (-2.0) *[constraint]*
(2) **Your Perception of the Task:** Your contribution might help your peers or community (8.0)
(3) **Who Asks you to Complete the Task:** A friend (4.7)
(4) **The Task Requires you to:** Collaborate with other people to complete the task (3.2)

We have improved the sum of the utility scores to 10.7, primarily building features to reduce task time and then creating social collaborations to increase unpaid contributors' intrinsic motivations based on the attributes from Figure 10.

7 Limitations

This research focuses on combining ideas from prior research, our efforts in developing our public system over 9 years, and a pilot study to select attributes and their associated levels to run the survey. To the best of our ability, these attributes and levels relate to a contributor's preference to select different crowd contribution tasks. While the study was conducted using our system, we systematically tried to select attributes that described typical crowd contribution tasks for paid crowdworkers and unpaid contributors alike. While this is the first attempt to combine this type of discrete choice experiment with human-computer interaction, we acknowledge that this is limited to a particular dataset on one topic: CS professors. While we have attracted hundreds of thousands of visitors, we do not have the same history as similar research efforts on datasets with a singular focus, such as Movielens [112]. We hope that future researchers will build and improve upon our methods with different communities and datasets.

Vandalism of data is a threat to the data quality for many public systems where people freely make contributions, ranging from Reddit [67] to Wikipedia and WikiData [36]. Our results for paid crowdworkers show multiple cases of people intentionally submitting bad data. For example, a submitted professor who is not a real person or whose profile was missing half its data. Our results for unpaid contributors showed no evidence of vandalism or laziness. The most common contribution mistake was not including where a professor received their Bachelor's degree. It is possible that Drafty's everyday unpaid contributors do not vandalize the dataset because of its value to their community. Because the data of Drafty is valuable specifically to the computer science community, the contributors in this study were likely very skewed towards users with interest and experience as computer scientists, who are also likely to be more technologically proficient than the average person. This may have skewed our results, so we suggest that future work look at other populations.

Future work should explore additional datasets and systems in combination with attributes and levels as contributors' motivations evolve. For example, future work could replicate our study design to answer the call to study the differences in motivation between novice and expert paid crowdworkers [35]. Likewise, every person who took our survey had some interest in Computer Science professors as they either freely visited Drafty or accepted a paid task on Prolific. This study design decision was intentional so their stated preferences from the survey could be compared with their real-world behavior within Drafty. Future work should study how the motivations might differ based on different datasets and even different datasets in the same system.

8 Conclusion

Our discrete choice experiment survey study features real-world users of our public system Drafty. We simultaneously study the preferences, motivations, and editing behaviors of Drafty's normal everyday unpaid contributors with paid crowdworkers we recruited from Prolific. We answer the call of prior research to provide external validity to the results of our discrete choice experiment [83], by combining participant's preferences in crowd contribution tasks from our discrete choice experiment with their real-world editing behaviors (accuracy per edit) from Drafty. By simultaneously studying the editing behaviors and motivations of unpaid contributors with paid crowdworkers, we have answered multiple calls for future research to study extrinsic and intrinsic motivators with real-world behaviors in-the-wild [13, 30, 88, 103]. While our study focuses on one specific public tabular dataset, Computer Science faculty, we developed our discrete choice experiment to provide insights that could influence and help other crowdsourcing and peer production researchers to build systems and tasks to contribute.

Our results show that pay level per task, estimated time to completion, someone's level of interest, and their contributions helping others are the strongest motivators for contributing. However, paid and unpaid contributors make different trade-offs when considering these motivators. Paid crowdworkers prioritize pay level and time to complete, while unpaid contributors are willing to complete longer intrinsically motivated tasks that pay less. Notably, participants who only made accurate edits within Drafty also preferred to take less pay for tasks that aligned with their interests and allowed them to contribute their specialized knowledge to help others. This aligns with prior paid crowdsourcing research [19] but expands this finding across the everyday users (non-crowdworkers) of our public system. The utility scores from our study can be used by people designing crowd contribution systems for paid crowdworkers and unpaid contributors. Hopefully, this moves us towards a fairer and more equitable future when designing crowd contribution systems, where incentives to contribute are balanced with user preferences.

The results of this discrete choice experiment are by no means a stopping point. This study could be conducted with users of other systems or recruited from popular platforms such as Wikipedia. Do those users have the same universal motivators as Drafty's users? Or do they make the same type of trade-offs when selecting tasks? These limitations are prime for future study as what motivates us evolves. Our pilot (2022) and main study (2023) reveal upward trends from paid and unpaid users who are more open to using AI within their crowd contribution tasks. Deploying discrete choice experiments within the contribution loop could help us develop future systems that integrate AI and other new features when needed and when the users feel comfortable using them.

Acknowledgments

We thank Michael Bernstein, Zoya Bylinskii, John F Hughes, and Malte Schwarzkopf for their timely feedback and intellectual support. This work is supported by the NASA EPSCoR RID Award #80NSSC22M0040. Finally, we thank the thousands of Drafty and Computer Science Open Rankings users for helping to contribute and maintain the tabular dataset of Computer Science Faculty profiles for the past 10 years.

References

[1] Bryan Alexander. 2020. The little spreadsheet that could, and did: crowdsourcing COVID-19, higher education, data, and stories. https://bryanalexander.org/research-topics/the-little-spreadsheet-that-could-and-did-crowdsourcing-covid-19-higher-education-data-and-stories/. (Accessed on 04/12/2021).
[2] Ali Alkhatib, Michael S Bernstein, and Margaret Levi. 2017. Examining crowd work and gig work through the historical lens of piecework. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. ACM, New York, NY, USA, 4599–4616.
[3] Martina Balestra, Coye Cheshire, Ofer Arazy, and Oded Nov. 2017. Investigating the motivational paths of peer production newcomers. In *Proceedings of the 2017 chi conference on human factors in computing systems*. 6381–6385.
[4] Yochai Benkler. 2016. Peer production and cooperation. In *Handbook on the Economics of the Internet*. Edward Elgar Publishing, Cheltenham, United Kingdom.
[5] Michael Bernstein, Mike Bright, Ed Cutrell, Steven Dow, Elizabeth Gerber, Anupam Jain, and Anand Kulkarni. 2013. Micro-volunteering: helping the helpers in development. In *Proceedings of the 2013 conference on Computer supported cooperative work companion*. ACM, New York, NY, USA, 85–88.
[6] Michael S Bernstein. 2013. Crowd-powered systems. *KI-Künstliche Intelligenz* 27, 1 (2013), 69–73.
[7] Michael S Bernstein, Greg Little, Robert C Miller, Björn Hartmann, Mark S Ackerman, David R Karger, David Crowell, and Katrina Panovich. 2010. Soylent: a word processor with a crowd inside. In *UIST '10 Proceedings of the 23rd annual ACM symposium on User interface Software and Technology*. ACM Press, ACM, New York, NY, USA, 313–322.
[8] Michael S Bernstein, Greg Little, Robert C Miller, Björn Hartmann, Mark S Ackerman, David R Karger, David Crowell, and Katrina Panovich. 2015. Soylent: a word processor with a crowd inside. *Commun. ACM* 58, 8 (2015), 85–94.
[9] Toine Bogers and Rasmus Nordenhoff Wernersen. 2014. How'Social'are Social News Sites? Exploring the Motivations for Using Reddit. com. *iConference 2014: Breaking Down Walls: Culture-Context-Computing* (2014), 329–344.
[10] Erin Brady, Meredith Ringel Morris, and Jeffrey P Bigham. 2014. Social microvolunteering: Donating access to your friends for charitable microwork. In *Second AAAI Conference on Human Computation and Crowdsourcing*. AAAI, Menlo Park, CA, USA, 8–9.
[11] Erin Brady, Meredith Ringel Morris, and Jeffrey P Bigham. 2015. Gauging receptiveness to social microvolunteering. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1055–1064.
[12] Robin Brewer, Meredith Ringel Morris, and Anne Marie Piper. 2016. " Why would anybody do this?" Understanding Older Adults' Motivations and Challenges in Crowd Work. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. ACM, New York, NY, USA, 2246–2257.
[13] Gordon Burtch, Yili Hong, Ravi Bapna, and Vladas Griskevicius. 2018. Stimulating online reviews by combining financial incentives and social norms. *Management Science* 64, 5 (2018), 2065–2082.
[14] Guillermo F Cabrera, Christopher J Miller, and Jeff Schneider. 2014. Systematic labeling bias: De-biasing where everyone is wrong. In *Proc. ICPR*. IEEE, Piscataway, NJ, USA, 4417–4422.
[15] Luca Cacciapuoti, Veselin B Kostov, Marc Kuchner, Elisa V Quintana, Knicole D Colón, Jonathan Brande, Susan E Mullally, Quadry Chance, Jessie L Christiansen, John P Ahlers, et al. 2022. The TESS Triple-9 Catalog: 999 uniformly vetted exoplanet candidates. *Monthly Notices of the Royal Astronomical Society* 513, 1 (2022), 102–116.
[16] Shuo Chang, F Harper, Lingfei He, and Loren Terveen. 2016. CrowdLens: experimenting with crowd-powered recommendation and explanation. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 10. 52–61.
[17] Yan Chen, F Maxwell Harper, Joseph Konstan, and Sherry Xin Li. 2010. Social comparisons and contributions to online communities: A field experiment on movielens. *American Economic Review* 100, 4 (2010), 1358–1398.
[18] Justin Cheng, Jaime Teevan, and Michael S Bernstein. 2015. Measuring crowdsourcing effort with error-time curves. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1365–1374.

[19] Aaron Clauset, Samuel Arbesman, and Daniel B Larremore. 2015. Systematic inequality and hierarchy in faculty hiring networks. *Science Advances* 1, 1 (2015), e1400005.

[20] Joanna Coast and Sue Horrocks. 2007. Developing attributes and levels for discrete choice experiments using qualitative methods. *Journal of health services research & policy* 12, 1 (2007), 25–30.

[21] Dan Cosley, Dan Frankowski, Loren Terveen, and John Riedl. 2007. SuggestBot: Using Intelligent Task Routing to Help People Find Work in Wikipedia. In *Proceedings of the 12th International Conference on Intelligent User Interfaces* (Honolulu, Hawaii, USA) *(IUI '07)*. ACM, New York, NY, USA, 32–41. doi:10.1145/1216295.1216309

[22] Paramita Das, Bhanu Prakash Reddy Guda, Debajit Chakraborty, Soumya Sarkar, and Animesh Mukherjee. 2021. When expertise gone missing: Uncovering the loss of prolific contributors in Wikipedia. *arXiv preprint arXiv:2109.09979* (2021).

[23] Victor De Boer, Michiel Hildebrand, Lora Aroyo, Pieter De Leenheer, Chris Dijkshoorn, Binyam Tesfa, and Guus Schreiber. 2012. Nichesourcing: harnessing the power of crowds of experts. In *International Conference on Knowledge Engineering and Knowledge Management*. Springer, New York, NY, USA, 16–20.

[24] Esra Cemre Su de Groot and Ujwal Gadiraju. 2024. " Are we all in the same boat?" Customizable and Evolving Avatars to Improve Worker Engagement and Foster a Sense of Community in Online Crowd Work. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–26.

[25] Dong Deng, Raul Castro Fernandez, Ziawasch Abedjan, Sibo Wang, Michael Stonebraker, Ahmed K Elmagarmid, Ihab F Ilyas, Samuel Madden, Mourad Ouzzani, and Nan Tang. 2017. The Data Civilizer System.. In *Cidr*.

[26] Greg d'Eon, Joslin Goh, Kate Larson, and Edith Law. 2019. Paying crowd workers for collaborative work. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–24.

[27] Anhai Doan, Raghu Ramakrishnan, and Alon Y Halevy. 2011. Crowdsourcing systems on the world-wide web. *Commun. ACM* 54, 4 (2011), 86–96.

[28] Daniel A Epstein, Fannie Liu, Andrés Monroy-Hernández, and Dennis Wang. 2022. Revisiting piggyback prototyping: Examining benefits and tradeoffs in extending existing social computing systems. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–28.

[29] Shaoyang Fan, Ujwal Gadiraju, Alessandro Checco, and Gianluca Demartini. 2020. Crowdco-op: Sharing risks and rewards in crowdsourcing. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–24.

[30] Claudia Flores-Saviaga, Ricardo Granados, Liliana Savage, Lizbeth Escobedo, and Saiph Savage. 2020. Understanding the complementary nature of paid and volunteer crowds for content creation. *Avances en Interacción Humano-Computadora* 1, 1 (2020), 37–44.

[31] Ujwal Gadiraju, Ricardo Kawase, Stefan Dietze, and Gianluca Demartini. 2015. Understanding Malicious Behavior in Crowdsourcing Platforms: The Case of Online Surveys. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 1631–1640. doi:10.1145/2702123.2702443

[32] Rishab A Ghosh, Ruediger Glott, Bernhard Krieger, and Gregorio Robles. 2002. Free/libre and open source software: Survey and study.

[33] Ilka H Gleibs. 2017. Are all "research fields" equal? Rethinking practice for the use of data from crowdsourcing market places. *Behavior Research Methods* 49, 4 (2017), 1333–1342.

[34] Daniel Haas, Jason Ansel, Lydia Gu, and Adam Marcus. 2015. Argonaut: Macrotask Crowdsourcing for Complex Data Processing. *Proceedings of the VLDB Endowment* 8, 12 (Aug. 2015), 1642–1653. doi:10.14778/2824032.2824062

[35] Benjamin V Hanrahan, Anita Chen, JiaHua Ma, Ning F Ma, Anna Squicciarini, and Saiph Savage. 2021. The expertise involved in deciding which HITs are worth doing on Amazon Mechanical Turk. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–23.

[36] Stefan Heindorf, Martin Potthast, Benno Stein, and Gregor Engels. 2016. Vandalism detection in wikidata. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. ACM, New York, NY, USA, 327–336.

[37] Benjamin Mako Hill and Aaron Shaw. 2020. Wikipedia and the End of Open Collaboration. *Wikipedia* 20 (2020).

[38] Chien-Ju Ho, Aleksandrs Slivkins, Siddharth Suri, and Jennifer Wortman Vaughan. 2015. Incentivizing high quality crowdwork. In *Proceedings of the 24th International Conference on World Wide Web*. ACM, New York, NY, USA, 419–429.

[39] John Joseph Horton and Lydia B Chilton. 2010. The labor economics of paid crowdsourcing. In *Proceedings of the 11th ACM Conference on Electronic Commerce*. ACM, New York, NY, USA, 209–218.

[40] Shih-Wen Huang and Wai-Tat Fu. 2013. 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*. ACM, New York, NY, USA, 621–630.

[41] Christoph Hube, Besnik Fetahu, and Ujwal Gadiraju. 2019. Understanding and mitigating worker biases in the crowdsourced collection of subjective judgments. In *Proc. CHI*. ACM, New York, NY, USA, 1–12.

[42] Hasan Humayun, Masitah Ghazali, and Mohammad Noman Malik. 2023. Is sustained participation a myth in crowdsourcing? A review. *European Journal of Innovation Management* (2023).

[43] Hasan Humayun, Mohammad Nauman Malik, and Masitah Ghazali. 2024. Analysis of Motivational Theories in Crowdsourcing Using Long Tail Theory: A Systematic Literature Review. *International Journal of Crowd Science* 8, 1 (2024), 10–27.

[44] Kazushi Ikeda and Michael S Bernstein. 2016. Pay it backward: Per-task payments on crowdsourcing platforms reduce productivity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 4111–4121.

[45] Lucie-Aimée Kaffee, Kemele M Endris, and Elena Simperl. 2019. When humans and machines collaborate: cross-lingual label editing in wikidata. In *Proceedings of the 15th International Symposium on Open Collaboration*. 1–9.

[46] Keiko Katsuragawa, Qi Shu, and Edward Lank. 2019. PledgeWork: Online volunteering through crowdwork. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–11.

[47] Warut Khern-am nuai, Karthik Kannan, and Hossein Ghasemkhani. 2018. Extrinsic versus intrinsic rewards for contributing reviews in an online platform. *Information Systems Research* 29, 4 (2018), 871–892.

[48] Johannes Kiesel, Martin Potthast, Matthias Hagen, and Benno Stein. 2017. Spatiotemporal analysis of reverted wikipedia edits. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 11. AAAI, Menlo Park, CA, USA, 122–131.

[49] Juho Kim et al. 2015. *Learnersourcing: improving learning with collective learner activity*. Ph. D. Dissertation. Massachusetts Institute of Technology.

[50] Aniket Kittur, Ed H. Chi, and Bongwon Suh. 2008. Crowdsourcing User Studies with Mechanical Turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. ACM, New York, NY, USA, 453–456. doi:10.1145/1357054.1357127

[51] Aniket Kittur and Robert E Kraut. 2008. Harnessing the wisdom of crowds in wikipedia: quality through coordination. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. ACM, 37–46.

[52] Aniket Kittur, Jeffrey V Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. 2013. The future of crowd work. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*. ACM, New York, NY, USA, 1301–1318.

[53] W Kuhfeld. 2001. Multinomial logit, discrete choice modeling. *An introduction to designing choice experiments, and collecting, processing and analyzing choice data with SAS. SAS Institute TS-643* (2001).

[54] Natalie Kupferberg and Bridget McCrate Protus. 2011. Accuracy and completeness of drug information in Wikipedia: an assessment. *Journal of the Medical Library Association: JMLA* 99, 4 (2011), 310.

[55] Marjan Laal and Seyed Mohammad Ghodsi. 2012. Benefits of collaborative learning. *Procedia-social and behavioral sciences* 31 (2012), 486–490.

[56] Tracy Xiao Liu, Jiang Yang, Lada A Adamic, and Yan Chen. 2014. Crowdsourcing with all-pay auctions: A field experiment on taskcn. *Management Science* 60, 8 (2014), 2020–2037.

[57] Jordan J Louviere and Emily Lancsar. 2009. Choice experiments in health: the good, the bad, the ugly and toward a brighter future. *Health Economics, Policy and Law* 4, 4 (2009), 527–546.

[58] Eddy Maddalena, Luis-Daniel Ibáñez, Neal Reeves, and Elena Simperl. 2023. Qrowdsmith: Enhancing paid microtask crowdsourcing with gamification and furtherance incentives. *ACM Transactions on Intelligent Systems and Technology* 14, 5 (2023), 1–26.

[59] Lindsay J Mangham, Kara Hanson, and Barbara McPake. 2009. How to do (or not to do)… Designing a discrete choice experiment for application in a low-income country. *Health policy and planning* 24, 2 (2009), 151–158.

[60] Andrew Mao, Ece Kamar, Yiling Chen, Eric Horvitz, Megan E Schwamb, Chris J Lintott, and Arfon M Smith. 2013. Volunteering versus work for pay: Incentives and tradeoffs in crowdsourcing. In *First AAAI conference on human computation and crowdsourcing*. AAAI, Menlo Park, CA, USA, 94–102.

[61] Adam Marcus. 2013. How I Learned to Stop Worrying and Love the Crowd. https://www.youtube.com/watch?v=FL9Q43zO1BQ (Online; accessed on 2020-04-20).

[62] Catherine C Marshall, Partha SR Goguladinne, Mudit Maheshwari, Apoorva Sathe, and Frank M Shipman. 2023. Who broke Amazon Mechanical Turk? An analysis of crowdsourcing data quality over time. In *Proceedings of the 15th ACM Web Science Conference 2023*. 335–345.

[63] Winter Mason and Duncan J. Watts. 2009. Financial Incentives and the "Performance of Crowds". *ACM SigKDD Explorations Newsletter* 11, 2 (2009), 77–85. doi:10.1145/1600150.1600175

[64] Arpit Merchant, Daksh Shah, Gurpreet Singh Bhatia, Anurag Ghosh, and Ponnurangam Kumaraguru. 2019. Signals matter: understanding popularity and impact of users on stack overflow. In *The World Wide Web Conference*. ACM, New York, NY, USA, 3086–3092.

[65] Marc Miquel-Ribé, Cristian Consonni, and David Laniado. 2021. Wikipedia Editor Drop-Off. *Wiki Workshop at The Web Conference 2021* (2021).

[66] Laurent Moccozet, Camille Tardy, Wanda Opprecht, and Michel Léonard. 2013. Gamification-based assessment of group work. In *2013 International Conference on Interactive Collaborative Learning (ICL)*. IEEE, 171–179.
[67] Carrie Moore and Lisa Chuang. 2017. Redditors revealed: Motivational factors of the Reddit community. *Proceedings of the 50th Hawaii International Conference on System Sciences* (2017), 1–10.
[68] Marçal Mora-Cantallops, Salvador Sánchez-Alonso, and Elena García-Barriocanal. 2019. A systematic literature review on Wikidata. *Data Technologies and Applications* 53, 3 (2019), 250–268.
[69] Meredith Ringel Morris, Jeffrey P. Bigham, Robin Brewer, Jonathan Bragg, Anand Kulkarni, Jessie Li, and Saiph Savage. 2017. Subcontracting Microwork. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. ACM, New York, NY, USA, 1867–1876. doi:10.1145/3025453.3025687
[70] Meredith Ringel Morris, Jaime Teevan, and Katrina Panovich. 2010. What Do People Ask Their Social Networks, and Why?: A Survey Study of Status Message Q&A Behavior. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. ACM, New York, NY, USA, 1739–1748. doi:10.1145/1753326.1753587
[71] Babak Naderi. 2018. *Motivation of workers on microtask crowdsourcing platforms.* Springer, New York, NY, USA.
[72] Robbie T Nakatsu, Elissa B Grossman, and Charalambos L Iacovou. 2014. A taxonomy of crowdsourcing based on task complexity. *Journal of Information Science* 40, 6 (2014), 823–834.
[73] Senthilselvan Natarajan, Subramaniyaswamy Vairavasundaram, Sivaramakrishnan Natarajan, and Amir H Gandomi. 2020. Resolving data sparsity and cold start problem in collaborative filtering recommender system using linked open data. *Expert Systems with Applications* 149 (2020), 113248.
[74] Phuc Nguyen, Ikuya Yamada, Natthawut Kertkeidkachorn, Ryutaro Ichise, and Hideaki Takeda. 2020. MTab4Wikidata at SemTab 2020: Tabular Data Annotation with Wikidata.. In *SemTab@ ISWC*. ACM, New York, NY, USA, 86–95.
[75] DOE Office of Science. 2022. Argonne Computer Scientist Works on Power and Processor Efficiency While Diversifying Workforce. https://www.hpcwire.com/off-the-wire/argonne-computer-scientist-works-on-power-and-processor-efficiency-while-diversifying-workforce/. (Accessed on 11/01/2022).
[76] Jasper Oosterman, Archana Nottamkandath, Chris Dijkshoorn, Alessandro Bozzon, Geert-Jan Houben, and Lora Aroyo. 2014. Crowdsourcing Knowledge-intensive Tasks in Cultural Heritage. In *Proceedings of the 2014 ACM Conference on Web Science* (Bloomington, Indiana, USA) *(WebSci '14)*. ACM, New York, NY, USA, 267–268. doi:10.1145/2615569.2615644
[77] Bryan Orme. 2002. Interpreting conjoint analysis data. *Sawtooth Software Research Paper Series* (2002).
[78] Alexandra Papoutsaki, Hua Guo, Danae Metaxa-Kakavouli, Connor Gramazio, Jeff Rasley, Wenting Xie, Guan Wang, and Jeff Huang. 2015. Crowdsourcing from Scratch: A Pragmatic Experiment in Data Collection by Novice Requesters. In *Proceedings of the Third AAAI Conference on Human Computation and Crowdsourcing (HCOMP)*. AAAI, Menlo Park, CA, USA, 140–149.
[79] Dennis Pilz and Heiko Gewald. 2013. Does Money Matter? Motivational Factors for Participation in Paid- and Non-Profit-Crowdsourcing Communities. In *Wirtschaftsinformatik*. https://api.semanticscholar.org/CorpusID:7858381
[80] Wikimedia Project. 2021. Wikipedia and Wikidata Tools - Meta. https://meta.wikimedia.org/wiki/Wikipedia_and_Wikidata_Tools. (Accessed on 07/10/2021).
[81] Wikimedia Project. 2022. Why is Wikipedia losing contributors - thinking about remedies. https://en.wikipedia.org/wiki/Wikipedia:Why_is_Wikipedia_losing_contributors_-_Thinking_about_remedies. (Accessed on 06/02/2022).
[82] Sihang Qiu, Ujwal Gadiraju, and Alessandro Bozzon. 2020. Improving worker engagement through conversational microtask crowdsourcing. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–12.
[83] Matthew Quaife, Fern Terris-Prestholt, Gian Luca Di Tanna, and Peter Vickerman. 2018. How well do discrete choice experiments predict health choices? A systematic review and meta-analysis of external validity. *The European journal of health economics* 19 (2018), 1053–1066.
[84] Alexander J. Quinn and Benjamin B. Bederson. 2011. Human Computation: A Survey and Taxonomy of a Growing Field. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 1403–1412. doi:10.1145/1978942.1979148
[85] Sam Ransbotham and Gerald C Kane. 2011. Membership turnover and collaboration success in online communities: Explaining rises and falls from grace in Wikipedia. *Mis Quarterly* 35 (2011), 613–627.
[86] Vithala R Rao. 2014. *Applied conjoint analysis.* Springer Science & Business Media.
[87] Katharina Reinecke and Krzysztof Z Gajos. 2015. LabintheWild: Conducting large-scale online experiments with uncompensated samples. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*. ACM, New York, NY, USA, 1364–1378.
[88] Jakob Rogstadius, Vassilis Kostakos, Aniket Kittur, Boris Smus, Jim Laredo, and Maja Vukovic. 2011. An assessment of intrinsic and extrinsic motivation on task performance in crowdsourcing markets. In *Fifth International AAAI Conference on Weblogs and Social Media.* AAAI, Menlo Park, CA, USA, 321–328.
[89] Niloufar Salehi, Lilly C Irani, Michael S Bernstein, Ali Alkhatib, Eva Ogbe, Kristy Milland, et al. 2015. We are dynamo: Overcoming stalling and friction in collective action for crowd workers. In *Proceedings of the 33rd annual ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1621–1630.
[90] Saiph Savage, Chun Wei Chiang, Susumu Saito, Carlos Toxtli, and Jeffrey Bigham. 2020. Becoming the super turker: Increasing wages via a strategy from high earning workers. In *Proceedings of The Web Conference 2020*. 1241–1252.
[91] Saiph Savage, Andres Monroy-Hernandez, and Tobias Höllerer. 2016. Botivist: Calling volunteers to action using online bots. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, New York, NY, USA, 813–822.
[92] Linda See, Alexis Comber, Carl Salk, Steffen Fritz, Marijn van der Velde, Christoph Perger, Christian Schill, Ian McCallum, Florian Kraxner, and Michael Obersteiner. 2013. Comparing the quality of crowdsourced data contributed by expert and non-experts. *PloS one* 8, 7 (2013), e69958.
[93] Danqing Shi, Xinyue Xu, Fuling Sun, Yang Shi, and Nan Cao. 2020. Calliope: Automatic visual data story generation from a spreadsheet. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2020), 453–463.
[94] Robert Simpson, Kevin R Page, and David De Roure. 2014. Zooniverse: observing the world's largest citizen science platform. In *Proceedings of the 23rd international conference on world wide web*. 1049–1054.
[95] Anjali Singh, Christopher Brooks, and Shayan Doroudi. 2022. Learnersourcing in Theory and Practice: Synthesizing the Literature and Charting the Future. In *Proceedings of the Ninth ACM Conference on Learning@ Scale*. ACM, New York, NY, USA, 234–245.
[96] C Estelle Smith, Bowen Yu, Anjali Srivastava, Aaron Halfaker, Loren Terveen, and Haiyi Zhu. 2020. Keeping Community in the Loop: Understanding Wikipedia Stakeholder Values for Machine Learning-Based Systems. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–14.
[97] Igor Steinmacher, Marco Aurélio Graciotto Silva, and Marco Aurélio Gerosa. 2014. Barriers faced by newcomers to open source projects: a systematic review. In *Open Source Software: Mobile Open Source Technologies: 10th IFIP WG 2.13 International Conference on Open Source Systems, OSS 2014, San José, Costa Rica, May 6-9, 2014. Proceedings 10*. Springer, 153–163.
[98] Pedro A Szekely, Daniel Garijo, Jay Pujara, Divij Bhatia, and Jiasheng Wu. 2019. T2WML: A Cell-Based Language to Map Tables into Wikidata Records.. In *ISWC Satellites*. ACM, New York, NY, USA, 45–48.
[99] Carlos Toxtli, Siddharth Suri, and Saiph Savage. 2021. Quantifying the invisible labor in crowd work. *Proceedings of the ACM on human-computer interaction* 5, CSCW2 (2021), 1–26.
[100] Shaun Wallace. 2022. Asynchronous Collaboration Systems for Evolving Information. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. ACM New York, NY, USA, 1–5.
[101] Shaun Wallace, Long Do, Alice Marbach, and Jeff Huang. 2020. *CS Open Rankings*. Retrieved January 21, 2025 from https://drafty.cs.brown.edu/csopenrankings/
[102] Shaun Wallace, Brendan Le, Luis A Leiva, Aman Haq, Ari Kintisch, Gabrielle Bufrem, Linda Chang, and Jeff Huang. 2020. Sketchy: Drawing inspiration from the crowd. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–27.
[103] Shaun Wallace, Alexandra Papoutsaki, Neilly H Tan, Hua Guo, and Jeff Huang. 2021. Case Studies on the Motivation and Performance of Contributors Who Verify and Maintain In-Flux Tabular Datasets. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–25.
[104] Shaun Wallace, Lucy Van Kleunen, Marianne Aubin-Le Quere, Abraham Peterkin, Yirui Huang, and Jeff Huang. 2017. Drafty: Enlisting Users to be Editors who Maintain Structured Data. In *Proceedings of the Fifth AAAI Conference on Human Computation and Crowdsourcing (HCOMP)*. AAAI, Menlo Park, CA, USA, 187–196.
[105] Jing Wang, Gen Li, and Kai-Lung Hui. 2018. Do Monetary Incentives Create a Spillover Effect on Free Knowledge Contribution? Evidence from a Natural Experiment. *Evidence from a Natural Experiment (June 25, 2018)* (2018), 1–24.
[106] Sylvain Weber. 2021. A step-by-step procedure to implement discrete choice experiments in Qualtrics. *Social Science Computer Review* 39, 5 (2021), 903–921.
[107] Galen Weld, Amy X Zhang, and Tim Althoff. 2024. Making online communities 'better': a taxonomy of community values on reddit. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 18. 1611–1633.
[108] Dennis M Wilkinson and Bernardo A Huberman. 2007. Assessing the value of coooperation in wikipedia. *arXiv preprint cs/0702140* (2007), 1–14.
[109] Chengyuan Xu, Kuo-Chin Lien, and Tobias Höllerer. 2023. Comparing Zealous and Restrained AI Recommendations in a Real-World Human-AI Collaboration

Task. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–15.
[110] Ming Yin and Yiling Chen. 2015. Bonus or not? learn to reward in crowdsourcing.. In *IJCAI*. 201–208.
[111] Xuefeng Zhang, Enjun Xia, Chao Shen, and Jiafu Su. 2022. Factors Influencing Solvers' Behaviors in Knowledge-Intensive Crowdsourcing: A Systematic Literature Review. *Journal of Theoretical and Applied Electronic Commerce Research* 17, 4 (2022), 1297–1319.
[112] Qian Zhao, Zihong Huang, F Maxwell Harper, Loren Terveen, and Joseph A Konstan. 2016. Precision crowdsourcing: closing the loop to turn information consumers into information contributors. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, New York, NY, USA, 1615–1625.
[113] Yudian Zheng, Guoliang Li, Yuanbing Li, Caihua Shan, and Reynold Cheng. 2017. Truth inference in crowdsourcing: Is the problem solved? *Proc. VLDB Endowment* 10, 5 (2017), 541–552.

A Qualtrics Method for Balancing Choice Sets

The algorithm generates random bundles for each attribute and level, then checks each choice set to ensure relative balance in the number of times someone sees each level. Qualtrics' algorithm does not force each level to appear the same number of times per participant. Qualtrics uses a randomized, balanced design approach to ensure the choice sets are varied and present all levels to each participant. This approach combines well with Hierarchical Bayesian estimation techniques, which Qualtrics uses to analyze participant's choice data.

Qualtrics' design ensures that the difference between the level seen the most per participant and the level seen the least is no more than a deviation of two. If Qualtrics fails to generate twenty choice sets that do not meet these criteria, it regenerates them again until the criteria are satisfied.

Please see Qualtrics white paper for more details on their methods https://www.qualtrics.com/support/conjoint-project/getting-started-conjoints/getting-started-choice-based/conjoint-analysis-white-paper/.

B Task Instructions for Paid Crowdworkers on Prolific

Prolific was used to post all paid crowdsourcing tasks. An example title used on the posts was "Help Build a Dataset of Computer Science Professors - University of Arizona." The instructions posted on Prolific for the paid crowdworkers include the following messages:

INITIAL MESSAGE You are invited to take part in a Brown University research study. Your participation is voluntary. :)

PURPOSE This study focuses on collecting information about specific computer science faculty members. Drafty is a public data system with thousands of computer science faculty profiles from the US and Canada.

PROCEDURES You have to add one new professor from the [UNIVERSITY NAME], not currently listed in Drafty. One row consists of a professor's:

(1) Full Name
(2) University (where they work at)
(3) Join Year (the year they started as a professor at that university)
(4) SubField (their primary research area)
(5) Bachelors (the university where they got their bachelor's degree)
(6) Doctorate (the university where they got their PhD)

Steps

(1) Open this link ([URL to faculty webpage for the university]) to visit this faculty listing page. Keep this page open.
(2) Visit Drafty using Prolific's study link (Open the study link in a new window)
(3) Compare the professors listed on Drafty with the webpage from step 1.
(4) Add one missing professor to Drafty. This is a professor listed on their webpage but not on Drafty.
(5) To add a new row, please select the white text "Add Row" in the blue bar at the top of the page. After adding a new row, you see a modal pop-up window with the Prolific completion code at the bottom.

*They must be a tenure-track professor who can solely advise a CS PhD student. They should have the title Assistant Professor, Associate Professor, or Full Professor.

COMPENSATION You will receive a base payment through Prolific for adding one new row of data to Drafty. You must add a new professor.

By recruiting Drafty's everyday visitors, we can compare their stated preferences for contributing to tasks with their real-world behavioral data collected within Drafty. For example, did they contribute data, or were they lurking? Thus, providing evidence of what motivates people who contribute to an existing public data platform such as Drafty.

B.1 Universities for Paid Crowdworkers on Prolific

University	Pay Level per hour
Case Western Reserve University	$8
George Washington University	$16
Georgetown University	$12
Illinois Institute of Technology	$12
Rensselaer Polytechnic Institute	$12
Temple University	$16
University of Arizona	$8
University of California, Davis	$16
University of Delaware	$8
University of Florida	$12
University of Houston	$16
University of Maryland, Baltimore County	$8
University of Oregon	$12
University of South Florida	$8
University of Tulsa	$16

Table 6: The universities used to create tasks on Prolific. There are five universities per pay level. The new universities added to Drafty as part of the study were the Illinois Institute of Technology, Temple University, and the University of South Florida. These were added to provide paid crowdworkers and unpaid contributors with around 25 new professors to add per university. The other universities listed were universities with the longest duration since a new professor was added in Drafty.

C Nonsensical Survey Responses

In crowdsourcing studies, nonsensical responses can most frequently occur in mandatory questions requiring text answers in the pre-survey. We reviewed all text survey responses and removed two participants with nonsensical survey responses. Here are some examples of the responses received:

(1) What motivates you to voluntarily contribute to public data systems like Wikipedia, WikiData, or Drafty?
   - The best way to answer questions about your contributions to the company is to give examples of what you have accomplished in the past, and to relate them to what you can achieve in the future.
   - There are two other key contributors to Wikipedia's success with attracting contributors, Hill's research suggests: Wikipedia offered low transaction costs to participation, and it de-emphasized the social ownership of content. Editing Wikipedia is easy, and instant, and virtually commitment-free.

(2) What motivates you to complete paid crowdsourcing tasks on platforms such as Prolific, Amazon Mechanical Turk, or UserTesting.com?
   - Describe the site first and explain what you did while you were there. reflect on what you learned during your visit. No additional research or information is needed.
   - the reason behind the not Drafty the data, is the collected data is not fully remove

D Additional Results

This section contains additional results not included in the main paper.

D.1 Additional Results: Alternative Method to Compute Trade-Offs

One method to compute trade-offs is comparing the preference share (i.e., total preference or utility) for a given crowd contribution task compared to a group's optimal crowd contribution task. Figure 9 shows how preferences and motivations shift as pay levels and task time are manipulated. The preference share for a task that pays $0 for unpaid contributors is 2 to 14 times greater compared to paid crowdworkers when comparing all estimated task completion times. This trend continues when comparing all everyday visitors (unpaid contributors) against those who only submitted accurate edits, where there is a 25–31% increase in preference share across estimated task completion times. For tasks that pay $8 per hour among those who only submitted accurate edits, the preference share for tasks that take 15 minutes or less to complete for unpaid contributors is 2 to 5 times greater than paid crowdworkers. While money can motivate accurate contributions, paid crowdworkers require larger pay levels per hour, while paying Prolific's minimum of $8 per hour to unpaid contributors should yield highly accurate edits. In summary, if pay per contribution could be introduced into a system like Drafty, that could help increase the number of accurate contributions. Next, we will analyze how much money or time can be changed to maintain the same level or preference (i.e., motivation) across different types of tasks.

D.2 Additional Results: Paid Crowdworkers Motivations Across Different Pay-Levels

The average utility scores for hypothetical tasks paying $8 per hour are relatively stable across paid crowdworkers regardless of how much we paid them to edit data on Drafty. However, these gaps increase when looking at the average utility scores for hypothetical tasks paying $12 or $16 per hour. There is a 6.3% increase in average utility scores comparing hypothetical tasks paying $8 between paid crowdworkers whom we paid $8 and $16 per hour. When comparing the same paid crowdworkers (paid $8 and $16 per hour), there is a

13.8% increase in average utility scores for tasks hypothetical tasks paying $12 per hour and a 13.9% increase for tasks hypothetical tasks paying $16 per hour. The more we paid crowdworkers, the more they preferred higher-paid tasks.