Understanding the Use of HIT Catchers and Crowd Knowledge Sharing: A Case Study of Crowdworkers on Amazon Mechanical Turk

By:

Haoyu Xie

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

in the

Information School

University of Sheffield

October 2023

Declaration

I, Haoyu Xie, hereby declare that the work presented in this thesis, titled "Understanding the Use of HIT Catchers and Crowd Knowledge Sharing: A Case Study of Crowdworkers on Amazon Mechanical Turk", is my own original work and has not been submitted for a degree at any other university or institution. Any literature, data, or works done by others used in this thesis are cited appropriately and listed in the bibliography.

- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.
- I have acknowledged all main sources of help.

Haoyu Xie

October 2023

Acknowledgements

First, I wish to express my heartfelt appreciation to my two esteemed supervisors, Dr. Suvodeep Mazumdar and Dr. Efpraxia Zamani, for their steadfast guidance and encouragement throughout my doctoral journey. Every collaboration with them was a joyful exploration that invariably enriched my insights. Their multidimensional evaluations and suggestions, encompassing the overarching research approach to the articulation of individual topics, have greatly improved my academic writing skills. It was through their persistent feedback and enlightening discussions that this thesis came to fruition. I am profoundly grateful for their mentorship!

I also extend sincere gratitude to my former doctoral and postgraduate supervisor, Dr. Alessandro Checco, who welcomed me into the realm of research, nurtured my good research habits with patience, and collaborated with me on various projects. His guidance also opened doors to numerous collaborative opportunities. I am very grateful to him for making my research rewarding and enjoyable!

These admirable individuals are not only my respected supervisors but also cherished friends who have inspired my life and career development. Their companionship has rendered my PhD journey an exciting and memorable experience. They have not only equipped me with the skills to navigate the research world but more importantly, instilled in me the essence of independent and critical thinking in both academia and life.

Most importantly, I owe a debt of gratitude to my parents for their unwavering financial and emotional support, which was the cornerstone of my smooth sail through this academic journey. I am deeply thankful for their patient waiting for me, their child who has been away from home for so long.

Lastly, I wish to thank the faculty of the Information School and my colleagues in the PGR office. During the time with them, I not only gained a sense of belonging, but also valuable friendships. Their company has undoubtedly enriched my doctoral journey, making it a cherished chapter of my life.

Abstract

Crowdsourcing platforms have become a vital component of the modern digital economy, offering a wide range of HIT (Human Intelligence Task) opportunities to workers worldwide. Meanwhile, crowdworkers' use of scripting tools and their communication with each other are continuously shaping the entire crowdsourcing ecosystem. This thesis explores the use of HIT catchers by crowdworkers and their sharing of skill-based knowledge that facilitates the popularity of such scripting tools. It is revealed that the use of HIT catchers affects the completion speed and HIT-worker diversity for the whole HIT group, while depriving job opportunities from others. This potentially undermines the stability of the platform under the current reputation system relying on numbers of approvals and approval rates. Subsequently, another study explored how work strategies under the use of HIT catchers, including HIT acceptance, backlog, and completion, affect HIT availability, completion time, and result quality. The study also found differences in work behaviours between workers using and not using HIT catchers. Finally, this thesis investigates the skill-based knowledge sharing behaviour of crowdworkers, which promotes the blooming of scripting tools including HIT catchers, to improve the fairness of work opportunities and mitigate its negative impact on HIT completion. Using PLS-SEM, we assess the factors influencing knowledge sharing in the domain of skills. The study reveals the significance of high performance expectation, low effort expectation, and the joy and satisfaction in motivating the crowd skill-based knowledge sharing. Overall, this study provides an in-depth exploration around these two types of collective behaviour, highlighting the important role of tool use and knowledge sharing in shaping the crowdsourcing ecosystem.

Table of Contents

List of Tables

List of Figures

List of Equations

List of Abbreviations

ABS	Agent-Based Simulation
AI	Artificial Intelligence
ALM	Application Layer Monitoring
API	Application Programming Interface
AVE	Average Variance Extracted
CB-SEM	Covariance-based structural equation modelling
CNN	Convolutional Neural Network
CR	composite reliability
DES	Discrete Event Simulation
ECDF	Estimator of the Cumulative Distribution Function
EE	Effort Expectancy
FC	Facilitating Conditions
FP	False Positive
FN	False Negative
GoF	Goodness of Fit
HIT	Human Intelligence Task
IDT	Innovation Diffusion Theory
IQR	Inter Quartile Range
KS	Knowledge sharing
KSI / KSB	Knowledge Sharing Intention / Behaviour
LLM	Large Language Model
MSE	Mean Square Error
MTurk	Amazon Mechanical Turk
P	Precision
PE	Performance Expectancy
PRE	Page Request Error
PLS-SEM	Partial Least Squares Structural Equation Modelling
R	Recall
RMS-theta	Root mean square residual covariance
SDS	System dynamic simulation
SRMR	Standardised Root Mean Square Residual
SVC	Support Vector Classifier
SVM	Support Vector Machine
TP	True Positive
UTAUT	Unified Theory of Acceptance and Use of Technology
TPB	Theory of Planned Behaviour
TRA	Theory of Reasoned Action
TAM	Technology Acceptance Model
SET	Social Exchange Theory
VC	Virtual Communities
REC	Reciprocity
REP	Reputation
REW	Rewards
SIT	Social Interaction Ties

SI	Social Influence
T	Trust
VIF	Variance Inflation Factor

List of Publications

Parts of this thesis, as well as research related to the topic, have been published, and a list of these publications is provided here:

- Xie, H., Checco, A., & Zamani, E. D. (2021, July). Design Principles and a Conceptual Framework for Crowd Teamwork Systems. In *PACIS* (p. 214).
- Xie, H., Checco, A., & Zamani, E. D. (2023). The Unintended Consequences of Automated Scripts in Crowdwork Platforms: A Simulation Study in MTurk. *Information Systems Frontiers*, 1-17.
- Xie, H., Maddalena, E., Qarout, R., & Checco, A. (2023). The Dark Side of Recruitment in Crowdsourcing: Ethics and Transparency in Micro-Task Marketplaces. *Computer Supported Cooperative Work (CSCW)*, 1-36.
- Xie, H., Mazumdar, S., Zamani, E. D. A Study on the Factors Influencing Skill-based Knowledge Sharing within Crowdworkers. Information Technology & People (in preparation, to be submitted).

Chapter 1 Introduction

1.1 Research Background

Micro-tasks, known as Human Intelligence Tasks (HIT), are small tasks more easily solved by humans than by computers, but require crowdsourcing due to the volume and size of the task, e.g., market research, image or video annotation, and training AI algorithms (Gadiraju et al., 2014; Liang et al., 2022; Sveen et al., 2020). Typical micro-tasks or HITs include market research questionnaires for a particular industry, or requests for participants to transcribe text from an audio recording (Difallah et al., 2015; Gadiraju et al., 2014). The last decade has witnessed a boom in micro-task crowdsourcing as an emerging work model (Connelly et al., 2021). Crowdsourcing platforms such as Amazon Mechanical Turk (MTurk) and Prolific allow requesters to hire online workers from all around the world at an affordable price (Heer & Bostock, 2010; Palan & Schitter, 2018). Crowdwork offers workers flexibility in scheduling, choice on where they work and working hours (Bohannon, 2016). It is used in a range of industries and academic fields including healthcare (Walters et al., 2018), medical images analysis (Petrović et al., 2020), behavioural accounting research (Brandon et al., 2014) and psychology (Gosling & Mason, 2015; Tam et al., 2021).

Within crowdsourcing platforms such as Amazon Mechanical Turk (MTurk), the job requesters, those who post the microtask, obtain low-cost online labour from a broad pool of human capital (Heer & Bostock, 2010; Sannon & Cosley, 2019). In the case of MTurk, for example, the crowdworkers[1] involved in a HIT usually spend a few minutes to a few hours on the task completion, submitting their survey responses or contributions, as requested by the job requester who posted the HITs. The requester ultimately pays the crowdworkers monetary rewards based on the factors including quality of output and the estimated time to complete the task (Litman et al., 2015; Xie, Maddalena, et al., 2023). Table 1.1 illustrates common reward methods across multiple crowdsourcing platforms. Here, we focus only on the types of rewards that have real market value,

[1] This terminology has been widely used in previous studies (Osterbrink & Alpar, 2021; Posch et al., 2022; Silberman et al., 2018). Another synonym, "crowd worker", is also popular in previous studies (Al-Qershi et al., 2021; Gadiraju, Checco, et al., 2017).

including monetary rewards and platform currencies that can eventually be converted to cash. It is worth noting that the crowdsourcing platforms that appear in the table offers small and discrete Human Intelligence Tasks, and therefore do not include freelancing platforms like Fiverr that carry out larger standalone tasks. While the table covers several crowdsourcing platforms, it is likely that owing to the rapid turnover of such platforms, the table is non-exhaustive (Leung et al., 2021). To ensure accuracy in categorization, Appen is not included in the table because it partners with multiple crowdsourcing channels that each have their own unique reward structures (Xie, Maddalena, et al., 2023). Instead, the channels that it partners with are treated as separate platforms in the table (such as InstaGC, ySense, Swagbucks and NeoBux). It is revealed that monetary rewards are more commonly used by the platforms included in the table than other two types of rewards including vouchers and virtual currency.

Unfortunately, in the case of MTurk, novice crowdworkers have to complete a large number of low rewarding HITs in order to increase the number of HITs completed from their worker profiles and, more importantly, achieve and maintain a sufficiently high HIT approval rate (Hara et al., 2018). Similar to the sellers' ranking scores in Amazon Marketplace based on feedback from customers, the HIT approval rate is an official measure of one crowdworker's quality of HIT completion over time. As it is commonly used by requesters as a filter when choosing their target crowdworkers, this metric often determines whether crowdworkers are eligible to receive high-rewarding tasks (Kaplan et al., 2018; Savage et al., 2020).

Table 1.1 Sample microtask platforms categorised by types of rewards.

Reward type	Platform
Monetary compensation	MTurk, Prolific, UserTesting, InstaGC, ySense, Swagbucks, NeoBux, Clickworker, Neevo
Gift card, vouchers	MTurk, UserTesting, InstaGC, Swagbucks
Virtual currency (cryptocurrency, platform currency, point)	InstaGC, Swagbucks, Prolific, Clickworker

However, the current working model of crowdsourcing platforms often encourages poor working conditions. First, often due to poorly designed, low-quality of HITs, it is difficult for crowdworkers to effectively measure the time required for completing a HIT prior to embarking on a new task. Secondly, in addition to the time spent on performing HITs, crowdworkers need to spend time on searching and identifying relevant HITs, reading reviews about requesters, reading related instructions, and potentially learning how to complete them, including learning how to interact with the customised HIT user interface (Martin et al., 2014; Sannon & Cosley, 2019; Toxtli et al., 2021). The additional time costs for such invisible labour are unpaid and often exceed the time spent performing the HITs (Chilton et al., 2010; Gadiraju, Yang, et al., 2017; McInnis et al., 2016). Furthermore, the way platforms are currently designed means that there are limited ways for workers and requesters to interact, making the former more likely to be treated unfairly (Fieseler et al., 2019; McInnis et al., 2016).

1.1.1 The Rapid Growth of the Gig Economy

The digital revolution has changed the norms that people exchange value of traditional labour in society (Bard et al., 2019; Imamov & Semenikhina, 2021). More specifically, the digital revolution has increased productivity while reducing the efforts required by individuals to perform specific tasks, ultimately leading to varying degrees of automation. The application of automation is gradually replacing traditional labour, such as automated driving or customer service. As a result, traditional norms of labour exchange are being disrupted by technological advances. Moreover, the labour market has been unprecedentedly expanded, whereby the worldwide spread of internet connectivity has allowed gig companies to recruit workers from all over the world, and especially from the developing countries (Alalawneh & Alkhatib, 2021; Graham et al., 2017; Uchiyama et al., 2022).

On the one hand, this leads to more flexible employment for workers: Jacques and Kristensson (2019) analysed the results of four large-scale surveys of MTurk workers conducted over six years, and found that the participants were no longer in full-time jobs and their estimated poverty levels had fallen below national average level. These trends show that gig workers become more flexible to choose their temporary jobs and maximise their earnings in a rapidly changing labour market.

On the other hand, this leads to segmentation of the labour market between the employers and gig workers (Rani & Furrer, 2019). Employers are more flexible in their allocation of labour, including the recruitment of gig workers, who are part-time contract workers, more often than standard, full-time employees. Workers in non-standard employment relationships are highly substitutable as their jobs have a lower skills requirement than those taken up by full-time employees and are more readily available in the global labour market (Wood et al., 2019). Coupled with the lack of legal protection for these gig workers, the bargaining power of employers is further strengthened (Piasna & Myant, 2017).

Based on the “2017 U.S. Freelance” study led by Upwork and Freelancers Union, there are over 57 million independent contractors, which is about one-third of the total US workforce (Dunn, 2020). Among those independent contractors, a large proportion of these are gig workers. Moreover, this number is increasing at a rate of about 18% per year (Kässi & Lehdonvirta, 2018).

Among the many types of gig workers, those who focus on completing microtasks (HITs) have less control over the rewards of their work than freelancers. It is important to note that the scope of work between microtasks and freelancer jobs is different. Microtasks are often small, repetitive which require human cognitive skills (Margaryan, 2019). They do not often require specialist skills and only need minimal training. In contrast, for freelancers, tasks often require multiple advanced skills, often encounter completely new problems, and require unique solutions (Blaising & Dabbish, 2022; Rani & Furrer, 2019). The unbalanced design of microtask platforms and the low bargaining power of crowdworkers means that they have to accept rewards determined by requesters, rather than bidding based on their skill level, quality of service, buyer ratings and accumulated reputation, as freelancers selling their skills on platforms like Fiverr do (Ke & Zhu, 2021; Maffie, 2020). This passive pricing system further increases the chances of exploitation of crowdworkers. Although freelancers are required to receive a reverse selection from buyers during the bidding process and that there is heterogeneity in buyers' willingness to pay for the same services. However, freelancers are still given the power to bargain on the basis of their personal bargaining power relative to that of the buyer (Ramadhiani & Adnan, 2023).

1.1.2 Roles Served by Crowdsourcing Platforms

This section explores the three core roles of crowdsourcing platforms in the microtask marketplace: recruitment agents, rule-setters, and mediums for negotiation. By breaking down these three roles, we can better understand how crowdsourcing platforms create order in the microtask marketplace, and how they regulate the behaviours of platform members to drive the marketplace forward.

1.1.2.1 As Recruitment Agents

Firstly, as a "recruitment agent", the crowdsourcing platform recruits and filters crowdworkers to complete microtasks for job requesters (Cui et al., 2021). MTurk, Prolific, Appen and other platforms inherited the ways of making profit from traditional recruitment agencies. They connect task requesters in need of Artificial Intelligence data with crowdworkers from across the globe, thus helping them to collect data or get paid for providing such data.

1.1.2.2 As Rule-setters

Second, they act as "rule-setters", providing a set of operational guidelines and standards for microtask transactions and assuming risk for job requesters. However, the fairness and transparency of such platforms for crowdworkers were criticised in previous studies (Borromeo et al., 2017; Xie, Maddalena, et al., 2023).

1.1.2.3 As Mediums for Negotiation

The crowdsourcing platform acts as an intermediary between the two parties (requester and crowdworker) for the negotiation. From the job requesters' point of view, these platforms keep the price low through their strong bargaining power and knowledge of industry price standards. But from the workers' view, they have not given sufficient consideration to crowdworkers' rights as they tend to attract more buyers instead of sellers just like the intermediaries in other industries (Chu & Manchanda, 2016; Fieseler et al., 2019). In other words, these platforms influence the unbalanced bargaining powers of the two parties in terms of unequal industry knowledge, or unequal supply and demands. This influence can be reflected from further reducing or amplifying the inequality in bargaining power between the two parties (Fieseler et al., 2019).

MTurk acts as a medium between job requesters and crowdworkers. Although from the workers' standpoint, MTurk brings workers flexible working hours. However, due to the large number of active workers on the platform and the fact that the vast majority of microtasks do not require high level skills, this results in workers having significantly less bargaining power than a limited number of requesters (Graham et al., 2017). The platform then leverages the difference in bargaining power between the two parties to help requesters reduce their expenditure on collecting data (Wood et al., 2019).

1.1.3 Fairness Considerations in Crowdwork

Crowdworkers have been revealed to be treated unfairly by the crowdsourcing platforms (Fieseler et al., 2019; Lascau et al., 2022; McInnis et al., 2016). The labour supply is in excess of demand (Wood et al., 2019). Wood et al. (2019) evaluated the job quality through the participated crowdworkers from Southeast Asia and Sub-Saharan Africa. 54% of the respondents of the study said that they did not have enough job opportunities. Wood (2019) also claimed workers got de-skilled by the overwhelmed competition from the globe and low bargaining power caused by crowdsourcing tasks. Since crowdworkers, as a type of gig workers (Schulte et al., 2020), are treated as independent contractors but not employees, they cannot benefit from traditional labour protection laws (Reynolds & Kincaid, 2023).

Due to the lack of legal protection, most platforms are not inclined to protect the rights of workers. For example, the MTurk policy provides little recourse to workers when they are treated unfairly (McInnis et al., 2016; *Participation Agreement*, 2020). Such legal and institutional gaps have led to Turkers [2] being perceived as invisible workforces, lacking sufficient transparency and communication with other platform members (Cohen et al., 2020). Furthermore, this difference results in information asymmetry and the power imbalance for both sides of the transaction. If we use MTurk as an example, the information asymmetry means the platform does not allow workers to choose ideal tasks based on the reputation score or reward level of job requesters, while the job requesters can filter the workers based on their personal information and task acceptance rates. The power imbalance here means the requesters could reject the workers' outputs without justification, while keeping their outputs without any penalties (El Maarry et al., 2018). TurkerView Bridge was developed with the aim to resolve disputes between requesters and workers (ChrisTurk, 2022). This platform acts as a third party, helping workers to send their appeals to requesters in order to claim the income they have been unfairly denied. In addition, an appeal system called 'Turkish Judge' was developed, inviting crowd workers as judges to rule on whether one worker’s submission had been fairly rejected by the requester when an appeal was launched by their peers (Cohen et al., 2020). In general, both ways shape the decision on the dispute by bringing in a third party and require additional effort from the worker or requester to

[2] It refers specifically to the crowdworkers on MTurk (Savage et al., 2020)

deal with the dispute. Moreover, a problem that is difficult to resolve is that third party does not always have sufficient authority or trust to allow both parties to a dispute and accept final judgement.

Furthermore, crowd workers cannot rate requesters with a reputation system just like the one used by the requesters. Moreover, workers cannot build strong relationships with each other through the platform to fight for better rewards. They lack collective bargaining on employment, wage agreement and unions (Bergvall-Kåreborn & Howcroft, 2014). These features are perceived as unfair by the workers. As a result, the workers who receive unfair payment feel they are undervalued by both the requesters and the platform (Fieseler et al., 2019). Moreover, Fieseler's study on working opinions highlights the initial cause of the power imbalance: the discrepancies in information transparency for workers and job requesters on the platform (Fieseler et al., 2019).

1.1.4 Crowdworkers' Unpromising Income

Due to the lack of sufficient employment opportunities locally, and the impact of political events such as the lockdown under the pandemic, a big portion of the crowdsourcing workers earn their living expenses by doing these low-paid microtasks (Popiel, 2017; Spurk & Straub, 2020). Moreover, during the pandemic, a large number of existing crowdworkers increased their micro-tasking hours, thereby compensating for the decline in income associated with the shift in traditional work patterns (Reynolds & Kincaid, 2023). Unfortunately, about 96% of workers on MTurk earn below the US federal minimum wage (Hara et al., 2018; Woodcock & Graham, 2019). Worse still, this type of income is highly unstable (Reynolds & Kincaid, 2023). Even though the average payment from requesters is $11.58/h, most of them just publish low-paid tasks which make most rewards fairly low. A study on the general crowdwork ecosystem shows that workers are exposed to being treated unfairly (El Maarry et al., 2018). In detail, 58% of the participants were disappointed with MTurk's overall effort in blocking wage theft and unfair requesters. Unpaid work is one reason for their low hourly wages. Bad task design, technical errors and workers being unfamiliar with new tasks can also result in low hourly wage (McInnis et al., 2016).

Badly designed tasks, technical errors, and interface design errors (McInnis et al., 2016; Paulino et al., 2023) made by the job requesters may confuse workers and result in extra time, efforts, and even failure of submission or increased risk of rejection (Gadiraju, Yang, et al., 2017). Even if they are finally paid, the workers may have spent additional unpaid time and efforts searching for preferred and quality jobs, learning how to do the tasks they are not familiar with, and waiting for the response of their questions from the requesters. All these problems contribute to poor work efficiency, resulting in low hourly wages based on the efficiency wage theories (Gumata & Ndou, 2017; Rani & Furrer, 2019).

The above are discussed in greater detail in the following sections.

1.1.4.1 Extra Time and Workloads Spent on Task Learning

When facing new tasks, workers will put extra time and effort into learning how to do new tasks. This includes reading instructions, learning how to interact with the task UI and input their answers, etc. To save time for completing more tasks, crowdworkers are not guaranteed to read the task description carefully (Göritz et al., 2021; Rothwell et al., 2016). On the other hand, in order to improve their attentions to work, Researchers tried to attract their attention by using more visual interactive elements to highlight the necessary information, including gamification to enhance the interactive experience of microtask interfaces (Paulino et al., 2022; Qiu et al., 2021). In addition, the requester will ensure that task outcomes are produced by workers while maintaining adequate attention by asking workers to pass the attention test (Göritz et al., 2021). However, these additional hours of work performing attention tests or interacting with non-task elements reduce workers' hourly income.

1.1.4.2 Extra Time and Workloads Spent on Task Searching

Workers choose the preferred tasks from the HIT batch list on the platform. It allows workers to sort batches by the number of microtasks each batch contains, the reward for each microtask, and estimated completion time. However, it is still challenging for workers to find the appropriate tasks due to insufficient features for task matching. Worse still, they have to spend even as much time and effort on searching for preferred tasks as on completing them (Kurup & Sajeev, 2020; Safran & Che, 2018; Toxtli et al., 2021). This also leads to the situation that some workers, in order to reduce the extra time and effort spent in searching for tasks and to get higher rewards, tend to spend less time browsing through the task pages in order to pick the ones they prefer and are suitable for (Ambati et al., 2011; Chilton et al., 2010).This in turn leads to workers taking on tasks in which they are not interested or skilled, which does not only decrease their work quality for doing low quality tasks but also lowers their willingness to work in the future.

1.1.5 Structure of HITs and Searching for HITs

The unfair treatment and lack of guaranteed hourly earnings faced by crowdworkers have been discussed in the previous sections. This in turn led to concerns about the design of the HITs and the process by which crowdworkers search for them. This subsection introduces the structure of the HITs published on the MTurk platform, the definition of relevant terms and the search methods built into the platform. These introductions provide the fundamental knowledge background for understanding the research included in the thesis that follows. In addition, the introduction to the built-in HIT search function allows to gain a better understanding of the reasons for the extra task search time and the reasons why workers choose to use the scripting tools to assist with task acceptance.

1.1.5.1 Structure of HITs Posted on MTurk

A job requester could post a HIT batch (or group) containing multiple HITs, and this HIT group has its own Batch ID[3]. Each HIT is a distinctive task with unique content that differs from other HITs, and each HIT has a corresponding HIT ID. In addition, to improve the accuracy of results for each HIT, the job requester could assign one HIT to multiple workers. In this case, one HIT could contain multiple assignments with a unique Assignment ID. To summarise, in order of content hierarchy, the three terms are: HIT group, HIT, and assignment.

1.1.5.2 Built-in Ways of HIT Searching on MTurk

The specific HIT search and filtering features that come with the MTurk platform are illustrated to help further understand the difficulty for workers to search for HITs, and the need for an optimised search feature in the platform.

Sorting HIT groups by criteria: In MTurk's default HIT groups list page, workers can sort all available HIT groups by the number of HITs contained in a group, the amount of the reward, and the time each group was posted (Figure 1.1). In the more advanced filter options, workers can also

[3] Tutorial: Understanding HITs and Assignments: https://blog.mturk.com/tutorial-understanding-hits-and-assignments-d2be35102fbd

qualify themselves to work on HITs that require a Masters[4] qualification (Figure 1.2). In addition, workers can exclude those with low rewards by setting a minimum reward. Obviously, filtering based solely on HIT rewards does not consider the true rewards per unit of time, nor the overall approval rate of the HIT results, and is therefore not sufficiently informative.

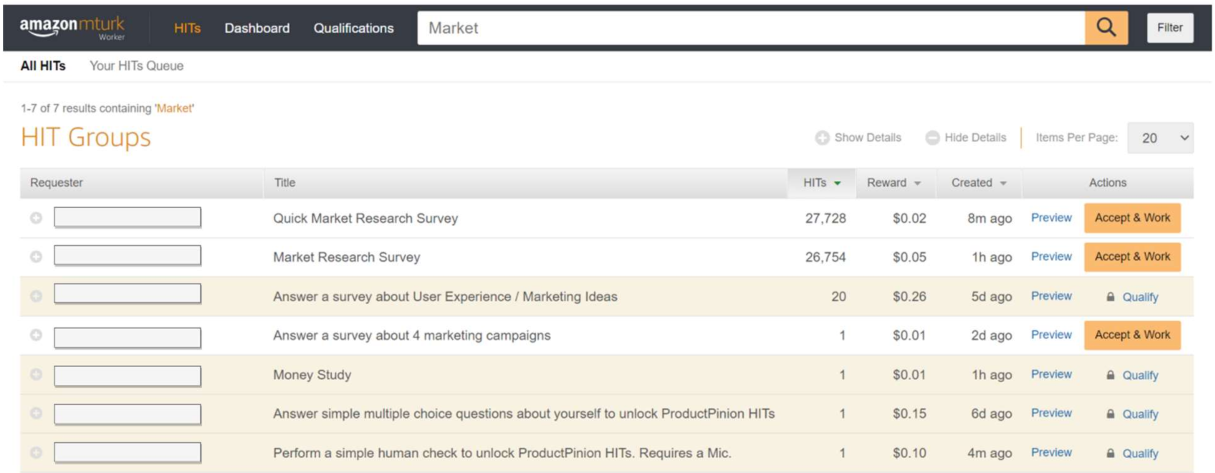

Figure 1.1 A list of HIT groups sorted by the total number of HIT included.

Search by keyword: In this search method, workers can obtain recommended HITs that have matching information in the HIT title, description, tags, and the name of the requester that posted the HIT. This search approach usually produces vague HIT recommendations, as not every description and label about the HIT is accurate (El Maarry et al., 2018).

[4] Simplified Masters Qualifications: https://blog.mturk.com/simplified-masters-qualifications-137d77647d1c

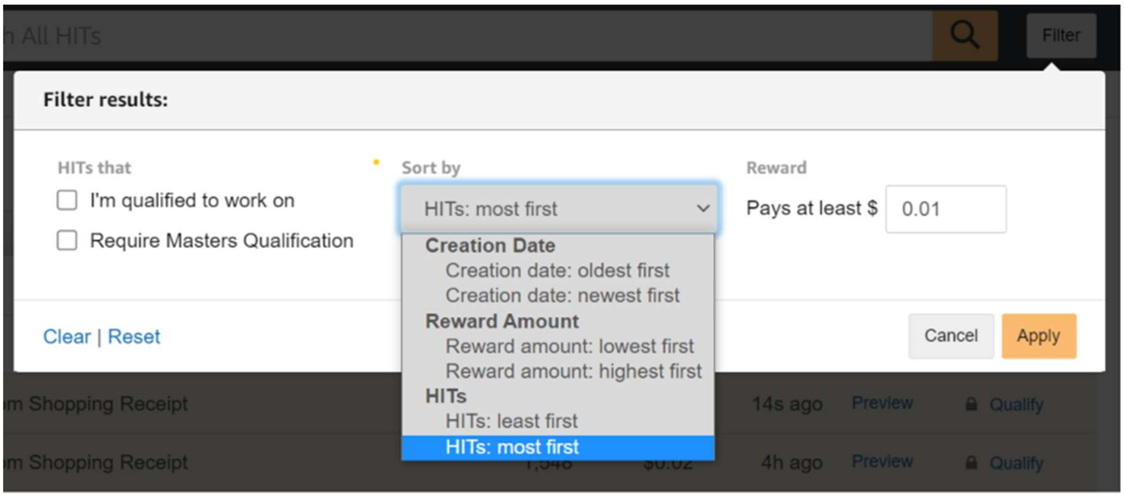

Figure 1.2 Screenshot of advanced filter options on MTurk

Search by specific requester: Workers can also get a list page of microtasks posted only by a specific requester by clicking on that requester's name on the HIT list page. This feature facilitates the monitoring of job opportunities by preferred requester as workers can more quickly notice newly posted HITs from a specific requester on this page.

Overall, the built-in HIT search functions do not meet the requirements of workers to obtain information about their ideal HITs in a competitive platform due to the lack of adequate filtering features, including the selection of HIT categories and monitoring of new HITs from specific HIT groups (El Maarry et al., 2018).

Furthermore, the way MTurk currently designed means that there are limited ways workers and requesters can interact, making the former more likely to be treated unfairly (McInnis et al. 2016). On the one hand, workers help each other to avoid being potentially exploited by requesters by sharing reviews about requesters and HITs on information-sharing platforms such as worker forums. On the other hand, crowdworkers seek help from browser extensions or scripts (e.g., Panda Crazy Max and MTurk Suite) to assist with their daily HIT work (Hellman, 2021; Ramirez, 2021) . The aim is to make the most of the time and effort invested on the platform; this may mean optimising the overhead of completing HITs, reducing the chances of unfair treatment by rogue requesters, identifying low-quality HITs, and gaining an advantage over other workers competing for the same higher-quality, higher-yield HITs (Irani & Silberman, 2013; Ramirez, 2023).

1.1.6 Impact of HIT Catchers

A HIT catcher is an algorithm executed on the worker's computer (usually in the form of a browser plugin, extension, or script) that allows the worker to automatically reserve HITs (Williams et al., 2019). Such algorithms send high-frequency acceptance requests to the endpoints used by the platform to allow workers to reserve HITs, increasing the likelihood of successfully reserving them (DonovanM, 2018; Hellman, 2023).

Numerous studies have revealed the positive impact of scripting tools on workers' earnings (El Maarry et al., 2018; Savage et al., 2020; Williams et al., 2019). However, based on the estimation by Robinson et al. (2019), it can be seen that the number of unique workers on MTurk has remained stable between 80,000 and 90,000 for each year between 2016 and 2018. In contrast, the number of users of the most popular HIT catchers, MTurk Suite (Uzor et al., 2021; Williams et al., 2019), is around more than 20,000[5], far less than the total estimated number of unique workers on MTurk. In other words, as the ecosystem represented by this technology is not well known by the majority of workers, the number of workers who are proficient in this skill and benefit from this ecosystem remains a minority (Savage et al., 2020; El Maarry et al., 2018).

With the increasingly popular use of HIT catching tools by the crowd workers, the competition among the workers appears to increase over time (Hanrahan et al. 2018). Workers without such tools cannot even see good HITs at all, since the most attractive HITs are instantly caught by the scripts that discover them first. Thus, HIT catching tools originally used to cope with the rapid velocity of the market appear to have led to this problem being further intensified (Hanrahan et al., 2018). The problems associated with this widespread use of HIT catching tools, although identified by researchers, have not been quantified in terms of their impact on platform members' job opportunities, work behaviours, HIT completion process and results. This is where this study aims to make a contribution. The answers to these questions could help to understand the impact of the tools on the platform, and the working conditions of the crowdworkers. Previous research has also uncovered that new Turkers left platforms more frequently than before due to a lack of quality HIT

[5] MTurk Suite in chrome web store: https://chrome.google.com/webstore/detail/mturk-suite/iglbakfobmoijpbigmlfklckogbefnlf

opportunities (Hanrahan et al., 2018; Kaplan et al., 2018). This rise concerns about the negative effect of diminishing new worker engagement on the diversity and quality of HIT results.

In other words, the use of HIT catchers potentially has a short-term impact and a long-term impact on result quality. In the short term, the majority of HITs in a newly published sought-after HIT group tend to be reserved by workers with HIT catcher skills (Hanrahan et al., 2018). This leads to a significant reduction in the fairness of HIT distributions, and therefore the biases generated by a few individuals could potentially affect many results. In addition, a decrease in the diversity of participants can affect the reproducibility of the study and the reliability of the results (Castille et al., 2019; Moss et al., 2020).

In the long term, the high frequency of new workers leaving the platform leads to an increase in the proportion of "professional participants". Specifically, as workers gain experience with microtasks in social research disciplines and become more familiar with the research processes carried out by requesters, they develop strong preconceived opinions that can interfere with task results (Conte et al., 2019; D. Hauser et al., 2018).

Existing research primarily focused on describing the crowdsourcing ecosystem represented by scripting tools including the HIT catcher, and the impact of scripting tool use has remained a preliminary exploration of mainly qualitative methods (El Maarry et al., 2018; Hanrahan et al., 2018; Savage et al., 2020). However, the impacts of the widespread use of HIT catching scripts on the worker population, tasks and platforms needs to be quantified in order to assess the extent of the deprivation of work opportunities for peers, the impact on HIT results, and the impact on speed of HIT group completion. More importantly, not enough attention has been paid on the impact that the openness of crowdsourcing platforms to third-party applications has on the working conditions of crowdworkers. This openness brings an enhanced experience of using the platform through the innovation of third-party applications (Wessel et al., 2017), but it also creates potential risks (El Maarry et al., 2018; Kaplan et al., 2018; Williams et al., 2019). Specifically, third party applications may collect and store sensitive data about workers and requesters. The breaching and misuse of data could lead to people’s loss of trust and confidence in the platform. Malicious scripts used to complete HITs automatically can directly decrease data quality. In addition, scripts that send too frequent HTTP requests can affect the stable operation of the platform server.

In overall, in crowdsourcing domain, the current findings around the impact of scripting tool use come mainly from qualitative research. There appears to be a gap in practical knowledge from a quantitative perspective. The impact of the openness of crowdsourcing platforms to third-party tools has not been given sufficient attention.

1.1.7 Knowledge Sharing within Crowd Community

Due to the limitation of platform designs, crowdwork is often treated as a form of work that is isolated and lacks interaction with coworkers (Gerber, 2021; Irani, 2015). In fact, workers engage in rich communication interactions outside of the platform in self-organised community environments, including online forums and channels in social apps (Gerber, 2021). These interactions greatly contribute to the dissemination of tasks and tools, resulting in rich collective behaviours (El Maarry et al., 2018). The emergence of self-organising communities is a form of self-regulation of worker groups, reflecting the efforts of crowdworkers to cope with unfair work systems and to improve their working conditions.

Specifically, due to the unstable income (Hara et al., 2018), unfair treatment by requesters (McInnis et al., 2016), and job stress (Wood et al., 2019), crowdworkers form different communities defend themselves and peers through knowledge sharing. The knowledge being shared includes job opportunities (Zyskowski & Milland, 2018), tools as well as work strategies (El Maarry et al., 2018), comments on requesters and microtasks (ChrisTurk, 2022), and more. On the one hand, They share knowledge of using tools in the crowd community, which allows a wide range of workers to benefit from scripting tools (El Maarry et al., 2018; Scholz, 2016). On the other hand, the impact of using scripting tools on the crowdwork ecosystem and the factors driving their knowledge sharing remain unclear.

Online communities play a key role in facilitating workers' mutual support (Garcia Martinez, 2017; Mason & Suri, 2012). Crowdworkers get involved in community management, finding career opportunities, and building social connections (Gray et al., 2016). Additionally, independent communities interact with each other at different levels (Yin et al., 2016). To improve their own working conditions, crowdworkers create, use and share tools to assess requesters (Irani & Silberman, 2013), obtain task suggestions, visualise task data (Hanrahan et al., 2015), manage completed tasks (Hellman, 2021) and facilitate communication with communities (ChrisTurk, 2022). Therefore, knowledge sharing is also an important way to facilitate the use of tools. A study on the microtask scripting tools by Williams et al. (2019) reveals that the HIT catcher, which is used to filter and automate the reservation of tasks, is popular within the crowdworker communities. However, through the study in Chapter 5, it was found that not every participant was using HIT catchers. This leads us to look further into the topic of knowledge sharing among workers.

It was also found that workers with higher incomes on MTurk used more tools and were more actively engaged in the community (Kaplan et al., 2018). This also suggests that deeper engagement in knowledge exchange provides workers with additional technical advantages. Thus, after examining the phenomenon and impact of workers' use of the HIT catcher, we further investigate what factors led people to share skill-based knowledge including utilising tools, which drove the prosperity of the entire crowd tooling ecosystem (El Maarry et al., 2018).

1.1.8 Key Terms and Concepts of HIT

This section introduces some key concepts that are important to the context of this thesis.

HIT availability: One HIT being available means the current HIT is visible in MTurk HIT list, so it can be accepted and completed by any crowdworker that meets the worker requirements and qualifications.

HIT backlog: It means that the HIT becomes temporarily unavailable due to the worker's particular behaviours. Causes of HIT backlogs include, but are not limited to, a HIT being received by a crowdworker, held in their HIT queue until this HIT expires, and then retrieved by MTurk server from the current worker's HIT queue. In addition, a HIT may be continuously previewed in the browser but not accepted by a worker, making it invisible to other workers and preventing them from accepting this HIT being continuously previewed.

HIT expiration: Each HIT has a time limit after it is accepted by a worker, and the time limit is named "Allotted Time"[6] by MTurk. If the current worker cannot submit a response to the HIT within the time limit, a HIT expiration event is triggered. This event would result in the task no

[6] FAQs: https://www.mturk.com/worker/help

longer being available to the current worker and would be withdrawn from the current worker's HIT queue by MTurk.

HIT group completion time: A requester normally publish a group of small and discrete HITs as a batch[7]. "HIT group completion time" means the total time spent to complete a whole HIT group / batch.

HIT-worker diversity: This indicates the overall equity of opportunities on doing HITs for each worker participated and the result diversity for the HIT group. The more fairly the HIT completions are distributed among the participants, the higher the HIT-worker diversity.

1.2 Research Aims and Questions

The thesis aims to study how the use of HIT catchers is impacting crowdwork strategies, result quality, workers' job opportunities, and the crowdsourcing platform. In studying this, the thesis also aims to understand how crowdworkers share skills-based knowledge that drives the popularity of HIT catchers. Therefore, two research questions are revealed:

RQ1: What are the impacts of the use of HIT catchers on HITs and crowdworkers? The research question is further divided into two sub questions:

RQ1.1: What impacts do HIT Catchers have on HIT-worker diversity, response quality, completion time, HIT availability and backlog?

RQ1.2: How does the use of HIT catcher impact the work behaviour and job opportunities of crowdworkers?

Knowledge sharing, as another collective behaviour, has contributed to the popularity of scripting tools (El Maarry et al., 2018). In addition, knowledge gap among workers influences their work strategies including the use of tools, which in turn contributes to their income gap (Kaplan et al., 2018; Williams et al., 2019). Therefore, Research Question 2 (RQ2) focuses on the factors that affect the sharing of skill-based knowledge among workers.

[7] Publish a batch of HITs: https://docs.aws.amazon.com/AWSMechTurk/latest/RequesterUI/PublishingYourBatchofHITs.html

RQ2: What are the key factors influencing the skill-based knowledge sharing within the crowd communities?

In answering the Research Questions, the work described in this thesis seeks to achieve the following Research Objectives (RO):

RO1: Review current literature related to HIT catchers, including research on the phenomenon and impact.

RO2: Investigate the MTurk platform as a case study, including an exploration of the mechanism by which microtasks are published, reserved, and completed by MTurk.

RO3: Reveal the impact of using HIT catchers via a simulation framework built from the study output of RO2.

RO4: Review existing methods of detecting worker behaviour that can be applied to study the impact of HIT catchers.

RO5: Design and develop an experiment by publishing image annotation tasks to assess the impact of the use of HIT catchers on worker behaviours, job opportunities, HIT dynamics, and results.

RO6: Develop a conceptual, measurement and structural model based on theories related to behavioural research and subjective perceptions of crowdworkers to study the factors influencing their skill-based knowledge sharing behaviour.

By reviewing current literature related to HIT catchers (RO1), we set a foundation for understanding the current impacts of these tools on crowdwork. This knowledge will directly inform RQ1. Investigating the MTurk platform gives a practical understanding of HIT state transitions and time required (RO2), which is essential to achieve RO3. By revealing the impact of HIT catchers through a simulation framework, we can assess the impact of using HIT catchers on HIT dynamics, HIT-worker diversity and more, which would directly address RQ1.1 and may touch on RQ1.2. Conducting an experiment using image annotation tasks (RO4, 5) can give practical insights into the real-world impacts of HIT catchers on worker behaviour and the results, thus addressing RQ1. Finally, RO6 is about understanding the factors that influence skill-based knowledge sharing behaviour among crowdworkers, making it the primary objective to answer RQ2.

1.3 Research Contributions

The contribution of this thesis are as follows:

1. This is the first study that holistically reviews the phenomenon and impact of crowdworkers' use of HIT catchers (Section 2.2).
2. This is the first time to systematically review the detection of crowdwork behaviours, and the correlation between behaviour traces and result quality (Section 2.4).
3. It is the first time to comprehensively review the crowd knowledge sharing and related theories to study this behaviour (Section 2.6).
4. This thesis extends our understanding of the impact of using HIT catchers by demonstrating how reputation systems from crowdsourcing platforms can contribute to the Matthew effect (Section 4.2), whereby those with effective use of HIT catchers can benefit at the expense of others with less technical advantage (Section 4.5).
5. Based on the technique of Application Layer Monitoring (ALM), we incorporate event data of microtask state changes, which extends the exploration of worker behaviour to non-task completion phases (Section 5.2.5).
6. A predictive model using Support Vector Classifier (SVC) is developed for assessing the quality of image annotation tasks based on worker behaviours (Section 5.3.7). This extends our understanding and methods for behaviour-based quality assessment of HIT results.
7. This thesis extends our understanding of the impact of HIT catchers on HIT results and worker behaviours. The use of HIT catchers results in significantly longer completion times for the entire HIT group and lower quality results for text generation tasks. Moreover, it leads to workers’ more frequent attention switches and reduced focus time during HIT completion (Section 5.3).
8. This thesis extends the literature of crowdwork strategies by investigating worker behaviours including using multi-devices, multi-HITing and potential irregularities quantitatively (Section 5.3.1).

9. This is the first time to study the factors affecting the skill-based knowledge sharing within crowdworkers using PLS-SEM. Performance Expectancy (PE), Effort Expectancy (EE) and Reward (REW) are revealed to influences the crowdworkers' Knowledge Sharing Intention (KSI), while EE also directly influences Knowledge Sharing Behaviour (KSB) (Section 6.5.4).

1.4 Thesis Structure

This thesis, titled "Collective behaviours within the crowd communities： the use of HIT catchers and knowledge sharing", is divided into seven chapters, of which chapters 4, 5, and 6 detail three studies.

Chapter 1 introduces the research context, aiming to establish the research questions and objectives. It further presents an overview of the research contributions and outlines the structure of the full text.

Chapter 2 presents a thorough literature review, assessing the current state of research in the fields of crowdsourcing, simulation, and virtual communities. It critically reviews the use of HIT catchers, worker behaviour, the methodologies of simulation experiments, and crowd knowledge sharing, thus identifying the gaps and challenges in these areas.

Chapter 3 discusses the philosophical underpinnings of the research, elucidating the ontology and epistemology that shape the research design. Here, a clear statement of the purpose of each study, reasons for the chosen data collection and analysis methods, and the intrinsic connections between each study are provided.

Chapter 4 answers the first research question, exploring the unintended consequences of the use of HIT catchers. This investigation is grounded in an approach involving manual measurements and simulations, providing significant insights into HIT catchers' impact on crowdwork.

Chapter 5 builds upon the previous findings, moving to investigate real-life scenarios of crowdwork strategies. This chapter unveils unique worker behaviours, their impact on job opportunities, data quality, and worker diversity, presenting a more realistic picture of the unintended consequences of HIT catcher usage.

Chapter 6 presents a study on crowdworkers' skill-based knowledge sharing behaviour. Based on a structural equation model, factors affecting knowledge sharing are analysed in terms of individual experiences with sharing tools and social benefit exchange.

Finally, Chapter 7 synthesises the findings of the thesis. It discusses the main findings of each research chapter, summarising the impacts of using HIT catchers, worker behaviours, and factors facilitating knowledge sharing among crowdworkers.

Through this structure, this thesis provides an in-depth understanding of the use of HIT catchers and crowd knowledge sharing, filling crucial gaps in the existing literature and providing a robust foundation for future research in this area.

Chapter 2 Literature Review

2.1 Introduction

Prior to conducting the research, the literature in the related field was explored and reviewed to understand the current research gaps and to generate a clear research direction. To ensure the relevance of the literature, we mainly searched in academic databases including Google Scholar, Web of Science, and IEEE Xplore. These databases contain many academic articles covering relevant areas such as microtask crowdsourcing, scripting tools, microtask work behaviour research and knowledge sharing behaviour research. To ensure that the literature review is relevant and comprehensive, we searched for keywords including, but not limited to, "crowdsourcing working conditions", "crowdsourcing scripting tools", "crowdsourcing knowledge sharing", "quality/algorithmic control in crowdsourcing", "crowdsourcing workers' behaviours/strategies", "application layer monitoring", "machine learning quality prediction", and "theories in knowledge sharing study".

Although HIT catchers are widely used among crowdworkers, research on their use is still limited (El Maarry et al., 2018; Kaplan et al., 2018; Williams et al., 2019). The literature review on the use of HIT catchers (Section 2.2) helps us to identify research gaps in this area, thus clarifying the value and contribution of this study. Subsequently, a literature review on the topic of microtask quality control (Section 2.3) helps us to understand the mechanisms by which crowdsourcing platforms assure result quality. The need to investigate behaviour-based quality control in this study is demonstrated by reviewing the limitations of common current quality control methods. In addition, there is a link between microtask quality control and knowledge sharing: crowdworkers may exchange tips on building reputation scores, or even share gold standard answers for specific microtasks, to help each other earn income faster (Checco et al., 2018). However, such knowledge sharing can affect the effectiveness of quality control methods discussed under this topic.

In Section 2.4 Microtask Work Behaviour, a review of the literature related to crowdwork behaviour detection and behaviour-based quality assessment provides a better understanding of how such behaviours can be detected, how workers interact with crowdsourcing platforms and how such interactions can affect task completion and data quality. The review of the Quality

Control and Microtask Work Behaviour helps us to understand how the use of HIT catchers affects worker behaviour and result quality of microtasks and has long term impacts on the development of both workers and crowdsourcing platforms through the current reputation system.

Simulation provides us with a way to predict the impacts of using HIT catchers. By reviewing studies using simulation (Section 2.5), we understand the strengths and limitations of different simulation frameworks, providing a comprehensive reference for subsequent experiments.

Microtask crowdsourcing involves not only the assignment and completion of tasks, but also communication between workers. A review of the literature related to crowd knowledge sharing (Section 2.6) enables an understanding of how knowledge about HIT catchers spreads within the crowd communities, what the influencing factors are, and what gaps there are in the current understanding of microtask knowledge sharing. Crowd knowledge sharing is relevant to all the other themes in the literature review. It represents the collective wisdom of a worker group and has potential impacts on tool use, job quality, and task completion strategies. As a summary, Figure 2.1 demonstrates the inner connection between the five topics in the literature review.

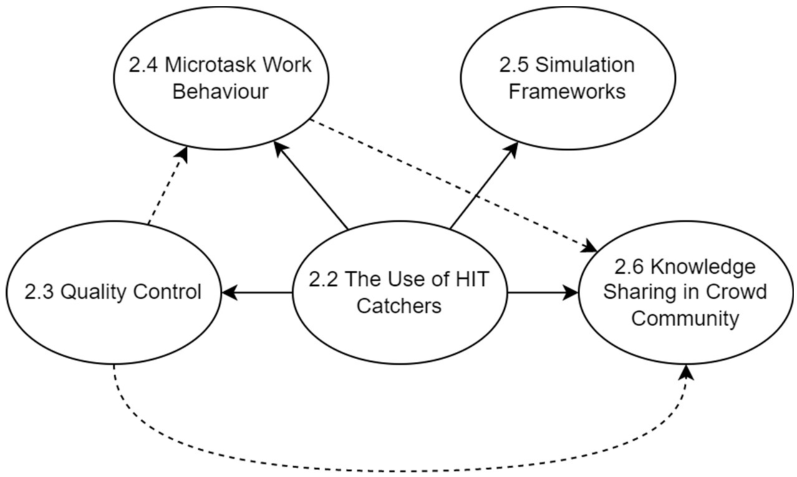

Figure 2.1 Concept mapping of topics in the literature review

2.2 The Use of HIT Catchers

The use of HIT catchers, or catching scripts, in crowdsourcing platforms allows crowdworkers to maximise their income and optimise their time by identifying and capturing microtasks that match their interests and skills (Uzor et al., 2021; Williams et al., 2019). These scripts help workers gain an advantage in securing high-quality HITs while reducing unpaid labour spent searching and filtering irrelevant tasks. However, despite the potential negative impacts of HIT catchers on platform ecosystems, research on their effects in the crowdsourcing field is still in its infancy (Hanrahan et al., 2018). Existing studies mainly analyse qualitative data, such as interviews and worker feedback, without quantifying the impact of script use on crowdworkers and platform growth.

2.2.1 Introduction to HIT Catchers

The main way crowdsourcing platforms like MTurk achieve HIT-Worker matching is through a search engine (El Maarry et al., 2018). Once a worker identifies a suitable HIT batch, they can preview one of them and decide whether to accept the job. The HIT will be then assigned to the worker for a fixed amount of time (also called Allotted Time[8]), after which the HIT will be put back in the market should the worker fail to complete it on time. This procedure of search and selection can become very tedious and time-expensive when working on microtasks. For this reason, crowdsourcing platforms devised some additional functionalities to increase the efficiency of the job assignment phase: workers can, at HIT completion, auto-accept the next HIT in the same HIT group. Moreover, workers can have a queue of up to 25 HITs reserved at any given time, allowing them to group the search and reservation phase and to ensure that a sizable amount of HITs are reserved before starting to work, thus reducing context switching (ChrisTurk, 2017). The worker community shares HITs that have been reviewed as high quality through third-party forums and review platforms (Irani and Silberman 2013). However, accessing and reserving high-paying jobs can still be difficult because the competition between workers can cause a lot of failed reservation attempts. For this reason, many workers use HIT catchers (Williams et al., 2019).

Catching scripts essentially simulate human behaviour. The script allows individuals to expedite the process of identifying and securing extremely limited items in an online platform, particularly when the demand for these items is high and exceeds the supply (Vancea et al., 2020). Similarly, to partially address the unpromising hourly income, and specifically to address the strong

[8] FAQs: https://www.mturk.com/worker/help

competition among crowdworkers (D'Cruz & Noronha, 2016), many of them have turned to using scripts such as PandaCrazy Max (Ramirez, 2023) and TurkerView (ChrisTurk, 2022) to save time from searching and help them filter microtasks (Irani & Silberman, 2013).

Table 2.1 categorises and compares HIT catching capabilities of popular tools. MTurk Suite and Panda Crazy Max rank highest in terms of the number of total installations. In particular, HIT catching tools include support for filtering, whereby HIT batches are filtered and displayed according to user preferences, allowing users to choose the target HIT batch directly from the filtered list to start catching HITs (Hellman 2021). Automatic catching has evolved as well. For example, users can now set an upper limit on the number of HITs they accept automatically, to prevent having too many HITs expire (which would increase the worker abandonment rate) and to allow the possibility to reserve other quality HITs manually (Ramirez, 2023; Watwani, 2023). Furthermore, some HIT catchers use load balancing to optimise acceptance frequency or dynamically adjust it for multiple HIT groups (Ramirez, 2023; Schultz, 2020). It can be revealed from Table 2.1 that most of the tools could catch specific HITs according to ID, adjust the catching frequency, catch multiple targets synchronously, and remind users. Interestingly, the function of catching according to keywords and other text descriptions is not popular within the tools listed here. Very few tools effectively integrate the advanced HIT search with the auto catching functionality. Instead, tools prefer to let users manually add desired HITs to the catching list. This table is also discussed in Section 5.2.5.1.

Table 2.1 Summary of popular tools containing HIT catchers[9].

Features of HIT Catching Function	Panda Crazy Max	Turkmaster	MTurk Suite	Mturk Engine	Stax	Turk Guru
Adding target HITs. It includes creating watchers or panda jobs manually by Group ID or directly from the MTurk HIT list and plugin search result.	✓	✓	✓	✓	✓	✓
Managing target HITs. It includes customisable watcher settings or panda job cards, multi-tab management, grouping for the target HITs and data import / export.	✓	✓	✓	✓	✓	
Adjustable catching frequency. It means user can change interval between each HIT catching attempt, and even catch HITs in dynamic frequency for a higher success rate.	✓	✓	✓	✓	✓	✓
Concurrent acceptance between multiple target HITs.	✓	✓	✓	✓	✓	✓
Limitations on the number of HITs being auto accepted. It could be the maximum number in the queue, or the maximum number accepted per day.	✓					✓
HITs queue monitoring.	✓			✓	✓	✓
Notification on the acceptance of HITs. It includes sounds, pop-ups, desktop notifications etc.	✓	✓	✓	✓	✓	✓
Auto-accept HITs directly by keywords, categories, etc.						✓
Number of users on Chrome Web Store by 21 June, 2023	10,000+		20,000+		1,000+	2,000+
Total Installations on Greasy Fork by 21 June, 2023	121,072	98,665		9,365		

Sources: (DonovanM, 2018; Hasan, 2018; Hellman, 2023; Ramirez, 2023; Schultz, 2020; Watwani, 2023)

A HIT group or HIT batch will keep its Group ID and URL unchanged if it is not revised by job requesters (ChrisTurk, 2017). When the requester re-publishes this HIT group, workers could accept HITs within this HIT group earlier than the platform HIT group list page via the customised URL. The existence of this backdoor-like access allows workers who monitor the target HIT group in advance via the URL to obtain the HITs much faster than others. By automating this monitoring behaviour, HIT catchers further facilitate access to specific HIT groups, which explains why many

[9] Three are three outdated tools included: Turkmaster was last updated on Jan 3, 2018; Mturk Engine was last updated on Jun 21, 2018; Stax was removed from Chrome Web Store on Jan 3, 2022.

high-quality HITs are snapped up before they even appear on the platform HIT list page (El Maarry et al., 2018; Williams et al., 2019). The very short retention time of quality tasks in the platform is an important issue that most workers have to face and has been widely discussed by workers in the community forums[10].

2.2.2 The Effects of HIT Catchers Use

In the crowdsourcing platform domain, these catching scripts help to capture micro-tasks that match crowdworkers' own interests and skills. This allows them to maximise their income while reducing the amount of time spent on searching through the HIT list and filtering out irrelevant HITs, which often results in unpaid labour. Such scripts help crowdworkers identify microtasks that offer higher rewards and are aligned with their skills and interests. Specifically, these scripts allow crowdworkers to catch microtasks posted by specific job requesters (based on their ID numbers) with high ratings or task-based specificity (Saito et al., 2019), where the requirements of the task align with their prior experience or personal interests (Dror et al., 2011; Geiger & Schader, 2014). In other words, these scripts support selective automatic catching of microtasks by the crowdworkers based on personal preferences. Therefore, such HIT catching scripts or HIT catchers can help crowdworkers optimise the time and effort spent on completing HITs, reduce the chances of being treated unfairly by malicious requesters, identify low-quality HITs, and gain an advantage when competing for the same high-quality HITs (Irani & Silberman, 2013; Ramirez, 2023). Research by El Maarry et al. (2018) also reveals its positive effects: catching scripts help users to mitigate the asymmetries of the market and the lack of HIT search capabilities in the native platform.

Experienced crowdworkers can earn up to $12 per hour with the help of such scripts (Newman, 2019), as they have better access to higher paying tasks and, in turn, their overall response approval rates on the platform improve once they complete them. One study showed that high-paying microtasks with high reputation scores were booked within seconds of being posted (Hanrahan et al., 2018). Thus, crowdworkers who have browser scripting skills and knowledge which are necessary for the use of a catching script have a significant advantage over those who do not,

[10] Here is a list of relevant posts: https://www.mturkcrowd.com/threads/where-have-all-the-hits-gone.6057/; https://www.reddit.com/r/mturk/comments/cjvnz4/hits_always_showing_up_as_there_are_no_more_of/; https://www.reddit.com/r/mturk/comments/pff53z/hits_disappearing/

making it difficult for less experienced ones to identify tasks with high or moderate rewards. Gradually, this leads to a high frequency of crowdworkers becoming frustrated and leaving the platform altogether, which is detrimental to the development of the platform (Hanrahan et al., 2018). In other words, the HIT catching tools originally used to cope with the high velocity of the market appear to have further intensified the problem they aimed at solving led to this problem being further intensified. The problems associated with this widespread use of HIT catchers, although identified by researchers (Williams et al., 2019; El Maarry et al., 2018), have not been yet quantified in terms of their impact on crowdworkers platform members and the output of the HITs.

Previous research has also uncovered that new Turkers exit platforms frequently due to a lack of HIT searching features (El Maarry et al., 2018). This means that fewer crowdworkers will be completing more HITs. Often, high quality HITs are caught by workers who use the most advanced and effective HIT catching tools. When these tools do not limit the maximum number of HITs that can be completed by a single worker, the diversity of outcomes in HITs decreases. The reduction in diversity, on the other hand, affects the reproducibility of studies and the reliability of results (Castille et al., 2019; Moss et al., 2020).

In addition, a large number of microtasks may be captured by scripts and sit idle in crowdworkers' queues until they are completed, or until they expire (also known as task abandonment) because the crowdworkers cannot complete all caught tasks within the time allotted by the job requester (Han et al., 2019). In the latter case, the completion of the entire batch is delayed. In turn, the impact of the use of catching scripts on batches' completion time and response quality has not been sufficiently quantified and studied.

Crowdworkers' motivations for using HIT catchers and its impact on work behaviour have been initially explored. Firstly, the need for the use of assistive tools in crowd work has been tied to the lack of full disclosure by requesters on official platforms including Mturk and the inadequate search function for HITs (El Maarry et al., 2018; Kaplan et al., 2018). Some studies have categorised the tools commonly used by crowdworkers and defined tools such as HIT catching scripts to help future researchers target specific types of tools (Williams et al., 2019; El Maarry et al., 2018). Furthermore, crowdworkers' use of HIT catching scripts could result in an increase of interruptions in attention at work and interference with daily life (Williams et al., 2019). While the existing studies have identified phenomena such as multitasking and optimisation of work, particularly among Super Turkers[11], using scripts and how these lead to increased earnings, they have mainly done so by analysing said phenomena based on qualitative data (interviews and feedback from workers), without quantifying the extent of the impact of crowdworkers' use of scripts. In other words, there seems to be an empirical gap in the prior research. Previous research has focused primarily on qualitative perspective. By far, no study has attempted to investigate the correlation between the use of HIT catchers and specific work behaviours. In addition, the impact of the use of the HIT catcher on HIT results and on the HIT group completion process has not been explored, and investigation of these empirical issues is important. The factors affecting data quality have always been an important research topic (Chmielewski & Kucker, 2020; Kennedy et al., 2020; Loepp & Kelly, 2020).

In summary, while the use of automated HIT catchers was originally intended to improve crowdworkers' access to their preferred tasks, their overuse can destabilise platform ecosystems, inhibit healthy platform growth and negatively impact crowdworkers who do not use or are not skilled at using scripts. The following section discusses why crowdsourcing platforms are open to the use of such automated scripts despite their potential negative impact on their ecosystem, and what measures they have taken to control script overuse.

2.2.3 Crowdsourcing Platforms' Perspectives on HIT Catchers

Crowdsourcing platforms, like most online platforms, are in essence marketplaces where buyers and sellers can meet and exchange services for an agreed fee (Mohammadi & Hashemi Golpayegani, 2021). Similarly, to other online platforms, in order to enhance the platform's functionality and therefore its overall competitiveness, crowdsourcing platforms provide complementors (also called third parties) with access to the platform, who create plug-ins, add-ons and other extensions (Wessel et al., 2017). This provides sufficient autonomy to third parties and encourages complementary innovations (Boudreau, 2010; Hein et al., 2020).

In crowdsourcing platforms, the use of automated microtask catching scripts is a reflection of platforms' openness to such complementors, where third parties are able to develop their scripts

[11]It refers to the crowdworkers earning higher income than the averages on MTurk (Savage et al., 2020).

in compliance with the platform's regulations (*Acceptable Use Policy*, 2018). These platforms explicitly allow the use of automated scripts so that crowdworkers are better able to search for and preview microtasks. These enhanced features makes platforms more attractive to crowdworkers by improving their workflow (Abbas & Gadiraju, 2022; Xie, Checco, et al., 2023).

However, the challenges observed in other platforms (such as crowdfunding and open-source platforms) are also present. Wessel et al. (2017), for example, discuss that a major challenge is identifying the balance between platform openness towards third parties and maintaining control. Being too open can potentially destabilise the ecosystem. In the case of crowdsourcing platforms, the stability of the ecosystem extends beyond retaining oversight of operations; openness needs to be balanced against the need for ensuring the fair treatment of crowdworkers and providing job requesters with high quality outputs, both of which feed into the healthy growth of the ecosystem. Amazon Mechanical Turk, for example, prohibits the use of scripts that send requests at an excessively high frequency and those that automatically accept HITs (*Acceptable Use Policy*, 2018). It is unclear, however, whether this ban is for the purpose of maintaining the operational stability of servers, for ensuring the fair treatment of all crowdworkers or for satisfying diversity in the data collected via HITs. In reality, while the imposed limitation on the frequency supports operational stability, it is unknown whether it can ensure fairness between those who use automated scripts and those who don't.

2.2.4 Crowdworkers' Perspectives on HIT Catchers

HIT catchers keep crowdworkers on the job, reduce their time spent on searching for HITs and let them not miss the highly rewarded HITs (Kaplan et al., 2018; Savage et al., 2020). These potentials for increasing their income become the main reasons for the general acceptance and endorsement of such tools by crowdworkers.

However, crowdworkers suffer from interruptions caused by HIT catchers as it interrupts them with automatic reminders when they are focused on performing a task or even in a non-working state (Williams et al., 2019). In addition, HIT catchers may also cause the crowdworkers to pause their work and think about which tasks to keep and which to abandon because they are reserving too many unfamiliar HITs (Williams et al., 2019). These disruptions inevitably increase the resistance of crowdworkers to HIT catchers.

Although there is very limited research on the attitudes of crowdworkers towards HIT catchers, we can still find evidence in the MTurk related forums. In addition to the interruptions caused by such tools, forum members expressed confusion about how to use the HIT catchers due to their technical difficulty. Such an example (paraphrased from original post) is shared here:

> *"Scripts are confusing to me because I have only recently started using them. Despite the fact that I have managed to get Hit Forker and Panda Crazy Max to cooperate, I am aware that I am still lacking one essential part of them: the ability to obtain HITS before they run out. I've read that you can have them waiting in your queue while you work on others, which is helpful because I don't know much about using scripts and anything that makes things easier for me helps."*
>
> – Crowdworker explained concerns about using HIT catchers [Turker Nation: Oct 16, 2022]

The confusion comes not only from a lack of basic knowledge of how to install browser plugins and load web scripts, but also from the frequent technical changes to browsers, websites, or scripts. This lack in knowledge potentially disadvantages crowdworkers in employing technical solutions to identify tasks that would be appropriate and sufficiently rewarding for them, thereby losing out on more knowledgeable peers (Savage et al., 2020).

2.2.5 Conclusion: Navigating the Complexities of HIT Catchers: Balancing Productivity, Fairness, and Ecosystem Health

Crowdworkers' employing HIT catchers reveal an intricate balance between efficiency and fairness. While HIT catchers allow workers to capture tasks quickly, thereby increasing their productivity, they also raise concerns about the fair distribution of tasks and access to tasks for all workers. This conflict emphasises the need to further investigate the systemic impact of HIT catchers on the crowdsourcing ecosystem. Furthermore, there is also a need to think about countermeasures to mitigate its negative impact.

2.3 Quality Control

This section reviews current quality control approaches for microtasks in terms of three topics: reputation system, consensus algorithm, and gold standard. The review includes specific quality assessment methods, applicable scenarios or advantages, and limitations for each approach.

2.3.1 Reputation System

Quality control mechanisms are critical to the successful operation of a crowdsourcing platform, and the main mechanisms for controlling the quality of work outcomes include algorithmic control based on worker reputation systems (Gol et al., 2019).

Based on current algorithmic control mechanisms on MTurk, requesters tend to set their posted HITs to be visible only to master workers with very high approval rates, in order to filter out inexperienced workers from the large labour pool (Waldkirch et al., 2021). Specifically, the platform assists requesters in filtering out workers who deliver high-quality results when posting tasks based on a reputation system consisting of the worker's HIT approval rate, the number of HITs completed, and relevant qualification labels (Sodré & Brasileiro, 2017). This reputation system is argued to provide a good estimate of workers' future performance, allowing job requesters to verify the qualifications of workers as soon as a task is posted, and blocking potentially malicious workers (Zhu & Carterette, 2010). In addition, as a form of informal control (Kirsch, 1997), the reputation system encourages workers to strive for higher ratings by completing more HITs and improving HIT approval rates, therefore getting more quality jobs.

However, issues regarding reputation systems are gradually being identified and discussed. (Loepp & Kelly, 2020) found when conducting research on MTurk that there was no significant difference in the quality of HIT results between master workers and regular workers, but instead master workers generated biased results in the regular psychometric tests because of their extensive experience in answering similar surveys. Furthermore, Wood et al. (2019) argued that the 'symbolic power' of platform reputation score is identified as an emergent market bargaining power, whereas workers lacking platform reputation suffer from a lack of income and a constant insecurity of being abandoned by the platform.

However, in recent years, problems with reputation systems on crowdsourcing platforms have surfaced and received widespread attention. A study by Loepp & Kelly (2020) revealed an interesting phenomenon: on the MTurk platform, master workers, who are generally recognised as representing high quality of work, did not differ significantly from regular workers in terms of the quality of microtasks completed. More notably, as master workers have extensive experience in completing questionnaires, this may have adversely affected the results of regular psychometric tests (Conte et al., 2019; Hauser et al., 2018). MTurk's reputation system overemphasises the pass rate of workers completing microtasks, leading to a tendency for workers to choose tasks that are less likely to be rejected to maintain a high reputation (Rzeszotarski & Kittur, 2011). This mechanism provides a way for those who may not have submitted high quality work to maintain a high reputation. In addition, Wood et al. (2019) further notes that the 'symbolic authority' of platform reputation has become an emerging market bargaining tool. Workers who lack platform reputation are not only limited in their income, but are also in a perpetual state of marginalisation and abandonment by the platform.

In summary, the ability of the current reputation system that crowdsourcing platforms relies on to objectively reflect the workers' HIT completion qualities, and the fairness of such algorithmic control mechanisms in treating new crowdworkers, has been increasingly questioned by research. However, the negative impacts of this control mechanism remain to be further explored, which have not been quantified in practice. Therefore, there is a gap in the empirical experience of research on the negative effects of this control mechanism.

2.3.2 Consensus Algorithm

Another quality control strategy to be introduced is the consensus algorithm. The core idea of this algorithm is to aggregate the feedback from a set of workers to arrive at a final prediction (Yanagisawa et al., 2022). One of the consensus computing methods is Majority Voting, which determines the final correct answer by simply aggregating the responses of multiple workers (Nordheimer et al., 2015). However, this method can be affected by responses with different levels of credibility. To address this issue, the Weighted Majority Voting method was proposed, which assigns weights to each answer based on the worker's historical performance, thus optimising the accuracy of the results (Zhang et al., 2017).

Although optimised consensus algorithms improve the accuracy of results, as they usually rely on multi-agent systems, this leads to redundant information and additional communication overhead in the system, which increases the overall cost of obtaining reliable results (Yang & Choi, 2021). To address this issue, Yanagisawa et al. (2022) proposed a dynamic microtask release model that aims to reduce the total number of responses while maintaining label accuracy, thereby effectively reducing the cost of result collection.

In addition, such algorithms can be affected by cyber-attacks, especially Sybil attacks (Wang et al., 2020). In this attack, the attacker floods the system with false information by creating multiple forged identities, thus disrupting the algorithm's judgement of the answer driven by the majority effect (Dong et al., 2022).

2.3.3 Gold Standard

Another widely adopted approach is the use of gold standard data to assess the quality of the results submitted by workers (Checco et al., 2018). Specifically, by comparing workers' answers with a set of predetermined gold standard answers, requesters could estimate the quality of each worker's answer and identify potentially malicious or inefficient workers accordingly (Burmania et al., 2016; Hettiachchi et al., 2021; Zhao et al., 2023). This approach is particularly suitable for microtasks that have explicit answers (Al-Qershi et al., 2021).

However, there are limitations to this approach. Firstly, the gold standard answers may conflict with workers' subjective interpretations and potential biases, which may further affect requesters' assessment of the data quality (Naderi et al., 2021). Second, there are additional costs associated with creating and maintaining these gold-standard questions, and they may no longer be applicable as the content of the microtasks changes and is updated, thus reducing their usefulness (González Pinto et al., 2019).

2.3.4 Conclusion: Evolving Quality Control in Microtasks: Bridging Traditional Methods and Behavioural Insights

The current landscape of quality control in microtask crowdsourcing, encompassing reputation systems, consensus algorithms, and gold standards, presents unique challenges and limitations. Reputation systems, while prevalent, may not always accurately reflect the true quality of work.

Consensus algorithms, though beneficial for accuracy, can lead to increased system overhead and vulnerability to cyber threats. Gold-standard methods, typically reliable, can conflict with subjective interpretations of workers and entail significant maintenance costs. These challenges highlight the critical need for innovative quality control methods that focus on worker behaviour analysis.

2.4 Microtask Work Behaviour

Worker behaviour detection in crowdsourcing platforms is an emerging approach in assessing the quality of task results. This detection is mainly performed at the application layer and is referred to as Application Layer Monitoring (ALM). This section reviews existing research on methods for detecting and analysing worker behaviour at the application layer and methods for assessing the quality of different types of microtask results.

2.4.1 Detection of Worker Behaviours

The monitoring of worker behaviour is usually carried out at the application layer such as browser pages, so it is also called Application Layer Monitoring (ALM) (Hirth et al., 2014). The behavioural data generated by the workers during HIT completion is also named "task fingerprinting" (Rzeszotarski & Kittur, 2011) or "behavioural traces" (Goyal et al., 2018). As an implicit measure of process quality and hence the quality of HIT result, ALM has three advantages over other common quality testing methods for HIT results:

Firstly, as the monitoring of worker behaviour is carried out via scripts hidden in browser pages, the presence of the monitoring behaviour is barely noticeable to the worker. It also means that the ALM does not require additional gold standard questions as an answer filter. Whereas gold standard questions have been found to be maliciously exploited by fraudulent workers, thus losing their role in data quality assurance (Checco et al., 2018; Gadiraju et al., 2015). Especially for tasks other than close-ended questions such as video annotation tasks, where it is difficult to design gold standard questions, quality assurance through ALM is essential (Mok et al., 2016).

Secondly, ALM monitors worker behaviour in real time, so it is possible to predict the quality of results based on worker behaviour in a short time before the results get evaluated. This also avoids additional costs for manual or third-party quality assessment of the results.

Finally, because ALM assesses quality in a way that is independent of the result data, it has the potential to prevent bias caused by quality assessment methods on the data itself, such as faulty gold standard questions that can misjudge the quality of the data.

In general, the data used in current research on the detection of worker behaviour mainly include action-based and time-based data. The action-based data includes mouse and keyboard actions, workers' interactions with interface elements, and focus events on the browser page. The time-based data includes the time spent completing the task, focusing on the task page, etc.

2.4.1.1 Action-based Detection

Firstly, the existing studies on the detection of cursor and keyboard operations include cursor trajectories with coordinates, mouse clicking/over/scrolling events, keypress, the ways of workers type answers in the text fields (Goyal et al., 2018; Hirth et al., 2014; Mok et al., 2016; Rzeszotarski & Kittur, 2011).

Regarding the processing of such action-based data, existing studies have calculated cursor speed and acceleration (Mok et al., 2016), the distance of cursor movement, the amount of scrolling (Rzeszotarski & Kittur., 2011) and other parameters from cursor trajectory information for subsequent analysis. The numbers and positions of clicking were also applied for analysis (Mok et al., 2016). In addition, by visualisation and correlation of behaviour traces and output, the difference in worker behaviour that provides different quality outcomes can be identified (Rzeszotarski & Kittur, 2012). By applying correlation analysis methods such as a random forest model using regression, labelling accuracy could be predicted based on worker behavioural traces (Goyal et al., 2018).

2.4.1.2 Time-based Detection

In addition to action-based data, existing studies have also described worker behaviour from a time-based perspective, including overall completion time (Mok et al., 2016). Time spent on separate events were also included such as time spent for answering each question, time for reading instructions, answering and considering (Hirth et al., 2014), time for making continuous judgements (Zhu & Carterette, 2010), time spent focusing on HIT (Goyal et al., 2018).

The study of time-based data allows for the effective classification of worker types. Specifically, Zhu and Carterette (2010) summarized three types of workers based on the variation in the time to complete continuous judgment during each task: normal, periodic, and interrupted workers. Whether a worker cheated was then predicted based on the worker classification. Al-Qershi et al. (2021) proposed a novel model based on time series and showed the behavioural features of workers and the different types through the model. The integration of time-based data analysis, compared with the action-based analysis, allows for a higher level of assessment of worker efforts.

In addition, the rapid development of deep learning models in recent years has opened up new possibilities for the study of worker behaviour. For example, In the study from Al-Qershi et al. (2021), a lightweight deep learning model CNN model was applied to evaluate the quality of results based on both action-based and time-based behaviour data. In a study by Gadiraju et al. (2019), for image transcription and information finding tasks, workers were categorised according to mouse operation behavioural traces, and high-quality results were obtained by proposing a supervised machine learning model for worker categorisation.

However, these studies of worker behaviour have not yet focused on the impact of the use of HIT catchers on worker behaviour and have not yet explored methods for detecting scripts. In addition, there is a lack of clear descriptions of the use of scripts in terms of behaviour, and the influences of their use are not yet clear.

2.4.2 Correlation Between Behaviours and Results Quality

Following a review of behavioural detection methods, this section reviews the existing studies on the correlations between worker behaviours and quality of HIT results.

Similar to the classification on the perspectives of detection of worker behaviours, current studies explored the correlations between behaviour and result quality from the perspectives of both time and action.

2.4.2.1 Time-based Factors

Regarding the time-based factors, task completion time and percentage of actual completion time compared to reported time are the factors associated with quality of results (Hirth et al., 2014; Rzeszotarski & Kittur, 2011).

Specifically, in the research made by Hirth et al. (2014), the HITs published in their study required workers to read the text and answer a number of multiple-choice questions. By analysing the time-related data they recorded, it was found that almost all workers whose HIT completion times were below the confidence threshold also provided data of a quality below the pass threshold. Setting a completion time threshold is therefore an effective way of assessing the quality of the results. In addition, the average amount of time a worker spends thinking about answering questions is an important indicator for assessing result quality, especially the time spent answering the last question. One reasonable interpretation is that the time spent thinking about answering the questions is a good reflection of the efforts spent by the worker in answering them. Moreover, the longer time spent answering the last question, the more it reflects the seriousness of the worker's attitude to work.

Rzeszotarski and Kittur (2011) published three types of HITs in their study, which included asking workers to identify nouns in a word list, add keywords to pictures, and read a text then answer reading comprehension questions. By analysing user behavioural events, it was found that workers often took on multiple HITs and put them on hold while working on others. In addition, there was a huge discrepancy between the HIT completion time reported by the platform and the time actually spent on working. For the word recognition and reading comprehension HITs, the tasks with large differences between reported and actual time spent on completion were inclined to be of poorer quality.

2.4.2.2 Action-based Factors

Regarding the action-based factors, the diversity of textual input, the degrees of interaction with task page interface was found to correlate with the quality of results.

It was discovered that more fields accessed, more unique characters typed in, more clicks, and more total time spent could predict higher precision scores of HIT results for content generation tasks like providing keywords for images (Rzeszotarski & Kittur, 2011). In a subsequent study by

Rzeszotarski and Kittur (2012), it was also found through behavioural traces that in the HITs asking workers to write a summary based on the video, those who typed in answers after watching the video as well as those who did not watch the video at all tended to provide lower quality results than those who typed while watching the video.

Al-Qershi et al. (2021), on the other hand, found through worker behaviour and HIT results data provided by the study from Goyal et al. (2018) that: for HITs that require workers to assess the relevance of documents and topics, statistics on mouse movements and browser tab focus change events can be a good indicator of workers' attitudes to work. It is worth noting that in this type of HITs, workers have to switch between the task page and the document page in order to read the document and fill in the conclusions. So frequent mouse movements and switching between pages reflect a good work attitude.

In summary, the behavioural requirements of workers vary considerably between different types of HITs, so there is no guarantee that one behavioural characteristic judgement could be used to assess the result quality of all HITs. However, factors such as task completion time, focus time, and text diversity are generally applicable to assessing the quality of results for most types of HITs.

Furthermore, although existing research has identified and categorised numerous worker behaviours associated with low quality results, these are limited to the workers’ interactions with task pages during the completion of HITs. In other words, the task acceptance behaviour of workers has not been included in the study of worker behaviour. Moreover, no factors have been formed to assess the quality of task results through data related to task acceptance behaviour. Although Rzeszotarski and Kittur (2011) found that there were workers accepting multiple HITs all together, the relationship between HIT over-acceptance and low-quality results has not been explored. Furthermore, no link has been built between the backlog of HITs and the use of HIT catchers.

2.4.3 Conclusion: The Uncharted Territory of Worker Behaviour and HIT Catcher Dynamics

The emerging focus on worker behaviour, including Application Layer Monitoring (ALM), offers a nuanced approach to quality assessment. Nevertheless, the current research has not extensively delved into the effects of HIT catchers on worker behaviour or explored detection methods for using HIT catchers. Task acceptance behaviours, including the handling of task backlogs, have not been incorporated into studies of worker behaviour. The potential of using such acceptance behaviours as a factor in quality assessment remains untested and underexplored. Future studies could broaden the scope to include these aspects, potentially leading to more comprehensive behaviour-oriented quality control strategies.

2.5 Simulation Frameworks

On the MTurk platform, requesters who publish microtasks often lack effective tools to analyse the impact of workers' use of HIT catchers. The process of multiple workers simultaneously executing microtasks is complex: in addition to multiple workers competing to receive a limited number of microtasks, workers also face complex situations such as reserving tasks but failing to submit them in a timely manner, leading to expired tasks being withdrawn and re-available for other workers. Ultimately, the time and quality of the results are influenced accordingly. However, how these factors affect the completion of HITs through HIT catchers by workers have not been studied from a quantitative approach (Fernández-Macías & Bisello, 2020; Williams et al., 2019).

Existing research uses Live, Virtual and Constructive (LVC) methods to simulate various network attack behaviours and test the performance of network systems (Bergin, 2015; Damodaran & Couretas, 2015; Varshney et al., 2011). In contrast to real, virtual refers to the simulation of real individuals or processes by programming. Specifically, ‘live simulation’ refers to real individuals interacting with real networked computers. In contrast, ‘virtual simulation’ involves virtual participants or network devices. This type of simulation includes real individuals interacting with simulated networks, or virtual individuals interacting with real networks. The last type is ‘constructive simulation’. In this type of simulation, the participants and network devices are both virtual, so the interaction between the two is completely virtual as well (Kavak et al., 2021).

Under the context of crowdwork, simulations can also be designed based on the three categories mentioned above. Researchers can launch real HITs on MTurk to conduct on-site simulations, or they can create a virtual microtask working environment and invite real participants to complete tasks, thus conducting virtual simulations. Additionally, constructive simulations can be conducted by constructing virtual workers and virtual HITs through programming to simulate the task completion process.

In the review of studies about worker behaviours, many of them are categorised as live simulations. This is because these researchers posted real HITs on MTurk and collected real behavioural data from participants. When the data collection method and participant recruitment method are free of significant bias, the data from live simulation is often the most authentic and reliable. However, its cost is also higher than other simulation methods as such method involves real financial rewards for crowdworkers to participate in the study. Moreover, due to its high cost and a large number of uncontrollable factors in real life, live simulation is not suitable for conducting multiple repeated experiments with the aim of exploring the correlation between factors.

Here is an example of virtual simulation: Fan et al. (2020) proposed a novel crowdsourcing reward distribution model that involves grouping workers into a collaborative team to share risks. To investigate the impact of different levels of information transparency and reward allocation models on workers' task completion behaviour and result quality, Fan et al. (2020) established a crowdsourcing platform called CrowdCO-OP, similar to TurkPrime (Litman et al., 2017), and conducted multiple simulations with controlled task types, reward distribution types, and reward information presented to workers through a user interface. Compared to constructive simulation, this simulation method allows for precise behavioural information by observing real workers, including the number of tasks completed, completion time, and result accuracy for each participant. However, similar to live simulation, there is a high experimental cost due to the need to build a simulated work environment and recruit participants. Additionally, it is not easy to expand the scale of the experiment and look for correlations by manipulating different variables.

Regarding the constructive simulation, Saremi et al. (2021) simulated the process of task completion using DES. The simulation included events such as workers arriving in the simulation environment, accepting tasks, submitting them, and ultimately passing or failing them. The quality score of a task is generated by assigning a random number (Saremi et al., 2021). It is worth noting that this study has added features for tasks and workers in the simulation, such as a unique task ID, arrival time, duration, task status, and a worker's reliability coefficient and number of victories. These features help to add more detailed rules to the simulation, such as higher reliability workers completing tasks with higher quality, thereby improving the simulation level. The probability of random events such as task arrival is constructed using a Poisson distribution. One significant advantage of using constructive simulation is that it can easily simulate a large number of workers and tasks, and variable parameters can be flexibly adjusted. In addition, its operating cost is much lower than simulation methods that require recruiting real participants.

Models are conceptualizations of research objectives, and constructive simulation expresses this model as an observable and understandable system (Turnitsa et al., 2010). In the process of simulating the research objective, the entities, operational processes, and complex interactions within the system are represented through programming using parameters and functions, and are used to answer research questions. Among various constructive simulation methods, discrete event simulation (DES) and agent-based simulation (ABS) are commonly used to simulate human behaviours (Brailsford et al., 2006; Siebers et al., 2014). Moreover, system dynamic simulation and hybrid simulation are also widely applied for macro-level and multi-layer simulation. In the upcoming sections, four main types of simulation framework have been extensively discussed, including their definitions, the logic of constructing models, the applicable research background, and the differences from other types of simulation methods (Table 2.2).

Table 2.2 A comparison of four simulation types

Simulation Type	Advantages	Disadvantages	Related Studies	Potential Use within Crowdwork Context
Discrete event simulation	The ability to model complex routing and sequencing rules, as well as many randomly occurring events (Ponis et al., 2013), resulting in an ordered queue of events (Siebers et al., 2014). Different variables can be controlled and manipulated to assess the impact on system performance.	Not good at simulating multipple individuals with autonomous behaviours	Carmen et al., 2015; Coppock, 2019; Smith and Srinivas, 2019; Van Lier et al., 2016	Simulating the behaviour of the MTurk server in managing the scheduling of HITs at scale. Exploring their correlation with worker behaviour, availability of HITs at different stages of the experiment, etc. by varying variables such as time allotted for each HIT and the size of HIT queue for each worker.
Agent based simulation	It allows to define the autonomous and interactional behaviour of each agent, observing interactions between individuals at a microscopic level (Baptista & Neves-Silva, 2021) (Herrera et al., 2020). The decisions and behaviours of individuals in a target system can be well defined.	Computational complexity is higher than DES. The results are also sensitive to parameter settings	Wojtusiak et al., 2012; Wagner and Agrawal, 2014; Bouarfa et al., 2013	By defining the autonomous behaviour of each individual worker, the competition between workers for microtask resources and their interactions with the MTurk server can be simulated.
System dynamic simulation	Capable of representing the causality of events in a system (e.g. dynamic regulation of temperature) (Majid, 2011). Suitable for macro-level simulations.	Inability to model real-life problems in detail at the entity level (Wakeland et al., 2004) SDS is poor at modelling detailed resource allocation problems and optimisation or direct prediction (Brailsford & Hilton, 2001)	Davahli et al., 2020; Duggan, 2016; Suryani et al., 2020	Modelling the microtask allocation process at a macro level.
Hybrid simulation	The ability to combine the benefits of different simulation types.	More difficult to build than other single types of simulations, as it requires the integration of multiple simulation frameworks.	Aringhieri, 2010; Djanatliev and German, 2013; Saremi et al., 2021	SDS can be used at the macro level to model the task allocation process and ABS can be used at the micro level to model the decision-making process of individual workers. Another idea is to use ABS for the construction of autonomous behaviour of workers. Also, the state change process of each individual HIT can be constructed in a top-down manner through DES.

2.5.1 Discrete Event Simulation

DES is a dynamic, stochastic, and discrete simulation technique (Banks et al., 2005). In discrete event simulation (DES), the behaviour of a system is modelled as a series of discrete events that occur over time (Ponis et al., 2013). One of the advantages of DES compared to other types of simulation models is its flexibility. In other words, DES can model systems with multiple entities and resources. These systems often have complex routing and sequencing rules, as well as many randomly occurring events.

DES has been applied in several fields, including resource scheduling in healthcare systems such as emergency department (Carmen et al., 2015; Maaß et al., 2020), and studying the generalisability of treatment effects (Coppock, 2019). Furthermore, in the logistics domain, DES was used to model the storage of goods and the sorting process (Smith & Srinivas, 2019), or the internal collaboration between multiple distribution centres (Van Lier et al., 2016). DES replicates the complex sorting rules in the preceding scenarios, assisting the researcher in identifying problems and making informed decisions. The MTurk platform, which is the focus of this study, is a complex system that includes many worker entities and microtask resources. The process of workers accepting and completing HITs involves complex sequencing rules and a lot of randomness. Therefore, from a flexibility perspective, using DES to simulate MTurk is an appropriate choice.

Furthermore, technically speaking, the system is centralised in DES (Majid, 2011). One of the advantages of using DES over other simulation techniques such as system dynamic simulation (SDS) or ABS is that it models the system as an ordered queue of events (Siebers et al., 2014). Therefore, DES can simulate the large-scale management and scheduling behaviour of MTurk server for HITs. Specifically, after each HIT is accepted by a worker, a series of status changes will automatically take place, such as expiring from the worker's queue, then being taken back by the MTurk server, and then available to other workers again after the cooling down period. In other words, specific status changes of each HIT occur at discrete points in time.

Meanwhile, in DES models, researchers can control and manipulate different variables to assess the impact on system performance. This enables researchers to test different scenarios and discover correlations between outcomes and variables to make decisions that optimise the system. In research on the MTurk platform, the correlation between the variables set in DES and worker behaviour or the availability of HITs in different experimental stages. Such independent variables include time allocated for each HIT and the size of HIT queue for each worker.

Both continuous simulation and DES are suitable for simulating stochastic dynamic models. However, the variables in continuous simulation are constantly changing with time. In other words, continuous simulation takes into account the effect of time on variables such as chemical reactions. In comparison, DES is more applicable to situations where variables change during events (Özgün & Barlas, 2009). In addition, DES takes into account the impact of events on the system and the interactions between events, making it more suitable for simulating server systems, queuing systems, or goods dispatch systems. Because the process of managing HITs by the MTurk server involves discrete events and resources, it is more appropriate to use DES in this study.

In summary, the flexibility of Discrete Event Simulation (DES) in constructing complex systems, the ease of tuning parameters, and the scalability in experimental size make it a good method for studying worker behaviour in MTurk platforms. It allows researchers to understand changes in worker behaviour and the impact on the MTurk platform system under different parameters in a controlled and reproducible manner.

2.5.2 Agent Based Simulation

Agent based modelling and simulation are used to model complex real-world systems consisting of autonomous and interactive individual agents (Baptista & Neves-Silva, 2021). Each agent represents an individual in the real world and has a set of characteristics and rules that they follow, enabling them to make decisions and interact with each other and their surroundings. This approach allows the researcher to observe interactions between individuals at a micro level in a simulation. In other words, each agent has its own thread of execution. Thus, the system is decentralised and built from the bottom up. The researcher defines the autonomous and interactive behaviour of agents at the individual level. Ultimately, ABS generates macro-systemic phenomena from the interactions between agents.

Axtell (2000) explained a number of reasons for using agent-based simulations (ABS). These include: the decisions and behaviours of individuals in the system can be well defined; secondly, agents can reflect the way individuals behave; and furthermore, the process of growth and change is dynamic and cannot be accurately predicted. Given the above reasons, in my study, to evaluate the impact of the work behaviour of those with and without HIT catchers on the worker group and microtasks in the MTurk platform, the decisions and behaviours of different types of workers can be defined based on existing research. Secondly, the task acceptance and completion behaviour of each worker can be constructed into agents. Furthermore, the number of tasks accepted per worker, the number completed, and the completion progress of the entire HIT group are dynamic and difficult to predict accurately. In summary, it is feasible to simulate the crowdworker behaviour through ABS.

The definition of agent has gained consensus based on previous research, which includes autonomy, interactivity, and identity uniqueness (Herrera et al., 2020; Jennings, 2000; Macal & North, 2009). Next, the three perspectives of identity characteristics, behavioural characteristics and behavioural motivations are explained in detail.

In terms of identity characteristics, agents are modular and independent. An agent is a discrete entity with a unique identity and has unique behavioural and decision-making capabilities. The requirement of discreteness for agents means that agents are heterogeneous and identifiable in terms of behavioural characteristics, individual parameters. In the context of my study, each agent representing a worker needs to have a unique id and be able to have unique behavioural capabilities depending on whether it uses HIT catchers or not. More importantly, each agent can make behavioural decisions based on individual characteristics such as their work progress, available space in the HIT queue, and technical ability.

In terms of behavioural characteristics, agents need to interact with other agents as well as with their environment. Therefore, the rules and functions by which agents interact with others or the environment need to be defined. In the MTurk context, agents representing crowdworkers need to compete with each other for the limited HIT resources and submit the results on time. This process involves competition between agents, and their interaction with the MTurk server, the body of the environment. Therefore, the rules for agents to compete for resources and the rules for changing the state of HITs in the server need to be clearly defined.

In terms of behavioural motivation, the behaviours performed by agents are autonomous and self-centred. In addition to this, agents may also be goal-driven, meaning that their behaviour is driven by a specific purpose and adjusts their behavioural habits according to changes in goals and whether they are achieved. More advanced agents can also learn and adapt based on their experiences, for example through machine learning (Wojtusiak et al., 2012). However, because modelling is usually more concerned with individual behaviour under established rules, making agents adaptive is not usually the main purpose of modelling (Macal & North, 2009).

In addition, a significant advantage of ABS is the ability to simulate emergent phenomena and allow researchers to make observations. For example, studies of crowd evacuation (Wagner & Agrawal, 2014) and air traffic (Bouarfa et al., 2013) have tended to use ABS because they both involve emergent phenomena. In my study, similar emergencies also existed, for example, a small number of workers may have accepted all the HITs within a HIT group, resulting in other workers losing all available HITs for a short period of time and thus having to be idle.

2.5.3 System Dynamic Simulation

System dynamic simulation (SDS) is used to understand the dynamic behaviour of complex systems over time at the aggregate level. It is used as a strategic planning tool for macro-level research objectives such as population health and development of ecosystems (Duggan, 2016; Majid, 2011). Compared to the previous types of simulations, system dynamics simulation emphasizes feedback loops and interdependencies between different components (Sumari et al., 2013).

SDS is constructed based on the causal relationships of events in a system. This simulation method describes the behaviour of the system as a number of interacting stock, flow and feedback loops in a causal loop diagram (Cordier et al., 2017; Mustafee et al., 2010). In the example of the home heating system shown in Figure 2.2, the system first sets a temperature target. The stock level, which represents the heat inside the room, helps the system to decide how much more heat should be added into the room to reach the aim temperature. This flow of adding heat and another flow of heat loss both change the amount of heat (stock level) inside this room. Once the aim temperature has been reached, the flow stops, and no more heat is added into the room. This mechanism of adjusting flow of heat based on room temperature is called feedback loop. The letter B in Figure represents the balance of stock level maintained by this feedback loop (Duggan, 2016).

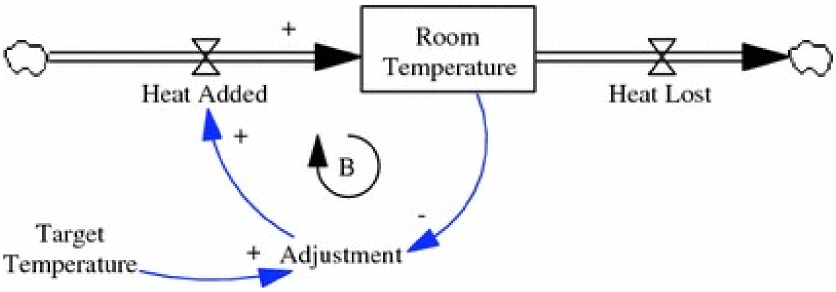

Figure 2.2 Home heating system as a sample SDS (Duggan, 2016)

In other words, a feedback loop is a closed chain of causality, starting from a stock level, through a set of criteria based on the stock level, to determine whether a flow should be generated to change the stock and thus meet the criteria. A positive feedback loop tends to reinforce or amplify the behaviour of a system. For example, rewarding an employee based on performance will result in better performance by that employee. In contrast, negative feedback loops tend to reduce or resist changes in stock level within the system. The home temperature system mentioned above includes a negative feedback loop. Another example is that one's health symptoms would go away after taking appropriate medication (Majid, 2011).

However, one of the limitations of SDS is the inability to model real-life problems in detail at the individual level (Wakeland et al., 2004). This in turn has led to SDS being less effective in modelling detailed resource allocation issues than DES (Brailsford & Hilton, 2001).

2.5.4 Hybrid Simulation

An increasing amount of research has been carried out to compensate for the limitations of a single simulation approach by combining multiple methods to model complex real-world systems. To further understand how appropriate simulation methods should be applied in specific research contexts, this section reviews several studies that used hybrid simulation methods and explains the reasons for applying particular methods in each study. Furthermore, how they have applied and combined these simulation methods are also discussed. Ultimately, a simulation modelling approach suitable for the crowdwork context is developed based on the literature review.

Hybrid simulation is a simulation approach that combines different modelling methods, such as discrete event simulation, system dynamics simulation, and agent-based simulation. Hybrid simulation can take different forms, combining datasets from different sources and rules by developing appropriate architectures and coupling strategies (Barbosa & Azevedo, 2017). Aringhieri (2010) developed a hybrid simulation model by modelling the emergency medical service centre workflow using DES while the ambulance and call centre behaviour using ABS. In the study by Djanatliev and German (2013), the system dynamic model was used to dynamically generate agents to simulate patients. The process of diagnosis and treatment of patients in hospitals was modelled through process-oriented discrete event simulation.

Saremi et al. (2021) modelled the operational processes of the Crowdsourced Software Development (CSD) platform. Specifically, the correlations between task arrival, worker reliability and task allocation were simulated at the macro level via SD. Subsequently, the life cycle of each task was simulated using DES. This simulation includes events such as worker arrival in the simulated environment, acceptance of the task, submission, and pass or fail of the final result. At the micro level, the decision-making process of the worker is simulated via ABS. This includes the registration of tasks based on individual profiles, task submission and whether their submissions win or not.

It is worth noting that the study modelled the behavioural decisions of agents through random numbers, such as the arrival of workers through a Poisson distribution, and the triggering of registration events of workers for tasks based on dynamic registration probabilities. Subsequently, there is often competition from multiple workers when registering for a task, the probability of an individual successfully registering for a task follows a Bernoulli distribution. Compared with my study, Saremi et al. (2021) modelled the task lifecycle process differently from the HITs posted on MTurk. A HIT that is abandoned by a worker is then reassigned to other workers under server scheduling and continues to flow through the marketplace until it is submitted or deleted. The simulation of HITs on MTurk involves a more complex process and the measurement of more system parameters.

2.5.5 Conclusion: A comparison of simulation approaches

To summarise, DES could model the interaction behaviour between a system and many individuals (Zhang, 2018). In contrast, ABS could simulate independent interactions between individuals and the environment or other individuals at the micro level, and is able to fully reflect the heterogeneity among individuals. However, the modelling of a large number of individuals results in higher computational costs. SDS allows researchers to observe system behaviour at the macro level (Davahli et al., 2020), but may not be as flexible as DES in describing causal relationships between model elements (Suprunenko, 2021), and is also unable to simulate microscopic interactions between individuals as ABS. In contrast to the above simulation approaches, hybrid simulation highlights the transition from modelling individual behaviours and interactions to observing broader system dynamics, thereby providing a comprehensive perspective.

However, current simulations may oversimplify worker decision-making processes and task dynamics, leading to gaps in understanding the diverse task acceptance and completion strategies. Future research should reveal more realistic worker behaviour patterns. Therefore, more realistic simulation models that can capture the diverse nature of crowdwork environments could be built.

2.6 Knowledge Sharing in Crowd Community

A virtual community is an online social organisation where members share information through communication to learn from each other or solve problems collaboratively (Chou, 2020). Members from all over the world generate the desire to join virtual communities and interact with other members based on common interests, goals, interests, etc. In turn, members gradually form social and emotional ties with each other as they interact with peers in the community (Lenart-Gansiniec, 2017). This in turn reinforces one's sense of identity as a member of the community and maintains active participation in community activities.

As mentioned in Chapter 1, the crowdworkers who perform microtasks to earn commissions face numerous challenges: poorly designed tasks, technical errors, and interface design errors made by job seekers (McInnis et al., 2016) can confuse workers, lead to extra time, effort or even submission failure or increased risk of rejection (Gadiraju, Yang, et al., 2017). Even if they are eventually paid, workers may have spent extra unpaid time and energy searching for preferred, quality work, learning how to do work they are unfamiliar with, and waiting for the requester to respond to their questions. All of these issues can lead to low productivity and result in low hourly wages based on efficiency wage theory (Gumata & Ndou, 2017; Katz, 1986).

Virtual Communities for Crowdworkers

As a result of these unfair labour practices, group workers have formed different virtual communities of exchange to defend their rights. Members from all over the world create a willingness to join virtual communities and interact with other members based on common interests, goals, interests, etc. (Osterbrink & Alpar, 2021). In turn, members gradually form social and emotional ties with each other as they interact with peers in the virtual community (Lenart-Gansiniec, 2017). Knowledge sharing is the act of transferring valuable content, where individuals spread the knowledge, experience and skills they have acquired to others (Zhang et al., 2017). Research on digital workers, among others, suggests that KS can positively impact worker well-being and performance by building trust among workers, passing on quality job opportunities (task information and new digital work platforms), mentoring to help others with micro-tasks, and providing moral support (Gray et al., 2016; LaPlante & Silberman, 2016). In the field of crowdsourcing, knowledge shared among workers includes what aids to use and how to use them, what requesters to look for and avoid when accepting a task, tips on performing a specific microtask, etc. (Gray et al., 2016).

The KS behaviour in turn reinforces one's sense of identity as a member of the community and maintains active participation in community activities (Lenart-Gansiniec, 2017). Unlike face-to-face interaction, communication between members of virtual communities is mostly through text, images, etc., and does not need to take place in real time, nor does it require a real identity. Moreover, thanks to the archiving of communication content on the community platform, communication from multiple parties does not need to occur simultaneously. Therefore, a virtual community with stable technical support can attract a large number of members with similar interests or goals and facilitate ongoing knowledge sharing between members (Hsu et al., 2007; Pi et al., 2013).

Knowledge Sharing

Knowledge sharing (KS), as a coordination of group behaviour, is necessary for people to create collective value, drive sustainable organisational development and gain individual competitive advantage (Kim & Park, 2017). Knowledge sharing has substantial impacts at both the organisational and individual levels, such as improving individual and organisational innovation (Al-Husseini & Elbeltagi, 2018), performance (Marouf, 2016), organisational learning (Park & Kim, 2018) and individual creativity (Lee, 2018). With the role of knowledge sharing, an organisation can become a learning organisation that can sustainably produce collective intelligence. This includes the acquisition, sharing, processing, and storage of knowledge (Shateri & Hayat, 2020). Platforms such as virtual communities cannot prosper without active sharing by knowledge contributors. The popularity of the ecosystem built up by third-party tools in crowdsourcing platforms also benefits from the active exchange of knowledge between workers (El Maarry et al., 2018). How to motivate platform members to share knowledge has been an important research issue (Hsu et al., 2007).

Knowledge sharing is also an act based on an exchange relationship in which participants have expectations of rewards such as pleasure (Xiao et al., 2017). Multiple members form the act of knowledge sharing by providing and acquiring knowledge. At the same time, they create new knowledge in the process (Lenart-Gansiniec, 2017).

Therefore, a review of research on KS occurring in VCs is presented next. This is divided into knowledge sharing behaviour within a wider virtual community, and voluntary KS behaviour among crowdworkers. It is worth noting that this study focuses on crowdworkers' voluntary knowledge sharing behaviours. In contrast, some microtasks require participants to share knowledge about skills and personal information for reward (Oelen, 2022), and this type of task-request oriented knowledge sharing behaviour is not focused on in this study.

2.6.1 Knowledge Sharing in Virtual Communities

The emergence of social media has changed traditional forms of knowledge dissemination, making it easier and faster (Alamir & Navimipour, 2016). Virtual communities are a popular form of social media for companies, unions, interest groups and other organisations (Lai et al., 2018). Members of online virtual communities exchange information and share knowledge in a new way through the internet. Such communities have a wider reach than traditional offline communities that require face-to-face interaction (Vahdat et al., 2020). In addition, the efficiency of group interaction is greatly enhanced by the removal of time and location constraints (Tang & Yang, 2005).

In contrast to knowledge sharing in traditional environments, in virtual environments, especially in virtual communities, knowledge sharing process is extremely dependent on the communication platform or technology (Oanţă, 2020). The technological strength of the virtual community therefore greatly influences the experience of members in the interactive act of sharing knowledge, which in turn influences their willingness to share and ultimately their behaviour.

2.6.1.1 Motivations for Sharing Knowledge in Virtual Communities

The existing literature examines individuals' motivation to engage in knowledge sharing in virtual communities from both extrinsic and intrinsic perspectives (Lai et al., 2018). Intrinsic motivation involves the psychological and spiritual aspects of people's satisfaction in participating in the activity. Specifically, self-efficacy (Glassman et al., 2021), trust (Tseng et al., 2019), enjoyment (Maharani, 2017), altruism (Lai et al., 2018) have all been found to be intrinsic motivational factors that influence members of online communities to share their knowledge. While extrinsic motivation comes from individuals' expected rewards from the outside, such as reputation, reciprocity, commitment (Deng & Guo, 2018, 2018; Fang & Zhang, 2019; Luo et al., 2021; Maharani, 2017). Interestingly, monetary rewards have been found by several studies not to be a positive influence on knowledge sharing behaviour (Fang & Zhang, 2019; Maharani, 2017).

From a socio-economic perspective, individual behaviours such as sharing knowledge are motivated by what is in their best interests (Nguyen et al., 2019). People are more likely to engage in knowledge sharing activities when extrinsic motivations in the form of tangible rewards exist (Fait & Sakka, 2021). Another important extrinsic motivation, reciprocity, has also been shown to be one of the main motivations for knowledge sharing (Nguyen et al., 2022). When a member gathers valuable knowledge from a knowledge contributor, that member also needs to share the knowledge he or she possesses in exchange for reciprocity and further encourages more members to participate in knowledge sharing.

It has been shown that perceived self-efficacy and perceived self-pleasure are two important intrinsic motivators for knowledge sharing (Nguyen et al., 2019). Perceived self-efficacy refers to individuals being confident enough to complete a task, which motivates them to be more willing to perform the task (Lai & Chen, 2014). Similarly, individuals with high levels of knowledge self-efficacy have strong self-motivation and are therefore more willing to share their knowledge (Ergün & Avcı, 2018). Furthermore, perceived self-pleasure refers to the pleasure individuals derive purely from the act of helping others, rather than expecting anything in return (Tønnessen et al., 2021). With this motivation, workers develop a willingness to share their knowledge and thus gain satisfaction and enjoyment. Unlike virtual teams within organisations, the crowd knowledge sharing is not about collaborating on one project. In comparison, they share knowledge to help peers better understand a particular HIT or requester, to get good job opportunity and to learn new skills.

2.6.1.2 Roles in Knowledge Sharing

Whereas the motivations are found not to have a definite level of influence on the KS behaviour, there are other factors that moderate their potential effect, such as the type of individual. As illustrated in Table 2.3, members involved in knowledge sharing include lurkers who only view knowledge, askers that raise questions and answers (posters) who share knowledge (Fang & Zhang, 2019; Hung et al., 2015), while those who share knowledge also include experts and general groups (Zhang et al., 2017). Specifically, the same motivations, such as the pleasure of helping others and the self-efficacy of knowledge, result in different effects on lurkers and contributors due to the limitations of individual experiences, e.g., lurkers have not felt the pleasure of helping others (Fang & Zhang, 2019). Furthermore, knowledge sharing by posters who regularly share knowledge requires sufficient interpersonal trust, whereas knowledge sharing by lurkers requires the influence of peers, reciprocity, and the perceived ease of use of the sharing medium (Hung et al., 2015; Lai & Chen, 2014). Another factor that may play a moderating role could be the different types of knowledge shared, which includes links to tasks, techniques and instructions for completing tasks (Di Gangi et al., 2022; Gray et al., 2016), evaluations or work experiences (Brawley & Pury, 2016) about the task and the requester (Osterbrink & Alpar, 2021), etc.

Table 2.3 Comparative Overview of Roles, Motivations, and Influencing Factors in Knowledge Sharing Activities.

Role	Description	Motivation	References
Lurker	Only browsing knowledge without actively participating in sharing or asking questions. Lurkers make up a larger proportion of online knowledge sharing participants.	Peer influence; Perceived ease of use of communication tools	(Fang & Zhang, 2019; S.-Y. Hung et al., 2015; Kang, 2022; Lai & Chen, 2014; M. Nguyen et al., 2023)
Asker	Ask a question seeking specific knowledge or a solution. Knowledge sharing is often triggered by Askers.	Seeking knowledge	
Poster	Answer questions and share knowledge and experience, which may include experts with	Happiness of helping others; Self-efficacy of knowledge;	

	specialised knowledge and generalist groups with limited personal experience.	Need for sufficient interpersonal trust; Reciprocity

2.6.2 Theories on Knowledge Sharing

In addition, much of the current research emphasises the study of social exchange factors, which are difficult to help researchers develop a comprehensive understanding of complex behaviour such as knowledge sharing. Existing studies have conceptualised the above mainly through Technology Acceptance Model (Assegaff et al., 2011), Theory of Reasoned Action (Almuqrin, 2022), Theory of Planned Behaviour (Fang & Zhang, 2019), Social Exchange Theory (Luo et al., 2021). However, knowledge sharing behaviour in virtual communities is influenced by more than just social and psychological factors such as reciprocity. Virtual communities are also socio-technical systems (Wan et al., 2017), and the process of knowledge sharing in a community encompasses the process of behavioural attitude formation and the acceptance of community communication technologies. It is therefore necessary to incorporate theories considering technology acceptance in the study (Marikyan & Papagiannidis, 2021a).

2.6.2.1 Technology Acceptance Model (TAM)

According to TAM, the process of technology acceptance involves three stages: external factors such as system design features trigger an individual's assessment of its perceived ease of use and perceived usefulness, leading to an effective response, including intention to use the technology, and ultimately influencing the behaviour of using the technology (Wicaksono & Maharani, 2020). Specifically, perceived usefulness refers to users' beliefs about how a technology can improve their overall performance including productivity and efficiency for a particular job. It also encompasses users' subjective views on how the technology can help them solve a problem. One’s evaluation of perceived usefulness could be influenced by their past experience with similar technologies, their perceived ability to learn and use the technology, and the technology's compatibility with their existing workflows.

Perceived ease of use, on the other hand, emphasises the effort/cost they put into performing the behaviour (Chen et al., 2011; Ibrahim & Shiring, 2022). For crowdworkers, these costs include the effort of registering with the virtual community, the time and effort spent finding the corresponding forum topic, the effort of sorting out their reflections on their experiences into a text, etc. From the perspective of technology acceptance, online knowledge sharing behaviour among crowdworkers can be seen as a process of individual adaptation and dependence on the sharing technology provided by the mediating platform. The theory therefore better explains the determinants of the acceptance of technologies that support knowledge sharing by crowdworkers.

2.6.2.2 Theory of Reasoned Action (TRA)

Theory of Reasoned Action (TRA) explains and predicts human behaviour based on an individual's attitudes, subjective norms, and behavioural intentions. The theory suggests that people's behaviour is influenced by their beliefs about the outcome of the behaviour and the subjective norms or social pressures associated with that behaviour. TRA has been applied in studies including smartphone use (Farhi et al., 2023), tax compliance (Hanum et al., 2020) and social marketing for health promotion (Rybina & Garkavenko, 2020).

TRA does not assume that humans use the information they have rationally and systematically (Hartanti et al., 2021). Instead, the theory suggests that the specific beliefs people hold about a behaviour rationally generate the intention to perform that behaviour (Procter et al., 2019). Such beliefs may arise from social pressures and subjective norms.

2.6.2.3 Theory of Planned Behaviour (TPB)

Theory of Planned Behaviour (TPB) is an explanatory model of behavioural intentions, which TPB assumes that individuals' behavioural intentions are controlled by attitudes, subject norms (SN) and perceived behaviour control (PBC) (Bosnjak et al., 2020). Specifically, attitudes represent subjective preference towards behaviour. Subject norms relate to the extent to which social pressure is perceived by an individual when conducting the behaviour. Perceived behavioural control is defined as one’s perceived ease of conducting a behaviour (Hagger & Hamilton, 2023).

Previous studies have examined the relationship between factors in the knowledge sharing domain based on TPB, such as Chennamaneni et al. (2012) who decomposed the three TPB belief constructs to identify underlying factors and examined the direct influence of PBC on KSB. In Ramayah et al.'s (2013) study, a sense of self-worth was found to influence SN factor, while SN had a direct effect on both Attitude towards Knowledge Sharing and KSB. In addition, PBC and Organizational Climate were also found to have a direct effect on knowledge sharing behaviour.

However, discrepancies in the findings of previous studies have resulted in the relationship between factors in TPB not being clarified. In addition, the TPB framework does not incorporate other important factors, including efforts of using the tools.

2.6.2.4 Innovation Diffusion Theory (IDT)

Diffusion of innovation theory was developed by EM Rogers in 1962 (Rogers, 2010). It originated from interpersonal communication to explain how an idea, behaviour or product gains momentum over time and spreads through a specific group of people or social system. The act of online knowledge sharing is disseminated by an independent individual to all members through a virtual community. As other members receive knowledge, it is first disseminated to themselves, who then make decisions about their own will and emotions. For example, individuals may choose to share this knowledge with more people because they benefit from the knowledge shared by others. The influence of these personal and social factors ultimately drives new individuals to develop knowledge sharing behaviour and to innovate on this behaviour, resulting in new ways of sharing knowledge, for example through private instant messaging channels (Slack, Telegram) or plugins (TurkerView). The potential result of this diffusion, according to this study, is that more and more crowdworkers embrace and begin to engage in knowledge sharing behaviour, driving the evolution and innovation of this behaviour.

2.6.2.5 UTAUT - an Integrated Approach

Although each theory provides a unique understanding of the target behaviour, they each have their own limitations, and thus reliance on a single theory limits our overall understanding of complex behavioural scenarios. Moreover, facilitating conditions influence behaviour directly rather than through intentions (Yu et al., 2021). Venkatesh has developed a unified theory of acceptance and use of technology (UTAUT) by integrating several theories of behavioural research, including those mentioned above (Venkatesh et al., 2003). Specifically, in the UTAUT model, Performance Expectancy (PE), Effort Expectancy (EE) and Social Influence (SI) directly influence users' behavioural intentions to use the system.

Performance Expectancy represents indicators of perceived usefulness, extrinsic motivation, outcome expectation, etc. from earlier models (Marikyan & Papagiannidis, 2021a). This refers to the degree to which a person believes that using the technology will help them achieve their goals or perform better. In this study, PE can be defined as the extent to which crowdworkers believe that the use of knowledge sharing tools will enable them to achieve better performance in the sharing or acquisition of crowdwork-related knowledge. This factor is influenced by the perceived usefulness of the technology, as well as the individual's confidence in their ability to use it effectively (Onaolapo & Oyewole, 2018). In general, the higher an individual's performance expectancy, the more likely they are to adopt a new technology. In other words, if the crowdworkers believe the knowledge sharing (KS) tools improve their KS experience from the perspective of effectiveness, speed and relative advantage, they would be more willing to use them.

Effort Expectancy (EE), on the other hand, refers to the degree to which a person believes that using the technology will be easy and require minimal effort (Quadri & Garaba, 2019). It is constructed based on TAM, MPCU, IDT driven perceived ease of use and complexity metrics (Chauhan & Jaiswal, 2016; Gupta et al., 2008) The meaning of EE in crowdsourcing research is similar to the interpretation regarding perceived ease of use in TAM. Venkatesh et al. (2003) considered EE as the level of ease associated with using an information system. It includes the extent to which crowdworkers expect that using KS tools will not require physical and mental effort (Onaolapo & Oyewole, 2018). It can be assumed from the theory that crowdworkers may be more inclined to use KS tools if they realise how easy it is to share and access knowledge using communication tools.

Social Influence, SI, is similar to the subjective norms, social factors and image constructs used in TRA, TPB, CTAMTPB, MPCU, and IDT, where people's behaviour is adjusted according to how they are perceived by others. In other words, this refers to the degree to which crowdworkers' peers or the platforms they work on support or encourage the use of the knowledge sharing tools. This influence often becomes significant in cases where the use of technology is mandatory, such as when a company mandates that employees communicate internally through a certain information system (Marikyan & Papagiannidis, 2021a). However, the impact of this factor needs to be re-evaluated for non-mandatory use of technology, such as joining a crowd community in this study.

Facilitating Conditions (FC) and Behavioural Intentions (BI), on the other hand, are considered to directly influence the usage behaviour of target groups. Specifically, Facilitating Conditions (FC) are defined as the extent to which an individual perceives that the organisational and technical infrastructure exists to support the use of the system (Onaolapo & Oyewole, 2018). In other words, FC are factors that make possible the use of KS tools for crowdworkers' KS behaviours. Indicators such as perceived behavioural control and compatibility play a large role in determining FC. The extent to which crowdworkers can effectively use KS tools for knowledge sharing and acquisition depends on the availability of collective resources (such as the number of forum members), the skills and infrastructure required to implement the functionality (such as linking HITs in the list to comments from other workers via plugins). This implies that crowdworkers' belief in the availability of community resources and technical infrastructure to support the effective use of KS tools may influence whether they use them.

2.6.2.6 Social Exchange Theory (SET)

In the context of crowdsourcing, the knowledge sharing behaviour is achieved by crowdworkers through the technology of virtual communities, channels, and social apps, and therefore involves the use of technology by individuals (Kaplan et al., 2018). However, knowledge sharing itself also involves an exchange of benefits between individuals, such as gaining prestige through sharing knowledge, or simply enjoyment. In a study on willingness to share knowledge in Chinese virtual communities, based on SET, Luo et al. found that social relationships, reputation and reciprocity had significant effects on knowledge contributors' willingness to share. (Luo et al., 2021) Specifically, in the process of knowledge exchange, people tend to establish and maintain long-term relationships with others. The costs and rewards of the period become the determinants of subsequent behaviour.

According to social exchange theory, interpersonal exchange behaviour depends on reciprocal responses from others (Yoshikawa et al., 2018). The theory emphasizes the importance of resource exchange and social relationships between individuals (Stafford, 2017). As the initiators of the exchange, crowdworkers give resources, but do not necessarily receive them in return. This reward further influences the person's future knowledge sharing intention and behaviour. For other individuals who receive the knowledge shared by the initiator, they will also be influenced to varying degrees to develop the intention to share knowledge and eventually act on it (Luo et al., 2021).

2.6.3 Knowledge Sharing among Crowdworkers

Knowledge sharing among crowdworkers often involves costs including time and effort, and SET can help to understand the behavioural intentions by considering the costs and expected benefits that workers experience in their knowledge sharing behaviours (Jahan & Kim, 2021). SET also emphasises the social support and relationships that people build through exchange behaviours (Cook, 2015; Gray et al., 2016; Ihl et al., 2020; Margaryan, 2016). Therefore, in the context of knowledge sharing by crowdworkers on MTurk, Social Exchange Theory (SET) could be helpful in understanding the factors that motivate individuals to share their knowledge with others (Li, 2015; Wang et al., 2015). In addition, from the perspective of technical tools, theoretical models such as Technology Acceptance Model (TAM) or Unified Theory of Acceptance and Use of Technology (UTAUT) could be used to explain and predict the crowdworkers' use of such communication technologies (Khalid et al., 2023; Matli & Wamba, 2023; Wallace & Sheetz, 2014).

It has been found that crowdworkers commonly share work-related knowledge through social platforms such as virtual communities (VC) (Kaplan et al., 2018; Osterbrink & Alpar, 2021), which is the main way in which workers engage in knowledge acquisition and sharing. Knowledge sharing (KS) as a form of collaboration is initiated by crowdworkers to improve their work experience, such as greater efficiency and fewer unfair rejections. Among crowdworkers, knowledge sharing includes sharing comments about HITs and requesters, their work completion statistics, and even sharing task suggestions (Figure 2.3, Figure ***2.4***) (Brawley & Pury, 2016; Osterbrink & Alpar, 2021). From being treated unfairly by requesters to sharing comments and reputation scores about requesters, crowdworkers collaborate to avoid malicious requesters, gain more bargaining power, and be rewarded more fairly (Brawley & Pury, 2016; LaPlante & Silberman, 2016). The quality of task output will also be improved through more effective communication methods between workers and requesters (McInnis et al., 2016). The following (paraphrased to avoid deanonymization) posts highlight how crowdworkers seek support via the virtual communities:

"Question: Is it just me who has issues with his hits? It either says they aren't 'ready' yet or just loads indefinitely. I've never been successful in opening one of his. I've also tried different browsers."

"Answer: He has numerous HITs and accounts on Mturk, so I'm not sure which ones you're referring to. Even so, I do not do them. It's pointless to do them if they don't work for you. You could send a message to inquire."

– Crowdworker shares their confusion to avoid a malicious requester [Mturk Forum: Sep 30, 2021]

"Question: Is there something I missed? Since 2017, I've been doing MT. I limit myself to hits that I am qualified for, usually at a minimum of 35 cents. Suddenly, I'm only finding Noah Turk hits that, for some reason, never load for me, and a few other random hits. I wasn't doing many hits per day...maybe 5-15, but in the last 3 weeks or so, if I'm lucky, I might find 1 or 2 that I can do. I have gone several days in a row with no workable hits. Is anyone aware of what is going on?"

"Answer: On MTurk, this is typically the slowest time of year. Things generally slow down around the holidays and don't really pick up again until mid-late January."

– A forum member got answers by posting their question [MTurk Crowd: Jan 8, 2022]

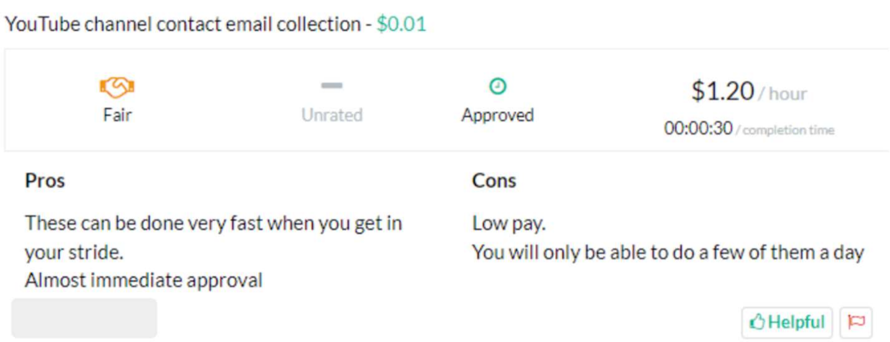

Figure 2.3 HIT feedback shared on TurkerViewJS (ChrisTurk, 2022).

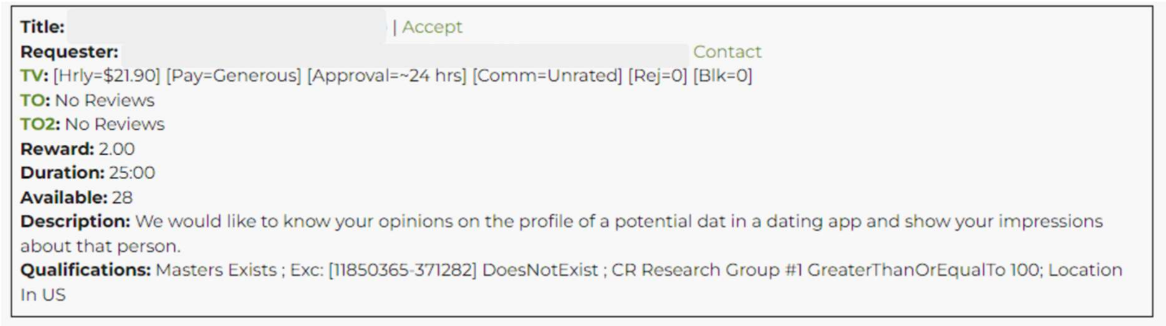

Figure 2.4 Crowd worker shares a well-paid HIT opportunity [Mturk Forum: June 29, 2022]

Previous research has found that workers access and share information about the task in many groups on Reddit and Facebook or independent websites such as MTurk Crowd, MTurk Forum and Turker Nation to share and discuss the HIT with other workers (LaPlante & Silberman, 2016). A review of representative forums and knowledge-sharing tools is presented next, including the types of knowledge included and the way they are shared.

2.6.3.1 Forums and Tools for Knowledge Sharing

Turkopticon provides aggregated rating and feedback for job requesters and HITs. This reputation mechanism allows workers to choose tasks with better rewards and from more 'reputable' job requesters. This also means, however, that job requesters with low reputation scores cannot access high-quality workers and that false or abusive comments cannot be effectively avoided.

Crowd-Workers (Callison-Burch, 2014) provides workers with a quantitative evaluation of the requester through sharing workers' task completion records, therefore, to calculate the hourly rate, payment time, rejection rate and reasons for each requester. Compared with the qualitative evaluation on requesters from Turkopticon (Irani & Silberman, 2013), Crowd-Workers makes the evaluation more objective and measurable with more quantitative information.

TurkerNation (Zyskowski & Milland, 2018) is an online forum that allows crowdworkers to discuss HITs, requesters and even daily life in the community. This community allows workers to socialise in the chatroom and tries to organise the workers into groups for task information sharing. Such online forums encourage the communication between requesters and workers as groups to improve the efficiency and influence of communication. Moreover, workers get a sense of social belonging and self-identity.

Unlike the forums above, TurkerView provides users with a convenient and efficient way to share task-oriented information based on community information sharing by embedding into the MTurk HIT pages in the form of a plugin (ChrisTurk, 2022).

Furthermore, it was also stated in Williams et al.'s (2019) study that crowdworkers would also communicate with their subordinate private teams through messaging applications such as Discord and Slack. Through joining these private teams, their motivation to work with teammates are also boosted.

In general, while crowdworkers use plugins such as Turkopticon and TurkerView to quickly access the reviews about tasks or requesters, they lack the same level of trust in information easily obtained from plugins compared to information shared in forums or private channels. This is because the information lacks vouchers from acquaintances and workers trust information obtained through personal effort more (Gray et al., 2016; LaPlante & Silberman, 2016). The forums or sharing tools mentioned above were summarised according to the type of knowledge shared, as shown in Table 2.4.

Table 2.4 A summary of types of knowledge shared for each of the forums / sharing tools mentioned.

Name of tool / forum	Ratings of HITs and requesters	Feedback of HITs and requesters	HIT completion records (reward, reason for rejection, etc.)	Great HITs and other income opportunities	Knowledge for scripting tools	General news & social discussion
Turkopticon	✓	✓				
TurkerView	✓	✓	✓			
Crowd-workers			✓			
Turker Nation		✓		✓	✓	✓
MTurk Crowd		✓		✓	✓	✓
MTurk Forum		✓		✓	✓	✓

Based on the current phenomenon of workers seeking answers from others or sharing their knowledge of tasks through third-party forums and other channels it is evident that: allowing workers to share knowledge with peers about HITs more effectively while ensuring sufficient credibility will potentially alleviate conflicts between requesters and workers caused by poor communication (Callison-Burch, 2014; Irani & Silberman, 2013; Zyskowski & Milland, 2018).

These forums and tools are information systems that contain user identities, content categorisation, message posting and replying, user reputation systems, search functions, and community maintenance modules. Crowdworkers could share and access different types of knowledge through these information systems. Each information system has a unique knowledge management architecture and usage rules, such as different ways of categorising content and sharing knowledge. Each of these unique designs impacts the worker's experience of sharing knowledge (ChrisTurk, 2022; LaPlante & Silberman, 2016). In this thesis, such forums and tools are generically referred to as knowledge sharing tools (KS tools).

2.6.4 Knowledge Sharing for Task Completion

Knowledge sharing is not only intended to enhance one's own working conditions, but is also an indispensable means of accomplishing specific tasks. For example, simple tasks in citizen science often combine the knowledge of many voluntary knowledge contributors into a complete knowledge output. The knowledge could be obtained from observation, measurement or categorisation from each individual (Crowston et al., 2018; Dunn & Hedges, 2014; Ponciano & Brasileiro, 2015).

In complex citizen science, participants face more difficult collaborations, such as manuscript transcription translations or encyclopaedic knowledge contributions. This is because participants need to compare their knowledge with that of previous authors and add to or modify their contributions (Ferran-Ferrer, 2015; Yang, 2021). Similarly challenging tasks include crowd software development and text editing (Bernstein et al., 2015; LaToza et al., 2015). Such knowledge sharing behaviours are mainly initiated by the job requester or the platform on which the task is posted, with the aim of meeting the skill requirements of the project or improving the quality of the work (Kulkarni et al., 2012; LaToza et al., 2015). As these are knowledge sharing behaviours in order to fulfil the task requirements, they are not the focus of this study.

2.6.5 Previous Studies on Crowd Knowledge Sharing

In the discipline of information systems (IS), the behaviour of crowdworkers who participate in knowledge sharing through communication tools is receiving increasing attention (Ihl et al., 2020; Tang et al., 2019; Wu & Gong, 2020; Yan et al., 2021). This section provides an overview of current studies about crowdworkers' knowledge sharing behaviour, including current research progress, key factors influencing behaviour and the influence of knowledge sharing.

2.6.5.1 Types of Knowledge Being Studied

The study by Gray et al. (2016) describes through case studies how crowdworkers find and share information about tasks and requesters through face-to-face, communication software, online forums, etc. In addition, the knowledge being shared that current studies focus on include task links, personal insights on completing specific tasks, tips and guidance, or simply social content (LaPlante & Silberman, 2016; Tang et al., 2019). Moreover, reviews or work experiences about the task and the requester are also popular types of knowledge being shared (Brawley & Pury, 2016; Osterbrink & Alpar, 2021). Advice on how to become more efficient in microtask work is also included (Di Gangi et al., 2022).

2.6.5.2 Motivations of Knowledge Sharing

Previous research has identified a range of key extrinsic motivations for crowdworkers' voluntary participation in knowledge sharing, such as a perceived sense of belonging, a belief in increased self-esteem, a belief in reciprocity, and a belief in the possibility of influencing the requester (LaPlante & Silberman, 2016; Osterbrink & Alpar, 2021). At the same time, several studies have suggested an important set of intrinsic motivations for crowdworkers' involvement, such as personal 'openness' and friendship among colleagues (Brawley & Pury, 2016), altruism in perceived satisfaction in helping other crowdworkers (LaPlante & Silberman, 2016; Osterbrink & Alpar, 2021).

Factors that negatively affect behavioural intentions have also been studied, such as perceived cost and the limited binding effect of sharing comments in forums on requesters (Sedighi et al., 2016). The loss of knowledge power does not significantly affect the intention to contribute comments (Ye & Kankanhalli, 2017). Previous research has also found that crowdworkers' own anonymity, worker fragmentation and insecure working conditions inhibit motivations such as reciprocity and self-esteem that need to be based on stable interpersonal relationships (Osterbrink & Alpar, 2021).

2.6.5.3 Influence of Knowledge Sharing

The role of knowledge sharing among crowdworkers is also evident: crowdworkers share information to build trust, improve skills and gain moral support (Ihl et al., 2020; LaPlante & Silberman, 2016). In addition, knowledge sharing helps to increase job satisfaction, which in turn reduces crowdworkers' intention to leave (Brawley & Pury, 2016). Studies have also found that the result quality could be improved by allowing crowdworkers to freely discuss specific tasks and then update their answers (Chang et al., 2017; Drapeau et al., 2016; Tang et al., 2019).

However, knowledge sharing also has a negative effect, with unregulated malicious messages leaving communities such as Turkopticon plagued by harassment, insults, sexism, and unfounded accusations (Silberman, 2015).

2.6.6 Conclusion: Expanding Horizons in Crowdworker Knowledge Sharing: Beyond Basic Reviews to Technology-Driven Insight

Existing research on knowledge sharing behaviour among crowdworkers is still in its early stages. It primarily focuses on discovering and describing such behaviour (Gray et al., 2016; LaPlante & Silberman, 2016). Systematic categorization of shared knowledge content is yet to be undertaken (Brawley & Pury, 2016). Although there have been analyses on the factors that influence knowledge contributions among crowdworkers, said analysis typically focuses on a single community and the categories of knowledge are limited to reviews about microtasks or requesters (Osterbrink & Alpar, 2021). We need to examine more types of knowledge and categorise community members in order to investigate the effect of the same motivation on the knowledge sharing behaviour of different types of members. This is because people who are good at performing different types of work may have different attitudes and habits towards knowledge sharing behaviour. As such, further exploration of this will help researchers understand the impact of different motivations on the knowledge sharing behaviour of crowdworkers, and their linkages with task types.

More importantly, there is currently very limited exploration of workers' knowledge sharing behaviours from the perspective of technology use. The review on forums and tools also found that the unique technological features of these knowledge sharing tools as information systems (the way they interact, the way they categorise content) affect the crowdworkers’ experience of using them. This, in turn, could also affect workers' willingness and behaviour of knowledge sharing. By considering the technological factors, the design of such knowledge sharing tools could be improved to facilitate workers' willingness and behaviour of knowledge sharing.

2.7 Summary

This chapter examines the current research progress in crowdsourcing, simulation, and virtual communities relevant to the research questions, identifying the challenges faced in these areas. The discussion begins with an overview of HIT catchers and research on the impact of using such tools. It then reviews worker behaviour research, focusing on detection methods, behaviour-based quality assessment methods. Subsequently, the chapter reviews research using simulation experiments, and clarifies the scenarios to apply different simulation frameworks. The final section presents a review of research on knowledge sharing among crowdworkers, including influential factors and theories applied to study knowledge sharing.

Finally, it is revealed that: (1) Although there are studies mentioning HIT catchers, they have only explored the phenomenon of use and potential impacts from a qualitative perspective. Current research still lacks a quantitative approach through empirical data to explore how it affects worker behaviour, quality of results, and the completion process of the HIT group. The long-term impact to the whole platform has also not been extensively discussed. (2) Current crowdsourcing platforms rely excessively on reputation systems, resulting in the quality of microtask results not being assessed accurately enough. (3) Current research on behaviour-based quality assessment is only explored towards specific types of microtasks and is not generalised sufficiently. In addition, task acceptance behaviours, including task backlogs, have not been included in worker behaviour studies, and the feasibility of using acceptance behaviours as a quality assessment factor has not yet been tested. (4) There is a lack of understanding of the factors affecting crowd knowledge sharing from the perspective of technology use. In Chapter 3 , we shape several research designs around these research gaps and explains how our studies bridge the current research gaps.

Chapter 3 Methodology

3.1 Introduction

The literature review revealed numerous research gaps that provide a clear direction for this study: (1) Most of studies regarding HIT catchers were conducted from a qualitative perspective and lack quantitative studies based on empirical data. (2) The potential risks arising from the over-reliance on reputation systems in crowdsourcing platforms were not sufficiently discussed. (3) Current behaviour-based quality assessment studies are limited to specific types of microtasks, while there is a relative lack of research on task acceptance behaviour. (4) Although knowledge sharing plays a key role in the crowd community, there is still limited understanding on skill-based knowledge sharing.

To address these research gaps, this thesis first provides an initial exploration of the impact of HIT catcher use through a simulation study based on empirical data, and therefore bridging Gap (1) and (2). Subsequently, the impact of HIT catcher on work behaviours, HIT group completion processes, result quality, job opportunities was further explored through an experiment using real-world microtasks to bridge Gap (1). In addition, this study extends our understanding of behaviour-based quality assessment methods around Gap (3). Finally, for Gap (4), knowledge sharing in the crowdsourcing domain was explored through a factor analysis study based on participants' subjective evaluations. This chapter also presents the ontology and epistemology that underpin the thesis.

3.2 Research Purpose

Before proceeding with the philosophical assumptions that underpin the thesis, it is important to explain the research purpose, which can be used as a guide into the next sections.

The purpose of a study can be interpretive, explanatory, exploratory or descriptive (Saunders et al., 2009; Walsham, 2006). In other words, the research purpose might be exploring or explaining a particular topic or question (Wiesenberg et al., 2020), to interpret or understand how things work

(Walsham, 2006), to test a hypothesis (Fofana et al., 2020), or to describe a phenomenon (Kaplan et al., 2018).

When a researcher wants to better understand a topic, descriptive research is often required. Descriptive research aims to characterise the target phenomenon, such as the attitudes and habits of a particular population regarding a particular behaviour (Ulfha et al., 2019). It does not seek to explain or answer questions about how or why a population or phenomenon behaves as it does. Descriptive research is often combined with exploratory research and is used to help researchers expand in an unfamiliar area and discover deeper research questions (Hunter et al., 2019).

Exploratory research is usually based around a relatively unknown area of research and requires the formulation of new hypotheses or the development of new ideas for future research (Yang, 2021). Exploratory research can be used to generate new ideas or to find new ways to approach a problem. Some of the common methods often used include surveys, interviews, focus groups and observational studies (Oyong & Ekong, 2019; Wiesenberg et al., 2020).

Interpretive research is a type of research that involves the interpretation of data to draw conclusions. It focuses on understanding and interpreting the meaning of human experience and behaviour (Walsham, 2006). Interpretive research usually uses qualitative methods such as interviews and observations to help the researcher understand a concept or phenomenon from the perspective of the person experiencing it. This type of research is also well suited to the study of phenomena that are difficult to observe directly.

This study focuses on two types of crowd collective behaviours: the wide use of HIT catchers and the sharing of skill-based knowledge. First, we aim to describe the impact of the use of HIT catchers on metrics such as HIT completion speed and data diversity by collecting empirical data. In addition, we need to explore the impact of the use of HIT catchers on crowdwork strategies and microtasks. This is an open-ended question that aims to provide insights and explanations on how the use of HIT catchers affects various aspects of the crowdsourcing ecosystem. Thus, this thesis adopts a combined descriptive and exploratory approach for HIT catcher related studies.

Regarding the sharing of skill-based knowledge, we aim to answer what factors drive crowd skills-based knowledge sharing. This is an open-ended question aimed at exploring and explaining these factors rather than just describing the phenomenon. Therefore, we adopt mainly exploratory approach for knowledge sharing study.

3.3 Philosophical Assumptions

The importance of the philosophical assumptions that identify and support research approaches has been repeatedly emphasised (Duffy & Chenail, 2009). An understanding of the philosophical assumptions can help researchers to identify the limitations of different research methods and thus choose the most appropriate strategy to address the target question (Easterby-Smith et al., 2012).

The philosophy of research includes the researcher's beliefs about the nature of reality, the role of research in discovering knowledge, and the methodological strategies that should guide us in conducting research (Saunders et al., 2009). By comparing and reflecting on research philosophies, researchers can see other possibilities that can enrich their own research capabilities. In addition, the researcher's confidence in the chosen methodology and findings is enhanced by a deeper understanding of the philosophical assumptions (Holden & Lynch, 2004).

Philosophical assumptions matter for research because they provide a framework for understanding the research question (Mingers, 2003). Philosophical assumptions can guide what and how the research can be conducted. They also guide how the output can be interpreted. Philosophical assumptions come from ontology and epistemology, and the understanding of how knowledge is produced depends on the researcher's understanding of reality. Ontology is concerned with what is reality, while epistemology guides what and how a researcher can know about reality (Bryman, 2012). Ontologies and epistemologies lead to the choice of particular research methods. This section delves into the consideration of ontology and epistemology, which underlie the adoption of positivism as both a philosophical stance and a research framework for this thesis (Bryman, 2012). It is worth noting that the study on crowdworkers' use of HIT catchers and the study on their sharing behaviour towards skill-based knowledge are two relatively separate topics, but they are strongly related in their research motivation and together they address the aim of this research. Furthermore, the research methods applied to these two topics differ due to requiring different ways of accessing knowledge.

3.3.1 Ontological Considerations

Ontology is a branch of philosophy concerned with the understanding of reality and its existence (Gray, 2021). In the Social Sciences, it is used to study the nature of social reality and to develop theories about how that reality is organised (Bryman, 2012). The central question is whether the target phenomenon of study can be treated as an objective reality that does not require additional constructs. Constructivism highlights the dynamic nature of social phenomena shaped by ongoing interactions among actors (Fruggeri, 2021). This perspective is favoured in studying phenomena of using scripting tools and sharing knowledge, relying on qualitative methods for interpretation. It could bring us contextual insights with rich qualitative data. However, compared to constructivism, objectivism could provide generalisable findings with more empirical validity using statistical analysis based on quantitative data across large population (Jonassen, 1991). Moreover, objectivism relies on the standardised methodologies that can be replicated, generating more consistent and less biased findings. Therefore, this thesis uses objectivism as the research ontology.

Objectivism emphasises that the social phenomena cannot be influenced by the researcher and constrain the behaviour of members within the social organisation. Objectivists believe that the meaning of the world exists objectively, and separate from human perception (Jonassen, 1991). We can recognise it through a scientific and objective method, which in turn can be represented by a theoretical model. In terms of ontology, the use of HIT catchers by crowdsourced workers and the sharing of knowledge by workers through different strategies are both objective phenomena and the effects they cause are objective and do not depend on the consciousness of the researcher, nor are they altered by the subjective preferences of the observer. In other words, both social phenomena are discoverable realities and exist independently of the researcher (Cohen et al., 2007; Pring, 2004).

The crowd knowledge sharing behaviour is a real-life social phenomenon whose behavioural patterns, motivations and influencing factors can be observed and measured through empirical research. When studying this collective behaviour, we are attempting to capture and understand this objective reality rather than subjective interpretations or constructions. Objectivism is therefore more appropriate for this research. The factors affecting this behaviour can be measured and analysed through questionnaires, which aim to capture objective reality and produce findings that are generalisable rather than based on the researcher's subjective interpretation. Furthermore, in objectivism perspective, crowd knowledge sharing behaviour is independent of the researcher's awareness and perception. This means that the researcher should take a neutral stance to avoid subjective bias and ensure the objectivity and reliability of the findings.

3.3.2 Epistemological Considerations

Epistemology helps researchers to understand how knowledge is created (Saunders et al., 2009). More specifically, epistemology indicates how we believe knowledge can be accessed. From the perspective of positivist epistemology, an objective reality can be studied to generate absolute knowledge. In other words, social reality exists without regard to how different people perceive and interpret it. They believe that a single, universal truth may be found through scientific investigation. The counterpart to this is interpretivism, which emphasises the need to answer research questions about the social sciences with a different research logic from that of the natural sciences, one that reveals the uniqueness of human society (Bryman, 2012).

3.3.2.1 Positivism

The belief of positivism is that knowledge can and must be developed objectively and that the views of the researchers or participants do not influence its development. The role of the researchers is limited to collecting data and interpreting measurable results, while holding an objective attitude and remaining separate from the participants in the phenomenon (Saunders et al., 2009). This characteristic matches those of the workers' HIT catchers use and knowledge sharing phenomena that are the focus of this study. The researcher can maintain an objective attitude by collecting and analysing the data as the behaviour occurs, which in turn leads to objective facts that can be generalised. Specifically, the use of HIT catchers by crowdworkers objectively influences the access to work by users and non-users, as well as the speed of completion of the overall HIT batch. These effects are real and long-standing, independent of the researcher, and the magnitude of these effects is objective. Positivism relies on deductive methods to test a priori developed hypotheses, which are usually stated quantitatively, and functional relationships can be drawn between explanatory factors (independent variables) and outcomes (dependent variables) (Park et al., 2020). In this thesis, I use positivism to measure the influences and make generalisations about the basis of my findings regarding the influences of the use of HIT catchers and the factors affecting the crowdworkers' knowledge sharing behaviour.

For positivism, the significance of the research is to test theoretical hypotheses and provide material for theory development (Bryman, 2012). Current research focuses on understanding the relationship between the factors in the constructed model and the knowledge sharing behaviour of crowdworkers, while the relationship between factors and behaviour can be considered as single objective realities that can be measured directly using quantitative methods. Therefore, quantitative methods are appropriate to address the research questions in this study.

3.3.2.2 Interpretivism

As an opposing research paradigm to positivism, interpretivism emphasises the importance of understanding, interpreting the subjective experiences and meanings of individuals and groups. Reality needs to be interpreted because it is constructed. Interpretivism focuses on understanding a person's subjective experiences, perspectives, and cultural contexts, rather than observing only objective phenomena. It reveals the complexity of the social sciences when studying human behaviours, as opposed to the positivist approach which emphasises objectivity and measurement (Crotty, 1998; Johnson & Duberley, 2000; Willis et al., 2007). For example, to understand the attitudes of crowdworkers towards knowledge sharing, it is necessary to consider the values, traditions, and community culture of the entire crowdworker community to better understand the factors that motivate or discourage them from sharing knowledge. In interpretivism, the researcher needs to interact and communicate with the participants to understand their subjective experiences, thoughts and feelings. This includes qualitative research methods such as interviews and questionnaires to obtain as much information and detail as possible to help the researcher understand their social and cultural context and to accurately interpret their behaviour and experiences.

Through interpretivism-based research methods such as interview and focus group, it is possible to learn about each participant's subjective perceptions of the impact of utilising the script and the elements impacting knowledge sharing behaviour in answer to the research questions of this thesis. However, these independent explanations cannot be generalised directly to a wide range of worker groups. In addition, the impact of the HIT catcher on the results and the extent to which different factors influence knowledge sharing behaviour cannot be objectively measured.

3.3.3 Conclusion

In conclusion, this thesis uses positivism to objectively assess the impact of HIT catcher and the factors that influence knowledge sharing behaviours, forming objective facts that can be generalised. In addition, positivism helps to test the theoretical hypotheses formed based on the literature review and provides a basis for the development of the existing theories.

3.4 Research Design

In this thesis, we embark on a comprehensive exploration of our research topic through three distinct studies. These studies are:

A Simulation Study on the Unintended Consequences of HIT Catching Tools in Crowdsourcing Platforms: This study, detailed in Chapter 4, delves into the effects of HIT catchers on MTurk. Through simulation based on empirical measurement, it investigates how the use of HIT catchers affects overall completion speed, data diversity, job opportunities of crowdworkers, and the potential consequences for the entire platform. The term "unintended consequences" refers to the unforeseen or unexpected outcomes that arise from the use of HIT catching tools on crowdsourcing platforms. These consequences include potential negative impacts on task availability, task completion speed, data diversity, and the overall health of the crowdsourcing ecosystem.

Crowdwork Strategies with the Aid of HIT Catchers: Presented in Chapter 5, this study explores the strategies workers employ when using HIT catchers. It examines how these tools influence HIT access dynamics, HIT opportunities, worker behaviours, and result quality. The study also investigates the behaviour-based quality assessment for image annotation HITs.

Factors Influencing Crowd Knowledge Sharing Behaviour: Presented in Chapter 6, this study assesses the factors that drive skill-based knowledge sharing, from the perspectives of using communication tools and social exchange.

Figure 3.1 reveals the connections between these distinct studies. Specifically, both Chapter 4 and Chapter 5 focus on the role and impact of HIT catchers in the crowdsourcing platform. The findings from Chapter 4 set the stage for understanding the broader implications of HIT catcher use, while Chapter 5 provides a deeper dive into the practical influence caused by using these tools. Due to the unequal technical advantages resulting from the knowledge gaps, which in turn affected participants' utilising HIT catchers in the experiment, we then explored the factors affecting skill-based knowledge sharing in Chapter 6. Crowd knowledge sharing is not just influencing tool adoption, but also significant in community growth, and the overall health of the crowdsourcing ecosystem (Kaplan et al., 2018; Williams et al., 2019; Yin et al., 2016).

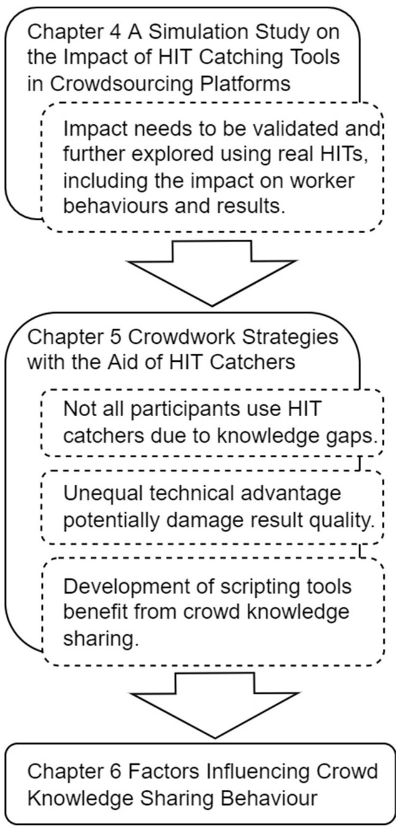

Figure 3.1 Connections within studies.

In summary, Chapters 4 and 5 provide a comprehensive understanding of the impact of HIT catchers on crowdsourcing workers and microtasks, answering the first research question (RQ1). Chapter 6 delves into the factors influencing knowledge sharing among crowdworkers, addressing the second research question (RQ2). Together, these studies offer a wide view of the collective behaviours within crowd communities, from the tools workers use to improve working conditions to knowledge sharing which makes tools popular. Each study is summarised next including the research methodologies applied.

3.4.1 Study 1: A Simulation Study on The Unintended Consequences of HIT Catchers

The aim of this study is to conduct an initial exploration of the influence of workers' use of HIT catchers on the processes and outcomes of HIT groups and workers. As highlighted by the literature review regarding the use of HIT catchers (Section 2.2.2), existing literature does not provide a clear definition of their impact, as well as a quantification of the impact. One potential reason for this could be the technical challenge of obtaining experiment data, including tests of whether workers use HIT catchers, monitoring the real-time status of HITs, and monitoring their work strategies. In addition, researchers have generally studied the HIT catchers as a small part of a microtask assistive tools, examining multiple scripting tools as a whole, thus lacking a more focused perspective at the use of the HIT catcher (El Maarry et al., 2018; Kaplan et al., 2018; Williams et al., 2019). Finally, the impact of the HIT catcher on workers is only vaguely conclusive from a qualitative standpoint, but how this is accomplished and the theory underlying the phenomenon have not been quantified.

In the first study of this thesis, to conduct a quantitative exploration of the impact of the HIT catchers use, a hybrid simulation model based on measurements of the platform was applied to simulate a proposed HIT completion scenario under different numbers of HITs, workers and proportion of two types of workers.

The methods used in this study for data collection and analysis are described in this section. More details about this exploratory simulation study and these methods are presented in Chapter 4.

3.4.1.1 Data Collection Method

The methodology of this study needs to generate data that can help expose and quantify the impacts of using HIT catchers. To achieve this, the HIT state changes between being published and being submitted were first made explicit by observing the HITs through the platform. Subsequently, multiple groups of HITs were published on MTurk Sandbox for observation experiments. During the experiments, the time spent by the HITs to enter different states after being published was measured with the help of customised web scripts and a browser extension. Such HIT states include the acceptable state, visible state, expired state, etc. Key measurements include: the time spent from a HIT being accepted to being assigned into this worker's HIT queue by the MTurk Sandbox server; the time spent from a HIT being expired from one's HIT queue to being re-published by MTurk.

Finally, based on the measurements associated with the HITs, as well as the work strategies defined by the researcher, a simulator was constructed to model the HIT group completion process for different numbers of HITs, numbers of workers, and proportions of workers using HIT catchers. During each simulation, the HITs accepted and completed by each worker agent, the time spent for HIT completion were collected. Furthermore, regarding the HIT group, diversity of participants involved in HIT completion, total number of HITs accepted and completed by workers using and not using the HIT catcher, and total completion time for the whole HIT group were also collected.

The use of simulation has the following advantages over collecting data through the real HITs:

1. Customised scenarios can be tested safely via simulation. Specifically, as this study focuses on the exploration of the unintended consequences, the size of the experiment and the proportion of workers under both types need to be moderated, and other factors that may bias the results of the experiment need to be controlled.

2. The uncertainties caused by complex factors in the real world such as the uncontrollable proportion of HIT catcher users, their working efficiency and/or replicability of expected behaviours across different population groups (HIT catchers and non-HIT catchers) all contribute to the difficulty to effectively control these variables over multiple rounds of experiments in real world situations. This makes it difficult to obtain significant observations that can be comparable across different population groups. Therefore, as an initial scoping study, generating experiment data through DES was determined to be a more appropriate approach than collecting data directly through posting HITs directly.

3.4.1.2 Data Analysis Method

Based on the data generated in the simulator, our analyses focus on two sets of measures, HIT group-related and worker-related. Specifically, the analysis includes the total completion time of the HIT group and the HIT-worker diversity for each full run of the simulation. In addition, the analysis includes the total number of HITs accepted and completed by the two types of workers, which in turn allows for a comparison of their job opportunities. The equation for the calculation of HIT-worker diversity is given below:

$$Diversity = 1 - \frac{\sum_{i=0}^{n}(2i-n-1)x_i}{n\sum_{i=0}^{n}x_i} \tag{3.1}$$

This equation is derived from the equation for Gini coefficient of inequality, which is widely used in the field of economics (Cowell, 2011). x_i stands for the array in ascending order that stores the numbers of finished HITs for each worker. 1-Gini Coefficient is applied to represent the overall equity of opportunities on doing HITs for each worker and the batch diversity (Buchan, 2022; Chien et al., 2018). The more fairly the HIT completions are distributed among the worker group, the higher the HIT-worker diversity. In other words, the larger the ratio, the more fairness of catching HITs for each worker, and the higher diversity of the batch results since they come from a wider range of workers.

3.4.1.3 Key Findings

It was found from the study that with the increase of the simulation scale (number of HITs and workers increased by the same percentage), the total completion time for the HIT group becomes longer, while the proportion of the total number of people involved in completing the HITs becomes lower. Under a fixed simulation scale, the technical advantages of HIT catcher workers have led to more job opportunities for them, regardless of their proportion of all the workers involved. In addition, the findings revealed the existence of a tragedy of the commons (Greco & Floridi, 2004): the over-acceptance of HITs by HIT catcher workers took other workers' job opportunities. Meanwhile, the HIT expirations due to the HIT catcher workers' excessive backlogging deprive themselves of the opportunities to complete the abandoned HITs later. Worse still, HIT abandonment rate is a qualification used by requester to filter workers (Hara et al., 2018). Excessive HIT abandonments would keep them from getting more job opportunities. In addition, the overall completion of the HIT group has been slowed down. These findings are further elaborated in Section 4.5.

However, these findings are based on the results of data from a simulator containing assumptions of work behaviours. Therefore, the next study (described in Chapter 5) further validates and extends the findings regarding the effects of the HIT catcher in a real scenario. In addition, more jobs stem from HIT catcher workers' technical advantages, the cause of this unequal technical advantages is the information gap (Hanrahan et al., 2021; Irani & Silberman, 2013). Therefore, more research is required to identify the factors that affect the communication of this type of information between workers to reduce the information gaps and, consequently, the negative effects of unequal competence persistence.

3.4.2 Study 2: Crowdwork Strategies with the Aid of HIT Catchers

This study aims to validate the findings of the simulation study in a real-life scenario. Furthermore, to explore the diverse work strategies and the impact of HIT catchers' use in more details, an experiment was designed by publishing image annotation HITs in MTurk.

The experiment monitored the status of HITs being published, thus providing a complete reproduction of the process by which the HIT group was completed. In addition, the use of HIT catchers was detected by checking whether specific client-side HIT catchers were installed, thus quantifying the influence on the job opportunities of different types of workers. Therefore, the correlations between workers' use of HIT catchers, their work behaviours and result qualities were investigated.

3.4.2.1 Data Collection Method

Regarding participant selection: we chose MTurk as a recruitment platform because it offers a broad and diverse worker sample. In addition, we did not set a specific reputation score requirement for participants. Therefore, any worker registered on MTurk was eligible to participate in our experiment, which helped us collect data on diversity. We posted a HIT group containing 1000 microtasks on MTurk. For participants who submitted answers, we provided appropriate compensation (explained in Section 5.2.1). All participants were clearly informed about the purpose of the experiment, the procedure and the expected completion time before starting the experiment. They all had the right to withdraw from the experiment at any time without any consequences.

During the experiment, participants were asked to label specific objects in the images and to provide textual feedback. Specifically, the street view images of the annotation HITs came from Cityscapes (Cordts et al., 2016). Participants were asked to annotate the objects from the images with different categories such as human and vehicle. Furthermore, they were asked to provide their subjective perceptions about the images and general feedback on the HITs. More details about the design of HITs are available in Section 5.2.1.

HIT responses including annotations and textual feedback provided by the participants were collected. In addition, a web script was loaded after the HIT page being rendered from the participants' browsers. The data collected through this web script include basic information of browser and operating system being used, participants' IP addresses, whether specific HIT catchers were installed, the visibility of the browser tab showing HIT pages at different timestamps, and the timestamp of when the HIT pages were opened on the participants' browsers. Finally, the state logs for each published HIT during the experiment was collected via the AWS SQS queues tracker [12] using MTurk API [13]. More details are elaborated in Section 5.2.5 (Monitoring Techniques).

3.4.2.2 Data Analysis Method

The first thing that needs to be analysed is the quality of image annotation HIT responses, which includes the accuracy of the annotation of the items in the image and the diversity of the textual response content. Creative tasks often require unique and imaginative content from workers, and having a response with high textual diversity is important (Teevan et al., 2016). Both the annotation quality and the diversity of textual responses were obtained to help the comparison of the performance of crowdworkers using and not using detected HIT catchers. More details about result analysis are explained in Section 5.3.5 and Section 5.3.6**Error! Reference source not found.**.

For each HIT that was completed, the worker's focused time, unfocused time, and backlog time on that HIT were calculated separately. Subsequently, based on the time measurements related to

[12] Amazon SQS: https://aws.amazon.com/sqs/
[13] Boto3 Documentation: https://boto3.amazonaws.com/v1/documentation/api/latest/reference/services/mturk.html

these HITs, the HIT acceptance strategy, attention switching during HIT completion, HIT abandonment behaviour, and working on multiple HITs simultaneously were detected and analysed for each participant. Furthermore, the numbers of HIT completion were also compared between two types of workers.

Based on the above analysis and the installation status of specific HIT catchers for each participant, the differences in terms of HIT acceptance, HIT abandonment, work engagement, result quality and job opportunities were compared between the two types of workers.

To understand the impact of HIT catchers' use on the completion process of HIT group, the status of each HIT was tracked and analysed using event records including HIT reservation, expiration, and abandonment obtained through the AWS SQS queues tracker. In this way, the complete process of reservation and backlog of all HITs in the experiment can be visualised and interpreted.

3.4.2.3 Key Findings

In summary, Study 2 analysed data collected from a real scenario to understand the diverse work strategies among worker groups. It validated the findings from Study 1 (Section 3.4.1.3) about the impact of the use of HIT catchers on HIT groups and crowdworkers, especially the tragedy of the commons due to excessive HIT backlogging and abandonment.

This study further expanded our understandings of HIT catchers' use on HIT state dynamics, the quality of the results, and the impact on job opportunities for different types of workers. Furthermore, the participants' diverse work strategies were explored in detail. It was revealed that the workers using HIT catchers backlogged HIT longer, spent less time actively working on the HITs, and spent less time focusing on the HIT page. Furthermore, HIT catcher workers were found to work in parallel on multiple HITs more often.

Regarding the result quality, non-HIT catcher workers were observed to achieve higher annotation quality and completeness on average. Furthermore, they also provided more diverse textual responses with more efforts. Textual diversity refers to the variety and range of vocabulary and sentence structures used within a response. It is an indicator of the richness and complexity of a textual response. These findings are further elaborated in Section 5.3.

3.4.3 Study 3: Factors Influencing Knowledge Sharing Behaviour within Crowdworkers

As another collective behaviour within the crowd community, knowledge sharing drives the popularity of scripting tools (El Maarry et al., 2018). Furthermore. It was revealed from Study 1 and 2 that workers' use of the HIT catcher varied considerably and that the impact of the HIT catcher on the number of HITs completed by the user, and therefore on earnings, was obvious. Previous research has found that a common reason why workers do not use or not skilled at scripting tools is a lack of access to these knowledge (Kaplan et al., 2018; Williams et al., 2019). In other words, a knowledge gap potentially exists between high-income and average workers.

More importantly, it was revealed from Study 1 that a higher portion of HIT catcher workers (a fairer technical advantage) could mitigate the negative impact of HIT catcher use. In Study 2, it was found the HIT catcher workers' lower average annotation quality, efforts for writing text and textual diversity are potentially due to unfair distribution of work opportunities. Therefore, to improve the fairness of work opportunities and mitigate its negative impact on HIT completion, diversity, output quality. I decide to study how to fill their knowledge gap of tooling practice and increase the popularity of HIT catchers. Therefore, Study 3 explores the factors that influence workers' skill-based knowledge sharing behaviours, and thus considers how the knowledge gap among them can be improved.

3.4.3.1 Data Collection Method

First, a conceptual model for assessing the relationship between workers' knowledge sharing intentions, behaviours, and influencing factors was constructed based on UTAUT and SET, and hypotheses about the relationship between the factors were generated based on this model. Then the survey questions were designed to assess the latent constructs within the conceptual model. The latent constructs include exogenous variables such as performance expectance, and endogenous variables such as knowledge sharing behaviour.

Regarding participant selection: our sample was drawn from MTurk and there were no specific reputation score requirements for participants, as well as other criteria, including location, age, and qualifications. Therefore, any worker registered on MTurk is eligible to participate in our experiment, especially new workers with low reputation scores due to unfamiliarity with work skills. This helps to provide sufficiently rich and diverse research sample.

Survey results were collected by publishing HITs containing a link to a Google Form on MTurk. The information collected via survey included: participants' demographic information, frequencies, approaches, and knowledge types of knowledge sharing behaviours. Furthermore, their subjective perceptions on each observed variable were collected via Likert Scale questions (Jamieson, 2004). As a complement, statistics and textual responses of participants' perceptions and strategies of using HIT catchers were collected at the end of the survey. More details of data collection are explained in Section 6.4.

3.4.3.2 Data Analysis Method

A descriptive analysis was conducted regarding participants' demographics, knowledge sharing behaviours, and their use of HIT catchers. This includes not only a quantitative description of their choices, but also a summary of the qualitative content based on the textual responses provided by the participants.

The main objective of the analysis section was to use structural equation modelling to test the hypotheses generated in the construction of the theoretical framework to assess the relationship between the factors, knowledge sharing intentions and behaviours. Specifically, after preliminary tests of sample data using outlier test, normal distribution test and multicollinearity test, a measurement model was initiated based on the conceptual model.

Regarding the initial measurement model, the internal consistency of latent constructs, factor reliability of the observed variables, convergent validity and discriminant validity of the model were tested. Through these tests, the observed variables and latent constructs not meeting the requirements were removed from the measurement model, and therefore the structural model was formed.

The explanatory power, predictive ability and overall fitness were examined around the structural model to ensure the validity of the influential relationships analysed through the model.

Finally, a PLS-SEM analysis of the latent factors influencing knowledge sharing intentions and behaviours based on the structural model was conducted, which includes both direct and indirect effects. While interpreting the influential relationships between the latent factors, the textual responses provided by the participants were combined to improve the understanding of the causes of the influential relationships between the factors.

Further details of data analysis are available in Section 6.4.3 Data analysis and Section 6.5 Results.

3.4.3.3 Key Findings

This study found a significant direct effect of reward on knowledge sharing intention and a significant indirect effect on sharing behaviour. In other words, crowdworkers want to share skill-based knowledge to gain the enjoyment and satisfaction from sharing, and the new knowledge gained from the communication. In addition, most participants worried about their technical advantages being diminished by sharing knowledge, and this concern reduces their willingness to share knowledge.

Effort expectancy, which is the efforts to use the communication technologies, not only influences knowledge sharing intention, but also influences sharing behaviour directly. Finally, it was found that participants perceived the enjoyment, satisfaction, and knowledge from others bring higher impact to eventual sharing behaviour than the ease of use, speed, and the effectiveness of the communication technologies.

3.5 Ethical Considerations

Ethics approvals were obtained before conducting the studies from the Research Ethics Committee at the University of Sheffield (Appendix A Ethical Considerations). The participants were all crowdworkers from the MTurk, and they were invited to complete the Human Intelligence Tasks specifically design for our studies. The Information Sheets introducing the experiments and Consent Forms (Appendix A Ethical Considerations) were provided to participants prior to accepting the HITs. Participants were asked to confirm that they agreed with the forms, how the researcher use, store, and delete the data they provided after the experiment.

Payments to participants were made based on time spent completing each task. The maximum completion time for each task was calculated via a pilot study, ensuring that the hourly wage at least equals the UK minimum wage. The payment did not violate the anonymity of participants because the crowdsourcing platform (MTurk) acts as s financial intermediary. Therefore, the researcher did not have any personal information related to payments about the participants. All non-malicious responses were paid in one to two days. Details of rewards were further discussed in each study.

Participants were not asked in the study for personally identifiable information, such as email addresses. However, to ensure the authenticity of the results submitted by the participants, their Worker IDs were collected. In addition, the IP addresses of the devices used by the participant were collected in Chapter 5 to study their work strategies. These IDs and IP addresses were deleted after the validity check of responses and data analysis to ensure confidentiality of personal data.

3.6 Conclusion

This chapter summarises the design, data collection and analysis methods of the studies included in this thesis. The three studies provide a more comprehensive understanding of the impact of the use of HIT catchers, and the sharing behaviour of skill-based knowledge that leads to the widespread use of HIT catchers.

The first two studies found that:

1. The use of HIT catchers, while gaining more opportunities for users, had negative effects, including delaying the completion of HIT groups, lowering the quality of results, and increasing reputational inequality among workers.

2. The cause of unequal technical advantages in using scripting tools is the information gap (Hanrahan et al., 2021; Irani & Silberman, 2013). In the simulation study (first study), as more workers use HIT catchers, their negative impact on the HIT group and overall work opportunities decreases accordingly.

Therefore, the third study investigates how people share knowledge and what factors influence this behaviour. By facilitating the knowledge sharing among crowdworkers, their information gaps could potentially be reduced, and they could have more equal technical advantages on using HIT catchers.

The following three chapters discuss each of these three studies in experimental order.

Chapter 4 A Simulation Study on the Unintended Consequences of HIT Catching Tools in Crowdsourcing Platforms

4.1 Introduction

Previous research has emphasised the positive effect of HIT catchers on workers' job opportunities, however, it has also discussed workers' complaints that opportunities for quality tasks in the platforms tend to exist only for a very short period of time, and that it is therefore increasingly difficult to obtain good tasks. There is a paradox: on the one hand, HIT catchers are intended to enhance their job opportunities and are already necessary for individual workers (El Maarry et al., 2018; Williams et al., 2019). On the other hand, workers who use and don't use tools seem to struggle with a lack of good task opportunities due to the tools' disturbance of the platform's task resources (Hanrahan et al., 2018). This suggests that the use of HIT catcher by worker groups appears to have unintended consequences, which are not clearly defined and quantified in the existing literature. Therefore, this chapter aims to explore and quantify the potential unintended consequences due to the use of HIT catchers among the crowdworker group.

With this aim, an approach based on manual measurements and simulations of the target scenario has been used for experiment data collection and analysis. Specifically, the time required for a HIT to make a transition between critical states (such as from being published to being acceptable) has been measured from the MTurk platform via web page scripts[14], and then the behaviours of HITs being accepted and processed[15] have been modelled with SimPy: a discrete-event simulation (DES) framework in Python. The analysis methods applied to the data collected from the DES have then been revealed. The results of the study have finally been discussed in the end.

Through this chapter, it has been disclosed that workers' use of the HIT catchers substantially increases their job opportunities in the short term, while depriving other crowdworkers' job

[14] Web scripts written specifically to collect time associated with state changes of HITs.
[15] After being accepted, a HIT may go through multiple transitions of states, such as being expired from a worker's HIT queue and being recalled by MTurk. The definition of HIT states is explained in section 4.3.1.

opportunities. This in turn slows down the completion of the entire HIT group and wastes the future job opportunities for those using the HIT catchers, potentially creating a tragedy of the commons in the end. In addition, the experiment has also found that: the larger the scale of the simulation, the higher the benefit of using the HIT catcher, as the HIT catcher users were penalised less for HIT abandonment; and the larger the scale of the simulation, the lower the HIT-worker diversity[16], as the average number of HITs completed by HIT catcher users was higher. Ultimately, based on the simulation results, the benefits of using HIT catcher were highest when there were 20% of workers using HIT catchers, while HIT-worker diversity was lowest, and the HIT group had the longest total completion time.

4.2 Theoretical Framing

Within a crowdsourcing platform environment, crowdworkers are described by their approval rating, which indicates the number of successfully completed HITs. Once a crowdworker completes a HIT, the job requester will examine the output and if it is satisfactory, they will approve it and reward the worker, and the worker's approval rating will increase; if the output is not approved, the worker is not paid and their approval rating will decrease. In addition, many HITs require a minimum approval rating. This means that there are cases whereby crowdworkers with low approval ratings, even due to being newcomers to the platform, are prohibited from accepting the said HITs by design (Brawley and Pury, 2016).

This approval system bears resemblance to the typical ranking and reward systems observed elsewhere (e.g., online marketplaces, h-index), whereby the ranking of individuals (e.g., merchants, researchers) is said to reflect expertise and mastery in particular types of HITs (Matherly, 2019). However, studies have shown that such systems are prone to bias, whereby rankings may push individuals to adopt strategic behaviours that do not necessarily support the flourishing of the ecosystem (Shen et al., 2015), and may lead to participants (human or otherwise) receiving unfair treatment (Gao & Shah, 2020).

[16] This indicates the number of workers completing the HIT group. The greater the number of participated workers, the greater the HIT-worker diversity.

In this study, we are interested in exploring, but crucially, quantifying the impacts enacted using automated catching scripts as materialised through the use of such ranking systems. We thus frame our empirical study drawing from Merton's Law of Unintended Consequences, also known as the Matthew effect (Merton, 1968), which suggests that those who enjoy greater visibility receive greater rewards, whereas those who are less visible, they receive disproportionately lower rewards and less recognition for the same performance. The Matthew effect is well documented and recognised in areas such as scientometrics and sociology for the examination of hierarchical systems (Fralich & Bitektine, 2020). However, much less is known regarding the extent to which the Matthew effect promotes or restricts equal opportunities for participation and reward within an environment governed by the presence of automated scripts.

4.2.1 The Matthew Effect: Competence and Reputational Persistence

When talking about the reward system for scientific contributions, Merton argued that society tends to honour those with greater reputation and visibility, irrespective of the degree of their contribution to a particular piece of work (Merton, 1968). Similarly, applications for research and development grants put forward by less known consortia and companies are often denied funding, because decisions are often influenced by the candidates' award history (Van Looy et al., 2004). Those who are less visible, and newcomers are disadvantaged in both cases, but this does not imply that they are less capable. Rather, it is perceptions regarding competence and reputation that persist and influence decision-making (Antonelli & Crespi, 2013).

Antonelli and Crespi (2013) have argued that competence persistence is an expression of virtuous Matthew effects, whereby the resources at one's disposal allow them to increase their overall competence, and thus their outputs, such as income, are simply higher.

In the case of crowdwork, automated scripts allow crowdworkers to secure microtasks that meet certain criteria with a frequency and duration that far exceeds that of a manual workflow. Crowdworkers who leverage these scripts can gain a higher skilled status compared to their peers, gaining a competitive edge (Kaplan et al., 2018). This status is often achieved at the expense of workers with lower approval ratings or not skilled in using scripts (Reschke et al., 2018). More importantly, the use of automated scripts influences one's income. Super Turkers, those who use multiple scripts together to access high quality HIT referral channels, gain extremely high earnings

(Savage et al., 2020). High rewarding HITs are finite and scarce, and scripts enable Super Turkers to locate them and reserve them in a matter of seconds when they do become available.

This is competence persistence at work. Super Turkers invest more time and effort in using automated scripts for HIT filtering (Kaplan et al., 2018), whereby HIT filtering allows them to identify HITs that match their skills and competencies, which they can tackle successfully, on time and receive the relevant reward. Those who do not or cannot use such scripts have fewer opportunities to identify enough HITs relevant to their skills, and complete fewer and fewer HITs overall. Worse still, because of fewer opportunities, workers often opt to complete lower quality HITs, posted by less (or non) reputable job requesters to increase their overall income. This is an additional risk for workers, because less reputable job requesters often reject without an explanation the submitted output, in which case the worker is not paid and the approval rating gets further reduced (Deng et al., 2016). As such, non-script crowdworkers become trapped in a vicious cycle caused by the lack of technical competence, the latter being irrelevant to the actual HITs.

Since these scripts are open for public download, theoretically, crowdworkers should have the same technical advantage, in another word, the same competence persistence. However, workers have different technical advantages due to their work strategies, different access to technical information and advanced script features (DonovanM, 2018; Hasan, 2018; Hellman, 2023; Ramirez, 2023; Schultz, 2020; Watwani, 2023).

There is still little research on the use of scripts by workers based on their background and experience, especially in terms of examining worker behaviour at the micro level. Experienced workers are more likely to be aware of new scripting tools with rich information channels, to know how to use them and locating and securing high-reward tasks (ChrisTurk, 2022; Hanrahan et al., 2021; Irani & Silberman, 2013). This consequently leads to novices spending more time searching for tasks, which in turn creates a clear income gap between Super Turkers and average workers (Savage et al., 2020). This gap is, of course, due to a combination of factors such as task search time and completion time. In general, the information gaps lead to unequal competence persistence among them. This triggered the researcher's interest in the phenomenon of knowledge sharing among workers, which in turn led to an analysis of the factors influencing this behaviour in the third study.

Reputation persistence refers to vicious Matthew effects, whereby it is posited that one's track record is testament to their skills and abilities, and is thus used as the evidence base for their selection for future employment, funding (Antonelli & Crespi, 2013), and for the purposes of our study, for HITs. For example, with regards to scientific contributions, it is not only the discovery itself that affects the popularity of an academic discovery, but also the status of the scientist who made it (Merton, 1968).

In crowdsourcing platforms such as Mechanical Turk, this track record is embodied in the approval rating. Drawing from the economics of big data, that relate to search costs and information asymmetry (Yan et al., 2015), we posit that a job requester is more likely to assign a HIT to a crowd worker with higher approval ratings and a higher number of completed HITs, because these scores allow them to filter out the excess of workers whose quality of work is not guaranteed, and because, consciously or not, HITs by workers with higher scores will be perceived as more trustworthy.

However, approved HITs contribute towards approval ratings, whereby the higher the approval rating the higher quality HITs the crowd worker can access in the future. As such, job requesters' filtering behaviour is critical for workers, because it may permit or prohibit access to future higher earnings. At the same time, however, it inevitably filters out new workers who may compete equally quality-wise but less well quantity-wise, due to not using automated scripts; thereby prohibiting them from securing better rewarding HITs in the future, and potentially driving them out for crowdwork altogether. As such, differences in the reputational persistence further trigger unintentional consequences.

4.3 Data Collection Methods

In this study, quantitative data was collected by adjusting the scale of the simulation (total number of HITs and workers involved) and the proportion of workers using HIT catchers. The generated data results have been stored and integrated in a csv format. Since csv files are plain text, their interaction with scripts is simpler. More specifically, they can be stored or imported into multiple data types easily, which is important for the analysis procedure.

4.3.1 Background of the Simulated Scenario

In the theoretical background, it was revealed that the additional gains of workers with higher skill positions are achieved at the expense of workers with lower skill positions. In the simulations, workers using HIT catchers were designed[17] to consistently fill their HIT queues without regard to their ability to complete all accepted HITs in time. The question arises as to whether, as suggested in the theoretical framing of the simulation study, HIT catcher workers gaining additional HIT opportunities cost non-HIT catcher workers their opportunities (as depicted in Figure 4.1)? And how does this affect the number of HITs completed? This study is a preliminary exploration of whether this theoretical phenomenon holds true in practice by simulating the crowdwork processes in the context of one HIT group and a specific number of workers. Within Figure 4.1, "Success" means one HIT being accepted successfully. "Rejected" means a worker's HIT acceptance request gets rejected. Each small blue block pointed by the "Request HIT" arrow represents the server latency from receiving client request to responding.

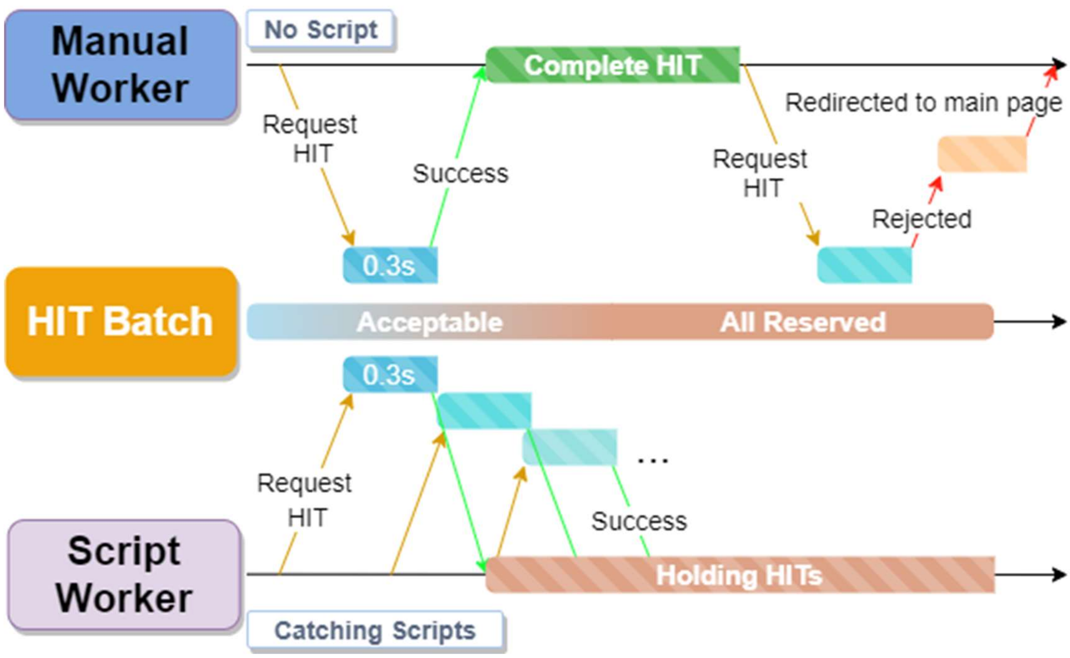

Figure 4.1 Workflow diagrams for both types of workers (Xie et.al, 2023).

[17] In the real situation, workers have a variety of HIT acceptance strategies together with other scripting tools. However, in this experiment, only the greedy accepting strategy is implemented to facilitate the building of the simulator.

The hypothesis to test in the simulation is that: the success rate of catching HITs for each crowd worker is influenced by the persistence of their technical skills under the condition that everyone has the same access to these HITs. In real practice, the factors for measuring workers' technology skill persistence are complex, including how to use tools more effectively, or combine multiple tools to improve productivity. Modelling these complex technology use behaviours requires a large amount of actual behavioural data as a reference, both in terms of behavioural logic of specific types and the estimated proportion of the total number of participants. Given the excessive workload, in this study, workers' technical skill persistence is constructed as the ability to use HIT catchers.

First, the definition of the different states of the HIT needs to be clarified to help design the DES framework. Here is a list of five key HIT states together with their explanations:

HIT publication: This refers to a job requester publishing a HIT group containing a number of HITs to workers on MTurk interactive interface or API, which will then appear on MTurk job search page and be found by workers.

HIT acceptance: The act of a worker clicking on the accept task link on MTurk task page to reserve the target HIT in the HIT queue of their worker account. Alternatively, HIT acceptance can be achieved by sending the HIT ID (also known as Batch ID or Project ID) of the target HIT group directly to MTurk server.

HIT reservation: The state in which a HIT is held in its HIT queue after it has been accepted by a worker. This state lasts until the worker completes and submits the result, or until the HIT expires and is then removed from the current worker's HIT queue by the platform.

HIT expiration: As explained in Section 1.1.8, Each HIT is given a time allotted by the platform after it has been accepted by a worker. If the current worker is unable to submit a response to the HIT within the time allotted, HIT expiration is triggered and the HIT is then removed from the current worker's HIT queue by the platform.

HIT submission: After being completed and submitted to the requester by the crowdworker, one HIT reaches the HIT submission state. In practice, the HIT result will then get approved or rejected by the requester. In simulation, all HITs reaching this state are terminated and removed without further state changes. Because this study focuses only on the process from HIT publication to submission.

4.3.2 Design of Simulation Model

To investigate the impact of different types of work behaviour on crowdworkers' job opportunities, the speed of completion and quality of outcomes of the HIT group, crowdworkers were modelled as agents and their autonomous behaviours were constructed according to whether they use HIT catchers or not. This approach follows agent-based modelling. It is also necessary to build the state change process for each individual HIT in a top-down approach through DES. The specific HIT state change process includes the current HIT being successfully accepted and completed. In another case, the HIT could be withdrawn from one's HIT queue if it is not completed on time. Then this HIT expires and becomes available again to others after a cooling down period.

In addition, the scheduling process of the MTurk server for all HITs needs to be simulated using DES. Similar to the limited number of staff in the healthcare system, HITs are a limited resource for worker groups in the MTurk platform. It should be clear that when multiple workers are trying to accept the same HIT simultaneously, the MTurk server needs to assign the HIT to the one who made the accept request first and reject the accept requests from others. In summary, this study requires a combination of DES and ABS. Specifically, worker behaviour should be modelled through ABS and state changes of HITs should be modelled through DES.

The behaviour of the requester was not modelled in simulation, as this study only focuses on the process from the publication of HIT group to when the contained HITs are all submitted, and does not involve the subsequent process of the requester approving and rejecting the HITs based on the quality of the results.

As this study focuses on the exploration of the effects of unintended consequences for workers' use of HIT catchers, the size of the experiment, the proportion of worker types, the size of the HIT group and the control of other factors that may bias the results of the experiment need to be adjusted. Therefore, generating experimental data via DES is more appropriate to the specific requirements of this study than collecting data directly via posting HITs. In addition, as mentioned in the research design section, compared with using simulations, if the experiments were conducted directly on the MTurk platform, the data collected would be subject to a number of confounding factors, such as the number of workers online at the same time, which varies greatly depending on the time of day and the time zone in which the workers are located, making it difficult to ensure a set of workers who could be continuously available to reserve HITs. In addition, the popularity of the HIT, which is influenced by several factors such as the type of HIT, the reward level and the reputation of the requester, would also create uncertainty as to whether the workers will attempt to accept the HIT.

This section constructs the simulation environment of interest to this study at three levels: conceptual, specification and computational (Leemis & Park, 2006). At the conceptual level, it is necessary to find out to what extent a realistic situation needs to be modelled, and to specify which events, states of agents, etc. should be included in the model. At the specification level, it is necessary to specify how specific events are to be simulated and what the different time intervals are. Ultimately, in the computational level, the simulation scenario is implemented based on the models generated in the first two levels.

4.3.2.1 Conceptual and Specification Level

In this section, the phenomena to be modelled for this study are firstly explained in text. Then they are modelled according to the interaction events between MTurk, worker agents, and HITs. The result is a discrete event model that simulates the management of the state of each HIT by MTurk, and an agent-based model that simulates the autonomous behaviour of each worker in the experiment. Subsequently, measurements were made on the MTurk Sandbox for the time spent for the HIT to transition between different states. Consideration of the time intervals between these key events can improve the realism of the simulation at the computational level. In the end, the discrete event and agent-based models were combined to form a hybrid simulation model for each type of worker. By constructing the phenomena at these two levels, I can understand the complete workflow of the simulation system, including how the interaction events of the workers and MTurk may affect the state of the HITs.

4.3.2.1.1 Process-Oriented Approach in Discrete Event Modelling for Server Behaviours

In previous research, DES was often designed using flow charts showing the interaction between source, process, decision, queue and delay (Patel et al., 2020). The HIT (source) in Figure 4.2 is first published on MTurk by the requester and then its own state is changed under the influence of different worker behaviour events or system events. Eventually, the HIT is submitted by the worker and thus leaves the flow chart.

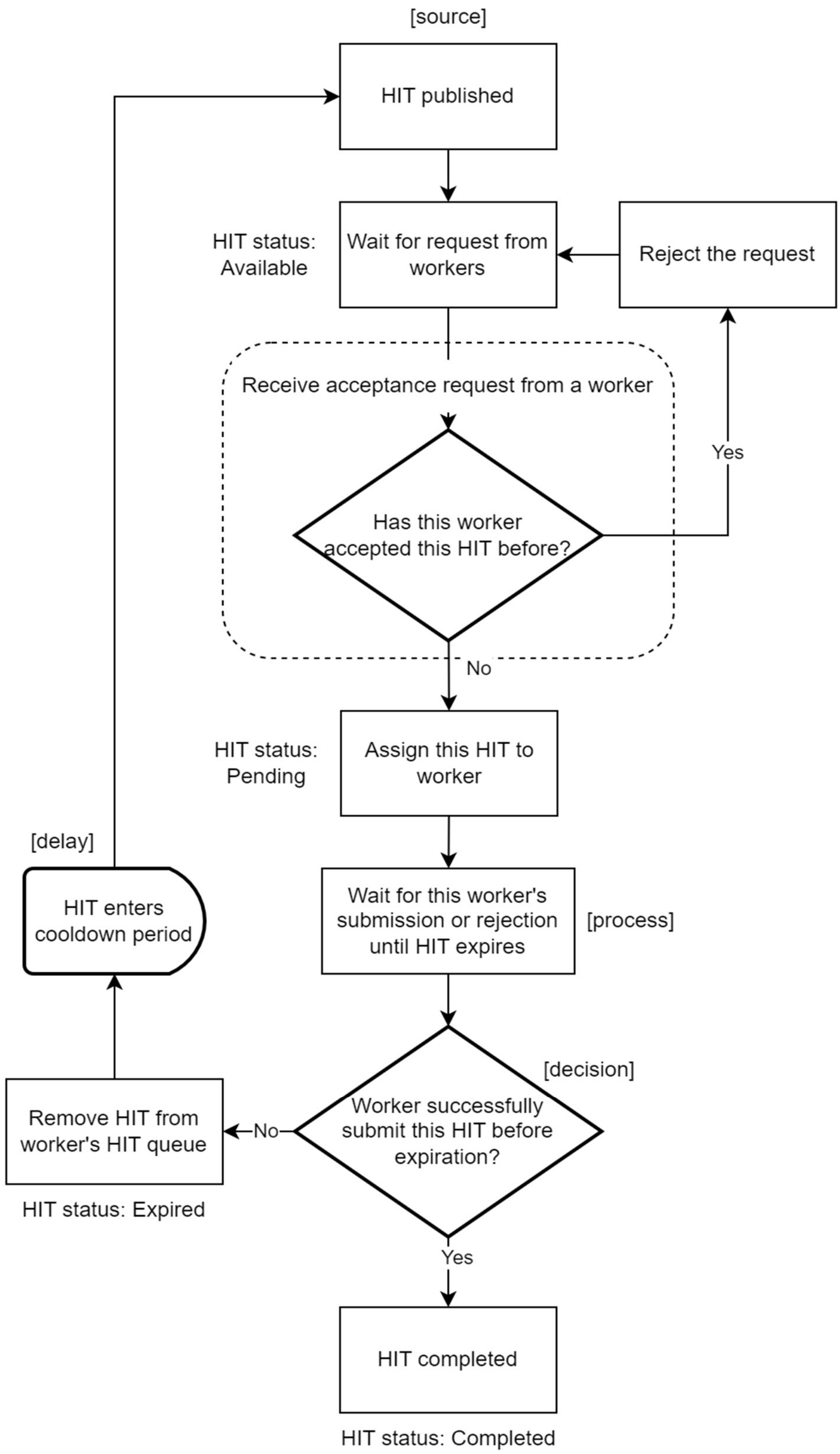

Figure 4.2 Flowchart of the MTurk server management process of each HIT in DES.

This study uses a top-down approach to model the HIT management process of the MTurk server into discrete events. Specifically, MTurk continuously manages the status of a HIT after it has been published by the requester. Meanwhile, MTurk waits for the HIT acceptance request from the worker group. When there are requests from multiple workers, MTurk takes the first request and assigns this HIT to this worker's HIT queue. At the same time, MTurk changes the status of this HIT to "pending", while other HIT acceptance requests sent later are rejected.

After this HIT has been successfully accepted, the status of this HIT is updated to "submitted" if it is submitted by a worker within the time allotted. Also, as this simulation only focuses on the worker behaviours, the interaction of the requester with the HIT (approve or reject HIT results) is not included in the simulation. Therefore, simulation ends the status management of this HIT after it has been submitted.

However, if the HIT is not submitted by the time allotted, MTurk will change the HIT status to expired and remove the HIT from this worker's HIT queue. Upon entering expired status, this HIT then enters a cooldown period and becomes available to others again afterwards.

The time intervals involved in this flow chart include:

1. The delay between the HIT being posted by the requester and actually being accepted or seen by the worker.

2. The network and server response delays that exist in the worker's interaction with MTurk, the duration of the cooldown period after the HIT expires, etc.

3. The time it takes for the HIT to be assigned to one worker after being successfully accepted.

4. The duration of the cooldown period after the HIT expires.

To increase the realism of the simulation results, these time intervals need to be accurately measured and implemented into the simulation program based on the design of the process.

4.3.2.1.2 Individual-Oriented Approach in Agent-Based Modelling for Worker Behaviours

To simulate the independent and autonomous behavioural threads of each worker within the experiment from a micro level, this study constructs workers as agents and implements the simulation of each worker's behaviour through agent-based simulation (ABS). A common approach to modelling ABS is to use the state transition model (Borshchev, 2013), which could exhibit different states within each agent, transition between the states, and the events that trigger those transitions.

In the proposed agent-based model, the agents representing the crowdworkers interact with the MTurk server through autonomous behaviour, thus causing changes to the status of the limited number of HITs in the simulation. Meanwhile, the states of agents are transformed in response to different events. Figure 4.3 shows all the states that an agent representing a crowdworker has during the simulation. The significance of distinguishing between these states is: compared to HIT catcher workers, those without HIT catchers often spend extra time on manually accepting HITs. In addition, workers without HITs have to spend time not being paid in waiting for future job opportunities that might arise. The difference in time spent by these workers outside of completing HITs reflects the impact that the use of HIT catcher has on their work behaviour and efficiency.

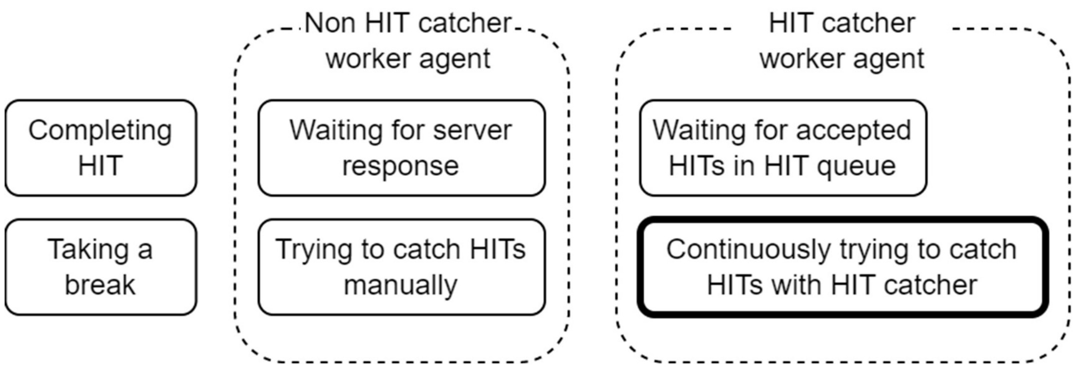

Figure 4.3 Illustration of available states for each agent.

Figure 4.4 and Figure 4.5 illustrate the state transition process for both types of workers in the simulation experiment. Specifically, for those not using the HIT catcher, they follow the rule of working on one HIT immediately after it has been successfully accepted. They will try to accept the next HIT only after the current HIT has been submitted or expired. What is notable is that when they focus on accepting HITs, they have to stop other states, including taking a break and working (Figure 4.2.4). In comparison, the HIT catcher workers can continuously try to catch HITs with HIT catcher without changing their current states (Figure 4.2.5). From the diagram it can be revealed that each HIT catcher worker has two complete state transition loops. Unlike non-HIT catcher workers, the condition for triggering their “Doing HIT” state is the existence of HITs in their own HIT queue. In summary, the HIT catcher workers do not need to spend extra time catching HITs during the transition between "Doing HIT" and "Taking a break" state.

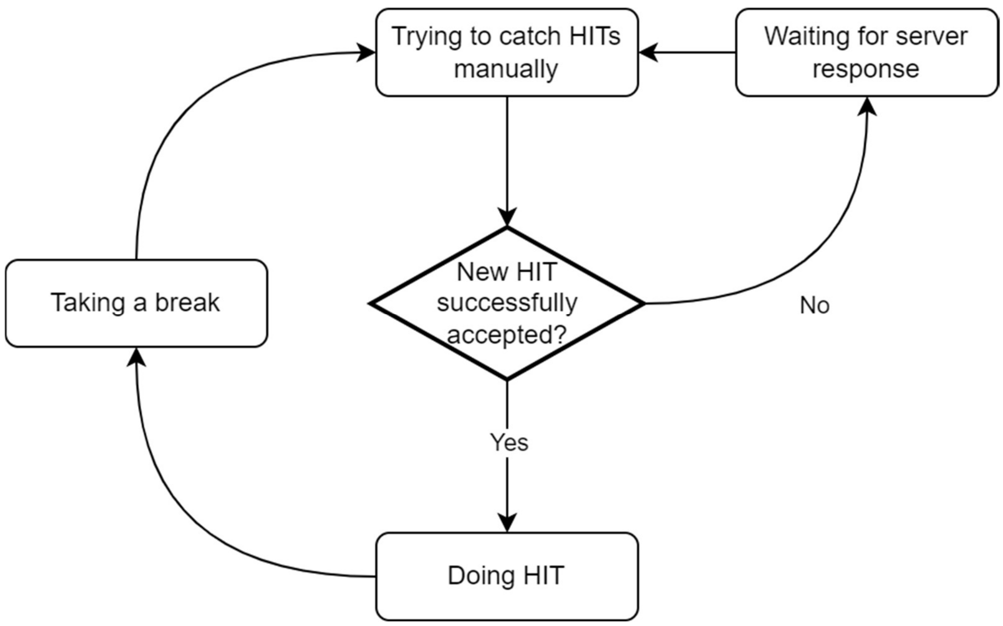

Figure 4.4 State transition model of the non-HIT catcher worker agent.

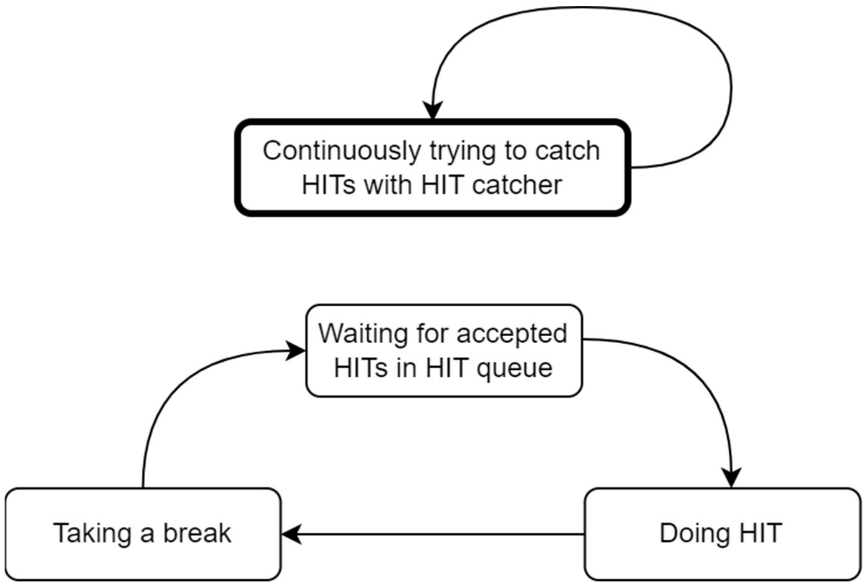

Figure 4.5 State transition model of the HIT catcher worker agent.

It is important to note that the two state transition models above cannot fully represent the work strategies of the two types of workers in the real world. Specifically, non-HIT catcher workers can choose to start doing HITs after successfully accepting multiple of them, or to catch the next HIT right after a previous failed attempt. HIT catcher workers can choose to stop the automatic HIT catching after a limited number of HITs have been successfully accepted, thus giving themselves plenty of time to submit them before their expiration. In other words, in real situations, workers could be flexible and change their work strategies to maximise their job opportunities and productivity. This limitation is further elaborated in Section 7.5.1. In contrast to a thorough reconstruction of the real situation, the focus of this experiment is to conceptualise the work strategies and thus reveal the impact of the differentiated technical advantages of the two types of workers on the job opportunities, the diversity of outcomes and speed of completion of the HIT group published by the requester.

4.3.2.1.3 Measurement of HIT Status Change Time

Before the development of the simulation framework, the durations between different key events throughout the life cycle of the HITs (time for a HIT to become visible, time for an expired HIT to be published again, etc.) were estimated on the MTurk Developer Sandbox (*Developer Sandbox*, 2021).

The MTurk Developer Sandbox is a mirror of the production platform which allows requesters to test their microtasks before publishing on the production site (*Using the Sandbox*, 2021). Other than the monetary transfer being disabled, the MTurk Developer Sandbox has the same functionalities as the MTurk platform.

Specifically, in order to measure the time interval between important events, including events such as a HIT group gets posted by a requester and actually becomes visible on a platform page, a Google Chrome extension[18] was developed to enable continuous acceptance of HITs with specific frequency (as shown in Figure 4.6). In the top section, the researcher can enter the HIT group Project ID of the target HIT group and choose to accept one HIT from the group immediately. Meanwhile, this extension could automatically record the time the HIT reserving request sent from

[18] Source code for extension: https://github.com/howrudoing/HIT-catcher-for-simulation-study/

client and the time the target HIT was successfully added to the researcher’s HIT queue. In the middle section of the interface, it allows the researcher to continuously reserve HITs from the target HIT group with a specific frequency at a particular moment in the future.

The purpose of this feature is to facilitate the researcher's use of other web scripts[19]. More specifically, page content monitoring script can be used in other tabs in the meanwhile to record the process by which multiple HITs become visible in the platform's HIT list after being published by the requester. In addition, this function is also used to estimate how soon HITs can be reserved by workers after being published. This feature helped us discover an interesting phenomenon: HITs could be reserved by workers based on the project ID before becoming visible in the platform. This finding is shown in detail in the next section about time measurement.

By interacting with the bottom section of the interface, the extension could automatically refresh the current HIT list web page and search for the target HIT group, with the aim of helping the researcher understand at what time the target HIT group becomes visible to all workers in the platform.

[19] Script used to search for HITs via list page: https://github.com/howrudoing/Scripts-for-thesis/blob/main/auto_refresh_search.js

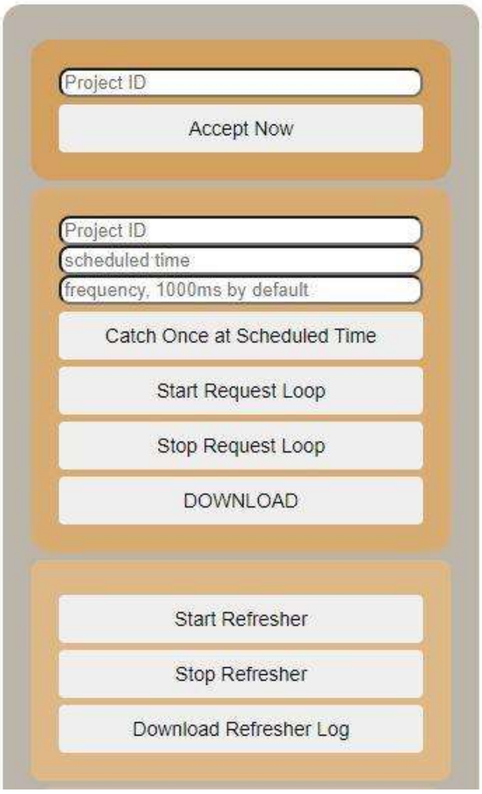

Figure 4.6 User interface for the HIT catcher that helps measure event durations.

Before presenting the definitions and measurements of the different time intervals, it needs to be clear that: all the timestamps generated in web script and browser extension being used in this experiment represent the number of milliseconds since January 1 1970 UTC by calling the function *Date.now()*[20]. With the help of this function, the timestamps generated in different web scripts and extensions can be compared to obtain the final time interval between important events.

Specifically, based on previous testing using the web scripts, a total of eight important time intervals have been addressed, and they need to be considered in order to increase the quality of DES. The definition of these time intervals, the purpose of the measurements and the measurement methods are explained in turn:

---

[20] Function Introduction: https://www.w3schools.com/jsref/jsref_now.asp

Interval from HIT Expiration to HIT Being Removed from the Worker's HIT Queue

This interval means how long it takes a HIT to be removed from the worker's HIT queue and make room for another new HIT. From the initial testing using web monitoring scripts, it has been found that the HIT does not disappear from the worker's HIT queue immediately after its expiration, so the exact length of this delay needs to be specified for a more appropriate design of DES.

To measure this time interval, a worker account has been used to accept as many HITs as possible to fill the HIT queue. At the moment that one of the accepted HITs expired while other HITs are still active, this worker tried to catch a new HIT and the timestamp for successful acceptance has been recorded. This procedure was repeated 10 times for an accurate result by averaging. It could also minimise the potential errors due to unstable bandwidth.

Interval from HIT Expiration to HIT Being Acceptable Again

This time interval is critical for the construction of DES, as the expiration of HITs is more common for HIT catcher workers who are used to over-accepting HITs, and getting an accurate time between HIT expiration and being acceptable again helps to model the transition of the HIT states more accurately.

When measuring how long a HIT becomes available again for one worker after expiring from another one's HIT queue, the bisection method was applied (Table 4.1). More specifically, one HIT group containing one HIT has been published, and Worker A reserved that HIT until it expires. Meanwhile, Worker B attempted to accept that HIT right after it expired from Worker A's HIT queue. In this measurement process, the researcher logged in as requester, Worker A and B from three devices connecting to the same network. The HIT catching behaviour was performed automatically using script to avoid errors caused by human behaviour.

Table 4.1 Summary of data collected for interval measurement between HIT expiration and re-activation. Rows framed and highlighted in grey indicate cases where Worker B failed to accept the expired HIT at the given time.

Time of delay (ms)	Index	Time the task expires	Time task request sent	Time of successful catch	Interval (ms)	Interval (s)	Network Latency(ms)
10000	1	1585170290698	1585170300698	1585170301600	10902	10.902	902
5000	2	1585170465718	1585170470718	1585170470931	5213	5.213	213
7500	3	1585170649439	1585170656939	1585170657371	7932	7.932	432
6250	4	1585170807668	1585170813918	1585170814340	6672	6.672	422
5625	5	1585171447454	1585171453079	1585171453819	6365	6.365	740
5312.5	6	1585251375514	1585251380827	1585251381679	6165	6.165	853
5000	7	1585251741177	1585251746177	1585251747557	6380	6.38	1380
5156.25	8	1585251958843	1585251963999	1585251964243	5400	5.4	244
5234.375	9	1585252216048	1585252221282	1585252222494	6446	6.446	1212
5195.3125	10	1585252443636	1585252448831	1585252449267	5631	5.631	436

Server Latency for a Worker to Accept One HIT After the Expiration of the Previous

This duration has been investigated to determine whether MTurk penalises workers for HIT abandonment by increasing the delay in HIT acceptance. In addition, a worker receiving a HIT from the same HIT group as the expired HIT may be different from a worker receiving a HIT from a different HIT group than the expired HIT. The average interval obtained from 10 rounds of measurements was 0.469s (catching a new HIT within the same HIT group) and 0.664s (catching a new HIT from a different HIT group) respectively. These values approximate the MTurk server response time. These two intervals show that in both cases the worker does not suffer this type of penalty due to HIT abandonment.

Time Spent for HIT Group From Being Published to Being Acceptable

The purpose is to compare it with the next time interval, thus helping to understand the technical advantages for the workers of the act of skipping the HIT search page and receiving the newly posted microtasks directly through the script.

The measurement of this metric involves the simultaneous use of a requester account and multiple worker accounts with the help of scripts. The main idea is to try to receive these HITs while the task is posted and to record the time interval between the HIT group being posted and the HIT being accepted, thus helping to understand the time interval between the HITs being posted and being acceptable.

Firstly, a customised web script was applied to help the researcher as a job requester publish the target HIT group at a particular moment in time. At the same moment, 10 worker accounts in MTurk Sandbox were used to apply all acceptable HITs from the target HIT group continuously via the experiment browser extension automatically. The reason for using 10 worker accounts for this experiment is that in MTurk Sandbox, each worker is allowed to reserve up to 10 HITs at the same time (unlike 25 HITs limit in MTurk). Therefore, after publishing a HIT group containing 80 HITs, it can be ensured that they are all reserved right after being available. Meanwhile, this extension kept track of the timestamp on the successful HIT acceptance. Finally, the time interval between each HIT being posted and being acceptable is obtained by comparing the timestamp when the worker successfully accepted each HIT with the timestamp when the HIT group was published by the requester. By aggregating these time intervals, a complete description of the changes in the HITs of the entire HIT group over two specific states can be obtained.

Figure 4.7 illustrates the point in time when each HIT was successfully accepted by a random worker in a HIT group containing 80 HITs, from the moment the whole HIT group was published. Interestingly, the scatterplot from the three experiments shows that the MTurk platform appears to have published the 80 HITs in three stages. While the earliest acceptable HIT appeared more than ten seconds after the HIT group was published, the latest acceptable HIT appeared one minute after the HIT group was published. The reasons for this significant delay, which clearly cannot be explained by network transmission, also deserve to be explored in more depth by future researchers based on the governance of the platform and the equity of job opportunities.

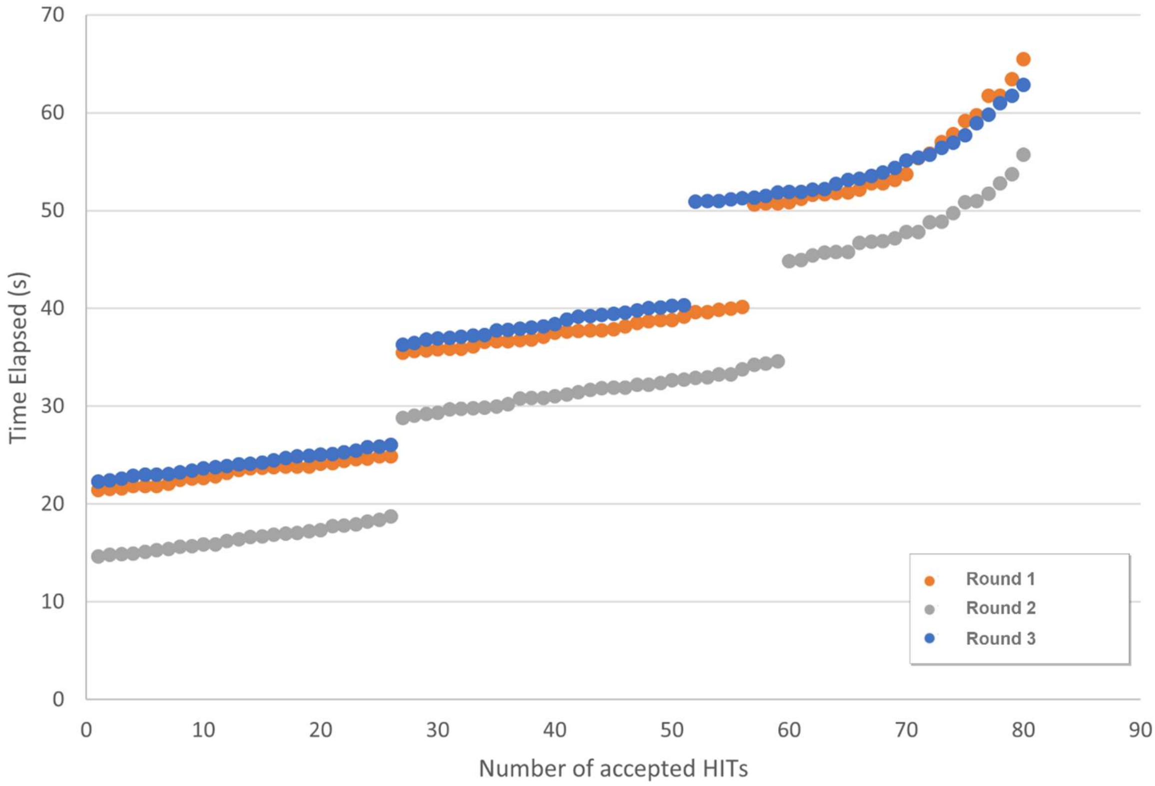

Figure 4.7 Distribution of numbers of accepted HITs over time after publishing a HIT group containing 80 HITs for 3 rounds.

Time Spent for HIT Group From Being Published to Being Visible

The idea of measuring this metric is to continuously refresh the HIT search page containing the target HIT group while it is published, thus simulating the behaviour of workers searching for the target HIT on the HIT search page. At the same time, the time interval between the publishing of a HIT group and each observation of the search page containing information on the number of HITs in the target HIT group is recorded, thus helping to understand the change in the state of HITs from being published to being discoverable by workers.

To record the number of visible HITs, the HIT list page containing the target HIT group was refreshed continuously through the Chrome extension for experiment. Meanwhile, the number of visible HITs was monitored using a customised web script to draw the HIT visible curve[21].

[21] Script used to search for HITs via list page: https://github.com/howrudoing/Scripts-for-simulation-study/blob/main/auto_refresh_search.js

To begin with, a customised web script has been applied to help publish the target HIT group at a particular moment in time. At the same moment, Worker A refreshed the MTurk HIT search page continuously. Every time the page was refreshed, the number of HITs inside this HIT group and the timestamp of observation were recorded via a web script. Ultimately, the time passed since the publication of the experiment HIT group was obtained by using the timestamp recorded for each refresh and the timestamp when the HIT group was published. This in turn helps to understand the number of HITs that can be discovered by workers from the page at different times after its initial publication (Figure 4.8).

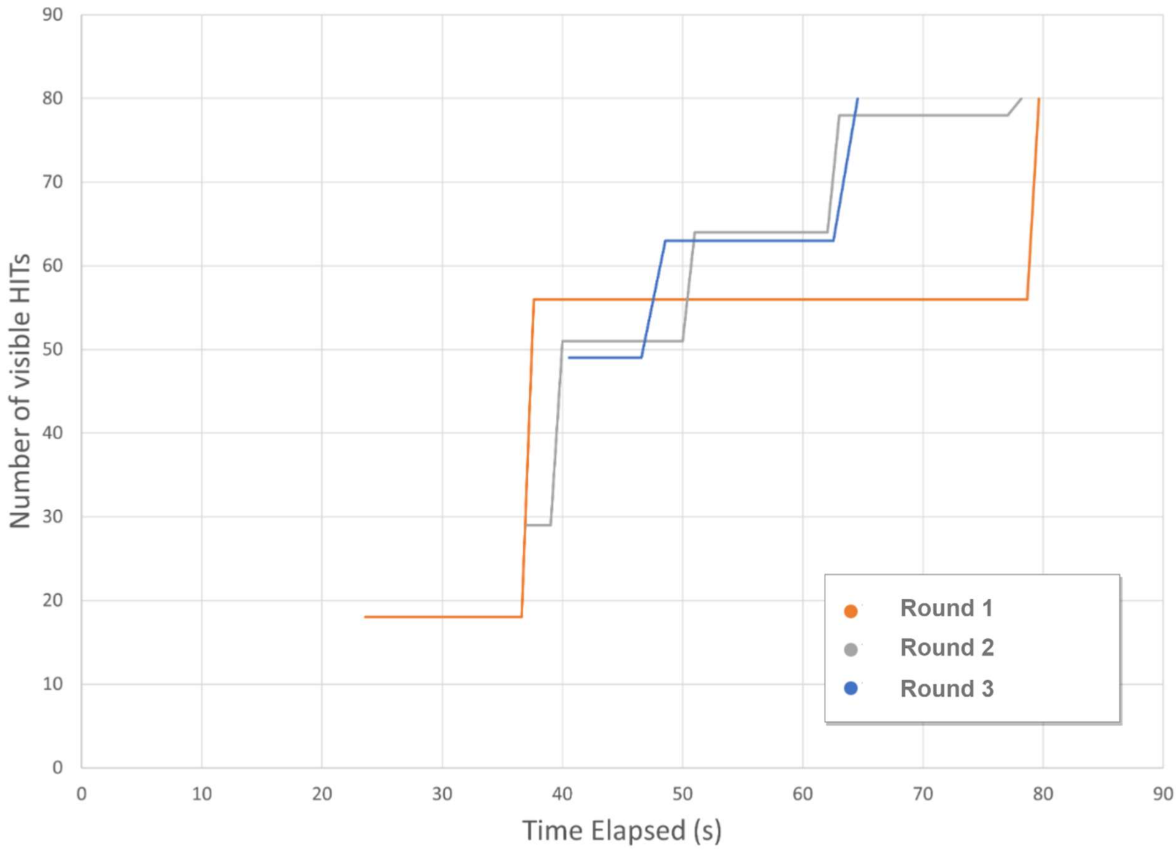

Figure 4.8 Distribution of numbers of visible HITs over time after publishing a HIT group containing 80 HITs for 3 rounds.

Similar to the time lag between HITs being published and being acceptable, the platform still made all 80 HITs visible in multiple stages, rather than all being visible to workers at the same time. From the above measurements, it can be observed that after a requester posts a HIT group, these HITs do not immediately appear as being acceptable / reservable in the platform, but are delayed by tens of seconds to a minute before being visible in full to workers from the platform. This raises the hypothesis that the new HITs being published become acceptable before being visible.

To further test this hypothesis, a total of 20 HIT groups with different numbers of HITs have been used to compare the time it took for the HITs to go from being published to being visible and acceptable as a whole (Figure 4.9). As can be recognised, the time taken by HIT groups from publication to overall visibility generally far exceeds the time it takes for their HITs to be acceptable. This significant time difference provides a technical advantage for HIT catchers to bypass the HIT list page and accept HITs directly by sending HTTP requests to the server, which provides an idea to answering the recurrent suspicion in previous research that good HITs are fleeting and cannot even be found on the search page (Williams et al., 2019). Although the two-time intervals mentioned above were not ultimately applied in this simulation framework, it still helps the researcher understand the significant impact of the use of HIT catchers on work opportunities by comparing the time differences between the two.

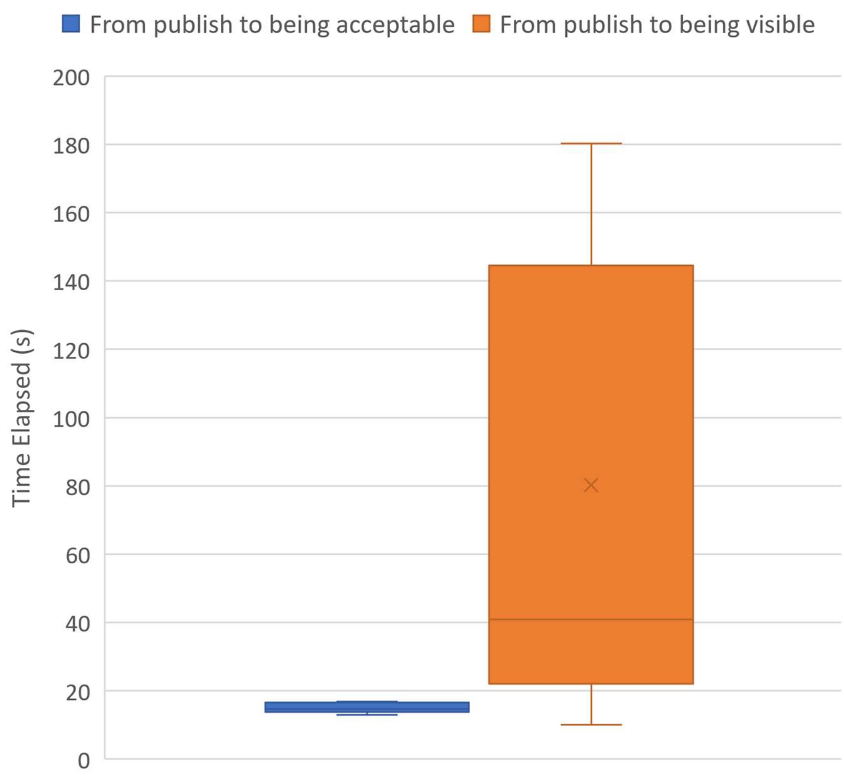

Figure 4.9 A comparison of the time spent for HIT groups to become acceptable and visible from being published.

Minimum HIT Catching Interval Without Page Request Errors (PREs)[22]

Understanding this interval helps to determine the frequency of HIT acceptance for HIT catcher workers in DES. Although in practice, workers using HIT catchers have very diverse HIT acceptance strategies, such as trying to reserve from multiple HIT groups at the same time or using scripts to automatically reserve the next few HITs after completing a fixed number of HITs. In contrast to the above, this DES focuses on a more basic case where only one HIT group exists. In addition, the HIT catcher workers' HIT acceptance strategy is designed to be extremely greedy, in other words, to get the most HITs in the shortest time possible. Interestingly, during each round of measurements, the beginning 7 to 10 requests sent from client were always valid and did not cause PREs regardless of the frequency. Therefore, to estimate the optimal catching frequency with less bias, the first 10 requests will not be counted for each round.

Pinging Worker Sandbox Server

Ping or latency is the duration of time it takes for a small data set to travel from a worker's device to the MTurk server and then return to the worker over the Internet. This latency should always be considered when measuring the intervals between events from MTurk server. Since the devices involved in this study were all used at the same time in the same network environment, the latencies within the measurements due to bandwidth are very similar. Therefore, the network latency was treated as a fixed value in the simulation and was already included in the measurement of other intervals.

MTurk Sandbox Server Response

Unlike the ping mentioned above, the time it takes for MTurk Sandbox server to respond to any request from a worker's device not only include the delay incurred in the transmission of the message over the network, but also the time taken by the server to generate a response. This time interval can be seen as the minimum time required for a worker to make any interaction with the MTurk server, such as clicking into a HIT preview page or checking personal details. Furthermore, this parameter helped to verify the existence of a delay penalty for a particular worker in any cases.

[22] Definition of Page Request Error: https://forum.turkerview.com/wiki/Page-Request-Error-PRE

In summary, the time intervals between these important events are required to be measured accurately to ensure a reasonable simulation of the worker activities through DES. As illustrated in Table 4.2, the time intervals between important events were measured 10-20 times each in MTurk Sandbox, thanks to the browser extension and web scripts specifically designed for this experiment. For intervals with low standard deviations such as the “Minimum HIT catching interval”, they were measured 10 times. However, for those with high standard deviations such as “HIT from expiration to being acceptable again”, 20 times of measurement were taken for more reliable average values.

Table 4.2 Summary of key event durations on MTurk Sandbox (all times are in seconds).

	Average Interval	Std Dev	Mean-Std Dev	Mean+Std Dev	Range	Number of measurements
HIT from expiration to being removed from HIT queue	1.81	0.30	1.51	2.11	0.60	10
HIT from expiration to being acceptable again	21.38	4.82	16.56	26.20	9.64	20
Worker accepting one HIT after the expiration of another HIT	1.01	0.35	0.66	1.36	0.70	10
Batch from being published to being fully or partially acceptable	15.90	5.82	10.08	21.72	11.64	20
Batch from being published to being fully or partially visible	67.03	55.11	11.92	122.13	110.21	20
Minimum HIT catching interval (without PREs)	1.00	0.01	0.99	1.00	0.02	20
Pinging Worker Sandbox Server	0.22	0.07	0.15	0.29	0.14	161
MTurk Sandbox Server Response	0.68	0.40	0.27	1.08	0.81	12

4.3.2.1.4 Overview of Hybrid Simulation Models

Two types of workers have been defined in this simulation: non-HIT catcher workers who look for work opportunities manually, and HIT catcher workers who accept target HITs automatically via HIT catchers. This simulation framework defines the workflow of two types of workers and sets up time intervals that include events such as executing each HIT and taking a break when the HIT is completed, thus facilitating my construction of the simulator. By combining the discrete event model and agent-based model mentioned in the beginning of the simulation model design section, and subsequently by measuring the time intervals between HIT events in the MTurk Sandbox, the hybrid simulation models for workers with and without HIT catchers were developed as shown in Figure 4.10 and Figure 4.11.

In specific, after the simulation gets started, non-HIT catcher workers attempt to accept the target HITs by opening the HIT preview screen in the HIT group list page. In contrast, HIT catcher workers accept the target HITs on an automated and continuous basis at a frequency of as fast as once per second, using a known HIT ID and the HIT catcher. For each worker, if there is an acceptable HIT in the target HIT group and the worker's own HIT queue is free, a HIT can be successfully accepted and stored in the worker's HIT queue waiting to be completed.

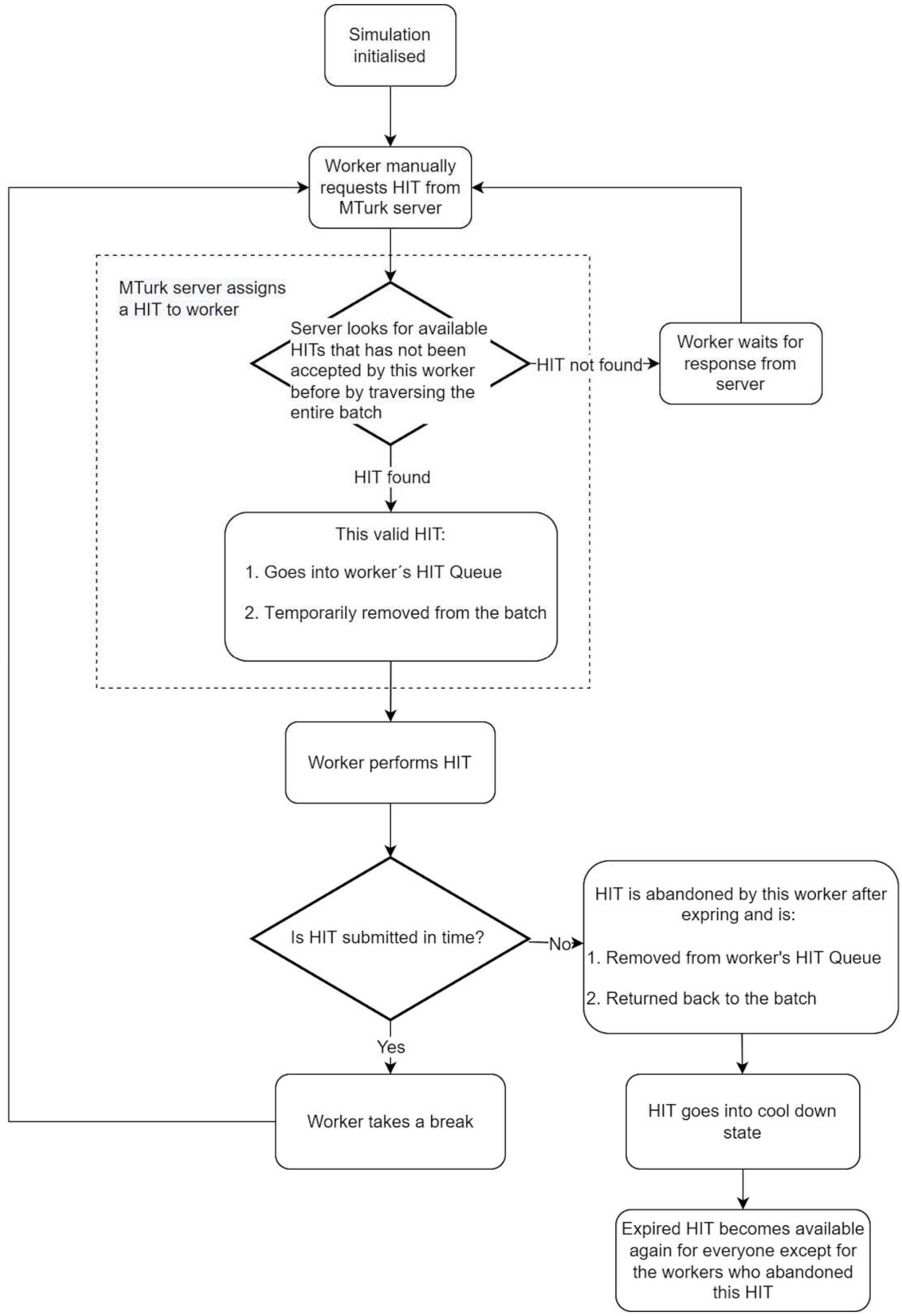

Figure 4.10 Hybrid simulation workflow diagram for non-HIT catcher workers.

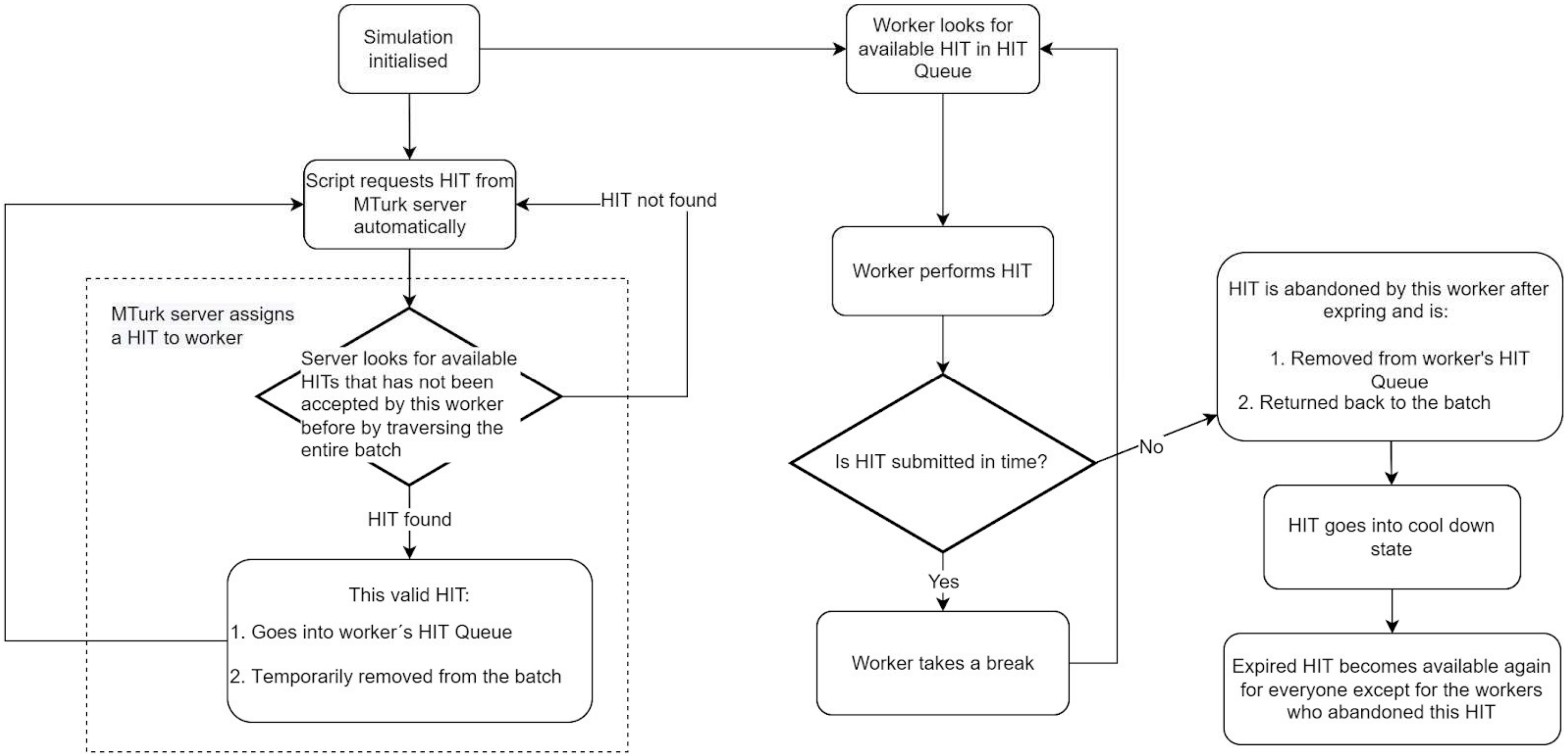

Figure 4.11 Hybrid simulation workflow diagram for HIT catcher workers.

As an important rule of MTurk, a worker who has passed the trial period could reserve up to 25 HITs in their own HIT queue at the same time (ChrisTurk, 2017). In this DES, for those non-HIT catcher workers who do not use HIT catchers, they will start executing the HITs they have just successfully accepted. They will then accept a new HIT after the previous one has been completed. For HIT catcher workers, thanks to the continuous HIT acceptance feature from the HIT catchers, they can focus on executing the HITs that have been successfully accepted by the script and stored in the HIT queue, especially those with sufficient time allotted before expiration.

In this simulator, each HIT can be reserved by one worker for up to 1 minute and usually takes half a minute to complete. Due to inevitable fluctuations in productivity, there is a chance that each worker fails to meet the deadline, and this feature is also reflected in the simulation through fluctuations in completion time. In detail, the exact time that one worker spends on completing one HIT is a random number within a truncated exponential distribution with the mean value of 30 seconds (Araman et al., 2019). For those HITs that are not submitted on time will be withdrawn by the platform from the worker's HIT queue and enter a cooling down period with an average value of 21.38s. In practice, it would take one crowdworker a few seconds to dozens of minutes to complete one HIT[23]. In this study, 30 seconds was chosen as the average HIT completion time in order to allow HITs to be completed faster, thus reflecting more clearly the impact of the two work strategies on the individual's job opportunities and on the completion process of the HIT group.

Apart from the time measurements in Table 4.2, the expected HIT completion time, and the worker acceptance strategy, in order to initialise the simulation model, (i) the percentage of workers using scripts, (ii) the number of workers per batch should also be defined. Regarding (i), no studies have yet been conducted to estimate the number of people using HIT catchers on MTurk. Moreover, this study aims to construct a simulation framework to quantify the unintended consequences of using catching scripts at a relatively micro level. Therefore, the total number of workers (ii) is limited to a maximum of 50.

Two groups of simulations were designed to investigate the variation in the impact of HIT catcher use on both types of workers and the whole batch at different experimental sizes and

[23] A Simple Formula for Predicting the Time to Complete a Study on Mechanical Turk: https://www.cloudresearch.com/resources/blog/a-simple-formula-for-predicting-the-time-to-complete-a-study-on-mechanical-turk/

percentages of HIT catcher workers. The results of the simulation depend on several random effects, such as whether workers have available HITs to perform all the time, etc., so multiple runs need to be conducted to estimate the expected scores (Seneta, 2013). In this study, 10 runs for each simulation were sufficient since the results clearly reveal the trend of variation and the boxplots are not overlapping.

In addition to the randomness of the worker's HIT completion speed and the rest time between jobs, the main factors that influence the worker's behaviour are the acceptability of the HITs and whether the worker's own HIT queue is available for the worker to receive new HITs. The acceptability of an individual HIT is not only a matter of whether the HIT can be re-posted in the platform, but also whether it has been accepted by the current worker. According to the technical rules of the platform, workers cannot re-accept previously abandoned assignments again, but they can still accept other assignments from the same HIT group.

4.3.2.2 Computational Stage

In this study, SimPy was used to construct the hybrid simulation model constructed above (Müller et al., 2021). Although it emphasises its adaptation to DES, it is still technically feasible to implement agent-based models. In SimPy, each worker agent can be initialised asynchronously using *env.process()*[24]. SimPy uses *env.timeout()*[25] to reflect time advancement, such as a worker taking a break, or spending time to complete a HIT. However, this function itself does not advance the simulation clock. Specifically, when a worker agent calls this function, that agent waits until the simulation clock advances to that time. While the agent is waiting, it does not block other agents' behaviours. In summary, SimPy can technically implement the proposed hybrid simulation model.

4.3.2.2.1 Implementation of Worker Behaviours

In the hybrid models, HITs being published in MTurk are passive objects as they are managed by MTurk and influenced by worker behaviours. In contrast, workers are active objects as they generate their own behaviour autonomously to interact with MTurk and HITs (Siebers et al., 2010). In the experiments, each HIT catcher worker agent can be initialised with the ScriptWorker class. Several functions are defined within each worker class to construct the HIT completion and acceptance behaviour (Figure 4.12). Similarly, non-HIT catcher worker

[24] Simply Process Interaction: https://simpy.readthedocs.io/en/latest/simpy_intro/process_interaction.html
[25] Simply Core Event Types: https://simpy.readthedocs.io/en/latest/api_reference/simpy.events.html

agents have similar functions, but with different approaches of implementation within each behaviour function.

```
class ScriptWorker:
  def __init__(self, env, worker_ID):···

  def RunScriptWorker(self):···

  def catchHIT_Loop_ScriptWorker(self):···
  # It simulates the process a script worker finishes the tasks one after another
  def completeHIT_Loop_ScriptWorker(self):···
  # HIT acceptance request for Script worker:
  def catchHIT_ScriptWorker(self):···

  def completeHIT_ScriptWorkers(self):···
#     print('completeHIT function FINISHED for Worker %s at %.2f' % (self.worker_ID, self.env.now))
  def cooldown(self, HIT_ID):···
```

Figure 4.12 A list of functions defined within a HIT catcher worker agent class.

Firstly, each worker agent is given a unique ID when it is initialised during experiments. Meanwhile, the id of HITs accepted and completed, number of attempts to catch HITs, and the times of failed attempts due to HIT abandonment are all recorded for data analysis afterward (Figure 4.13).

```
class ManualWorker:
  def __init__(self, env, worker_ID):
    """
    Store all necessary parameters for one worker instance
    """
    self.env = env
    self.worker_ID = worker_ID
    self.accepted_HITs = []
    self.finished_HITs = []
    self.HIT_queue = []
    self.Catching_Attempts = 0
    self.Catching_Failure_Due_To_Abandoning_Tasks = 0
```

Figure 4.13 Variables assigned and created when initialising a non-HIT catcher worker agent instance.

In the ScriptWorker class, two async threads are implemented for each agent to automatically accept HIT using HIT catcher and manually complete them (Figure 4.14). In contrast, the non-HIT catcher worker does not perform both behaviours simultaneously because they need to accept HITs manually.

```
 def RunScriptWorker(self):
#     print('Script Worker named %s is initialized at %.2f.' % (self.worker_ID, self.env.now))

    self.env.process(self.catchHIT_Loop_ScriptWorker())
    self.env.process(self.completeHIT_Loop_ScriptWorker())
    return
```

Figure 4.14 Async processes within each HIT catcher worker agent

It is worth noting that when a worker attempts to accept a HIT, they first check whether there is an empty slot in their HIT queue. Subsequently, given the presence of available HITs in MTurk, all the HITs that are eligible for this worker to accept are determined based on currently available HITs and their HIT acceptance history (Figure 4.15). Then a catching attempt is made towards one of the available HITs. One potential cause of failed HIT catching attempts is the competition from others. Someone else could have just accepted the same HIT meanwhile this worker is waiting for server response. However, if this catching attempt fails in the presence of available HITs, we can say this is one failed attempt due to worker's abandoning too many HITs.

```
# Check if there are available HITs:
if HITS_AVAILABLE:
  self.Catching_Attempts += 1
  # First create attempt_list by duplicating the HITS_AVAILABLE,
  attempt_list = HITS_AVAILABLE.copy()

  # Then remove all the accepted_HITs from the attempt_list,
  for i in self.accepted_HITs:
    if i in attempt_list:
      attempt_list.remove(i)

  # Then traverse all the IDs within this attempt_list.
  for HIT_ID in attempt_list:
    yield self.env.timeout(SERVER_RESPONSE_TIME)
    # Try to add this task into this worker's queue. Someone else might grab it during the server response time.
    try:···

    except Exception as e:···
  else:
      print('Worker %s: Existing tasks are all accepted before.' % self.worker_ID)
    # Record this as Catching_Failure_Due_To_Abandoning_Tasks
    self.Catching_Failure_Due_To_Abandoning_Tasks += 1
```

Figure 4.15 Illustration of the server process when a worker agent tries to accept a HIT.

The behaviour of the two worker agents after the successful acceptance of a new HIT has been implemented in different ways. Specifically, regarding the non-HIT catcher worker agents, they would try to complete this HIT by calling *self.completeHIT_ManualWorker()* right after adding a new HIT into their HIT queues (Figure 4.16). In the simulation, each time a worker agent starts doing one HIT, the system generates a random number as explained in the overview of hybrid simulation model section. The time allotted for each HIT was set to 1 minute, which means a HIT can only stay in one's HIT queue for 1 minute. Limiting the time allotted to 1 minute is to avoid a long backlog of HITs being reserved by workers. Then the HIT is determined to be successfully submitted if the time spent by the worker to complete the HIT is less than the time allotted for the HIT (1 min). If the worker fails to submit the HIT in time, this HIT would enter a cooldown period by running *self.env.process(self.cooldown(HIT_ID))* and then becomes temporarily unavailable to all worker agents.

```
# Record this successful attempt with timestamp
Attempt_List_ManualWorker[int(self.worker_ID)-1].append([HIT_ID,round(self.env.now,2)])

# print('HIT %s goes into Worker %s\'s HIT queue at %.2f.' % (HIT_ID, self.worker_ID, self.env.now))

# Check if worker can finish it within expiration time
complete_process = self.env.process(self.completeHIT_ManualWorker())
result = yield complete_process | self.env.timeout(TIME_ALLOTTED)
# print('*********************** Result generated*******************')

if complete_process in result:
  # print('HIT %s completed from Worker %s\'s HIT queue at %.2f.' % (HIT_ID, self.worker_ID, self.env.now)
  try: ...
  except Exception as e: ...

else:
  # print('HIT %s failed to be submitted by Worker %s before its expiration at %.2f.' % (HIT_ID, self.work
  self.env.process(self.cooldown(HIT_ID))
# Break the traverse loop to stop accepting HIT from this attempt

# print('Worker %s: catchHIT function finished.' % self.worker_ID)
return
```

Figure 4.16 Non-HIT catcher worker start doing HIT immediately after accepting this HIT.

In comparison, the HIT acceptance and completion behaviours are implemented in two async threads. Therefore, after a successful HIT acceptance, the timeout countdown will be turned on immediately. If this HIT has still not been submitted at the end of the countdown, it automatically enters a cooldown period (Figure 4.17).

```
# Record this successful attempt with timestamp
Attempt_List_ScriptWorker[abs(int(self.worker_ID))-1].append([HIT_ID,round(self.env.now,2)])

# print('HIT %s goes into Worker %s\'s HIT queue at %.2f.' % (HIT_ID, self.worker_ID, self.env.now))

# Go through the expiration process
yield self.env.timeout(TIME_ALLOTTED)
# print('Worker %s: Time is up for HIT %s, will check if it has been completed.' % (self.worker_ID, HIT_ID))

try:
  # Check if worker already submitted HIT
  for i in self.HIT_queue:
    if HIT_ID == i[0]:
      # print('HIT %s expires from Worker %s\'s queue at %.2f.' % (HIT_ID, self.worker_ID, self.env.now))
      yield self.env.process(self.cooldown(HIT_ID))
except Exception as e:
  print('Error occurred at the end of catchHIT_ScriptWorker()',str(e))
# End this for loop
return
```

Figure 4.17 HIT completion behaviours are not included in the HIT catching function for HIT catcher workers.

Regarding the HIT catcher workers, their HIT completion behaviour can be simulated via an independent thread as they do not need to be interrupted by HIT acceptance behaviour during work (Figure 4.18). When there exists at least one HIT they can work on, the time spent for completing this HIT can be generated by calling *getCompletionTime()*. Since this worker's HIT queue may contain more than one HITs, the worker will choose to complete those HITs that have enough time left to complete by comparing their remaining time.

```
def completeHIT_ScriptWorkers(self):
    """
    It simulates the process that worker finishes a task from HIT queue.
    """
    global Timestamps_Of_Successful_Submission_ScriptWorkers
    # print('completeHIT function STARTED for Worker %s at %.2f' % (self.worker_ID, self.env.now))
    # Check if there is available HIT to finish
    if self.HIT_queue:
        # print('Here is the queue for Worker %s' % self.worker_ID)
        # print(self.HIT_queue)
        # Work on the HIT that was accepted first or first check the time allotted for each HIT.
        submit_time = getCompletionTime()
        # print('It will take Worker %s %s to complete one HIT.' % (self.worker_ID, submit_time))
        # Go through the queue to look for the HIT that has enough time to complete
        for i in self.HIT_queue:
            # If this HIT has enough time remaining to complete:
            # TIME_ALLOTTED - (current time - time of acceptance) > submit_time
            if TIME_ALLOTTED - (self.env.now - i[1]) > submit_time:
                yield self.env.timeout(submit_time)
                # Once finished, this HIT will be removed from queue
                try:...
                except Exception as e:...
                break
            else:
                continue
        else:...
```

Figure 4.18 HIT completion behaviours for HIT catcher workers.

It can be revealed from the definition of cooldown function in Figure 4.19 that a HIT starting the cooldown period will be removed from the worker's HIT queue in the beginning. After the completion of the cooldown period, this HIT is added to the list of all available HITs and thus can be accepted by other workers.

```python
def cooldown(self, HIT_ID):
  """
  Simulate the process that one HIT cools down after getting expired from one's HIT queue
  """
  global HITS_AVAILABLE
    print('Worker %s: Before cooldown, HIT_queue is:' % self.worker_ID)
    print(self.HIT_queue)
  # Remove it from queue
  try:
    # Look for the element in HIT_queue with the input HIT_ID
    for i in self.HIT_queue:
      if HIT_ID == i[0]:
        self.HIT_queue.remove(i)
        yield self.env.timeout(COOLDOWN_TIME)
        HITS_AVAILABLE.append(HIT_ID)
        # print('HIT %s has been returned from Worker %s\'s queue to batch pool after cool
        return
  except Exception as e:
    print('An error occurred when coolling down HIT %s' % HIT_ID, str(e))
```

Figure 4.19 Simulation of HIT cooling down period.

4.3.2.2.2 Data Flow Within Simulations

The following diagrams illustrate the data flows involved in the acceptance and submission behaviour of workers in the simulation (Figure 4.20, Figure 4.21). Specifically, when a worker agent successfully accepts a HIT, the HIT's ID information is removed from a global variable list called *HITS_AVAILABLE* that stores all available HITs and added to the HIT queue of this agent. Meanwhile, this HIT is also recorded in this agent's accepted HITs list, thus preventing them from accepting this HIT again after abandoning it. In addition, the simulator continuously records the total number of HIT acceptances for both types of workers, thus calculating their cumulative number of acceptances at different times.

If this worker agent fails to accept one HIT due to HIT abandonment (HIT expired in their HIT queue before), this failed attempt will be recorded. The number of failed HIT catching attempts due to HIT abandonment for each worker is helpful in evaluating the impact of the use of HIT catchers on their job opportunities. In other words, one worker might have fewer job opportunities after getting too many HITs expired from their HIT queue.

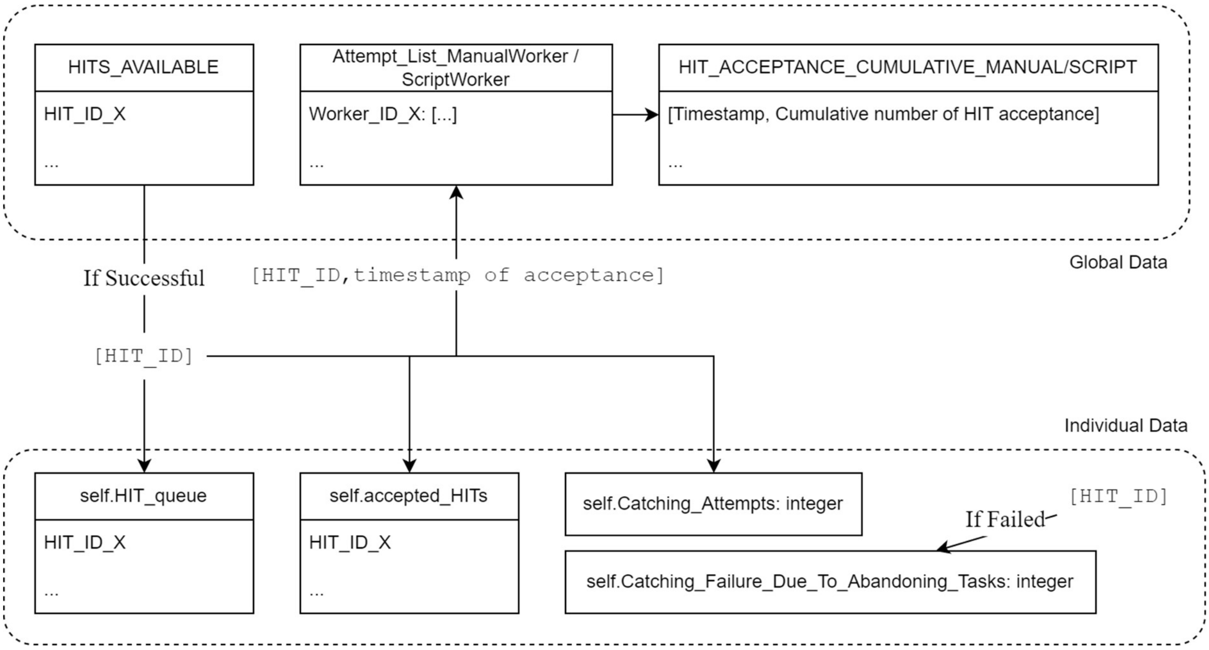

Figure 4.20 Data flow diagram of HIT acceptance

When a worker agent submits a HIT successfully, the ID of this HIT will be used to calculate the worker HIT diversity in the approaching period. Meanwhile, *HIT_ID* and the timestamp of submission are both appended to a list containing all successful submissions for two types of workers. Furthermore, the completion rate of the HIT group at different time points (*COMPLETION_LOG*) are calculated based on a list containing all HIT submissions (*SUBMISSION_LIST*). However, if the worker agent fails to submit a HIT before its expiration, The ID of this HIT will be appended to the *HITS_AVAILABLE* list after going through a cooldown period.

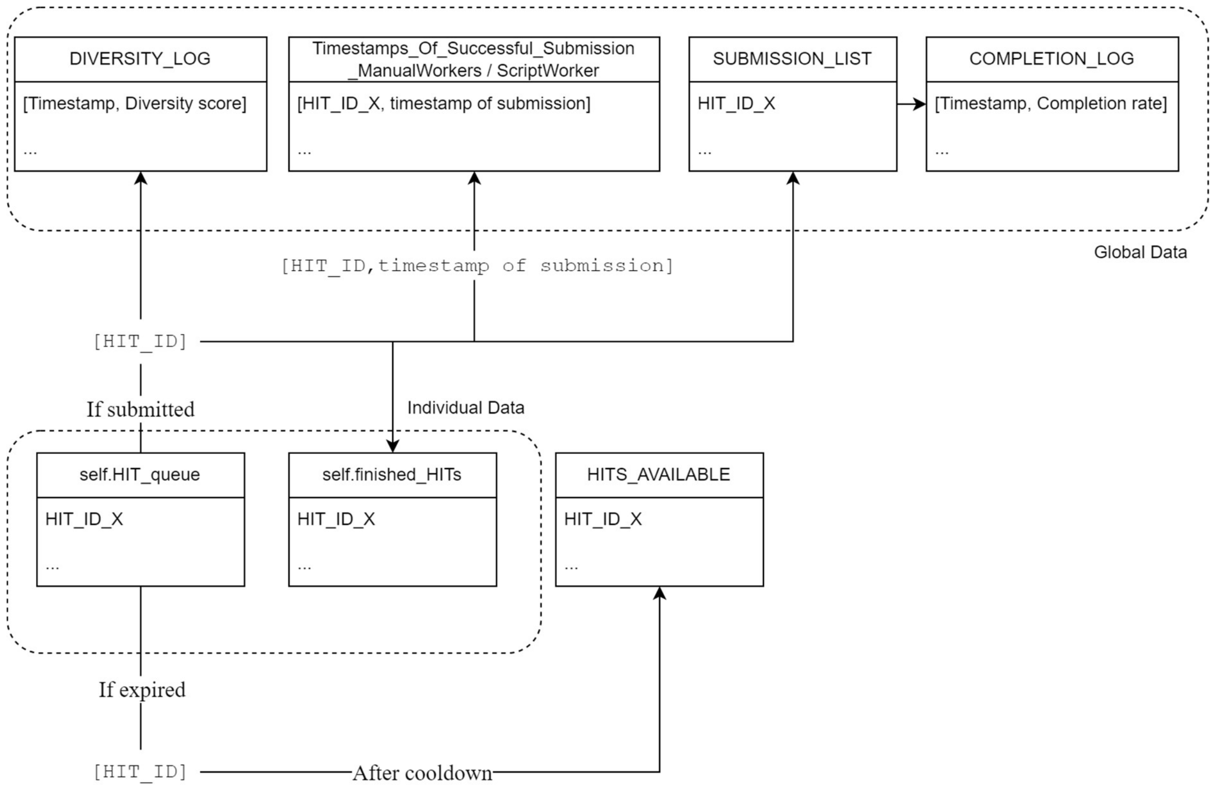

Figure 4.21 Data flow diagram of HIT submission

Through running the implemented simulation program, the output indicators related to the HIT groups can be generated, including the speed of completion at different time stages in the simulation, the total completion time and the response diversity. In addition, output indicators related to the workers could also be generated, including the number of HITs successfully received by each type of worker at different time stages in the simulation, and the number of successful completions. Furthermore, with the help of Google Collab programming platform, a user console has been created to facilitate the adjustment of parameters during the experiment (Figure 4.22).

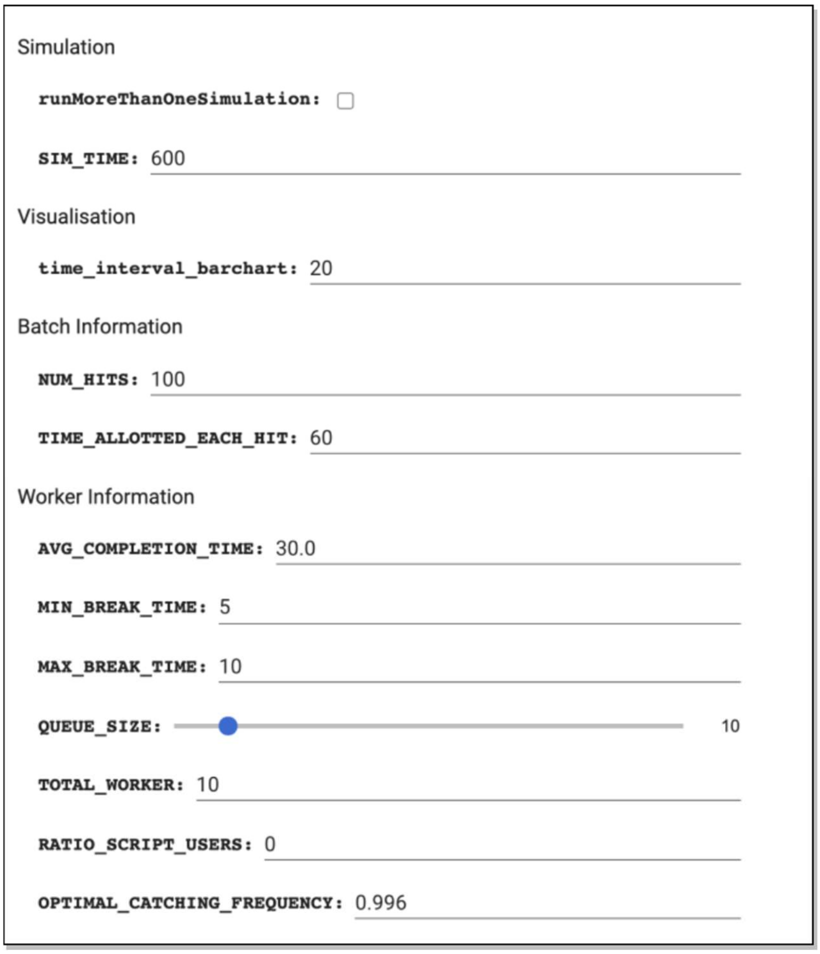

Figure 4.22 Example of the discrete event simulator configuration interface.

4.4 Sampling Method

4.4.1 Script Impact in Different Experimental Scales

The first group of simulations obtains data by gradually increasing the size of the experiment, which is increasing the number of HITs included in a batch and the number of workers by the same proportion (10:1), while controlling for the proportion of workers of both types and the ratio of the number of workers to the number of HITs. The aim is to explore how the unintended consequences of the use of the HIT catching tools on the workers and the HITs themselves change as the size of the experiment increases. More specifically, whether the gap between the job opportunities of the two types of workers is further magnified, whether the overall completion time was delayed, and whether the entire HIT group was completed with more workers involved.

4.4.2 Script Impact in Different Percentages of HIT Catcher Workers

The use of scripts gives workers a technical advantage, but what would be the impact on batch completion if there are more workers using scripts? To answer this question, simulations were conducted by increasing the percentage of HIT catcher workers from 0% to 90%, each time by 10%, by maintaining a total of 500 HITs and 50 workers. It is worth noting that scaling to 100% would result in no one being able to finalise the remaining HITs. As mentioned earlier, the technical rule of the platform is that workers cannot repeatedly accept the same HIT with the same Assignment ID. In other words, the remaining unacceptable HITs have been made to expire by all the workers using the HIT catchers. Therefore, this particular case was not considered in this simulation.

4.5 Data Analysis

4.5.1 Script Impact vs. Experiment Scale

This section is dedicated to explaining the data from the first simulation and it states that: as the numbers of HITs and the number of workers grows in a constant ratio of 10:1, the unintended influences of using scripts on the completion time of the HIT, the diversity of the data and the job opportunities for non-scripted workers are further amplified.

As can be illustrated in Table 4.3, the negative impact of the use of HIT catcher is gradually magnified as the scale of simulation increases. Fairness of catching HITs also decreases from around 0.88 to 0.65. The average number of HITs completed per non-HIT catcher worker also dropped from 4.8 to 2.1. In other words, as the batch size increases, the script deprives the average worker of more and more job opportunities. Based on this trend, we can gain a more tangible and quantifiable understanding of the impact of catching scripts on the platform in real environments with thousands of workers involved at the same time. In addition, since the standard errors of the non-HIT catcher workers' results are all less than 0.1, the variations of these results are not presented in the table.

Table 4.3 Summary of simulation statistics under 5 experimental scales

Batch Size	Total Number of Workers	Counts of HIT Acceptance per Worker		Counts of HIT Submission per Worker		Time of Batch Completion (s)	Worker-HIT Diversity
		Manual Worker	Script Worker	Manual Worker	Script Worker		
100	10	9.0	56.6 ± 1.2	8.4	11.1 ± 0.2	631.5 ± 15.2	87.9% ± 0.8%
200	20	6.9	52.7 ± 0.8	6.7	13.3 ± 0.1	727.9 ± 24.2	77.2% ± 0.8%
300	30	6.5	45.8 ± 0.9	6.3	14.3 ± 0.1	749.2 ± 13.7	72.4% ± 0.7%
400	40	5.4	38.2 ± 0.3	5.1	15.2 ± 0.1	773.4 ± 12.2	68.4% ± 0.6%
500	50	4.2	33.9 ± 0.3	4.0	16.0 ± 0.1	811.1 ± 14.3	65.2% ± 0.3%

Figure 4.23 Counts of HIT acceptance over time for both types of workers under the batch size of indicates the cumulative number of HIT acceptance for both types of workers when batch size is 100. Due to the reason that the HIT catcher workers caught HITs aggressively at the beginning of the simulation, many HITs expired from their queues before they even opened them, and they cannot re-accept them later anymore (unintentional HIT abandonment). What is worse, during the time they hold these HITs till expiration, the non-HIT catcher workers had to wait due to the lack of available HITs. As can be noted from Figure 4.23, within less than 100 seconds of the start of the simulation, the script workers have reserved the majority of the 100 HITs. Such a phenomenon is the tragedy of the commons for both the workers and the requesters. Specifically, it takes away the HIT opportunities from non-HIT catcher workers, extends the batch completion time, and wastes their own HIT catching opportunities due to unintentional HIT abandonment.

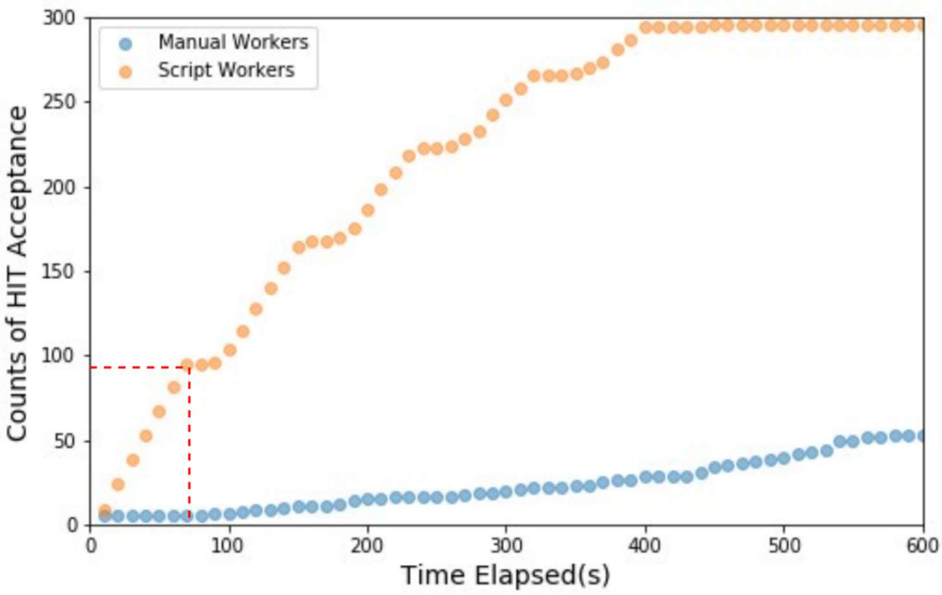

Figure 4.23 Counts of HIT acceptance over time for both types of workers under the batch size of 100.

However, with the increase of the experimental size, the ratio of HIT submission over acceptance for the HIT catcher workers increases (Table 4.3). As the number of total HITs increases, the penalty effect on workers for abandoning HITs becomes smaller. In other words, the impact of the tragedy of commons on HIT catcher workers reduces. This leads to an increasing gain from high-frequency HIT catching, and thus their ratio of HIT submission over acceptance goes up. In real life, if one HIT catcher worker can abandon many HITs without penalties, they tend to be aggressive in using catching scripts[26].

It can also be noticed from Table 4.3 that the average number of HITs completed by simulated HIT catcher workers gradually increases as the size of the experiment increases. This directly leads to a decrease in HIT-worker diversity due to the increasing proportion of HITs completed by HIT catcher workers, who make up half of the total workforce. Dennis et al. (2019) collected "disturbingly low-quality responses" in their experiments on MTurk and expressed concerns about the reliability of MTurk data. Whether or not the low HIT-worker diversity is one cause of the low data quality should be further investigated in future research.

[26] Reddit forum post: www.reddit.com/r/mturk/comments/g9tt14/how_to_catch_hits_more_quickly_on_hit_finder/

4.5.2 Script Impact vs. Percentage of HIT Catcher Workers

Figure 4.24 shows that the technical advantage of catching scripts for users is most significant when the percentage of HIT catcher workers reaches 20%, with an average of around 30 HITs submitted by each scripted worker, while it is only 5 for each non-HIT catcher worker. At the same time, the diversity of data is at its lowest, which is around 54% (Figure 4.25). This is because more than half of the total HITs are completed by only 20% of total workers. Meanwhile, due to the large number of HITs being reserved by a very small number of scripted workers, the non-HIT catcher workers cannot consistently catch the HITs, resulting in more than double the completion time (from 544.36s to 1136.96s) compared to if there were only non-HIT catcher workers in the simulation (Figure 4.25).

Interestingly, however, as more workers use scripts for automatic HIT acceptance, the diversity of data gradually returns. When 90% of all workers were using scripts, batch diversity returned to its initial level (88%), the same as it would have been with all non-HIT catcher workers (89%). The increase in diversity was accompanied by a gradual decrease in total completion time, reaching the second shortest time after the simulation with all non-HIT catcher workers at 70%. It indicates that the impact of catching scripts on the batch diversity decreases as it becomes more prevalent among all workers. However, it still has a great impact on total completion time due to consistently reserving too many HITs.

Regarding the batch completion time presented in Figure 4.25, when the percentage of HIT catcher workers is 0%, there is no one reserving multiple HITs at the same time with scripts. Therefore, the batch completion time is the lowest compared with other percentages because almost no one gets delayed in their work by a lack of acceptable HITs. When there are 10% of all the workers using catching scripts, the script has the greatest positive impact on the users. With a small number of script competitors and adequate number of acceptable HITs, each HIT catcher worker can accept as many HITs as possible without being affected by the platform's restrictions on accepting the same HITs repeatedly. But the drawback is that all the non-HIT catcher workers, who make up 90% of the total workforce, are affected by the difficulty of catching HITs and have to slow down their work.

As more and more workers use scripts to catch HITs, there are less workers who are affected by the difficulty of catching HITs. Therefore, the batch completion time reduces after the percentage of HIT catcher workers increases from 10%. However, when 70% of the total workforce are all using scripts, the use of scripts can have a far more negative impact on non-HIT catcher workers than their positive impact on the HIT catcher workers themselves, thus reducing the overall speed of batch completion.

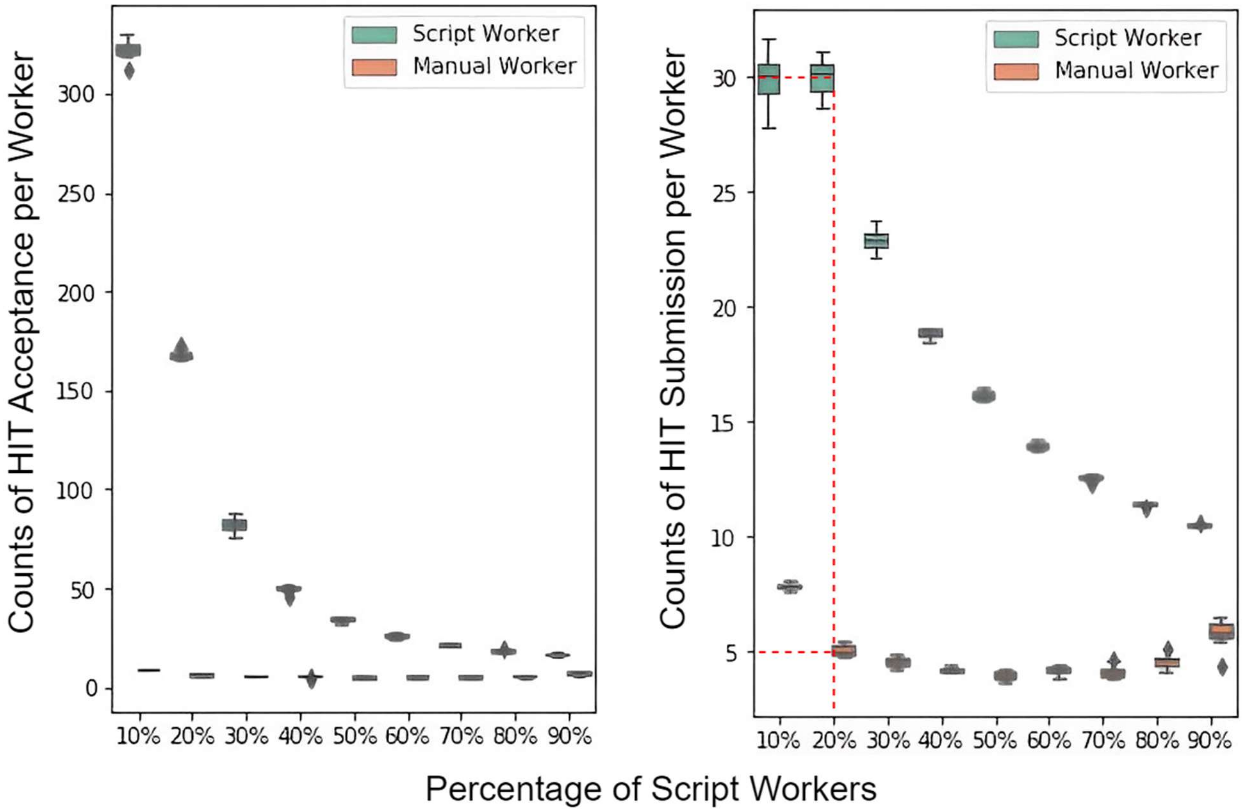

Figure 4.24 Worker related statistics under different percentages of HIT catcher workers (batch size = 500)

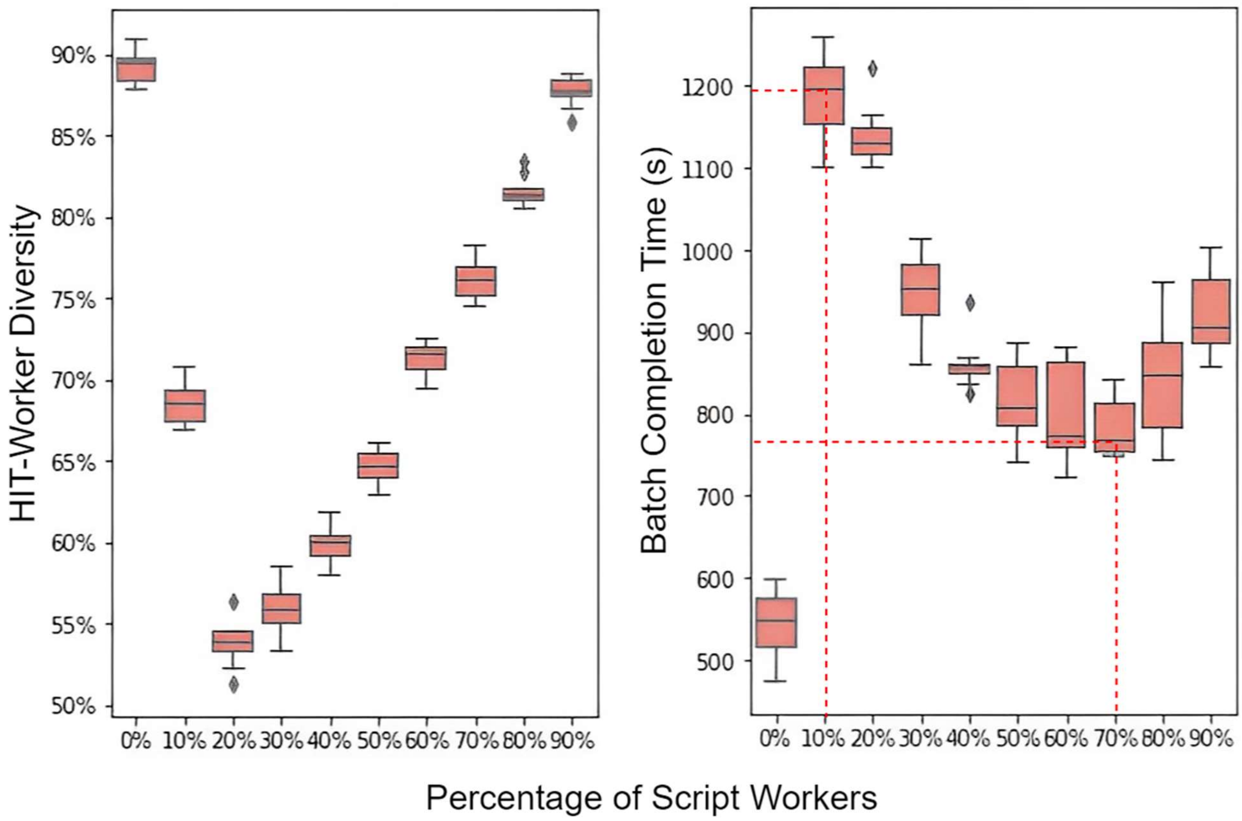

Figure 4.25 Batch related statistics under different percentages of HIT catcher workers (batch size = 500)

4.6 Findings

Based on the data analysis, we come up with the following key findings:

Script impact and experiment scale: as the number of microtasks and workers in a HIT group grows in equal proportions, HIT catcher workers gradually deprive those not using tools of their job opportunities. This not only resulted in longer overall completion time of the whole HIT group, but also in less HIT-worker diversity.

Script impact and percentage of HIT catcher workers: The technical advantage resulted in the most HIT completions only when very few workers (approximately 20% in this study) used scripting. This also resulted in the lowest diversity between HIT and workers and the longest completion times for the HIT group. However, as more workers used HIT catchers, the diversity of the data gradually recovered and the overall completion time decreased, but was still affected by the use of scripts.

Tragedy of the commons: Over-reliance on HIT Catchers can negatively impact workers, job requesters and the platform, leading to a "tragedy of the commons" situation.

4.7 Discussion

4.7.1 Theoretical Contributions

This study makes some important contributions to the crowdworking literature and platform studies.

First, previous crowdworking studies with an interest on the working conditions of crowdworkers, tended to explore the phenomenon from a regulatory perspective and explore how the lack of a clear regulatory framework leaves crowdworkers exposed to low wages, job insecurity and lack of opportunities for collective action or organising (Altenried, 2020; Gerber, 2021). Other scholars focussed on platform design and explored how platform features influence the power distribution between workers and job requesters (Fieseler et al., 2019; Irani & Silberman, 2013). To date, however, little attention has been paid towards analysing how the openness of the platforms to third party applications, such as automated catching scripts may create and further exacerbate less than ideal working conditions and inequalities among the crowdworkers themselves.

Wessel et al. (2017) have indicated that platform openness may be a source of innovation and may make the platform more attractive, but at the same time, it can be a source of risks, whereby insufficient control over third parties may destabilise the platform. Our findings extend our current knowledge with regards to how such openness may operate within a crowdsourcing platform context whereby the openness to the use of automated HIT catchers, provided by third parties, negatively influences the working conditions. We further quantify the impacts on the platform's participants in the short term and specifically show how the use of automated catching scripts impact the HIT acceptance strategies in the short term. Namely, we show that more than half of the total HITs may be completed by only one fifth of total workers, impacting batch diversity and significantly restricting the job opportunities of manual workers. In the longer term, script workers who have completed more HITs, will have improved their ability to use scripts and they will have enhanced their personal ranking. As such, competence and reputation persistence will lead to a continuously widening gap between script workers and manual workers, benefitting script workers at the expense of manual workers.

Second, our study enriches the growing body of literature on the impact of algorithmic control and working conditions. Gol et al. (2019) argue that the reputation system that feeds on HIT completion and approval is a good estimate for future performance, allows job requesters to verify crowdworkers' credentials and supports platforms to exercise the appropriate level of control for governance purposes. However, Wood et al. (2019) expressed concerns with regards to the consequences of such algorithmic control, where job requesters are able to identify 'quality' workers on the basis of the number of HITs completed. Indeed, job requesters tend to set very high acceptance criteria for the HITs they publish to filter out less experienced workers from the large labour pool as quickly as possible (Waldkirch et al., 2021). Our findings show that the reputation system that considers HIT completion and approval rates, is susceptible to the vicious impacts of the Matthew effect, where the extensive use of HIT catchers have adverse impacts on the opportunities allowed for manual workers and new workers in general. New workers in particular, are required to complete a high number of HITs as quickly as possible to attain an acceptable score in order to be later considered for higher quality HITs. In other words, they may find themselves completing a large volume of low-quality HITs, which are low rewarding and/or posted by less reputable job requesters (Savage et al., 2020), and thus risk completing HITs that may not be approved and thus not be rewarded (Kwek, 2020).

Based on our findings, we posit that this type of algorithmic control, over time, has a negative impact on the platform, as well. Newcomers to the platform become discouraged due to the indirect obstacles imposed by design (Brawley and Pury, 2016), and those who do not use HIT catchers, are more likely to abandon the platform altogether. Previous study has underlined that high turnover rate is indeed a threat for crowdsourcing platforms because job requesters may not be able to obtain high quality results for sufficiently low costs (Deng et al., 2016). In short, Matthew effect not only ultimately force newcomers and manual workers to leave the platform, but can potentially lead to the collapse of the platform itself.

4.7.2 Implications for Practice

Besides the theoretical implications, our study makes some important contributions to practice, as our findings can be used for considering platform openness and more crucially, informing platform design and automated HIT catchers.

This study furthers our understanding of the platform's task release mechanism, including the progressive publishing of microtasks within a HIT group. In addition, tasks that have expired will go through a cooling down period before being re-published.

This study also expands our understanding of the technical advantages from the use of HIT catchers. Specifically, HIT catchers not only help users to automatically accept potentially available HITs at a high frequency, but also give users higher permissions: HITs that have just been published can be accepted by HIT catchers in advance based on project ID before they become visible later on the platform.

Based on the above simulations, we also analysed potential countermeasures to reduce the negative impacts of tool use, such as adjusting the catching frequency and HIT queue length. It was found from these simulation results that: (i) the limitation on the script catching frequency could help improve the diversity of data and reduce the time of batch completion. (ii) As the batch size grows, the number of HITs that one worker can abandon increases, and the platform's rule against accepting the same HIT repeatedly has a less penalising effect.

As Hanrahan et al. (2018) explain, to avoid the HITs being over-accepted by the catching scripts, job requesters can limit the time allotted of each HIT. This would amplify the punishing effect of the platform's prohibition on workers accepting the same HIT repeatedly, thus prompting them to reduce the catching frequency used by their scripts and deterring them from securing an excessive number of HITs, which ultimately expire in their HIT queues. Of course, limiting the maximum number of completions per worker is the most direct and effective way to improve batch diversity.

Our study further shows that it is imperative for the sustainability and healthy growth of crowdsourcing platforms to identify ways on how third-party contributors can be encouraged to help improve the functionality of the platform while avoiding the unintended consequences on the platform’s stakeholders and the diversity of the data. Besides applying upper limits on HIT completion per worker, crowdsourcing platforms can draw inspiration and lessons learned from the strategies typically employed by online retailers in e.g., the sneaker and ticket industries, and potentially adjust the ways HITs are assigned. For example, the platform could first receive sign-ups from all workers interested in the batch over a period, and then use a lottery or equal distribution to assign the HITs. While such an approach would impact the overall completion time of the batch, the resulting delays would not exceed those currently observed due to non-completions, and the HITs could be assigned more fairly, thus ensuring better opportunities for the majority of the crowdworkers.

4.7.3 Limitation Due to Assumptions about Worker Behaviour

The simulation framework includes multiple assumptions (described in Section 4.3.2.1.2) regarding their work strategies, numbers of workers involved, time spent completing each HIT and time allotted for each HIT. Due to these assumptions, the experiment result contains biases that differ from the real situation, which is a common limitation of studies using simulations (Davis & Marcus, 2016). This requires more empirical research to provide a more accurate and detailed description and classification of worker behaviours. Thus, the simulation functions can be further enriched to be more in line with the diverse work strategies in real life.

4.8 Chapter Summary

The aim of Chapter 4 is to investigate the unintended consequences of the use of HIT catchers on users, other workers, and on the target HIT group, and thus indirectly on the crowdsourcing platform. The analysis reveals that HIT catcher workers gain far more work opportunities within a HIT group than non-HIT catcher workers, thanks to the technical advantages that scripts provide them with. Compared with the non-HIT catcher workers, the additional HIT submissions enhance HIT catcher workers' competence persistence, while the increase of this important statistic on their worker account might contribute in the long run to the accumulation of their reputational persistence. In other words, the technical advantage that scripts offer further increases HIT catcher workers' job opportunities in the long-term due to the Matthew effect (Figure 4.26). Moreover, consistent with our hypothesis, the additional gains for workers with higher-skilled ranking come at the expense of gains for workers with lower-skilled ranking. HIT catcher workers were able to access additional HIT opportunities, while depriving non-HIT catcher workers of equitable opportunities to access HITs.

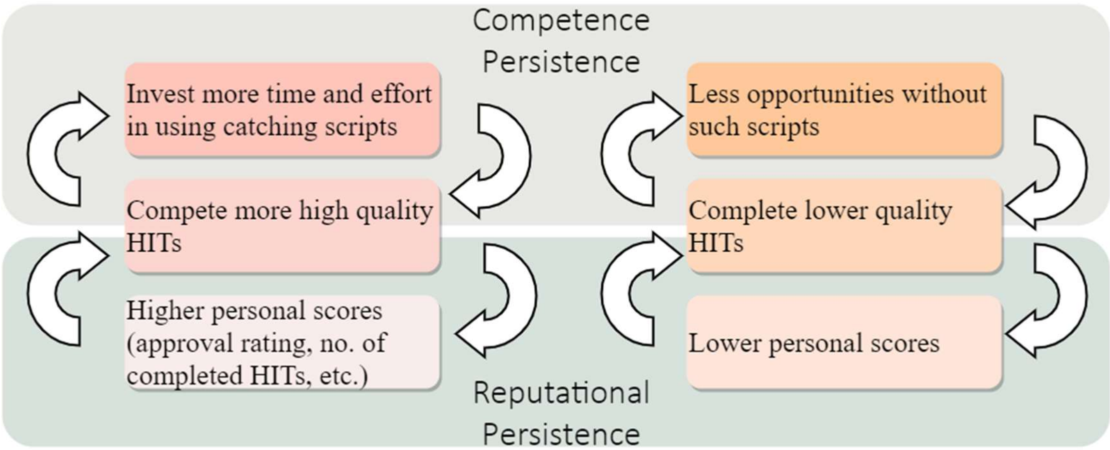

Figure 4.26 The virtuous (on the left) and vicious (on the right) Matthew Effect for crowd workers.

Our study also reveals the existence of the tragedy of the commons (Greco & Floridi, 2004). The over-acceptance of HITs by HIT catcher workers deprives non-HIT catcher workers of work opportunities, which in turn slows down the overall completion of the batch, thus leading to inevitable HIT abandonment, resulting to HIT catcher workers' damaging their own future work and skill building opportunities (Figure 4.27). Worse still, HIT abandonment rate is a qualification used by requester to filter workers (Hara et al., 2018). Excessive HIT abandonments would keep them from getting more job opportunities. The negative impacts of script use multiply as the scale of the batch increases, while the impact of the tragedy of the commons is magnified where the number of HITs waiting to be completed increase. However, we also found that a high prevalence of script use among workers can mitigate the impacts on batch diversity, because of the added competition between them.

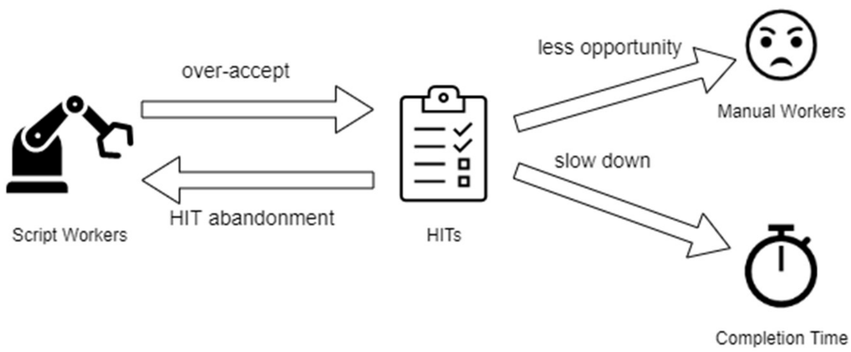

Figure 4.27 The tragedy of the commons caused by task over-acceptance of scripts.

The study in this chapter has explored and quantified the unintended consequences caused by the HIT catcher. More importantly, the data collection and analysis of this chapter is an important reference for the upcoming data collection from real HITs in the next chapter. In addition, this chapter has identified the use of HIT catchers is one cause of a reduction in the diversity of results. However, along with reduced data diversity, does the use of HIT catchers also have an impact on data quality? Furthermore, does the diverse working strategy of workers contribute to or mitigate the unintended consequences of the use of HIT catchers? These hypotheses cannot be tested by the simulation framework constructed in this chapter.

To conclude, our study shows that the use of automated HIT catching scripts can be beneficial for those crowdworkers that use them, but only in the very short term. Their excessive use leads to unintended consequences for all crowdworkers, the job requesters and the platform itself. Manual workers are left with few opportunities to increase their income and are likely to exit the platform. Progressively, job requesters may become disillusioned with the low diversity in completed HITs and the high turnover rates. In turn, these will impact the sustainability of the platform, as job requesters will be less likely to prefer it for posting HITs because supply will not be able to meet demand requirements. Our findings can inform script designers and crowdsourcing platforms to mitigate these unintended consequences, with the view not only to ensure the sustainability of the platforms but also ensure that crowdworkers enjoy better working conditions and equal opportunities.

The next chapter aims to validate the observations from the simulation study and to enrich the understanding of crowdworker behaviour under the influence of HIT catchers through experiments based on real HITs published on MTurk. These observations include the differences in job opportunities, abandonment behaviour towards HITs, the number of final tasks completed between HIT catcher workers and non-HIT catcher workers. In addition, the impact of the use of HIT catchers on the completion process to the HIT group will be further examined, including the speed of completion and the diversity of results.

Chapter 5 Crowdwork Strategies with the Aid of HIT Catchers

5.1 Introduction

The previous chapter details the simulation experiments conducted to identify the unintended consequences of HIT catchers on crowdworkers as well as HIT results. The aim of this study is to expand the understanding of crowdwork strategies through investigating a real-life scenario, including how they decide the number of HITs accepted and whether there is multi-task behaviour. This enriches the understanding of crowdwork behaviour and seeks to validate the unintended consequences due to using HIT catchers revealed in the previous chapter.

One challenge in advancing our knowledge on the topic is that platforms and requesters cannot easily detect when such tools are used. Another challenge is to quantify the impact that the use of automated catching techniques has on job opportunities, data quality and workers' behaviour. We identify and gain insight into the share of workers who employ such techniques and analyse their behaviour. Furthermore, we empirically measure their impact to requesters and platforms in terms of answer diversity, task completion times, and annotation quality for different types of workers. Moreover, we employ novel measuring techniques to reconstruct the task access, reservation, and completion dynamics. We observed that some workers use the platform in various and unexpected ways to overcome the race to the bottom of the gig economy. They use aggressive catching techniques to reserve tasks, do multiple HITs simultaneously to maximise their hourly wage, and even share accounts with multiple people and devices to increase the chance of task reservation. These behaviours in turn increase the tragedy of the commons effects on the task availability and therefore reduce worker diversity. Section 1.1.8 has explained the HIT related events, thus helping to understand the purpose of the study and the meaning of the data regarding the HIT status.

This study aims to answer RQ1 by publishing real HITs in the platform: What are the impacts of the use of HIT catchers on HITs and crowdworkers? More specifically, to what extent it affects the HIT-worker diversity and completion time of the HIT group? What are the effects on HIT availability, backlog, and expiration within the experiment? Furthermore, regarding the crowdworkers, how are their job opportunities get affected? What are the differences in work strategies between those using and not using HIT catchers? Finally, how does the use of HIT catchers influence the quality of HIT results? These detailed questions were answered through an experiment that makes use of novel monitoring techniques to correlate data from MTurk and the workers, allowing us to reconstruct the HIT reservation and backlog dynamics.

5.2 Methods

An experiment using real HITs posted on MTurk was designed and prepared for data collection from real crowdworkers. This section begins with an introduction to the HITs posted for the experiment, including the size of the HIT group, the HIT page layout, the content of the questions the workers were asked to answer, and the images used for annotation. Subsequently, the processing of the worker responses to the HITs and the calculation of the annotation accuracy are discussed.

Next, the content and collection methods for information about workers are illustrated, including how to detect whether a specific HIT catcher is installed and how to collect focus time on a HIT page. The information collection process regarding the HIT events is then explained. This includes the definition of different HIT events and how the information gets retrieved.

5.2.1 HIT Design

A HIT group containing 1000 HITs were designed through a pilot experiment, estimated to require a median time of 3 minutes to complete. The only difference between these HITs is the image for annotation. The web page conducting the HITs were designed and implemented by Dr Maddalena from University of Southampton and Dr Checco from the University of Sheffield. To measure the phenomena of HIT backlog, reservation, and expiration more easily at this relatively small scale, these experiment HITs were set to expire after 5 minutes (this time limit was revealed on the HIT page). In other words, each HIT that cannot be completed within 5 minutes was automatically removed from the current crowdworker's HIT queue and returned to the list of available HITs on MTurk. The experiment results are magnified by this extremely short HIT expiration time compared with the time applied in other studies. This makes the research phenomenon more visible and therefore facilitates the observation. The HIT itself includes an objective part, where the answer can be compared with the gold standard to get the annotation quality score. The HIT also includes a subjective part, which is useful to explore submission's diversity. This was done by calculating the diversity of the textual content of the subjective responses, which is an aspect of assessing the overall quality of task completion.

The task payment and acceptance rules have been designed so that the HIT group used for the experiment would follow all requirements of a sought-after HIT group (Savage et al., 2020). A common criterion is hourly pay should be over $8.29. In other words, this HIT group posted on MTurk could effectively attract HIT catchers by meeting their standards: a payment of more than $0.23 ($0.6 for each HIT in this study), an hourly wage above the median of $2 (estimated hourly wage was $12 for each HIT in this study). Moreover, there were no rejections of their HIT result submission. Finally, each HIT result was automatically approved 1 second after their submission by setting the auto-approval delay[27] before publishing the HIT group. This is because assessing the validity and accuracy of the task responses was an important factor to study, unlike other tasks, which only accept accurate responses.

This experiment received ethics approval (Application Reference Number: 041062) from the authors' institution on 27/06/2021 (Appendix A, Part 1). When starting, the HIT requires workers to read and accept a consent form. The HIT was presented to participants as depicted in Figure 5.1. The top of the page presents a collapsible panel with the instructions. Below the instructions, the working panel requires to label the city photos. The worker must select an item class first, and then click on the target items on the image canvas. Thus, a marker of the same colour of the item class label will be shown in correspondence of the click coordinates, and the items counter gets increased. A message is shown when the worker tries to label more than the maximum allowed limit of 10 items. Any of the added items can also be removed singularly, or as a group by clicking the "Remove all" button. Also, we include a distance control to avoid adding items too close or overlapping existing ones. Below the image annotation panel, the workers are asked to share a short sentence describing which feelings about the neighbourhood the image elicits. Finally, the worker is asked to provide some feedback about the HIT. To complete the task, the worker must provide at least one item in the canvas, and any non-blank text to the two text areas. Attempts of submission that do not satisfy these constraints trigger a notification.

[27] Approving and rejecting work:
https://docs.aws.amazon.com/AWSMechTurk/latest/AWSMechanicalTurkRequester/ApproveRejectWork.html

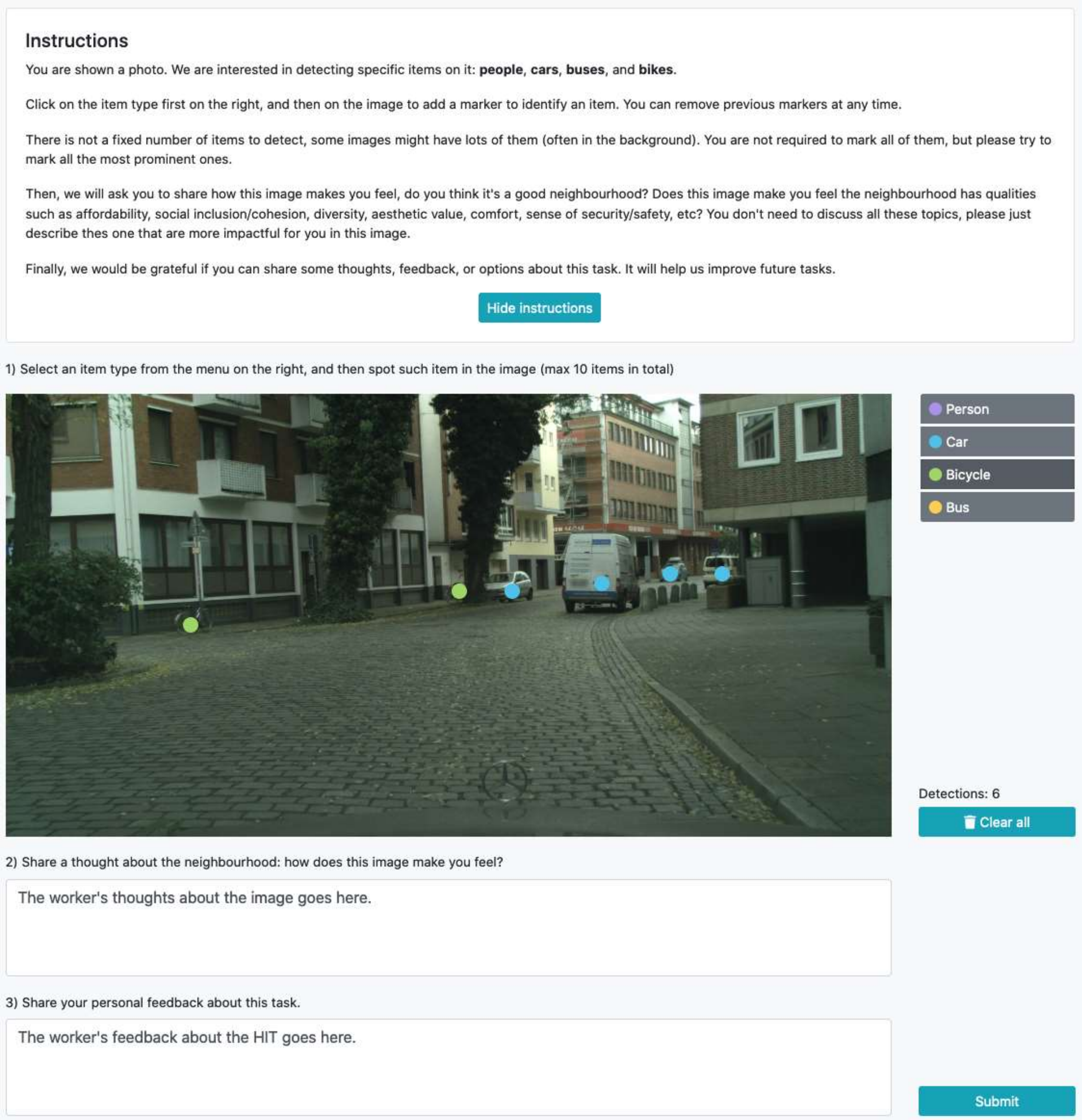

Figure 5.1 Interface of the HITs posted in experiment.

5.2.2 Image Content and Source

In the first part of the HIT, workers were asked to identify the positions of objects in street view images from Cityscapes (Cordts et al., 2016). Cityscapes is a dataset of images showing urban street scenes, taken from 50 cities, with labels from 30 classes. Annotation groups include human, vehicle, construction, nature, etc. Each annotation group contains more detailed categories, called classes. For example, the "nature" group includes "tree" and "terrain" classes, the "vehicle" group includes classes like "car", "bus", etc. In the provided dataset, each street view image (Figure 5.2) corresponds to two sets of annotations (coarse annotation and fine annotation). Each set of annotations contains a JSON file (Figure 5.5) and an image with polygons depicting different objects (Figure 5.3 and Figure 5.4). Specifically, the JSON file stores a set of coordinates of the polygons depicting the objects in the image. Furthermore, the image contains annotated polygons with different colours drawn according to the set of coordinates provided in the JSON file.

For all the 1000 HITs published in this study, each HIT contains a street view image randomly picked from the dataset. The fine annotation (Figure 5.3) was chosen as the gold standard for the annotation of each street view image. The coordinate data contained in the fine annotation JSON file enables accurate measurements of the annotation quality submitted by comparing whether the markers provided fall within the polygon range of each target object. To keep the complexity of image annotation low, four object classes were focused on: “person”, “car”, “bus”, and “bike”.

Figure 5.2 A random image of a street in Bremen (Cordts et al. 2016).

Figure 5.3 Finely annotated polygons with overlaid colours represent different classes of objects (Cordts et al. 2016).

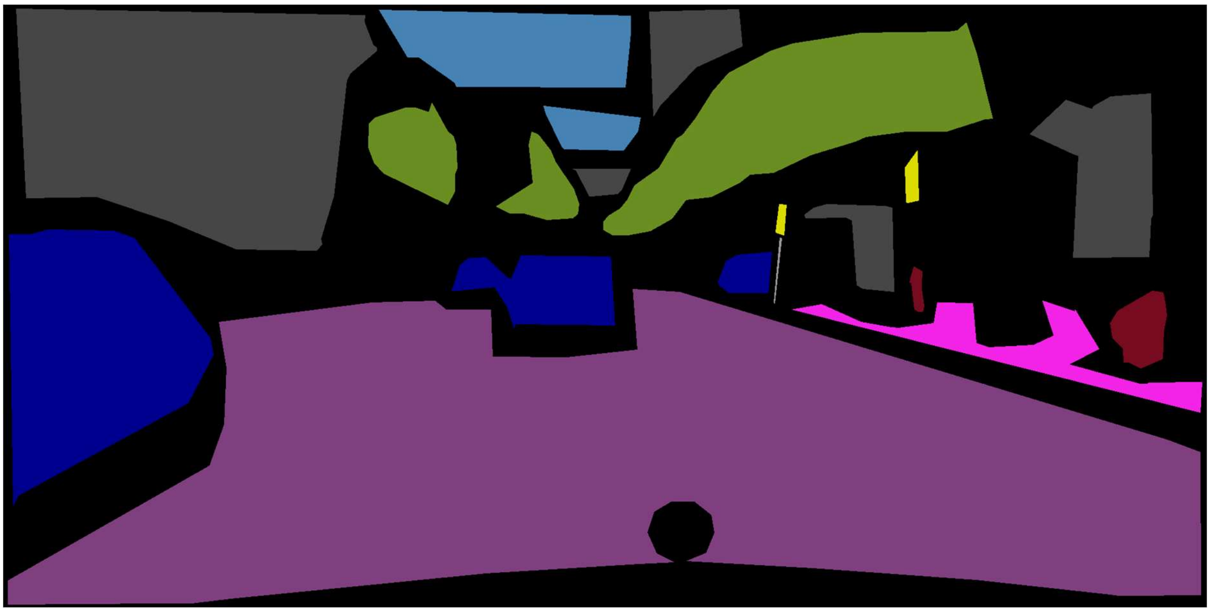

Figure 5.4 Coarsely annotated polygons with overlaid colours represent different classes of objects (Cordts et al. 2016).

```
{"imgHeight":1024,"imgWidth":2048,"objects":[
   {"label":"road","polygon":[[38,713],[32,713],[4,1011],[1825,997],[1368
      ,554],[1366,515],[1265,529],[1160,501],[579,591],[516,569],[298,565]
      ,[278,605],[351,611],[339,631],[339,647],[243,681]]},
   {"label":"sidewalk","polygon":[[1448,522],[1950,980],[2025,961],[2006
      ,888],[2049,675],[2034,670],[1902,573],[1907,550],[1943,515],[1943
      ,504],[1724,502],[1681,513]]},
   {"label":"car","polygon":[[1302,470],[1400,469],[1431,459],[1453,434]
      ,[1455,424],[1442,391],[1446,362],[1434,341],[1406,331],[1313,337]
      ,[1322,352],[1312,370],[1309,388],[1300,428],[1294,466]]},
   {"label":"vegetation","polygon":[[1372,10],[1349,13],[1311,30],[1278
      ,56],[1242,80],[1223,113],[1201,165],[1194,191],[1169,217],[1159
      ,236],[1138,266],[1130,275],[1121,278],[1098,278],[1091,273],[1083
      ,261],[1068,223],[1066,190],[1059,172],[1055,123],[1046,92],[1031
      ,61],[1031,38],[1026,24],[1345,3]]},
   {"label":"vegetation","polygon":[[935,9],[917,82],[894,135],[878,201]
      ,[863,239],[858,265],[858,296],[853,312],[845,328],[827,351],[803
      ,511],[733,503],[787,302],[774,275],[768,249],[762,143],[764,94]
      ,[762,72],[762,21],[765,11]]},
```


Figure 5.5 Illustration of coordinates representing object polygons in JSON format (Cordts et al. 2016).

5.2.3 Interpreting Worker Behaviours

We perform a descriptive analysis of the participants, including the browsers and systems they used. We also investigate the distribution of HIT submissions, multi-device usage, and HIT abandonment. Furthermore, we examine the dynamics of HITs, such as their reservation, expiration, and completion within the HIT group, and provide key trend interpretations. Subsequently, we analyse worker behaviours, starting with the strategies they employed to accept HITs, followed by the distribution of time spent on HITs. Additionally, we compare worker behaviours between those using HIT catchers and those without, considering factors including the number of HIT submissions, HIT focusing time, and multi-HITing behaviours.

5.2.4 Annotation Quality Evaluation

5.2.4.1 Definition of Variables

Precision and recall are commonly used evaluation metrics to measure the performance of a machine learning model or system in identifying relevant items from a dataset (Adnan et al., 2021; Juba & Le, 2019; Sajjadi et al., 2018). In this study, precision and recall can be used to evaluate the quality of the image annotations provided by the crowdworkers.

Specifically, precision measures the proportion of correctly identified relevant items out of all identification attempts (Powers, 2011). In the context of this study, precision measures the proportion of correctly labelled annotations out of all annotations made by the participant. A high precision score means the annotations tend to be correct and trustworthy.

Recall, on the other hand, measures the proportion of correctly identified relevant items out of all relevant items in the dataset. In other words, recall measures how well all relevant items could be identified in a dataset, regardless of whether there are any incorrectly identified items. A high recall score indicates that the system or model is effective at identifying most of the relevant items in the dataset. In the context of image annotation, recall would measure the proportion of correctly labelled annotations out of all the actual objects present in the image. A high recall score indicates that the workers have identified most of the target objects in the image.

By using precision and recall, both the accuracy and completeness of the annotations provided by the crowdworkers for each HIT could be measured. Next, the variables and equations used in the calculation of precision and recall are explained in order to prepare for the next operational steps.

TruePositive (TP) indicates the number of correct annotations under a specific object class, such as car.

FalsePositive (FP) indicates the number of false annotations. A FalsePositive is counted if there is no shape of a specific object in the position detected by the worker.

FalseNegative (FN) indicates the number of missed annotations. A FalseNegative is counted if there is one shape of object not detected by the worker. In addition, to keep the difficulty of the annotation within a reasonable limit, when there are more than 10 gold objects that need to be detected, $FN = 10 - TP$.

$$P = \frac{\sum_{i=1}^{n} TP_i}{(\sum_{i=1}^{n} TP_i + \sum_{i=1}^{n} FP_i)} \qquad (5.1)$$

For example, for a text search of a group of articles, the search precision is the number of results that match the search requirements divided by the number of all results returned (Powers, 2020). Similarly, in this study, the worker's accuracy for image annotation is the number of correct annotations ($\sum_{i=1}^{n} TP_i$) divided by the number of all annotation attempts ($\sum_{i=1}^{n} TP_i + \sum_{i=1}^{n} FP_i$).

$$R = \frac{\sum_{i=1}^{n} TP_i}{(\sum_{i=1}^{n} TP_i + \sum_{i=1}^{n} FN_i)} \tag{5.2}$$

In this study, the Recall for image annotation is the number of correct labelling made by the worker ($\sum_{i=1}^{n} TP_i$) divided by the total number of correct labellings ($\sum_{i=1}^{n} TP_i + \sum_{i=1}^{n} FN_i$) that should have been made in a perfect answer.

Even though the participants have been asked to detect up to ten items from each image, some images contain more than ten gold standard items. Since this would disadvantage the recall of workers who annotate gold images with more than ten items, a capped recall has been defined as:

$$R_{10} = \frac{\sum_{i=1}^{n} TP_i}{min(\sum_{i=1}^{n} TP_i + \sum_{i=1}^{n} FN_i, 10)} \tag{5.3}$$

F-score is a statistical measure used to evaluate the performance of a worker's annotation. It is the harmonic mean of precision and recall. Correspondingly, the capped $F\text{-}score_{10}$ is defined as:

$$F - score_{10} = 2\frac{P \times R_{10}}{P + R_{10}} \tag{5.4}$$

where P is the precision.

The variables related to the annotation quality measurement have been calculated following the steps illustrated below.

5.2.4.2 Quality Computation

First, as illustrated in the code example below, each worker annotation output was read as JSON format (Figure 5.6). Then the JSON file containing the preset annotation answers of this image was located based on the image address. Using the preset answer file, the polygon shapes representing each gold standard object in the image were drawn via *shapely.geometry*[28]. In this study, the Shapely Python package was used to draw areas of each object within images and measure distances between workers' annotations and each object. This step is used to determine whether the marks drawn by the worker on the image successfully annotated a specific object.

[28] Shapely Documentation: https://shapely.readthedocs.io/en/stable/index.html

```
Worker_Annotation_Output =
{
   "classifications":[...]
   "text-area-thoughts":"",
   "text-area-feedback":"",
   "worker_id":"0000",
   "assignment_id":"5678",
   "hit_id":"1234",
   "img":"bremen/bremen_000005_000019_leftImg8bit_blurred.jpg"
}
```


Figure 5.6 A sample worker annotation output.

Figure 5.7 is an example of worker annotation compared with the gold standard polygons. The solid colour shapes that overlay the items represent the gold standard, and the stars are the worker's annotated points. Red stars indicates that the annotator recognises the annotated objects as cars. Similarly, blue stars correspond to buses, and yellow stars correspond to pedestrians. In this image the *F-score*$_{10}$ of the worker was 0.95, indicating an excellent annotation quality.

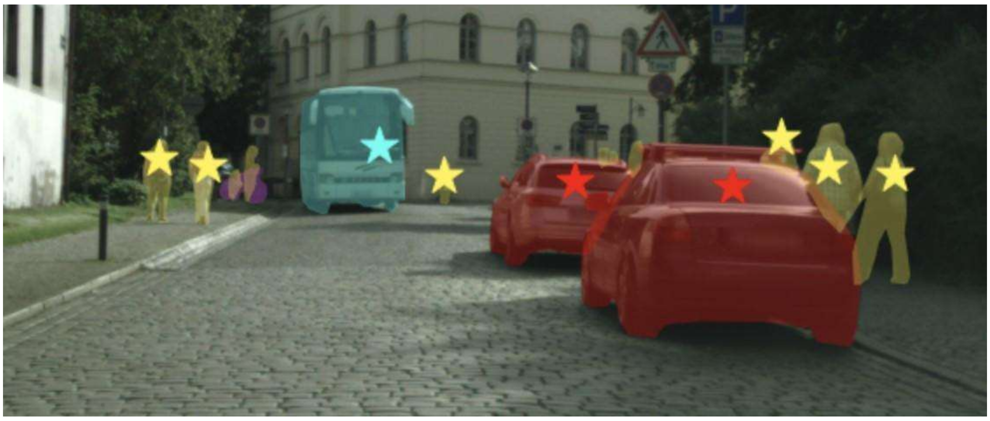

Figure 5.7 Illustration of how worker's annotations match the gold standard objects. Red stars indicate annotations of cars, yellow stars indicate pedestrians and blue stars indicate buses.

Using this case as an example of quality evaluation, the worker annotation result consists of a list of annotated point positions targeting a specific item class. In the illustration of the annotation results for one of the street view images (Table 5.1), *set_x* and *set_y* represent the pixel position in the image of the annotated point made by the worker. In contrast, *pos_x* and *pos_y* represent the percentage position of the annotated point in the image. In other words, using the first row of Table 5.1 as an example, it means the position of this annotated point is 73.71% of the image horizontal length and 46.88% of the image vertical length. Given the top-left corner of the image as (0,0), the coordinate of this annotated point is (73.71, 46.88). The overlapped annotations are handled using the matching algorithm explained in the next section.

In the official JSON file containing all the gold standard polygon coordinates, all coordinates are represented in a percentage of the image length and width. Therefore, (*pos_x*, *pos_y*) could be used to calculate the annotation quality.

Table 5.1 Illustration of the annotation results from worker response (all coordinate values are in pixels).

id	result_id	set_x	set_y	pos_x	pos_y	rel_pos_x	rel_pos_y	item_class
1	1	670.12	207.68	73.71	46.88	73.71	46.88	person
2	1	496.12	188.68	54.79	42.76	54.79	42.76	bus
3	1	129.12	235.68	14.9	52.97	14.9	52.97	bus
4	1	256.12	190.68	28.71	43.19	28.71	43.19	bus
5	1	306.12	185.68	34.14	42.1	34.14	42.1	bus
6	1	514.12	431.68	56.75	95.58	56.75	95.58	bus
7	1	648.12	248.68	71.32	55.8	71.32	55.8	bicycle
8	1	756.12	227.68	83.05	51.23	83.05	51.23	bicycle
9	1	621.12	215.68	68.38	48.62	68.38	48.62	bicycle

The pre-set gold standard polygon coordinates indicating objects within each street view image have been used to evaluate the quality of workers' annotation results. Based on each of the four item classes, a complete not oriented bipartite graph $G = (U, V, E)$ was modelled (Asratian et al., 1998). As illustrated in Figure 5.8 Stage 1, the graph can be separated into two types of vertices, namely U and V. In this study, U represents the set of gold standard objects having size m, V is the set of worker annotation points having size n, and E is the set of $m * n$ edges that connect every vertex in U with all vertices in V. In other words, the distances between each gold standard object and all the worker annotation points have been calculated and stored in E. The purpose of E is to decide whether the workers annotated correctly by comparing if a worker's annotation is in the shape of a gold standard object. The numbers on the edges in Figure 5.8 represent the distance between one gold standard object and one annotation point.

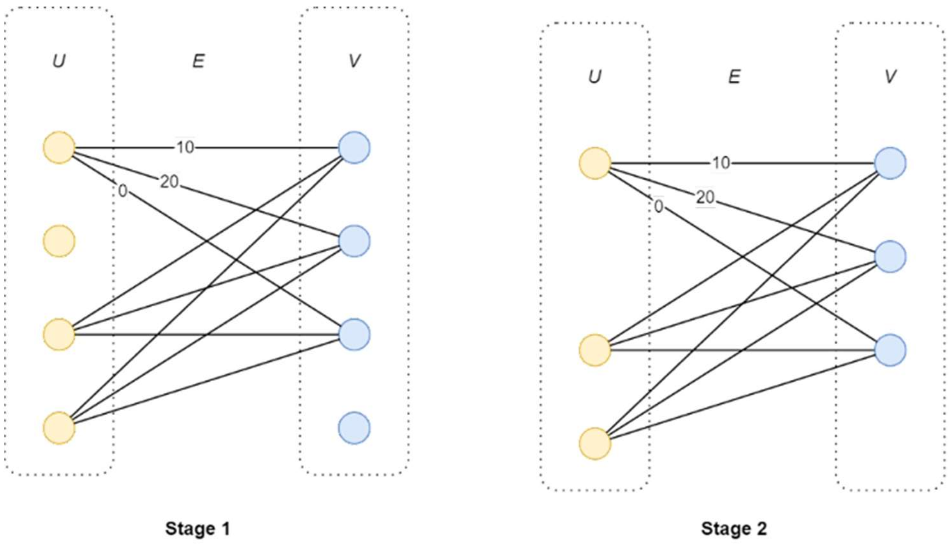

Figure 5.8 The quality checking process using bipartite graph (Stage 1 and 2).

Each e in E was weighted with the Cartesian distance between the detected object position and the shape of the gold standard objects that was drawn based on pre-set answers. The distance between each annotation and all gold items were calculated using *each_gold_item.distance (each_annotation)* [29] for each category. It is worth noting that such distance would be set to zero when the point of worker's annotation is inside the gold standard polygon, and progressively increases when the point moves away from it.

To reduce the misjudgement of worker annotation quality, a proximity distance threshold ε equal to 3% of the image size has been applied, and all the edges with distance greater than ε have been removed from E. This allows workers for a minimum margin of errors when not clicking exactly on the gold standard shape.

Next is the evaluation of annotation quality. To start with, the orphan nodes in U have been labelled as FN (False Negative). In other words, such gold standard objects did not have any matching annotation points from the worker response. Similarly, the orphan nodes in V have been labelled as FP (False Positive), which means there were no matching gold standard objects for such worker annotation points. In results, as illustrated in Figure 5.8 Stage 2, these orphan nodes have been removed from U and V before the calculation stage.

Then, on G, which represents a collection of edges connecting the nodes in U and V, a greedy version of the maximum weight matching strategy has been applied (Figure 5.9 Stage 3). More

[29] Shapely Documentation: https://shapely.readthedocs.io/en/stable/reference/shapely.distance.html#shapely.distance

specifically, the edges with the smaller distance between the nodes in *U* and *V* have been checked first. According to the proposed examination method, one edge would be marked as a valid match if it connects a worker annotation point that is within the shape of a gold standard object under a specific class.

After each valid match, the matched edges and connected nodes from both U and V have been removed (Figure 5.9 Stage 4). The worker annotation from that match has also been labelled as a *TP,* which represents a correct annotation.

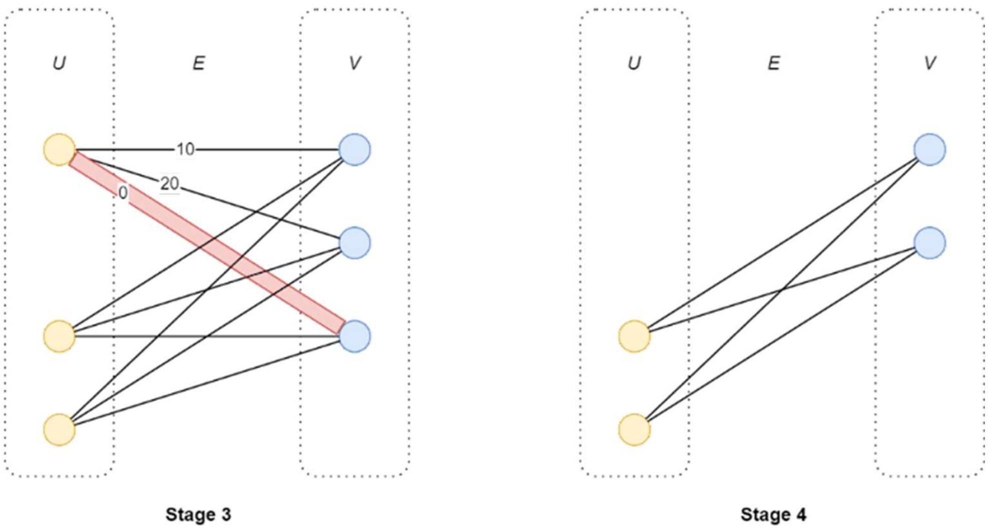

Figure 5.9 The quality checking process using bipartite graph (Stage 3 and 4) where red bar represents the edge with zero distance.

5.2.4.3 Integer Programming vs Greedy Matching

One special case that deserves further discussion is: there could be multiple worker annotations pointing to the same gold standard objects. To avoid all the annotations being labelled as TP, only one annotation from V was finally labelled as TP, and the remaining duplicated annotations were ignored and used to check if they match any other gold standard objects later (Figure 5.10).

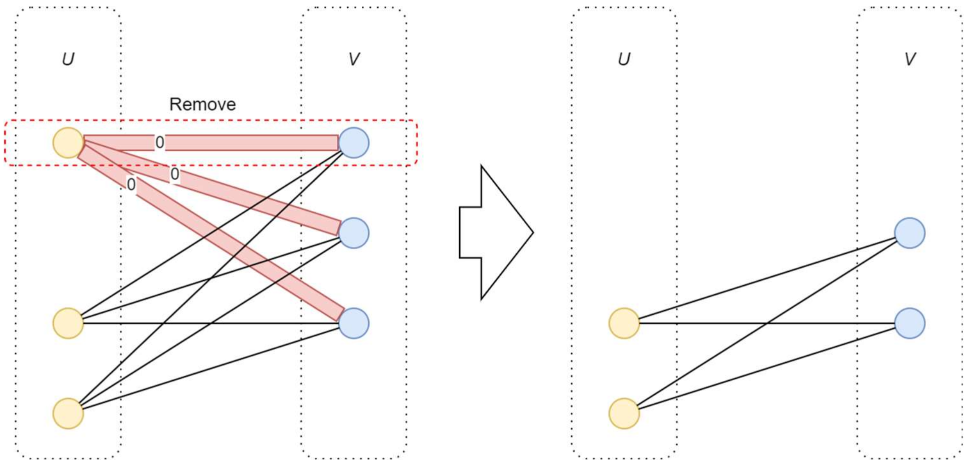

Figure 5.10 A case when multiple annotations match one gold standard object.

Finally, following the definition mentioned at the beginning of this section, the remaining nodes in V have been counted as *FP*, which means such gold standard objects were not successfully annotated. Similarly, the remaining nodes in U have been counted as *FN*, representing those annotations that did not point to any gold standard objects. However, this solution could result in an issue: the removed annotation from V could also be a perfect match for another gold standard object in U. Therefore, the greedy approach might not provide the optimal results for the annotation quality check.

In comparison, Integer Programming could be a more appropriate strategy for getting the maximum weight of all the matched edges in E. To start with, each edge has an associated weight w_e , the smaller the distance between a vertex in U (a gold standard object) and a vertex in V (a worker annotation point), the higher the weight of this edge connecting those two vertices. So, the ultimate purpose is to maximise the total weights of all chosen edges under the condition that each vertex in U and V only belongs to one edge in E. In other words, the goal is to try to match the gold standard objects with as many annotation points as possible under the constraint that both annotation points and gold standard objects must match with only one of each other in the end. As illustrated in Figure 5.11, three worker annotations correctly point to the gold standard object A as their distances are all 0. Following the greedy version of the maximum weight matching strategy, the first edge between vertex A and D would be labelled as a TP. However, it could leave gold standard object B without any correct annotations. Therefore, it makes more sense to remove the edge between vertex B and D following the Integer Programming solution in this case.

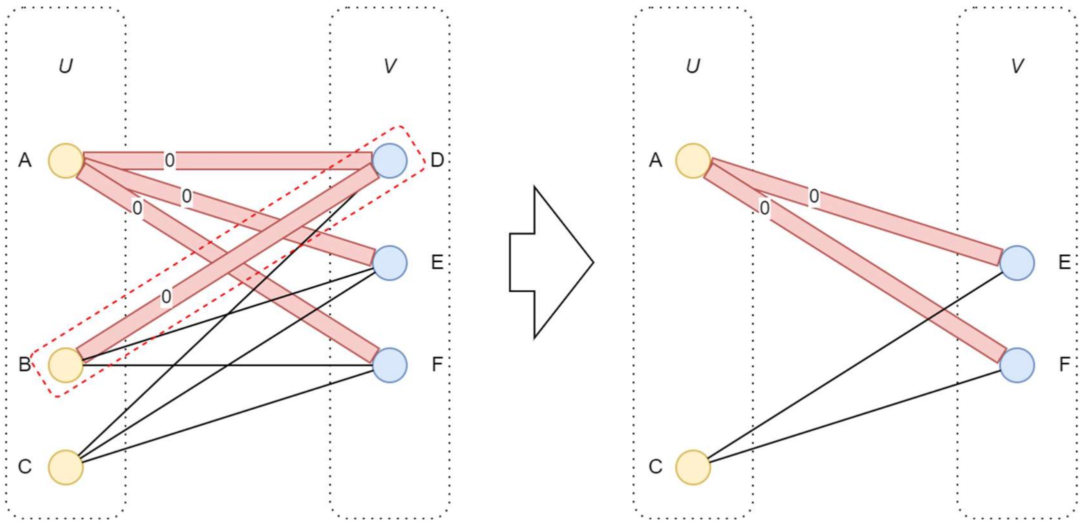

Figure 5.11 Integer Programming solution when multiple annotations match one gold standard object.

The above computation process has been applied for each of the target classes, including "car", "bus", "person" and "bicycle". As a result, the *TP*, *FN*, and *FP* scores have been computed for each class. The accuracy metrics of each worker annotation response including precision and recall have also been calculated based on these three groups of scores, and then used as an indicator of annotation quality.

5.2.5 Monitoring Techniques

Along with the collection of worker annotation on the street view images, multiple monitoring techniques were employed to help reveal HIT access dynamics, the phenomena and impacts of workers' use of HIT catchers.

5.2.5.1 Client side

The HIT loads some JavaScript able to detect: (i) whether the worker has MTurk Suite or MTurk Guru extension installed; (ii) the browser tab visibility properties, that allow inferring whether the HIT is currently being viewed or whether another browser tab is in focus; (iii) the timestamp of the completion of page rendering.

In Section 2.2.1, we categorised and compared the HIT catching capabilities of popular HIT catchers via Table 2.1. In terms of total installations, MTurk Suite and Panda Crazy Max rank highest. It is noted that only MTurk Suite and Turk Guru allow their detection by requesting the Chrome extension resources[30] through running JavaScript code[31] on a specifically crafted HIT web page. Specifically, only these two extensions define their *"web_accessible_resources"* with the wildcard *"/*"* in their manifest file, which allows the JavaScript code from clients to request for extension file resources (Mohammadi, 2019).

In addition, browser tab visibility properties were monitored and recorded by the script to help detect the duration of a worker's browser access to a HIT page. Specifically, through the HTML DOM EventListener functions, the worker's *'focus'* and *'blur'* events on the HIT page were logged along with their timestamps. The final HIT page activity data is shown in Figure 5.12. The total time spent by one worker on this HIT is the sum of the duration of multiple events from 'focus' to 'blur' or from 'focus' to tab closing. However, as there is a possibility that the worker closed the HIT web page during the completion process, the complete time finally spent on the current HIT were calculated by integrating all page event logs based on the worker's browser IP address and HIT_ID. In addition, the numbers of tab switching during each HIT completion were calculated. Both data could help to interpret the work strategies and work attitudes of different types of workers during the data analysis stage.

```
[
   {'event': 'focus', 'date': '2021-06-25T16:26:58.732Z'},
   {'event': 'blur', 'date': '2021-06-25T16:26:58.746Z'},
   {'event': 'focus', 'date': '2021-06-25T16:26:58.746Z'},
   {'event': 'blur', 'date': '2021-06-25T16:27:07.791Z'},
   {'event': 'focus', 'date': '2021-06-25T16:27:09.814Z'},
   {'event': 'blur', 'date': '2021-06-25T16:27:10.117Z'},
   ...
]
```

Figure 5.12 Sample browser tab event records from worker image annotation activities.

5.2.5.2 Server side

Using Amazon Web Services' Simple Queue Service (SQS), the events of HIT reservation, expiration, and abandonment for assignments were measured regardless of whether the HIT web pages were even opened by workers. Workers could continuously accept HITs without opening the HIT web page and abandon them later through HIT catching and queue management tools. By correlating the data from these two monitoring techniques, the HIT

[30] Manifest - Web Accessible Resources: https://developer.chrome.com/docs/extensions/mv3/manifest/web_accessible_resources/
[31] Script code used in experiment: https://github.com/howrudoing/Catching-Script-Study/

reservation and backlog dynamics were reconstructed to help understand the effects of the use of HIT catching techniques on HIT availability, backlog and expiration.

For example, the timestamps of *AssignmentAccepted* and *AssignmentSubmitted* for each HIT were obtained from the HIT status notifications obtained from SQS (Figure 5.13), and thus the total completion time of the HIT were calculated. By comparing the total completion time with the actual duration of HIT being performed which was collected from the client side, the total time that one HIT was reserved but not performed were further derived. This is regarded as the duration of HIT delay. This measurement is used to determine if workers using HIT catchers are more willing to delay their HIT completions. Specific procedures on combining data from multiple sources for HIT backlogged time calculations are explained in Section 5.3.3.3.


```
{
   "Events": [
    {
      "EventType": "AssignmentSubmitted",
      "EventTimestamp": "2021-06-28T18:41:39Z",
      "HITId": "3S1**********",
      "AssignmentId": "36U**********",
      "HITTypeId": "3D1**********"
    }
  ],
…
}
```


Figure 5.13 Sample HIT event data retrieved from SQS.

Based on the description of HIT event types from MTurk[32], one HIT can have the states described in **Table 5.2**. Other than the four states on the top, it is worth noting the introduction of three new states, which are substates of the HIT pending state. Thanks to the monitoring techniques described in the Client-side section, we distinguished pending HITs that are only in a worker queue (backlogged), from the ones that are being worked on (active), and subsequently for these whether the windows focus is on the HIT (focused). The method of calculating the durations of these three states is illustrated in Figure 5.14. The further understanding of the HIT pending states helps study the HIT reservation and completion dynamics.

[32] Use Mechanical Turk notifications: https://docs.aws.amazon.com/AWSMechTurk/latest/AWSMechanicalTurkRequester/Concepts_NotificationsArticle.html

Table 5.2 HIT states definition.

HIT state	**Description**
Available	published and available to be reserved
Pending	cannot be reserved (active or queued)
Completed	successfully completed by a worker
Expired	expired and will not be published anymore
Active	pending + a worker started to work on it
Backlogged	pending + is in a worker queue but is not active
Focused	active + a worker has the browser tab focused on it

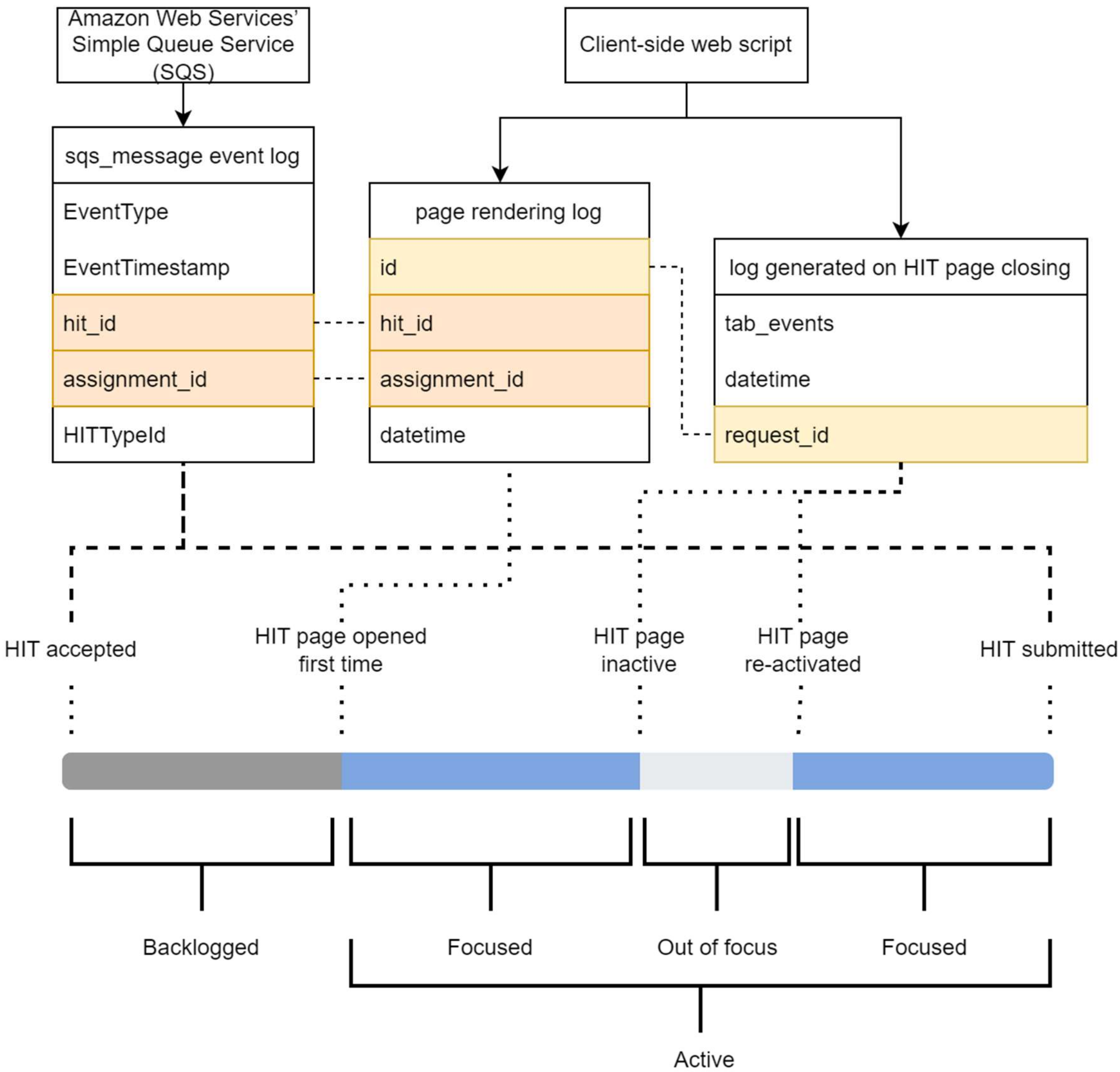

Figure 5.14 Calculation of different HIT pending status durations.

5.2.5.3 Monitoring HIT list page during experiment

A Python script using Boto3 API[33] was applied to record the number of HITs in the three states of available, completed, and pending among the 1000 HITs continuously. Therefore, the trajectory of the number of HITs in the three states during the experiment was then depicted.

5.2.6 Structure of Database Storing Raw Data

This section explains how data collected via different methods gets organised in the PostgreSQL database. The data storage and collection system used to conduct this experiment was realised by Dr Maddalena from University of Southampton. More specifically, five tables were created based on their unique sources of data (Figure 5.15). The *classification* table includes the image annotation results for each worker's HIT response. Each annotation made by a worker on a street view image automatically generates a row of records, therefore multiple rows of corresponding records exist for the same HIT. Moreover, records for the same HIT did not all originate from the same worker. This is because it is possible that the worker annotated the HIT but did not later submit it. Therefore, *result_id* was used to avoid confusion when matching annotation records with HIT submission records.

Similarly, the *result* data table stores worker's text responses for each HIT, including their descriptions of the current image and their feedback for doing this type of HIT. Each time a worker closed a HIT web page, a new row of records was created in the *result* table. This also means that multiple lines of records could be generated during the HIT completion process by a worker. More importantly, each row recorded the worker's focus activities on one HIT page before the page was closed as the *tab_events*.

The *user_agent* table stores the browser related data, including the installation status of two monitored HIT catchers, browser name, browser window size, client IP address, cookie and session token for authentication. The browser authentication information has been used to determine whether a worker with a unique IP address used more than one browser for crowdwork and whether multiple HITs were performed at the same time. Furthermore, *improta* within this table stores information including the browser and device operating system version used by the worker to access the HIT pages. Such information can be used to explore whether

[33] Scripts to track number of HITs under different states: https://github.com/howrudoing/Scripts-for-thesis/blob/main/track_available_HITs_sandbox_Alessandro.py

there are multiple workers from different IP addresses sharing one single worker account for crowdwork.

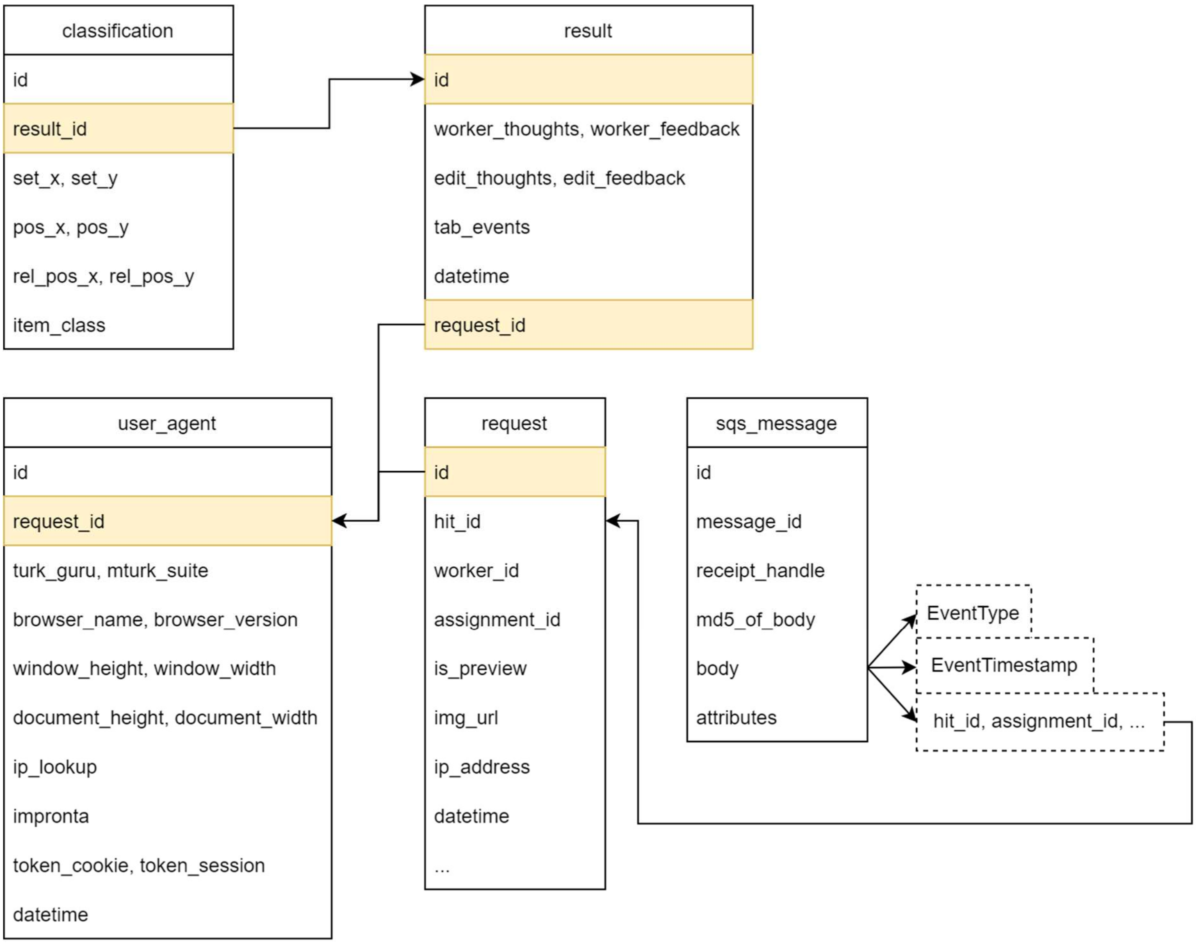

Figure 5.15 Five PostgreSQL database tables storing experiment data.

The *request* table stores the information gathered via web scripts when a worker opens one HIT page. It is worth noting that the *is_preview* variable helps to identify the worker´s purpose of visiting that HIT page (to preview or to perform). Although this parameter does not help detect whether workers are using tools that automatically accept HITs, it can help understand whether workers tend to skip the preview step after finding a desirable task, thus enriching our understanding of worker behaviour.

In addition, by matching *ip_address* and *worker_id* of multiple records within the *request* table, it could be determined whether there are multiple worker accounts matching one unique IP address. This situation could happen because multiple workers share the same PC and sequentially use their own accounts to work on HITs. However, by introducing the active times of the different HITs obtained in the *result* table, it is possible to exclude this case of multiple workers sequentially using the same PC for crowdwork, and thus confirm the existence of one worker using multiple accounts for microtasks at the same time on one PC.

In contrast, the *request* table can also be used to investigate whether there were people from multiple IP addresses logging into the same worker account at the same time. Since MTurk does not technically prohibit logging into the same worker account from multiple IP addresses, it is possible for this to happen. This phenomenon may be caused by the same worker logging on to different computers in sequence to carry out work on their account. However, by introducing the active time of the HITs in the worker's account obtained from the *result* table, and by overlapping the active time of multiple HITs to exclude the special case of the same worker working on different computers in sequence, it is possible to identify the existence of multiple workers sharing a single worker account on different computers to work on HITs.

Finally, the *sqs_message* table stores the MTurk notifications sent from Amazon Simple Queue Service (Amazon SQS[34]). These messages include the time and type of events happened on each HIT posted in this experiment. Such events include a HIT being reserved, abandoned, submitted, etc. The content and purpose of these records have been explained in the previous section regarding the monitoring techniques from the server side.

In summary, the records within each table can be linked together via their unique id number depending on the specific purpose of data analysis. These interlinked records help to build up an overview of the state changes of the HITs of this study. Moreover, the specific work behaviour of each of the workers involved in the study can be accurately described, which in turn helps to understand their diverse work strategies.

5.3 Data Analysis

This section explains how the analytical variables were generated from the raw data, and the insights through the interpretation of the results.

5.3.1 Overview of Participants

5.3.1.1 Multi-browser Use, HIT Attempt and Abandonment / Return

Before the analysis of participants who submitted HITs, we first conducted a description of the work strategy and HIT abandonment for the 576 participants who accepted and opened the HIT

[34] Amazon SQS: https://aws.amazon.com/sqs/

page. In other words, these 576 participants loaded the HIT page after accepting them, however they may not successfully submit the HITs in the end. Based on the 'impronta' information recorded in the *user_agent* table, the number of browsers and operating systems used by each participant was counted (Table 5.3, Table 5.4). It is revealed that Windows is the dominant choice for workers. In addition, mobile device operating systems such as Android and IOS are also used by workers to load HIT pages. This finding provides an empirical basis for previous research finding that workers use mobile devices to assist in their work (Williams et al., 2019). In terms of browser choice, Chrome and Firefox are the most used among the worker group, with a combined more than 99% usage rate.

Table 5.3 Usage ratio of different operating systems.

Name of Operating System	Count	Percentage
Windows	456	93.6%
Linux	11	2.3%
Android	7	1.4%
Mac OS	6	1.2%
Chromium OS	4	0.8%
Ubuntu	2	0.4%
iOS	1	0.2%
Total count of participants containing OS information	487	100.0%

Table 5.4 Usage ratio of different browsers.

Name of Browser	Count	Percentage
Chrome	426	86.9%
Firefox	60	12.2%
Opera	3	0.6%
Safari	1	0.2%
Total count of participants containing OS information	490	100.0%

The analysis of all 576 participants' device fingerprints revealed that at least 9 of them used multiple browser windows, monitors, or operating systems during the experiment. Each of these workers completed approximately 16.1 HITs, well above the average number of completions (≈ 5.85 HITs) for all 171 participants who had submitted HITs. Meanwhile, these 9 participants attempted an average of around 27.22 HITs (median = 27), compared to an average of 4.28 HIT attempts (median = 1) among the 576 participants. In other words, participants who worked with multiple browsers/monitors/operating systems got significantly more HIT opportunities.

These 9 participants abandoned / returned 11.11 HITs in average, with a median of 4. In comparison, the average number of HIT abandonment / return by each of the 576 participants was about 2.55, with a median of 1. It reveals that participants using multiple browsers were more likely to abandon or return HITs compared to the total sample group.

An extreme case is: one of the participants used at least 2 operating systems and 6 monitors (possibly from one or more devices) during the experiment. This participant submitted a total of 39 HITs during the experiment, which is far more than the average number of HIT submissions among all 171 participants who had submitted HITs (≈ 5.85 HITs). What is also surprising is that this participant attempted 41 HITs, far exceeding the average number of HIT attempts out of 576 participants (≈ 4.28 HITs).

In addition, as shown in Table 5.5, out of these 576 participants who accepted the experiment HITs, a total of 424 participants with unique *worker_ids* used the chrome browser, of which 204 were detected to be using HIT catchers, with a usage rate of 48%. Their average count of HIT attempts is around 4.81. In comparison, 220 participants were not detected using HIT catchers and their average count of HIT attempts is about 3.96. This means that participants who were detected as using HIT catchers tried more HITs. Overall, approximately 78% of the 576 participants only attempted 1-4 HITs.

The average number of HIT abandonment / return for participants using HIT catchers is around 3.12, while for those not using HIT catchers is around 2.82. It shows no obvious difference between the two types of participants on numbers of HIT abandonment / return.

Table 5.5 A comparison of HIT attempt and abandonment / return between two types of participants.

	No. of workers	Average No. of HIT attempted	Average No. of HIT abandonment / return
HIT catchers	204	4.81	3.12
No HIT catchers	220	3.96	2.82

Figure 5.16 shows a distribution of HIT attempt and abandonment / return counts. Specifically, about 95% of the participants attempted 1 to 10 HITs, while about 90% of the participants abandoned / returned 1 to 10 HITs. Furthermore, there is a huge difference between the total number of workers who ultimately submitted HITs (171) and those who attempted or abandoned HITs. Given the low difficulty of completing these HITs, it can be assumed that one potential reason for this widespread abandonment behaviour is that: HIT catchers reserved these HITs based on their filtering criteria without the user's awareness and backlogged them until expired.

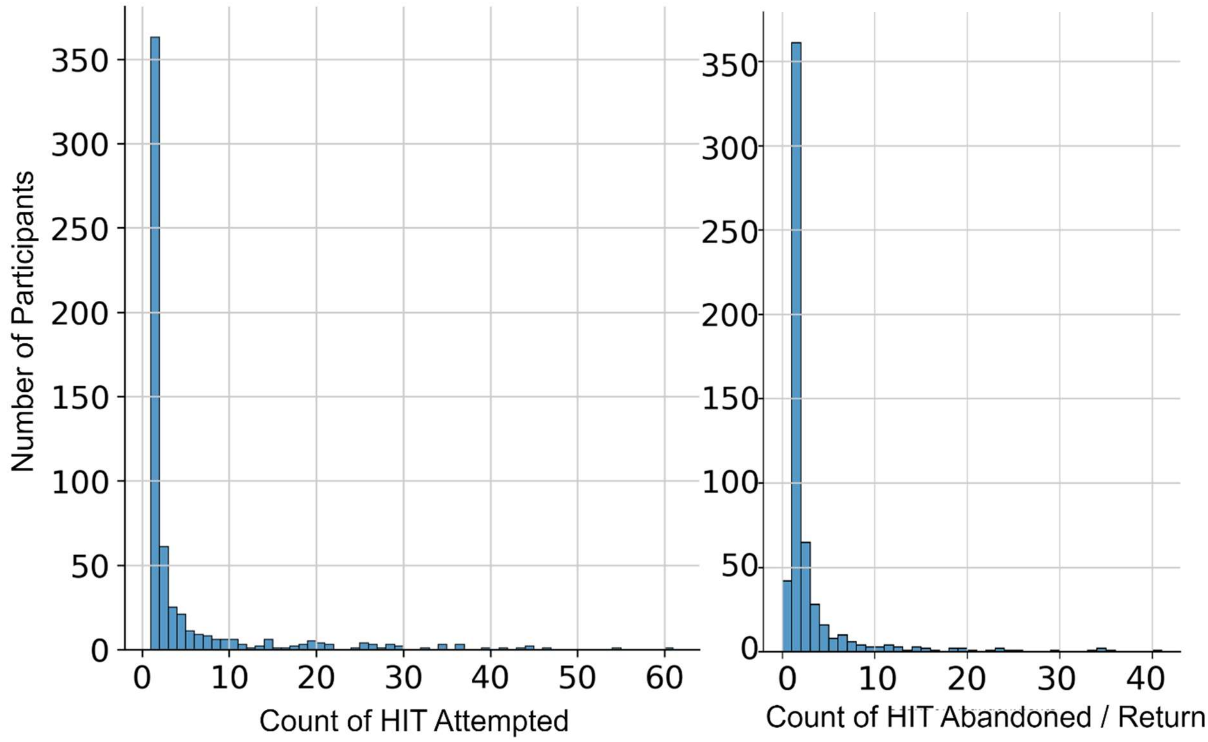

Figure 5.16 Distribution of count of HIT attempts and abandonments / return.

5.3.1.2 HIT Submission

In Table 5.6 we provide a summary of the workers and annotations distribution over different browsers. For workers using Chrome, it was also possible to detect their context switching behaviour between browser tabs, and the presence of HIT catching extensions.

Table 5.6 Number of HIT submissions and workers using different browsers.

	Total Count	Chrome					Firefox	Safari	Opera
		Total Count	None Installed	Turk Guru	MTurk Suite	Both Installed			
No. HIT Submissions	1000	518	203	49	210	56	473	9	0
No. Workers who submitted HITs	171	135	68	5	44	18	36	1	0
No. Workers who attempted HITs	576	424	220	78	189	63	60	1	3

Based on the data related to HIT submissions, a total number of 1000 HITs were completed by 171 unique workers. Please note that each unique worker has their own MTurk worker account, and it is possible for one unique worker to complete HITs via multiple unique browsers under different IP addresses. Therefore, the total number of unique workers could be less than the sum of each category as one worker can use multiple browsers during the experiment. However, another possibility is: the same worker account was shared by more than one unique person. In this case, the total number of unique workers may be greater than the sum of each category. In addition, while some workers were not detected using HIT catchers, it is still possible that they were using similar tools that we were unable to detect.

To minimise the impact of outliers and better reflect focused trends in the data, especially for data that may not be normally distributed, median was used for event time analysis rather than the mean. Based on the analysis of the time between HIT events, the median time for a worker between two HIT reservations that were eventually completed (exploited reservations) was 90 seconds, and the median time for a worker between two successful submissions was 86 seconds.

Regarding the equality of HIT submissions, the Gini coefficient for the HIT group among the 171 workers is 60.47%, which means the HIT-worker diversity[35] is 39.53%. Compared to 100%, which means a perfect equality for HIT submissions within the 171 workers, this indicates a high level of inequality. In particular, 76 workers managed to complete only one HIT. This HIT submission inequities are further explored in Section 5.3.3 HIT opportunities.

[35] The more fairly HIT submissions distributed among crowdworkers, the higher this diversity. Definition is explained in Section 3.4.1.2.

5.3.2 HIT Access Dynamics

The evolution of HIT reservation, expiration and completion were investigated in this section. Figure 5.17 describes the changing of the number of HITs with the states of available, completed and pending over time during the experiment. Apart from manual deletion by the researcher, the state each HIT was in during the experiment would always be one of these three. Therefore, the sum of the count of HITs in these states at each timestamp always equalled the total number of published HITs, which was 1000 in our case. It can be revealed that the number of available HITs was very low in the first 10 minutes, while the number of pending HITs rapidly grew in a few minutes. This indicates a dramatic competition to accept HITs and fill their HIT queues. Combined with Figure 5.18, it can be seen that within the first minute of the HIT group being published, more than 250 HITs were accepted, and the phenomenon of hundreds of HITs being accepted per minute was maintained for 5 minutes, accompanied by a large number of HITs being returned. This phenomenon was most likely since each HIT had an allotted time of only 5 minutes, which resulted in workers who over-accepted HITs not being able to complete so many in such a short period of time. As a result, they had to return those HITs that are about to expire, either manually or with the assistance of scripting tools.

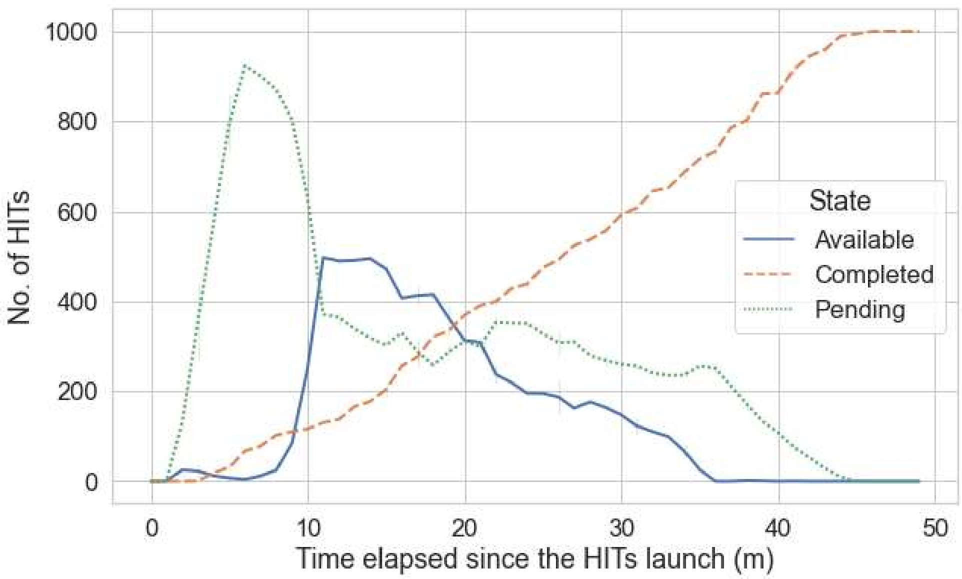

Figure 5.17 Dynamics of HIT status changes.

Nevertheless, after 5 minutes since the start of the experiment, the HITs held in the workers' HIT queues began to expire massively, with many HITs being abandoned. This means that even though workers were continually returning HITs that were too late to execute, there were still a large number of HITs not submitted in time, and thus expired and were forced to be retrieved by MTurk. Also, as Figure 5.17 demonstrates, near the 10th minute from the start of the experiment, a large number of previously expired and returned HITs were re-published by MTurk, resulting in a rapid increase in available HITs (Blue Line). It is worth mentioning when understanding the two figures that: Figure 5.17 shows the total number of HITs at different states at a given moment. In contrast, Figure 5.18 shows the number of the four HIT events that occurred within each minute.

Ten minutes after the start of the experiment, the number of HITs pending remained stable at around 300-400, while the number of HITs completed kept increasing linearly with time. Figure 5.18 shows that after the tenth minute of the experiment, the HIT acceptance, return, submission, and abandonment all remained at a steady rate until the end of the experiment. In contrast, the times of HIT being accepted remained stable over time in the frequency range of 50-100 per minute, slightly higher than the occurrence frequency of other HIT events. This means that workers generally tried to reserve or accept as many HITs as possible, regardless of whether they were ultimately successful in submitting them.

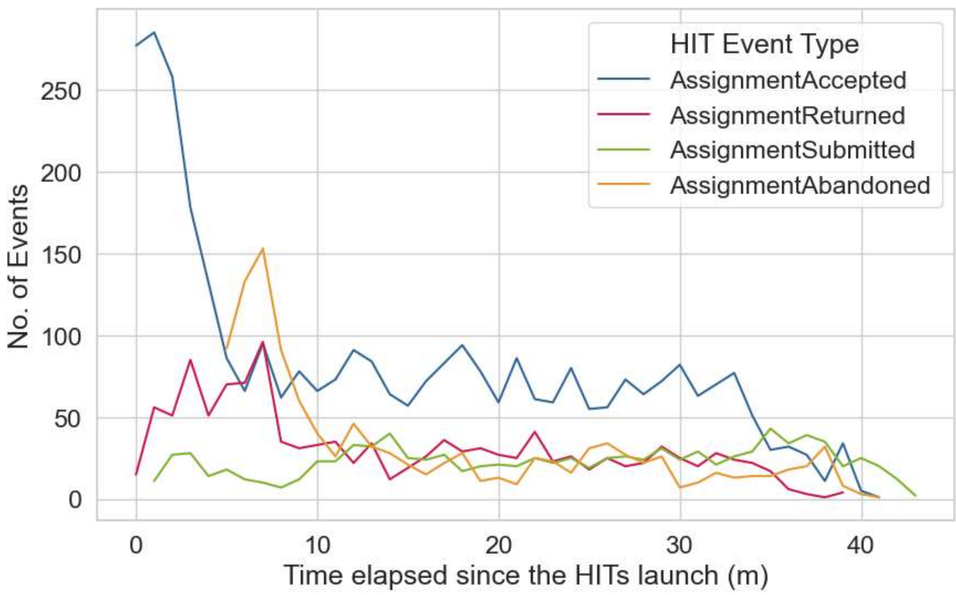

Figure 5.18 Counts of HIT events per minute.

In summary, after all 1000 HITs were accepted for a short period of time, they experienced a large number of expirations and returns within a short period of time. The number of available HITs then picked up quickly and they were accepted at a much lower rate than at the start of the experiment. One possible explanation is: after realizing that these HITs could only be reserved for 5 minutes, the crowdworkers adjusted their frequencies of acceptance for this HIT group, to avoid a large number of HIT expiration as a result of excessive HIT backlogging. The completion speed (green line in Figure 5.18) dropped continuously at the beginning of the experiment due to the large number of HIT backlogs. However, as a large number of expired and returned HITs were released again by MTurk, the completion speed gradually increased and was maintained until the end of the experiment.

It is worth pointing out that the HITs with pending state could be either backlogged or active[36], therefore it cannot be interpreted from Figure 5.17 alone how many HITs were backlogged[37] at any given time. This question is addressed in Section 5.3.4.3 (HIT backlogged time).

To describe the HIT reserving process more clearly, the unutilised HIT reservations are studied. Such unutilised HIT reservations did not result in a HIT submission, so they are also called unsubmitted reservations. Not surprisingly, most of the HITs published during the experiment were not successfully submitted after their first acceptance. In Figure 5.19.b, it can be seen that apart from more than a hundred HITs that were successfully submitted after being accepted the first time, the vast majority of HITs underwent multiple unsubmitted reservations before they were finally submitted. These HITs went through a number of being accepted but later being returned or abandoned, and these reservations without submission all slowed down the final completion of the entire HIT group. Figure 5.19.a shows that the total time spent on unsubmitted reservations for all 1000 HITs during the experiment was approximately 25 minutes.

[36] The HIT states are explained in Table 5.2

[37] It means a reserved HIT stays in one's HIT queue without being opened.

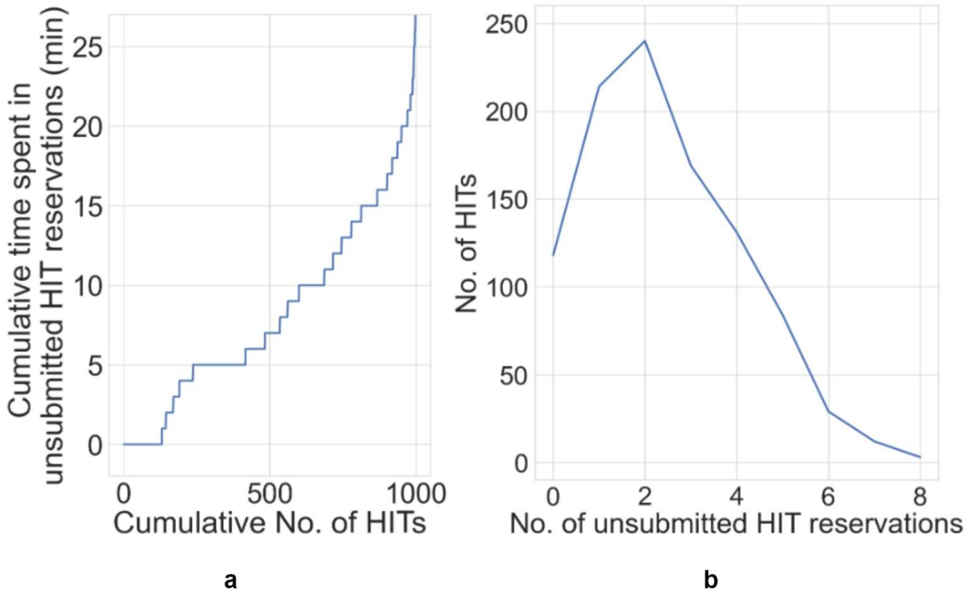

Figure 5.19 a: Cumulative time spent in unsubmitted HIT reservations. **b:** Count of HITs with different number of unsubmitted reservations.

5.3.3 HIT Opportunities

Due to the lack of worker information for each HIT event collected from the AWS SQS queues tracker[38], the average number of HIT reserved for each worker with and without the monitored HIT catchers cannot be calculated in this study. Fortunately, the numbers of HIT completion for each worker could be investigated by merging the related data from the *result, user_agent* and *request* tables.

It can be interpreted from Figure 5.20 that among all 171 workers with unique *worker_ids* who submitted their HIT results, 92 of them submitted just 1 or 2 HITs. Specifically, 76 workers only submitted 1 HIT during the experiment. Moreover, 9 of them submitted more than 20 HITs as an individual worker account. One of the workers even submitted up to 54 HITs. As interpreted in Section 5.3.1.1, the HIT-worker diversity is 39.53%, which is extremely low. Given the low level of difficulties in completing these HITs, the large imbalance in the number of HIT completions implicitly reflects the widely varying success rates in HIT acceptance.

[38] Amazon SQS: https://aws.amazon.com/sqs/

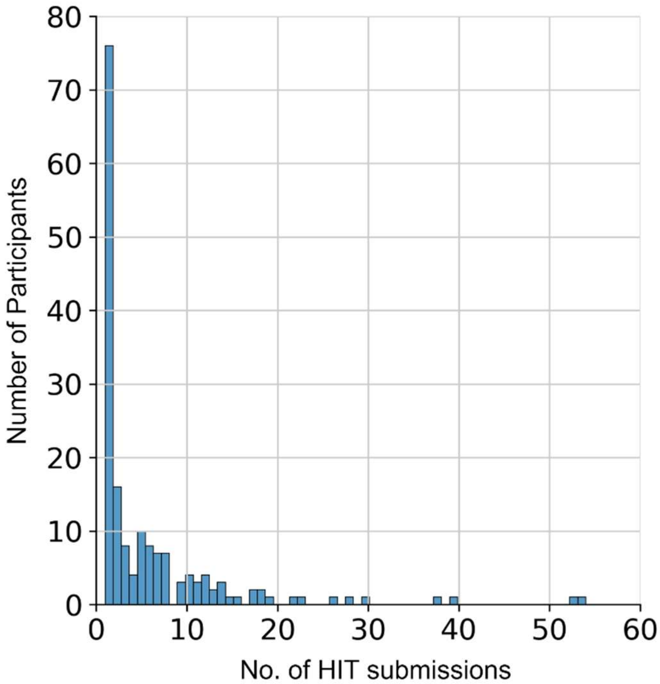

Figure 5.20 Worker counts with different numbers of HIT submission.

To find the correlation between the HIT catcher installation status and the number of HIT completion, a total number of 135 worker samples have been selected, as they completed HITs using Chrome browser based on the web script detection. As explained in the Monitoring techniques section, the customised web script could only detect whether the monitored HIT catchers were installed in the Chrome browser.

Among the 68 workers without monitored HIT catchers installed, they have completed an average of around 3 HITs each worker account, and a total of 199 HITs. In comparison, the remaining 67 workers with HIT catchers installed have completed an average of 5 HITs each worker account, and a total of 317 HITs. Except for data bias since all the HIT catchers cannot be detected from the browsers of workers, the difference in the number of completed HITs has already revealed the imbalance in participation due to the use of the monitored HIT catchers. In other words, there exists workers with other types of HIT catchers but still get identified as non-HIT catcher workers in this study, and the technical advantage that HIT catchers bring to workers could be much higher.

Besides the low number of HIT completion for the non-HIT catcher workers, the Gini coefficient of all the 1000 published HIT is about 0.605, showing a significant inequality on the number of HIT completion for each worker.

5.3.4 Comparison of Behaviours Between Two Types of Workers

The aim of this section is to investigate the work behaviour of participants, including how the relevant analytical variables are generated and how the differences between the behaviours of workers with and without the monitored HIT catchers can be understood in the light of these variables.

When comparing behavioural data for workers with and without the HIT catchers, only 518 HITs completed by the workers using the Chrome browser were studied, as this study only detected the focus status of the browser tabs and installation status of two HIT catchers of these workers through the return values of their Chrome browsers. It is important to note that this detection method has limitations, as the workers may be using other types of HIT catchers that cannot be detected through the means presented in the Method section. Furthermore, the fact that the monitored HIT catchers were installed does not mean that the worker was using it during the experiment. The proposed detection method provides a signal that can be used as a reference to understand the worker's behaviour in greater detail.

5.3.4.1 HIT Acceptance Strategies

Firstly, the worker's HIT acceptance strategies have been interpreted by whether they previewed the HIT before accepting it. Typically, workers can choose to either preview a HIT on MTurk's HIT list page or accept and execute the HIT directly with or without using automated tools. If a worker accepted the HIT directly without previewing the HIT page, this means that he chose to accept the HIT directly on the MTurk HIT list page or use an automated tool to do so. Therefore, this cannot help identifying whether the worker used HIT catchers, and the installation status of Turk Guru and MTurk Suite browser HIT catchers were still used to identify the type of workers.

The step is to determine whether any of the HITs received and submitted by each worker has been previewed by that worker beforehand. One difficulty, however, is that the *request* table in the database did not successfully record the *worker_id* and *assignment_id* for which the HIT was previewed, due to the client's access restrictions. However, in the non-preview state when the worker browsed the HIT page, all the above information was successfully recorded. In order to get information about the worker identity on the preview page, a method of identification based on IP address was attempted. Specifically, the real IP address from clients stored in the *user_agent* table have been updated to the *request* table (Figure 5.21). The ip_address in the *request* table was used as the id to identify the worker, and then the *worker_id* collected in the non-preview state was used to fill in the missing records containing the same *ip_address*. In this way, the timestamp of each worker previewing each HIT can be addressed (Figure 5.22).

After obtaining information about the worker identity who previewed a particular HIT, the timestamp recorded in the *sqs_message* table regarding when the HITs were accepted can be compared with the specific timestamp of when each HIT page was previewed by that worker, and thus determine whether each worker previewed the HIT page before accepting it.

The analysis of the HIT page preview records revealed a total of 400 HITs being previewed. Moreover, workers from 234 unique IP addresses previewed these 400 HITs. It is worth noting that not all workers from these 234 unique IP addresses submitted any HITs in the end, as some of them only previewed the HITs but did not successfully accept it or submit it.

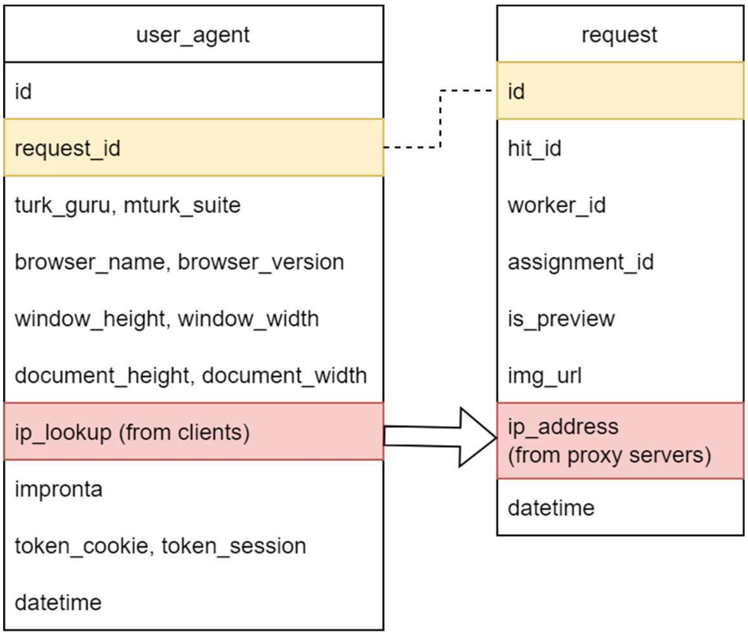

Figure 5.21 The process of updating IP address in request table.

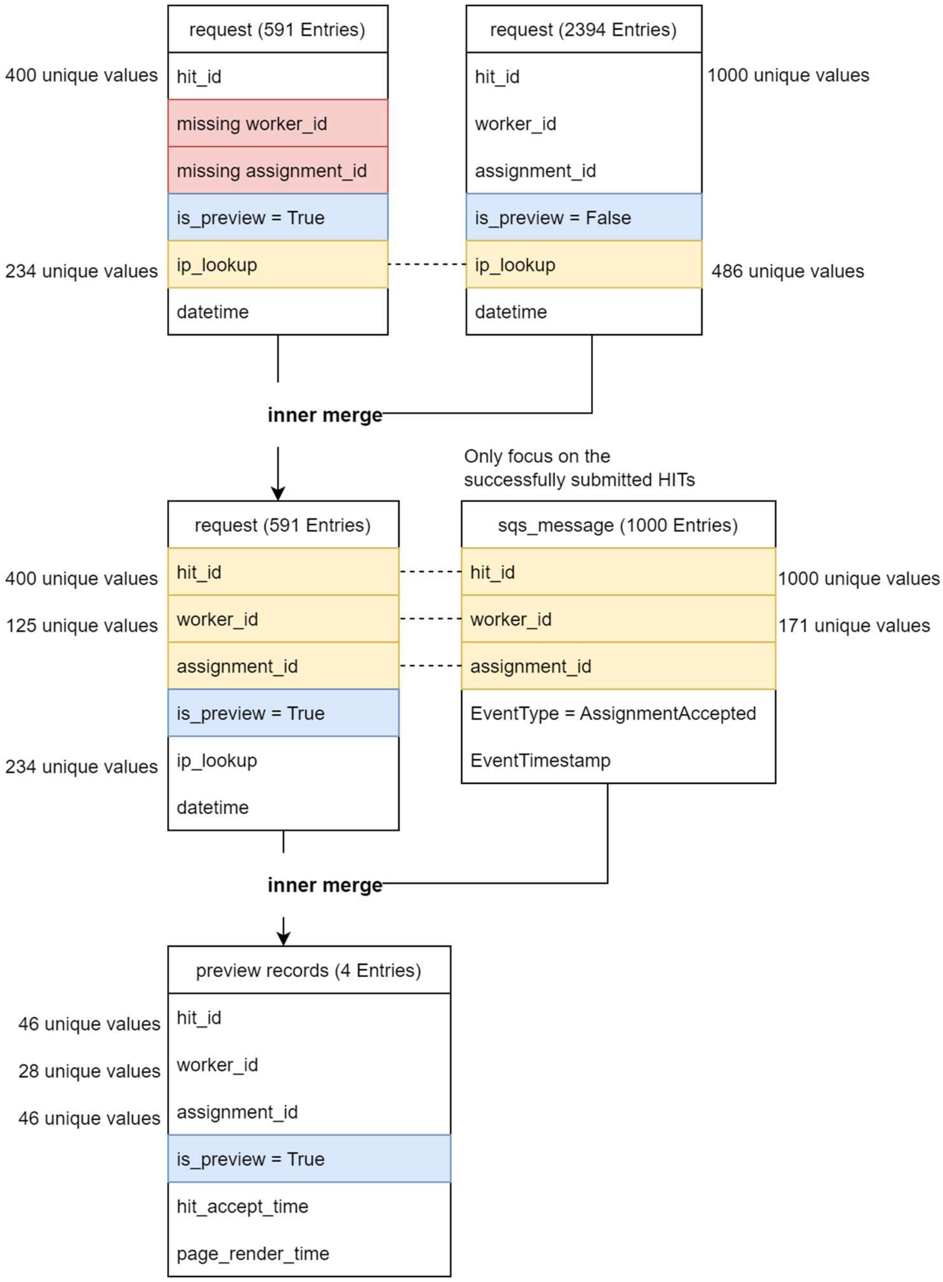

Figure 5.22 The process of finding the workers who previewed the HITs before HIT acceptance.

By comparing the records associated with the preview page with the records relating to the 1000 HITs that were successfully submitted, it was found that of the 171 workers who had successfully submitted HITs, only 28 workers previewed the HITs before accepting them. This means that only 28 worker accounts out of 234 IP addresses [39] who previewed HITs successfully accepted and submitted HITs. Given the low difficulty and high rewards of the

[39] The Worker ID from these IP addresses cannot be retrieved when they preview the HIT pages.

HITs used in this study, we presume that it is unlikely that workers would reject HITs after previewing them. This reflects the intense competition faced by the workers in accepting HITs. What is worth further discussion is the fact that for sought-after HITs, it is far more difficult to accept it after previewing the HIT page than to accept it directly by skipping the preview step via the HIT list page or using HIT catchers. This is because this HIT may have already been reserved by other workers during the preview. It was also found that all 125 worker accounts that previewed HITs came from as many as 234 IP addresses. This means that a large number of worker accounts were logged in from multiple devices during the experiment. This will be explained in detail in the simultaneous execution behaviour section.

5.3.4.2 Definitions of HIT States

There is a need to first identify and understand the different states of HITs in relation to worker behaviour, so that the analytical variables that need to be calculated in relation to work behaviour can be clarified and then their work behaviour can be studied.

As listed in Table 5.2, there are six main HIT states[40] to be explained. The last three HIT states are substates of the pending state. More specifically, the available state represents that one HIT has been successfully published to MTurk and become available to be accepted by workers. It is worth noting that one HIT being available does not mean it being visible. As explained in the previous chapter, the workers could accept an available HIT based on the Batch ID of the HIT group they belong to regardless of its visibility. Once this HIT gets accepted or reserved, it moves to the next state which is called pending.

In the pending state, this HIT has been assigned to a worker's HIT queue. Therefore, if the worker starts to work on this HIT by opening this HIT page, the HIT proceeds to the active state. In comparison, if the worker only reserves or accepts this HIT without even opening the HIT page, this HIT is identified as being in a backlogged state. Finally, if the worker is focusing on the HIT page by opening the tab from the browser, this HIT is identified as being in a focused state.

In addition to the Notifications of HIT events (stored in *sqs_message* table) obtained by Amazon SQS to determine the status changes of each HIT published in the experiment, the data collected from the client-side web page scripts (stored in *request* table) can also be used

[40] Understanding HIT States: https://blog.mturk.com/understanding-hit-states-d0bc9806c0ee

to determine which of the three pending states (active, backlogged, focused) each HIT was in at a specific timestamp. Next, the methods of determining the pending states are explained in detail and the different work strategies between workers are recognised on the basis of these state change data.

5.3.4.3 HIT Backlogged Time

Each worker might not open the HIT web page immediately after accepting it. Since the worker did not immediately start working on the HIT being accepted, the duration of a HIT being backlogged can also be interpreted as the duration of a HIT being delayed. By examining the differences in the amount of time different types of workers delay HITs, the impact of their work strategies on the speed of completion of the entire HIT group can be better understood.

To calculate the duration of a HIT being backlogged, the difference between the timestamp when the HIT was accepted and the timestamp when the HIT page was opened by the worker for the first time is required. Specifically, this time interval was obtained using the equation shown in Figure 5.23. In the equation, the Timestamp of a HIT web page being opened by this worker for the first time is also known as the timestamp of a worker actually started working on this HIT. This timestamp originates from the earliest *datetime* from the records with matching *worker_id* and *hit_id* within the *request* table. The duration of a HIT being backlogged is then obtained by calculating the interval between the timestamp when the page was first opened and the timestamp when it was accepted by the same worker. It should be noted that as the two sets of timestamps were collected from different approaches (one from the client scripts, the other from Amazon SQS), the time zones need to be normalised prior to the calculation.

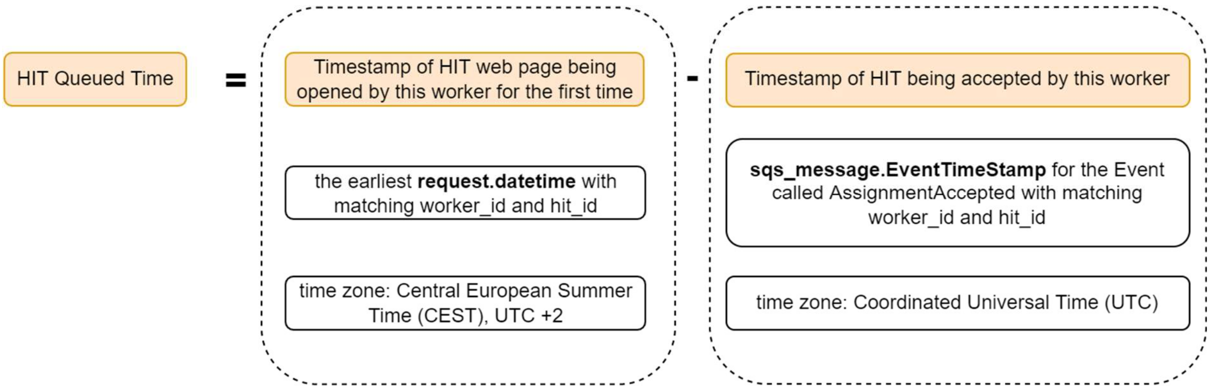

Figure 5.23 Equation of calculating the duration of HIT being backlogged.

In the end, the HIT backlogged time was calculated for each of the 1000 final successfully submitted HITs. As Figure 5.24 illustrates, approximately 800 HITs had their task pages successfully opened within 50 seconds of being accepted. The other 200 HITs were delayed for more than 50 seconds before being opened.

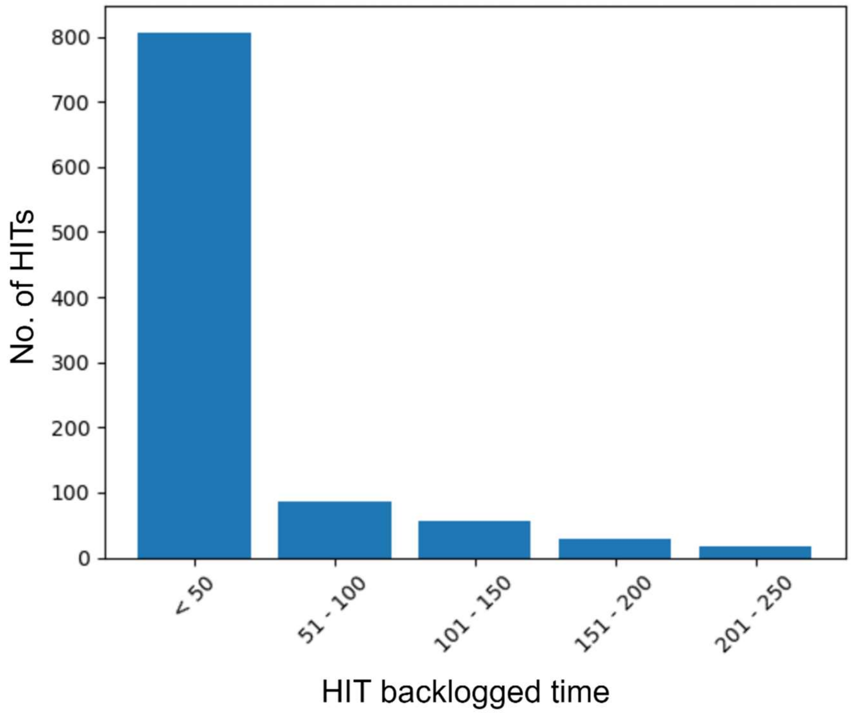

Figure 5.24 Distribution of number of HITs with different backlogged time.

An ECDF plot[41] could help understand the proportion of workers falling in each HIT backlogged time. Regarding the 135 workers using Chrome, it could be revealed from Figure 5.25 that under the same cumulative proportions of workers, those using HIT catchers had much longer HIT backlogged time than those not using HIT catchers. This difference is particularly significant when the proportions reach above 0.6 for both types of workers.

[41] Seaborn Documentation: https://seaborn.pydata.org/generated/seaborn.ecdfplot.html

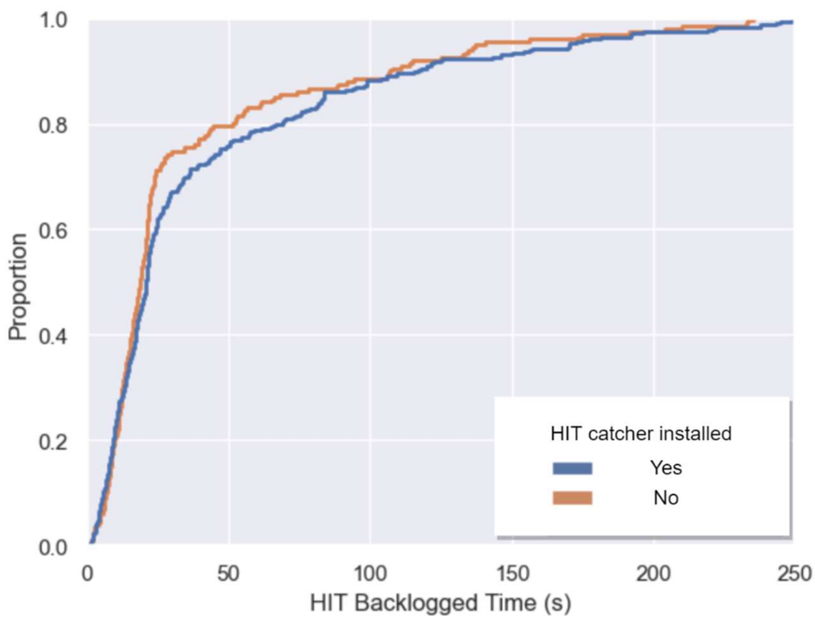

Figure 5.25 A comparison of cumulative backlogged time between HITs submitted by workers with and without HIT catchers.

To determine whether the use of HIT catchers influenced the amount of time each HIT got backlogged, we conducted an independent samples t-test. In this study, the sample sizes are robust enough to proceed with a t-test even if the data are not perfectly normally distributed (Poncet et al., 2016). This test compared the mean backlogged time between two groups: workers using HIT catchers and those without tools. The analysis generated a t-statistic of around 1.19, reflecting a difference in means that was more than one standard deviation apart. The associated p-value was around 0.24, indicating that the probability of observing such a difference, or one more extreme, under the null hypothesis was about 24%. Given that this p-value exceeds the conventional significance level of 0.05, we fail to reject the null hypothesis. There is insufficient statistical evidence to assert that the installation of HIT catchers has a significant impact on the time HITs are backlogged by workers. This finding suggests that while there may be a difference in the HIT backlogged time, this difference is not statistically significant at the commonly accepted threshold. It is important to note that a lack of statistical significance does not imply a lack of effect, but rather that the effect, if present, is not detectable within the variability of our data.

Table 5.7 Independent samples t-test on the impact of HIT catchers on HIT backlogged time.

Group	N	Mean	SD	t(df)	p
Tasks completed by HIT catcher workers	315	41.77	52.30		
Tasks completed by non-HIT catcher workers	203	36.45	47.58		
Combined	518			1.19(518)	.24

Figure 5.26 illustrates the percentage distribution of HIT counts across backlog time intervals by the two types of workers. Regarding the first duration interval, more than 60% of all HITs submitted by HIT catcher workers were backlogged less than 25s. In comparison, more than 70% of all HITs submitted by non-HIT catcher workers were backlogged less than 25s. This indicates that HIT catcher workers backlogged a larger proportion of HITs by more than 25s than non-HIT catcher workers.

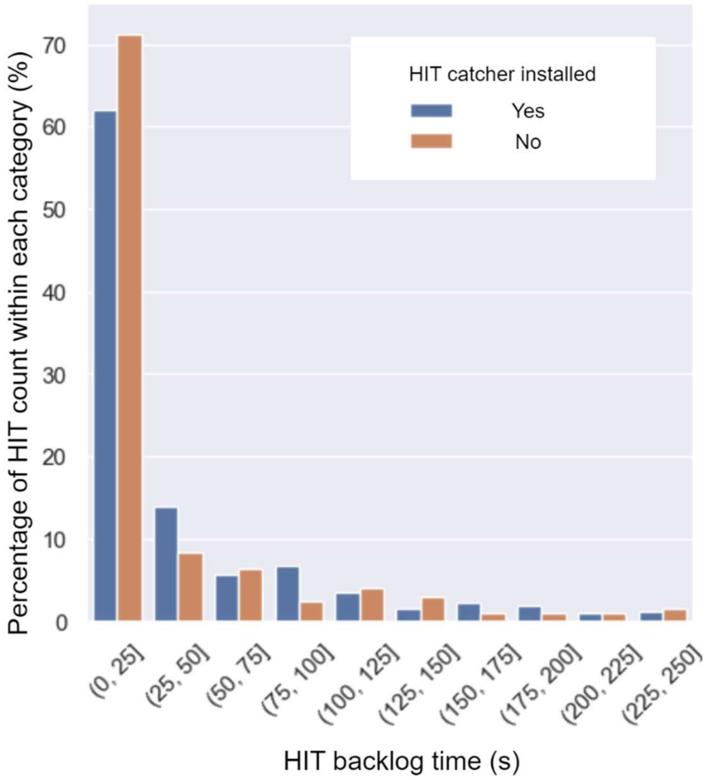

Figure 5.26 Distribution of percentage of HITs submitted by workers with and without HIT catchers regarding the HIT backlogged time.

5.3.4.4 HIT Focus Time

As explained in the HIT opportunities section, cleaned results have been stored in a PostgreSQL database containing five tables: *classification, result, request, user_agent,* and *sqs_message*. In order to calculate the time spent on doing HITs, a pandas dataframe has been created by joining the *result*, *user_agent* to the *request* table based on their unique ids (Figure 5.27).

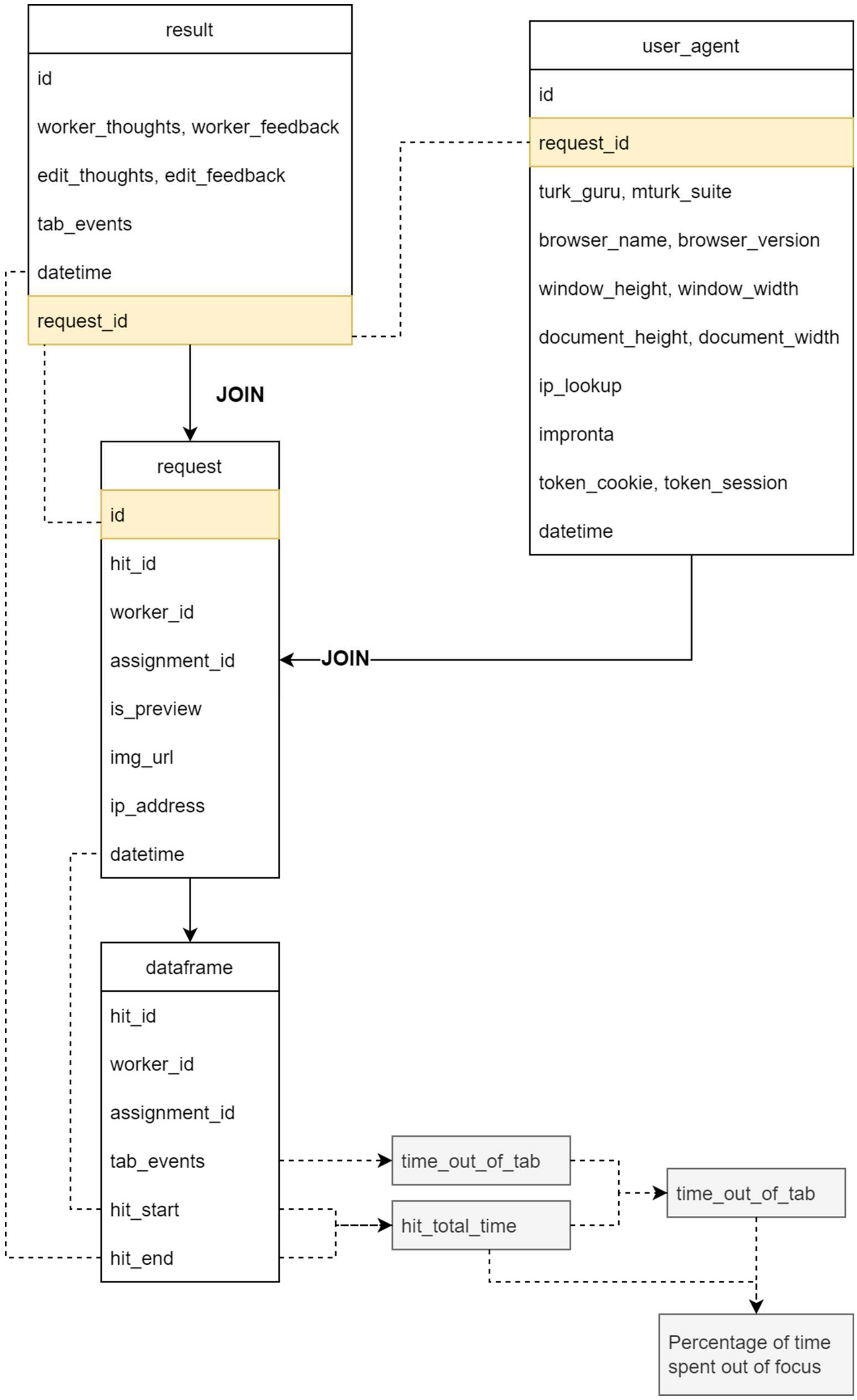

Figure 5.27 Calculation process of time spent on doing HITs.

First, *hit_total_time* has been calculated as the interval between *hit_end* and *hit_start*. In detail, *hit_total_time* represents the total time spent for the worker to submit a HIT since it was opened in their browser. As all data within the *result* table has been generated by the web script at the moment that a worker submitted the HIT response, *hit_end* represents the timestamp of the HIT submission. Similarly, *hit_start* represents the time a worker started doing this HIT because the data from the *request* table was generated when the worker opened the HIT page for the first time.

Subsequently, *time_out_of_tab* has been calculated. This variable represents the time a worker has spent out of viewing the HIT page during the HIT completion. In other words, a worker may switch to another web page which is irrelevant to the HIT they are currently working on. As a result, the actual working time could be much shorter than the overall time spent on HIT completion. This interval has been calculated as the sum of all time intervals between the moment a worker left the HIT page and the moment they started focusing on the HIT page. These two events have been identified via the web script and documented as the *tab_events* from the *result* table. In the same way, the time spent on focusing on the HIT has been calculated by subtracting the *time_out_of_tab* from the *hit_total_time.*

Finally, the empirical distribution function (ECDF) was applied to understand the differences in the amount of time the HIT catcher (with either of the two HIT catchers installed) and non-HIT catcher workers spent focused on the HIT page during HIT completion in this study. As indicated in Figure 5.28, the workers using at least one HIT catcher were more likely to spend less time focusing on the HIT page than those without using the HIT catchers being monitored in this study.

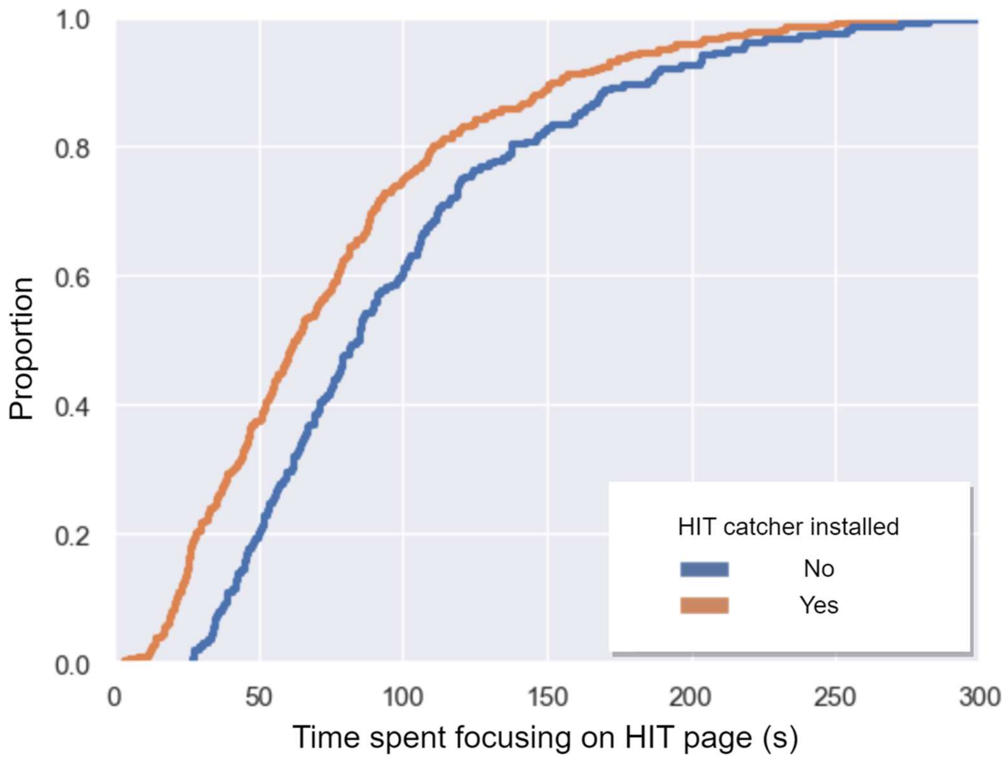

Figure 5.28 A comparison of average time spent focusing on the HIT page between the workers with and without the HIT catchers.

From Figure 5.29, it can be revealed that HIT catcher workers complete significantly more HITs than non-HIT catcher workers with a focus duration of less than 80s. In comparison, the HIT counts with a focus time of 100s and above is relatively similar for both types of workers. Combined with the distribution of percentage of HIT counts in Figure 5.30, it can be seen that a large percentage of HITs submitted by HIT catcher workers have a HIT focus time of less than 40s compared to non-HIT catcher workers. Furthermore, a greater proportion of HITs submitted by non-HIT catcher workers have a focus time greater than 120s. These differences in numbers and proportions reflect the tendency of non-HIT catcher workers to spend more time focusing on HITs.

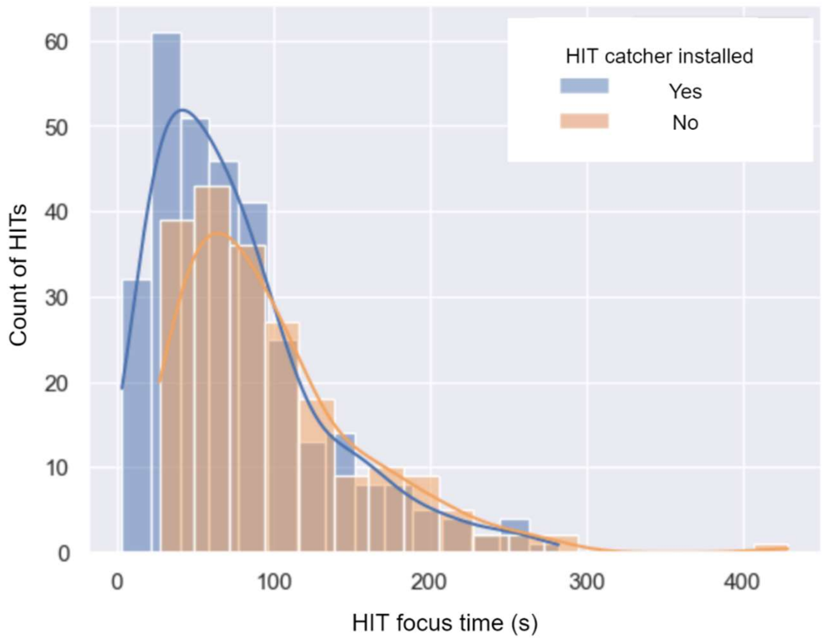

Figure 5.29 Distribution of HIT focus time for both types of workers.

To further explore the effect of workers' installation status of the two monitored HIT catchers on the level of focus during the HIT completion, a Two-way ANOVA test was used to investigate the effect of installing MTurk Suite and Turk Guru on the mean time workers spent actively focusing on HIT work (see Table 5.2 for definition of HIT focus state). Table 5.8 presents the results of the two-way ANOVA, which indicated significant effects of both MTurk Suite and Turk Guru on HIT performance duration.

Table 5.8 Effects of MTurk Suite and Turk Guru on HIT Performance Duration.

Source	SS	df	MS	F	p
MTurk Suite	39,583.49	1	39,583.49	13.23	.00003
Turk Guru	17,283.44	1	17,283.44	5.78	.016
MTurk Suite x Turk Guru	106,996.40	1	106,996.40	35.75	< .00001
Residual	1,538,045.00	514	2,992.17		

There was a significant main effect of MTurk Suite on the duration of HIT performance, $F(1, 514) = 13.23$, $p = 0.00003$, indicating that the installation of MTurk Suite significantly influenced the time workers spent actively focusing on HITs. Additionally, there was a significant main effect of the installation of Turk Guru, $F(1, 514) = 5.78$, $p = 0.016$, suggesting that Turk Guru also had a significant effect on HIT performance duration.

More importantly, the interaction effect between MTurk Suite and Turk Guru was significant, $F(1, 514) = 35.75$, $p < .00001$. This significant interaction indicates that the combined effect of having both MTurk Suite and Turk Guru installed is different from the sum of their individual effects on HIT performance duration.

The significant main effects and interaction suggest that while both MTurk Suite and Turk Guru individually contribute to changes in HIT performance duration, their combination leads to a different, more pronounced effect. These results support the notion that the integration of multiple HIT catchers may have a synergistic effect on worker performance.

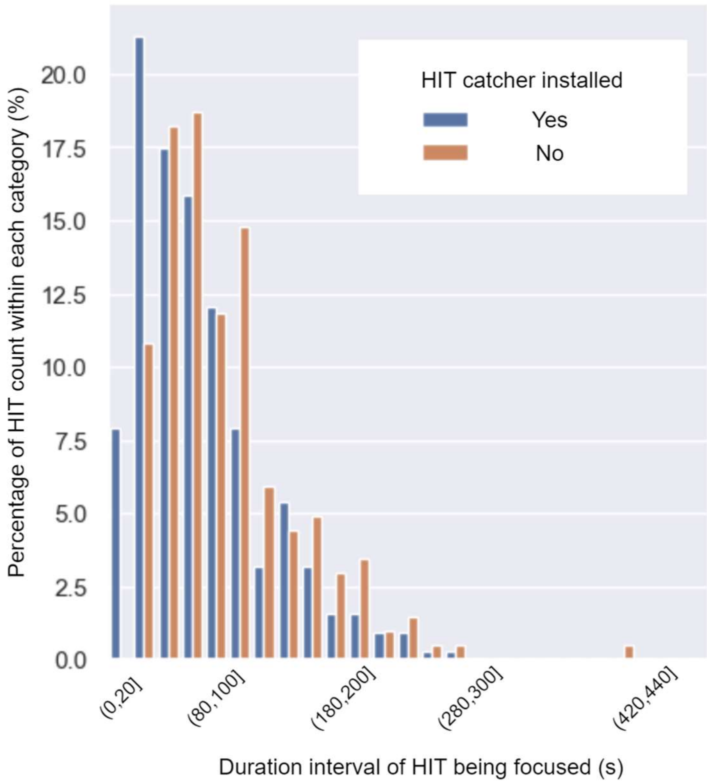

Figure 5.30 Distribution of percentage of HITs submitted by both types of workers regarding the HIT focus time.

Followed by the comparison of their focus time, the percentage of time spent out of the HIT page has been generated based on *hit_total_time* and *time_out_of_tab* for both types of workers. As illustrated in Figure 5.31, almost 40% of workers under both categories maintained focus on the HIT page from start till submission. By cumulating the worker proportions under different out of focus time percentages, it can be revealed that those with HIT catchers are more likely to spend time outside of the HIT page than those not using the HIT catchers being monitored in this study. On average, HIT catcher workers spent 13.6% more time out of focus when doing each HIT (16.15s vs 14.21s). Moreover, their average time spent focusing on HITs was less than those without HIT catchers by 22.3% (76.87s vs 98.95s).

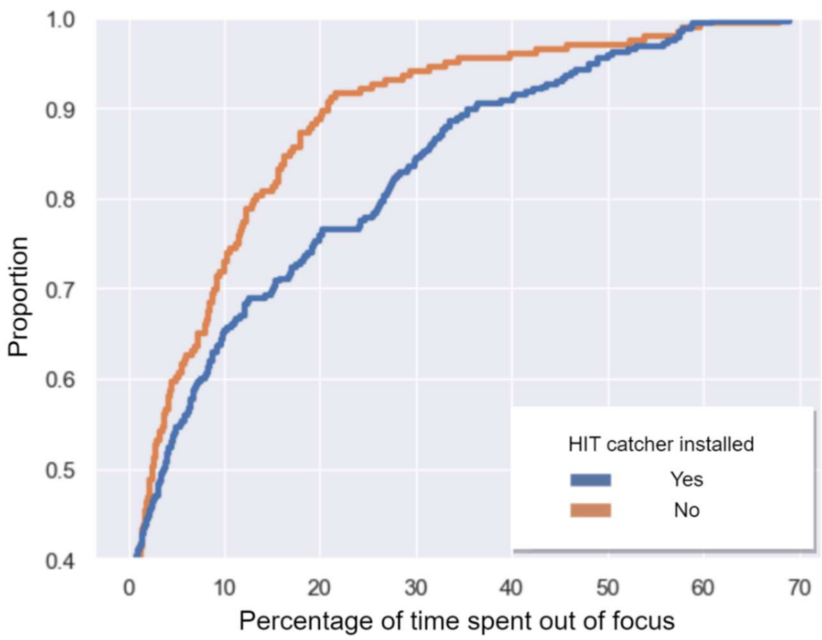

Figure 5.31 A comparison of average time spent not focusing on the HIT page between the workers with and without the HIT catchers.

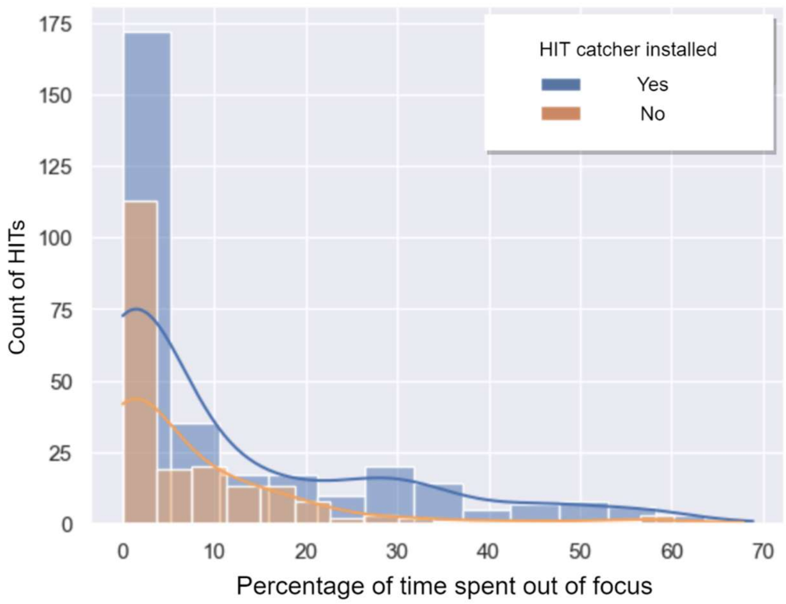

Figure 5.32 Distribution of time out of focus (%) for both types of workers.

To evaluate the effect of HIT catcher installation on participants' distractions at work, an independent samples t-test was performed comparing the percentage of time out of focus between tasks completed by workers with and without the HIT catchers. As revealed in Table 5.9, the analysis yielded a significant result, t = 3.14, p = 0.0018. This significant positive t-value indicates that workers with the HIT catchers tended to spend a larger percentage of their time out of focus compared to those without the HIT catchers.

Given the p-value is significantly less than the alpha level of 0.05, we reject the null hypothesis and conclude that there is a statistically significant difference in the time spent out of focus due to the HIT catcher installation. This suggests that the presence of the HIT catchers may be associated with increased distraction or multitasking behaviour among workers.

Table 5.9 Independent Samples T-Test on the Effect of HIT Catcher Installation on Workers' Distractions.

Group	N	Mean	SD	t(df)	p
Tasks completed by HIT catcher workers	315	11.95	16.25		
Tasks completed by non-HIT catcher workers	203	7.96	12.45		
Combined	518			3.14(516)	.0018

This finding is consistent with the finding in Figure 5.20 and Figure 5.21 that workers using the HIT catchers were more likely to spend less time on executing HITs and more time out of the HIT pages.

5.3.4.5 Multi-HITs: Doing Multiple HITs All Together

The significant correlation between workers' installation status of HIT catchers and their mean HIT active state time could be potentially due to their low work engagement, while the simultaneous execution behaviour could be a potential cause of their low engagement. To test this assumption, workers' simultaneous execution behaviours need to be examined. This behaviour is defined as one worker doing multiple HITs simultaneously. In other words, one worker would be defined as a multi-HITer if there is an overlap between at least two HITs being active. Table 5.10 has been generated after checking through all 135 workers who use Chrome web browser for HIT completion. It can be revealed that workers with the HIT catchers installed did multiple HITs simultaneously twice as much as those without it. In extreme cases, a very small percentage of workers even use multiple devices connected to the same account to maximise reservation and completion of HITs. Among the workers detected to be multi-HITers, an average of three HITs have been performed in parallel from a unique worker account.

These findings also confirm the observations made in the study from Ghosh et al. (2019) regarding worker behaviour.

Table 5.10 Multi-HITing behaviours under different HIT catcher installation status.

	Multi-HITer	Non-multi-HITer	Total
HIT catchers detected	12	55	67
	18%	82%	100%
No HIT catchers detected	6	62	68
	9%	91%	100%
Total	18	117	135

The phenomenon of multiple HITs being executed by the same worker account on multiple devices simultaneously raises the suspicion that there is a possibility of multiple workers using one account together. This irregular use of accounts has also been found in the forum discussion:

> *"I had a worker that worked on my account, so please refrain from criticising me about it so the matter doesn't become irrelevant (I know I shouldn't give my account to others). Nonetheless, he was earning (anonymised amount) each week before he changed my bank account and fooled me. I don't understand how he was doing it because he was obtaining qualifications that didn't seem to be available to anyone else." (Anonymous Turker, 2022)*

Interestingly, by examining all 4834 HIT page access logs stored in the *request* table, a total of 482 *worker_id* from 590 *ip_address* was found. It should be noted that all 4834 logs also include the HIT page visits not being successfully submitted, resulting in 2300 distinct *assignment_ids*. In comparison, the page access logs used in Section 5.3.4.1 HIT acceptance strategies only contain the last 1000 *assignment_ids* that were submitted. When examining the number of worker accounts corresponding to each IP, it was found that 14 of these IP addresses were logged into two worker accounts during the experiment. In contrast, when looking at the number of IP addresses linking to each worker account, 15 worker accounts were found to have been logged into from two or three IP addresses. This means that not only were there multiple workers sharing one worker account off-site, but there was also the phenomenon of one device logging into multiple worker accounts during the experiment. In addition to doing multiple HITs simultaneously (e.g., a worker using multiple devices to log in to the same account) and accessing the platform via VPN (Marshall et al., 2023; Zhang et al., 2022), these unusual phenomena could also be due to multiple individuals sharing the same worker account.

The reasons for sharing their own worker accounts with other individuals include earning more money or to help those without their own worker accounts. Specifically, many HITs published on MTurk only allow one submission from each worker account. It is possible for one individual to do such HITs multiple times with different worker accounts and thus improve their work efficiency. This involves a new approach of reward distribution, as those who lend out worker accounts often need a share of the rewards. In addition, the difficulty in obtaining a worker account could be a reason for account sharing among individuals. An applicant has to meet requirements before opening a worker account on MTurk, such as a proof of residence in a specific country. Those without such information would have to do HITs using the worker accounts from others.

With no doubt, sharing accounts among others could lead to further issues. From the workers' perspective, their accounts could face the risks of being suspended for violating the policies, which can result in losing all rewards. For the requesters, the data collected could be biased due to the duplication of worker identities.

In general, the number of HIT completions, HIT backlogged time, and focus time showed significant differences between workers with and without the HIT catcher during the experiment (Table 5.11). More precisely, the HIT catcher workers completed more HITs and spent 26.3% more time backlogging HITs in the HIT queue than the non-HIT catcher workers in average. In addition, their average time spent focusing on HITs was less than those without HIT catchers by 22.3%. Regarding the time spent out of focus, HIT catcher workers spent 13.6% more time when doing each HIT on average. In other words, HIT catcher workers were more inclined to backlog HITs (leave them idle in HIT queues), spend less time on doing HITs, and were less focused. Furthermore, HIT catcher workers contain a larger proportion of multi-HITers.

Table 5.11 Descriptions of HIT states.

	HIT catchers	No HIT catchers
No. of workers	67	68
No. of completed HITs per worker	5	3
Average time each HIT being queued (s)	41.77	33.07
Average time focusing on each HIT (s)	76.87	98.95

Average time out of focus for each HIT (s)	16.15	14.21
Average percentage of time spent out of focus for each HIT(%)	11.95	7.96

5.3.5 Accuracy of Image Annotation

Following the steps introduced in the Method section, the accuracy scores of image annotation results for each HIT have been calculated. As illustrated in Figure 5.33, the mean scores of the 135 participants who submitted tasks using the Chrome browser were categorized according to whether HIT catchers were detected. It can be revealed that the mean Recalls and F-scores of the annotation results from the workers with the monitored HIT catchers are generally lower than the workers without HIT catchers. However, no statistical significance has been found on the difference in accuracy between the annotation results by two types of workers. In other words, the workers using the monitored HIT catchers annotated the images with a similar quality to those not using HIT catchers. But the annotations were less completed (lower Recall score) due to a potential less engagement as revealed in Section 5.3.4.4 (HIT focus time).

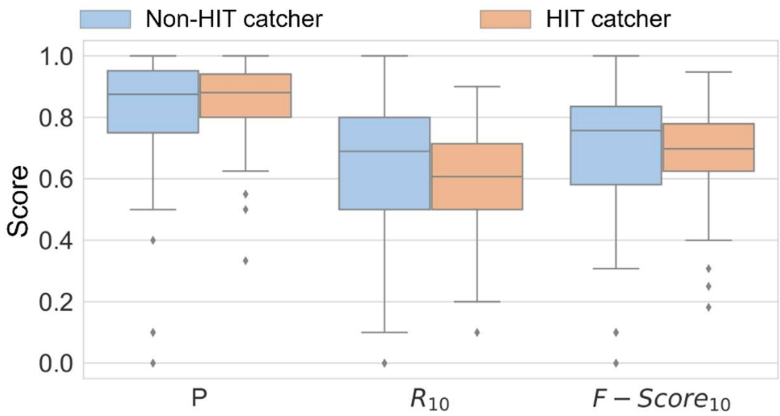

Figure 5.33 Average accuracy scores for two types of workers (67 HIT catcher workers and 68 non-HIT catcher workers).

However, if we focus on the scores for each annotation task, which is 518 HITs submitted via Chrome, the tasks accomplished with the HIT catcher have a higher average precision than those without it (Figure 5.34). Combined with the distribution of the number of HIT submissions for 67 workers using the HIT catcher (Figure 5.35), this difference on the average scores is probably caused by the fact that a small number of workers completed most of them with higher quality. Thus, focusing solely on scores for each annotation task affected our judgment of the overall performance of the worker group using the HIT catcher. In addition, this difference in scores means that we cannot predict the annotation quality simply by whether a HIT catcher is detected, but also need to combine it with other behavioural features.

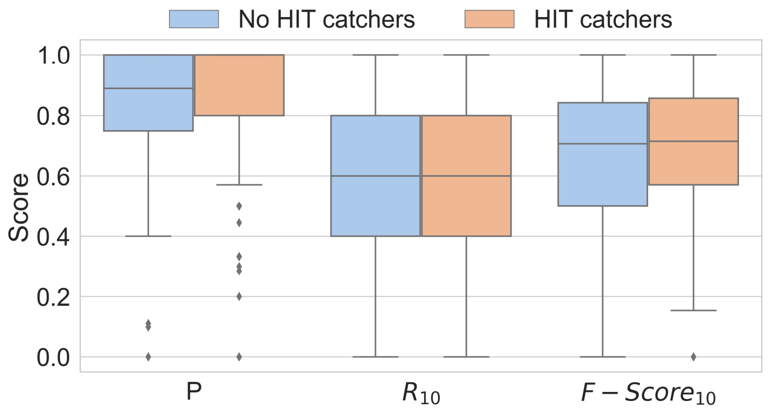

Figure 5.34 Average accuracy scores for image annotations (315 HITs completed with HIT catchers and 203 HITs completed without HIT catchers)**.**

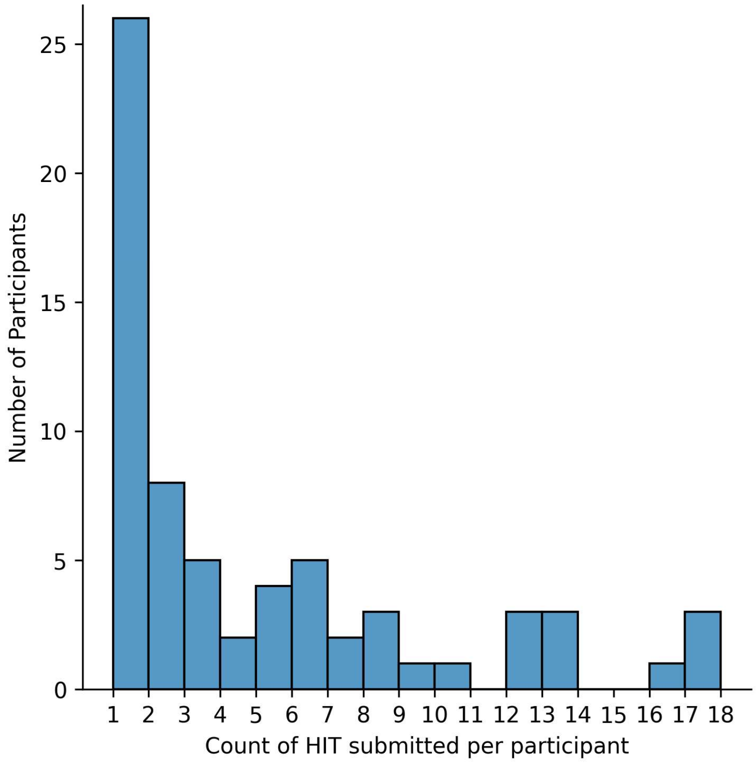

Figure 5.35 Distribution of the number of HITs submitted by each participant detected using HIT catchers.

In summary, the analysis of results reveals that non-HIT catcher workers outperform HIT catcher workers in terms of Recall and F-score, suggesting superior annotation quality and completeness.

5.3.6 Review of Textual Response

To provide a comprehensive reflection on the worker textual responses, word frequencies, diversity, and efforts made in textual responses are discussed. This analysis provides insights into the workers' engagement with the task, their understanding of the requirements, and the overall usefulness of their textual contributions.

5.3.6.1 Frequency of Words

A few examples of “thoughts” are: “Nice place to live. Lots of shades here”, “happy to have few cars on the way”, “safe and secure neighbour hood”, and “good”. Examples of “feedback” are: “Short and very easy instruction to follow. Also task layout is easy to follow”, “I like this task, as I can remain engaged mentally since it involves analyzing the contents of photos”, “The hit is interesting”.

The most common words in the “thoughts” are “good”, “looks”, “the”, “it”, “is”, “very”, and “feel”. This suggests that many workers express positive sentiments (like “good” and “nice”) and often use words that describe perceptions or appearances (like “looks” and “feel”).

Regarding their “feedback”, the most frequent words are “good”, “task”, “to”, “nice”, “easy”, “and”, “the”, “is”, “I”, and “none”. The prominence of words like “easy” and “nice” indicates that workers often comment on the nature of the task itself, possibly referring to its simplicity or pleasantness.

The prevalence of simple and positive words like “good”, “nice”, and “easy” in both their thoughts and feedback suggests that the responses are generally positive. However, the use of such generic terms might also indicate a lack of detailed or specific feedback. The repetition of such words across many responses might point to a certain level of repetitiveness in the content.

5.3.6.2 Diversity of Textual Response

The diversity of textual responses is important, especially in creative tasks such as novel writing (Teevan et al., 2016). As explained in the HIT design section, workers were asked to provide their subjective perception about seeing the street view images (named 'thoughts' in the results) and general feedback on the HIT (named 'feedback'). The HITs looking for workers’ textual description of their thoughts are often used to train natural language processing algorithms, in which case a high degree of response diversity is often required (Cho et al., 2019). On the other hand, their “feedback” works as a control field when investigating the diversity of their “thoughts”. This is because the “HIT feedback” for each worker may not change significantly when doing multiple HITs in the same HIT group for this study.

In natural language processing, the TfidfVectorizer[42] function is commonly used to convert text as vectors of features that capture the importance of each word or phrase. A similarity score was obtained by comparing the vectors of different responses. Moreover, the similarity score could be influenced by the length of response. However, other important factors including relevance or accuracy of response were not evaluated here.

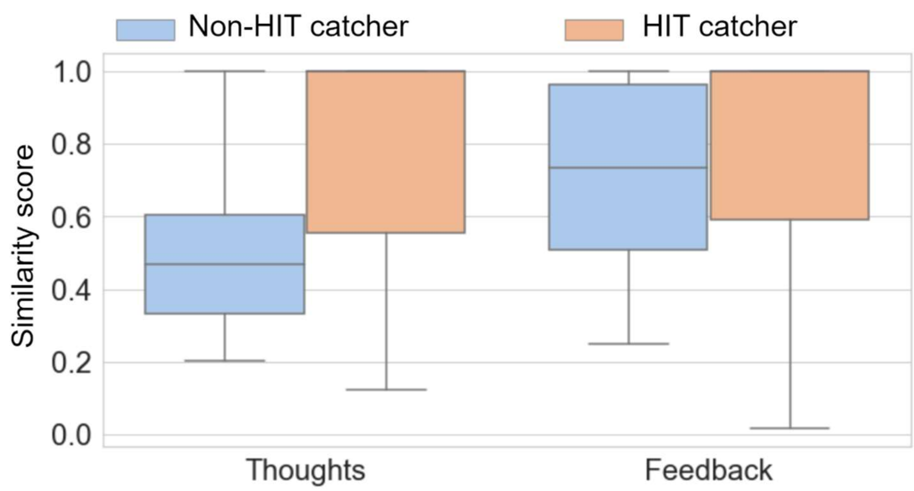

Figure 5.36 Average pairwise cosine similarity between worker TF-IDF text inputs.

In order to compute the diversity of the textual responses of a worker across the multiple HITs, the cosine similarity of the TF-IDF representation for each pair of responses was calculated using the TfidfVectorizer function from The sklearn.feature_extraction module. Then the mean similarity between all pairs is used as a similarity score for each worker. Before this computation, all workers that submitted only one HIT had been excluded. As revealed in Figure 5.36, for the ‘thoughts’ textual input, the HIT catcher workers showed a higher similarity (and thus lower diversity), while the non-HIT catcher workers showed substantially higher diversity, while both groups scored similarly for the less image dependent ‘feedback’ field.

5.3.6.3 Efforts Made in Textual Response

The definition of the number of text edits is that: each character added or deleted by the worker when completing the feedback is counted as a text edit. Character-level edits capture finer

[42] Scikit-learn Documentation: https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html

details in the editing process. While word-level edits provide a broad view of changes, character-level edits can reveal subtle but significant modifications like spelling corrections, punctuation adjustments, and small additions or deletions.

As the complexity of the question and the individual's familiarity with the topic could affect their editing behaviours, the number of text edits made by workers in the short answer question was not necessarily an indication of the completeness or accuracy of their textual responses. However, if one worker made multiple text edits, this may indicate that they are taking the time to carefully consider their answer and make any necessary changes. So, by studying the number of text edits in responses between the two types of workers, their attitudes and efforts spent when filling in content could be assessed and compared.

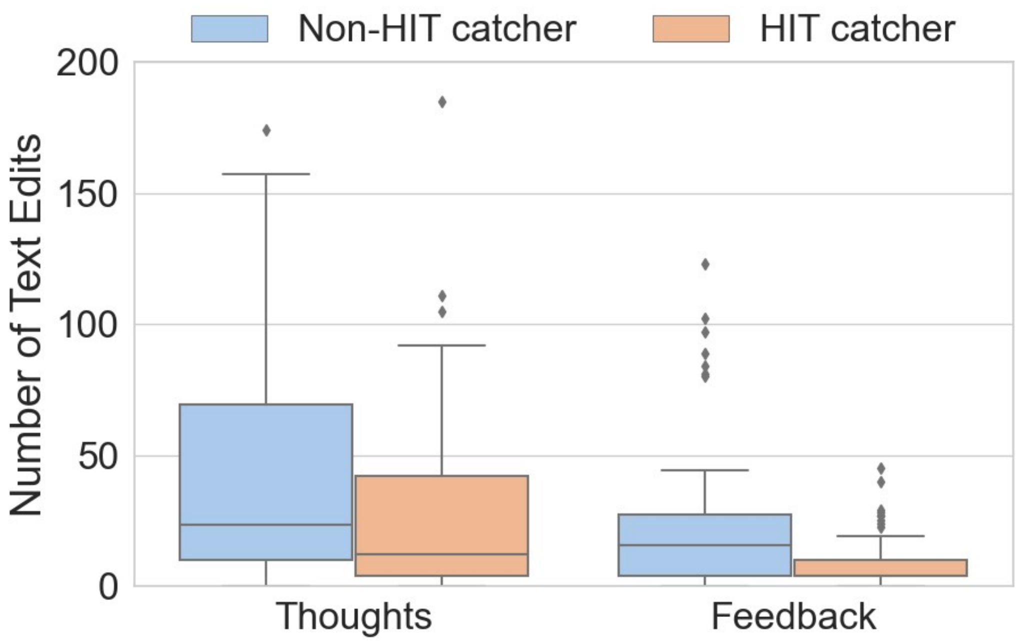

Figure 5.37 Comparison of numbers of text edit for both types of workers.

As illustrated in Figure 5.37, regarding the thoughts of street view images and feedback of HITs, non-HIT catcher workers made more text edits than HIT catcher workers overall. In addition, non-HIT catcher workers showed greater fluctuations in the number of edits than HIT catcher workers, reflecting the higher diversity of editing behaviours in the non-HIT catcher worker group.

The complexity of textual responses could also reflect the efforts each worker made in answering questions. More complicated responses usually require more cognitive effort, indicating deeper engagement with the work and a better understanding of the subject. Workers who provide extensive explanations are more likely to generate text with greater complexity. For short responses collected in this study, the average word length and syllable count were assessed to offer insights into the vocabulary complexity. Figure 5.38 and Figure 5.39 compare the average word length and average syllable count for two types of workers regarding their textual responses. For the word length, the median average word length for non-HIT catcher workers in their “Thoughts” and “Feedback” appears to be higher than for HIT catchers, indicating they use slightly longer words on average. The interquartile range (IQR) for non-HIT catcher workers, which shows the middle 50% of data, is slightly larger for HIT catchers in both “Thoughts” and “Feedback”, suggesting greater variability in the word lengths they use. Moreover, both types of workers have a broader distribution of average word length in their “Thoughts” compared to their “Feedback”.

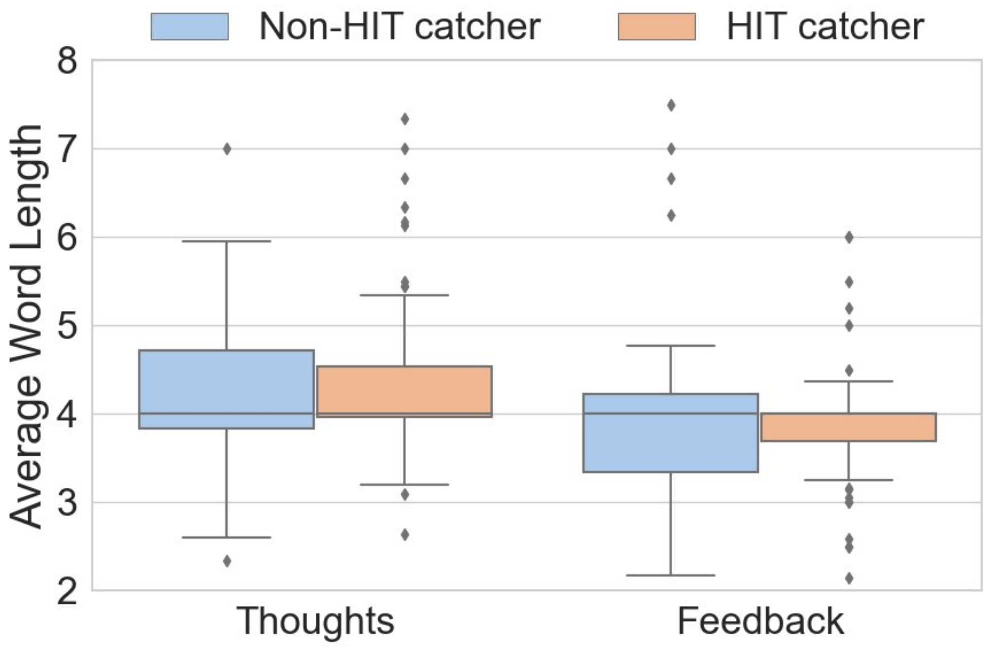

Figure 5.38 Comparison of average word length of textual responses for both types of workers.

Regarding their average syllable count as revealed in Figure 5.39, the median average syllable count for “Feedback” is marginally higher for non-HIT catcher workers. The IQR for average syllable count for HIT catcher workers in “Feedback”, suggesting most HIT catcher workers use a similar mix of syllable counts in their responses. In addition, the range of syllable counts for non-HIT catcher workers, particularly in “Feedback”, is wider, suggesting more variation in syllable use.

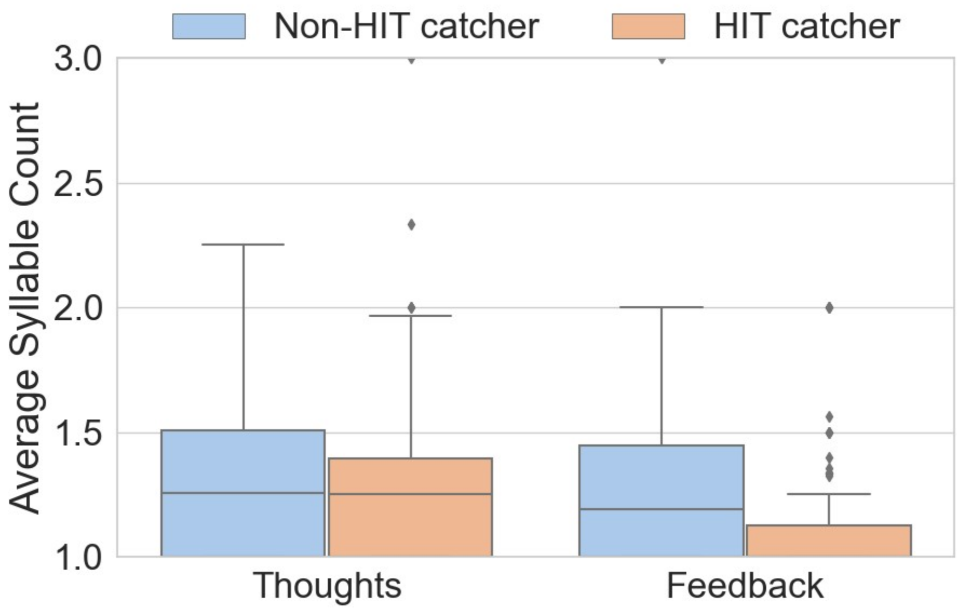

Figure 5.39 Comparison of average syllable count of textual responses for both types of workers.

In summary, non-HIT catcher workers exhibit more effort in text editing with higher diversity, as evidenced by their use of longer words and more diverse syllable counts in their responses. This is demonstrated in the average word length and syllable count metrics, where non-HIT catcher workers show a tendency to construct more complex responses, both in "Thoughts" and "Feedback". The data also indicates greater variability in the responses of non-HIT catcher workers, which may reflect a broader range of response strategies or differences in task engagement. Overall, these patterns suggest that non-HIT catcher workers may engage more deeply with tasks and produce more detailed textual responses.

5.3.7 Predicting Annotation Quality Based on Behaviour Data Using Support Vector Classifier

This section aims to predict whether the F-score of the annotation passes or fails by using the behavioural features of the participants and whether HIT catchers were detected. To achieve this goal, we trained a Support Vector Classifier (SVC) model for the quality classification.

Support vector machine (SVM) is a supervised learning algorithm, and for classification tasks especially binary classification problems, the goal of SVM is to find a hyperplane to best separate two classes of data points (Xia et al., 2015).

Compared to logistic regression, which is used for linear classification, SVC can solve non-linear classification problems by dealing with high-dimensional data through the use of kernel functions (Sheykhmousa et al., 2020; Silva et al., 2020). SVC allows the use of different kernel functions (Linear, Gaussian, Polynomial, Sigmoid) to accommodate different prediction methods (Azzeh et al., 2023). Due to the relatively small sample size of our training data (518 samples), complex models such as neural networks are more likely to be overfitted (Bornschein et al., 2020), and SVC could be a better option. Furthermore, we use the cross-validation method to avoid overfitting when finding the optimal model parameters (Ghojogh & Crowley, 2019).

5.3.7.1 Threshold for pass/fail classification of image annotation results

First, the threshold for pass-fail classification of results needs to be specified. Based on the Precision, Recall, and F-score line plots for the 1000 results that were submitted (Figure 5.40), it can be seen that Precision is generally higher than Recall, and there does not seem to be a clear trend of positive or negative correlation between Precision and Recall.

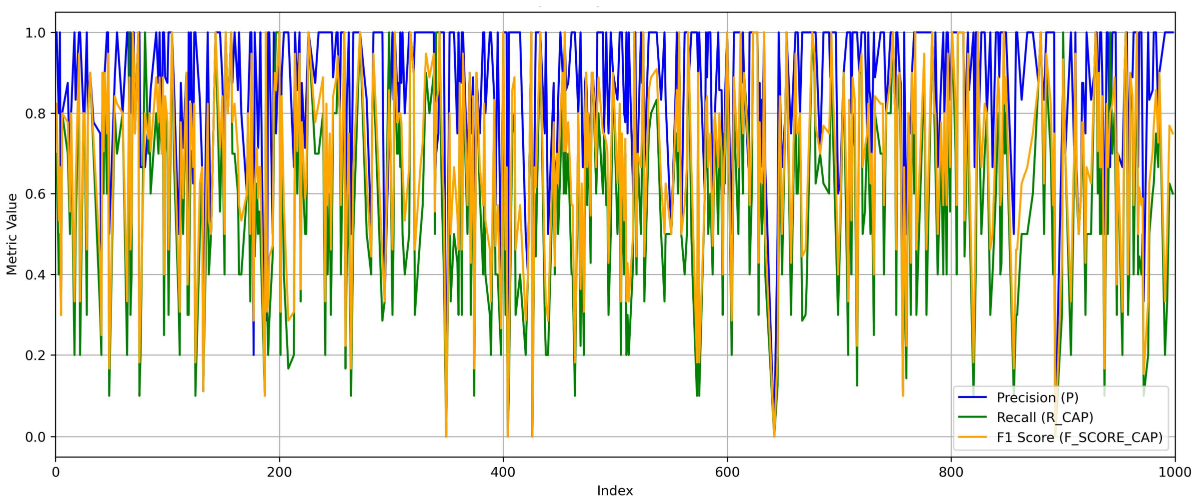

Figure 5.40 Precision, Recall, and F-score line plots for submitted results.

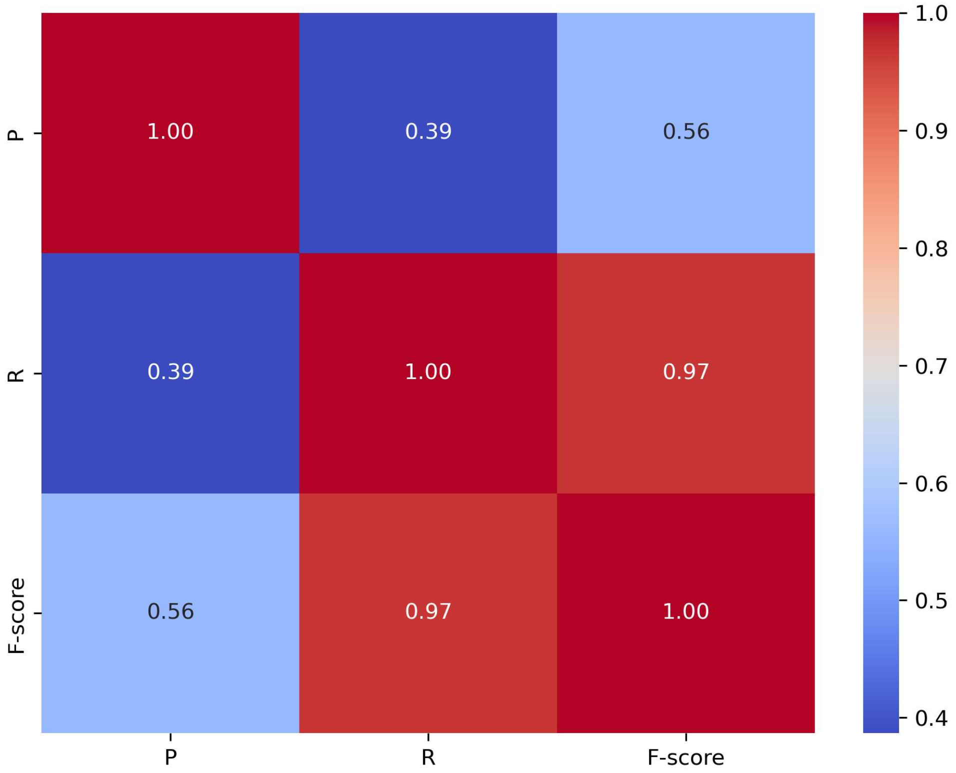

Figure 5.41 Heatmap of Precision, Recall and F-score.

Interpreting the correlations among these three metrics provides insights into the impact on other metrics when setting thresholds for one metric. Through observation of the heatmap (Figure 5.41), a notable correlation of 0.97 between recall (R) and F-score is evident, indicating a strong tendency for F-score to increase as recall improves. The correlation between precision (P) and F-score is moderate at 0.56, suggesting that precision moderately influences F-score, although not as strongly as the relationship between recall and F1-score. The correlation between precision (P) and recall (R) is comparatively lower at 0.39.

Table 5.12 Statistical Summary of Precision, Recall, and F-score.

	Mean	Median	Mode	Skew
Precision	0.863	0.9	1.0	-1.970
Recall	0.598	0.6	0.5	-0.186
F-score	0.677	0.706	1.0	-0.659

Image annotation results are widely used to train and fine-tune machine learning image recognition models. However, in diverse application scenarios, the emphasis on precision and recall varies across different models. For instance, in medical diagnosis, greater importance may be placed on recall to mitigate instances of false negatives and minimise the risk of misdiagnosis (Islam et al., 2020). In contrast, within recommendation systems, a greater emphasis might be placed on precision to minimize irrelevant recommendations. The F-score, which considers the harmonic mean of precision and recall, provides a more balanced reflection of both metrics. Therefore, we decide to employ the F-score as the parameter for setting classification thresholds.

From Table 5.12, the mean F-score is 0.677 and the median is 0.706. These values suggest a generally high level of annotation quality. A threshold of 0.5, being lower than both the mean and median, allows for the inclusion of a broader range of annotations, which can be particularly useful in scenarios where excluding too many data points (due to a high threshold) could be detrimental.

Moreover, in a negatively skewed distribution, the mean is less than the median. It indicates that the mean F-score is pulled down by a small number of lower scores. In other words, setting a threshold too close to the mean or median might exclude a significant number of annotations that are slightly below average but not necessarily of poor quality. A threshold of 0.5, being lower than both the mean (0.677) and median (0.706), allows for the inclusion of these data points.

Additionally, in the application of machine learning, it's essential to include a wide variety of data points to ensure that the models trained are robust and not overly tuned to only high-quality data. Setting the threshold at 0.5 serves this purpose well by ensuring that enough data points are included for a more generalised learning process.

As a result, to ensure inclusivity of data points and to mitigate the impact of outliers, we assigned the F-score of each HIT to the pass-fail category by setting a threshold of 0.5. That is, the annotations with F-score greater than 0.5 passed our quality standard. Figure 5.42 shows the distribution of the quality of all tasks, the percentage of F-score greater than 0.5 is: 75.98%.

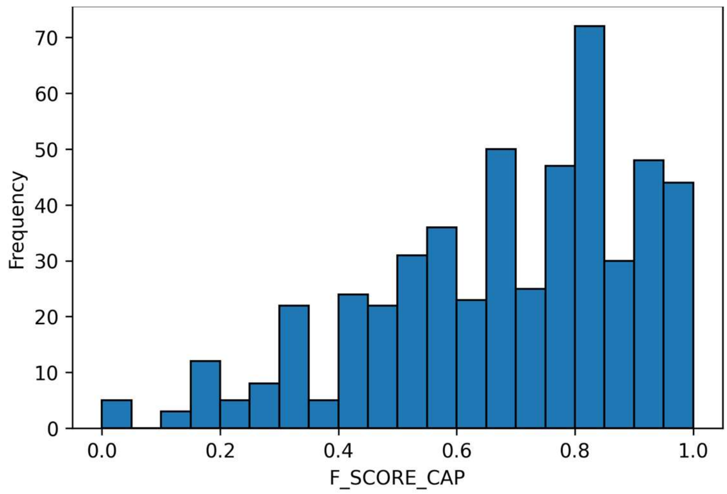

Figure 5.42 Distribution of F-score of 518 annotation tasks.

The training process of the SVC model is shown in Figure 5.43. Specifically, the features were selected and pre-processed in the beginning. The preprocessing included outlier removal, binary transformation of specific features and normalisation of the samples after splitting training and testing data. After preparing the training data, the optimal hyperparameters of the linear kernel were found through Bayesian optimisation. Finally, the predictive ability of the optimal model was evaluated.

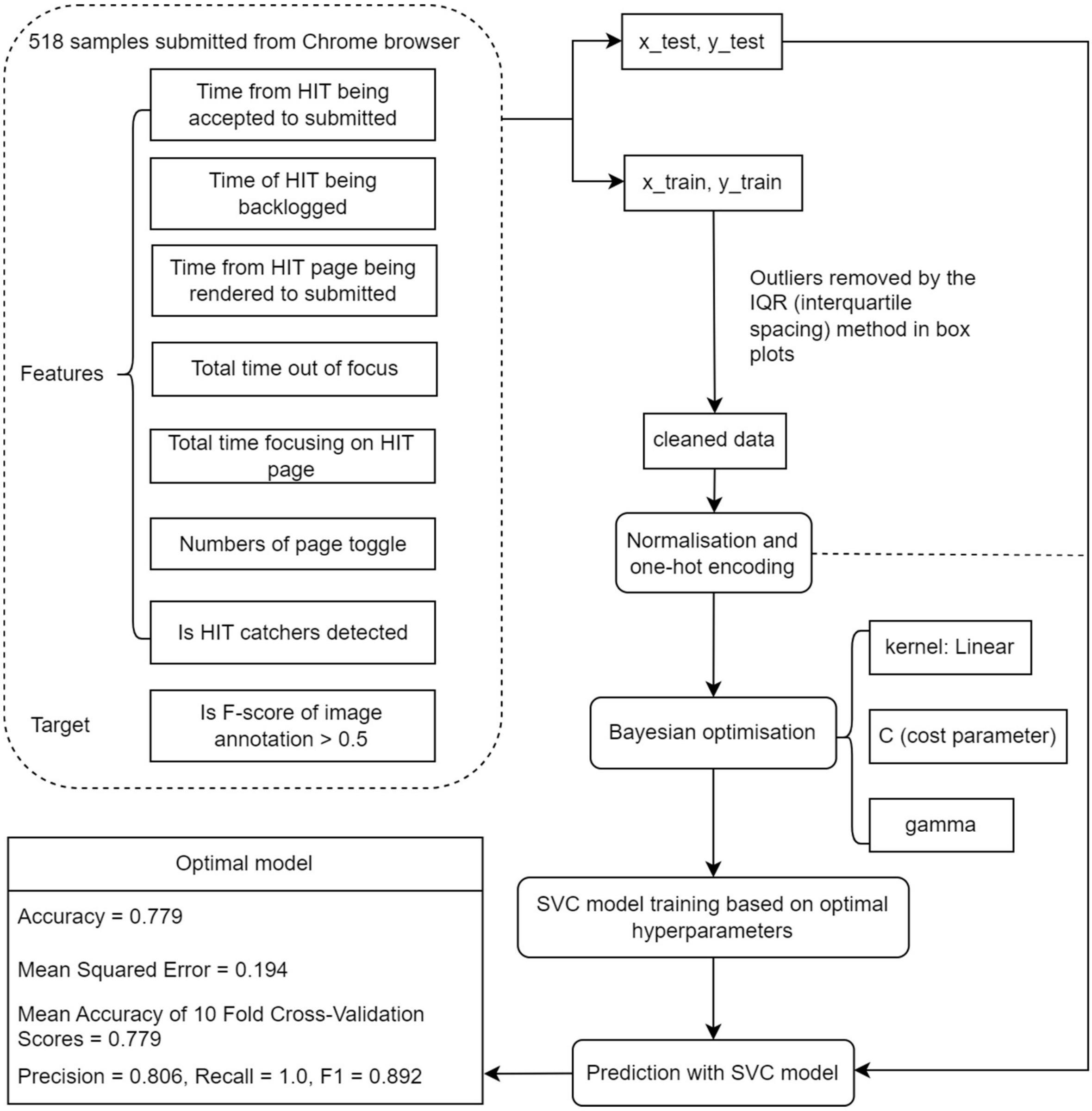

Figure 5.43 Overview of SVC model training procedure.

5.3.7.2 Data Pre-processing

This section illustrates the data preprocessing steps, including outlier removal, binary transformation of raw data, and normalisation.

Features for prediction were first extracted from the raw data associated with the 518 HITs completed via Chrome browser. Specifically, the features were included as shown in Figure 5.43. Subsequently, the HIT catcher installation status was transformed with one-hot encoding (0 and 1). In addition, the results were labelled as "not excellent" and "excellent" by 0 and 1 based on the previously set threshold. The dataset was then split based on features and target, with 80% used for training model and 20% for testing the predictions of the final model.

Outliers may negatively affect the decision boundaries of SVC, leading to a decrease in the performance of the model. To determine the range of outliers, the interquartile spacing range (IQR) method from the box plot was used as revealed in Figure 5.44 (Halder, 2019). The IQR is the difference between the 75th percentile (third quartile) and the 25th percentile (first quartile) of the data. The range of outliers was set to IQR multiplied by 1.5. The final sample size used for model training was 409.

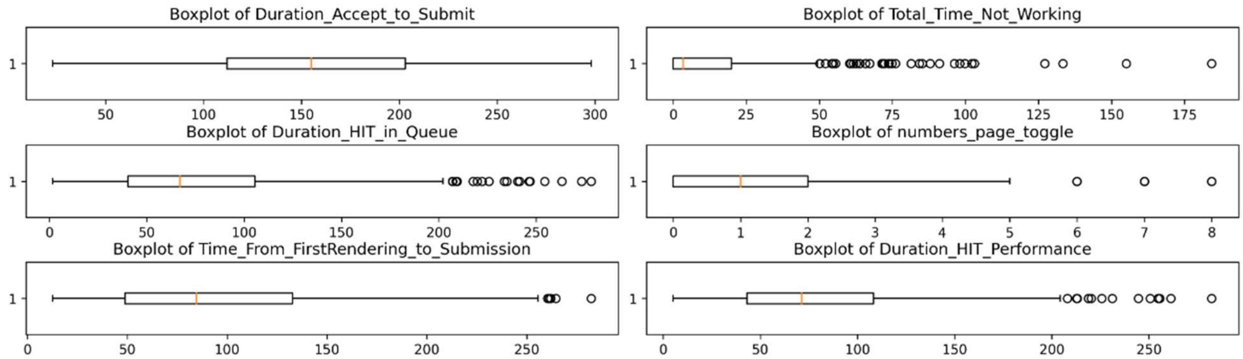

Figure 5.44 Outlier detection using IQR for features not binary converted.

Normalisation is done by linearly transforming the original data so that the converted values are mapped between [0,1] or [-1,1] without affecting the distribution or the relationship between the data (Zhang et al., 2022). First, SVC is very sensitive to the scale of different features. If one feature in the dataset has a much larger range than the others, then that feature may overly influence the model because it will be given more weight in the calculation of intervals and losses. In addition, normalising the data accelerates the training process of the model.

Next, the cleaned training samples were normalised with *MinMaxScaler()*[43]. Meanwhile, the same normalisation parameters were applied on the test set, thus maintaining data consistency between the training and test sets. It is worth noting that the test set should not be included when normalising the training set. Because the model's performance on the test set should reflect its ability to generalise on unseen data, introducing the test set for normalisation can lead to over-adaptation of the model to the test set, creating a false performance evaluation.

[43] Scikit-learn Documentation: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html

5.3.7.3 Model Selection and Bayesian Optimisation

First, for better model interpretability and thus understanding the effect of different features on the prediction results, linear kernel was chosen for SVC model training. Bayesian optimisation was then applied to find the optimal hyperparameters. This is a method used to find the best combination of hyperparameters by constantly sampling the parameter space and evaluating the performance of the model (Alibrahim & Ludwig, 2021; Treviso et al., 2021). The optimal hyperparameters found through Bayesian optimisation are C = 38.276487748089984, gamma = 0.0018061204818960854 for the linear kernel model.

5.3.7.4 Model Evaluation

The accuracy of the model trained by the optimal hyperparameters is about 0.779, which means that the model correctly classifies about 80.2% on the test data. The mean square error (MSE) obtained in the optimal model is 0.194. indicating that the model's predicted values have a smaller mean deviation from the actual values. A smaller MSE usually indicates a better model fit.

Cross-validation was also used to assess the predictive power of the optimal model. Its main purpose is to improve the generalisation ability of the model by dividing the training data into multiple subsets and then training/testing on each subset. Therefore, the model's performance on unseen data could be more accurately evaluated. Mean Accuracy of Cross-Validation Scores obtained after 10-fold cross validation is 0.779. This value indicates that the model correctly classifies 77.9% on average when making predictions on different subsets of data.

For a more comprehensive evaluation of the model, a confusion matrix was constructed based on the model predictions to assess predictive accuracy, sensitivity and precision (Düntsch & Gediga, 2019; Shen et al., 2020). The results are Precision = 0.806, Recall = 1.0, and F1 = 0.892. It reveals that the model has an 80.6% probability of being correct in predicting a passed annotation quality. In addition, Recall = 1.0 indicates that the model successfully captured all the samples with passed quality and did not miss any of them. F1 = 0.892 indicates that the model achieved a good performance in balancing the accuracy of prediction with the ability to identify samples with passed quality.

5.3.7.5 Feature Weight Analysis

The weights of different features on the prediction results in the generated linear SVC model were discussed. Figure 5.45 shows the weight of each feature in the model.

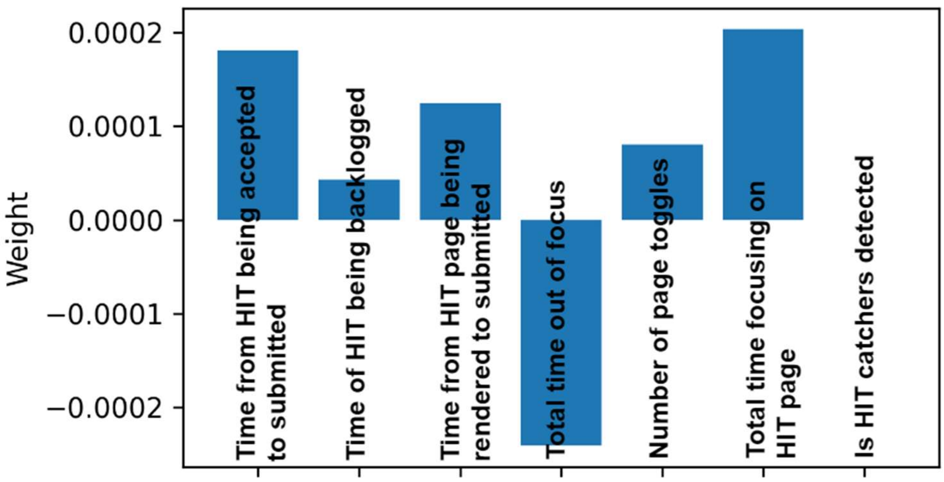

Figure 5.45 Feature weights of linear SVC model**.**

It is revealed that total time focusing on HIT page has the most positive impact on the prediction results, followed by time from HIT being accepted to submitted. in other words, the more time a worker spends on completing an image annotation, the more guaranteed the quality of annotation. Time of HIT being backlogged has the least impact on the predicted results due to the low weight score. It reveals a low correlation between task backlogging and the pass/fail of the annotation quality. Whereas time out of focus has a significant negative impact on the predicted results. This implies that lack of focus may lead to failed annotation quality. Interestingly, whether HIT catchers were detected showed minimal impact on the predicted results and is therefore its weight score is not presented in the figure. This means that by detecting whether HIT catchers are used or not does not seem to be a valid indicator for predicting the pass/fail of annotation quality. Finally, number of page toggles had a very low positive impact on the predicted results. This could be an implication that what causes participants to switch pages was not just because they did not focus on the task, but also because they were seeking guidance through other pages.

5.3.7.6 Conclusion

This subsection successfully predicted the pass/fail of the image annotation quality using F-score based on the participants' behavioural features through SVC. In addition to the significant positive effect of task completion time on quality prediction similar to that found in previous research (Hirth et al., 2014), a negative correlation caused by inattention on the quality of the results was also found.

It was also found that whether HIT catchers were detected or not as well as the HIT backlog time had extremely low weight scores in the prediction model. Combined with the comparison of annotation quality in Section 5.3.5, it can be observed that the cause of the lower annotation quality of workers who were detected using HIT catchers was not directly due to the use of the tool, but more likely due to other features such as time focusing on HITs.

5.4 Findings

Here is a list of findings summarised from the previous analysis section:

HIT Access Dynamics: During the experiment, most HITs experience multiple unsubmitted reservations before final submission, and all of these HIT backlogs slowed down the final completion of the entire HIT group.

Job Opportunities: Workers using HIT catchers, on average, completed more tasks than those not using them. This indicates that HIT catchers provide a competitive advantage in accessing and completing tasks.

HIT Backlog Time: Moreover, HIT catcher workers left HITs idle in the HIT queue for longer durations. This suggests that while they might be quick in reserving tasks, they don't start working on them immediately.

HIT Focus Time: Workers using the HIT catcher focused on the task page for a shorter average time and spent more time unfocused. This indicates that they were more easily distracted while completing tasks.

Multi-HITers Proportion: Among HIT catcher workers, there was a higher proportion of multi-HITers. This indicates that these workers are more likely to handle multiple tasks at once, which might be facilitated by the use of HIT catchers.

Multiple Devices and Browsers: There were instances of individual workers using multiple devices and browsers. This behaviour might be a strategy to maximise task access and completion.

Multiple Accounts on a Single Device: There were instances of multiple worker accounts logging into one device. This potential violation could be a strategy used by workers to further increase their task access.

HIT Result Analysis: Non-HIT catcher workers showed higher annotation quality and completeness than HIT catcher workers. Moreover, non-HIT catcher workers spent more effort on text editing overall, and with more text diversity.

Behaviour-based Quality Prediction: With SVC, behavioural factors could be successfully used to predict whether the F-score of the image annotation result is qualified or not. Effective behavioural factors included HIT completion time, focused and unfocused time, and number of focus switching. It was also found that the correlation between the use of HIT catchers and HIT backlog time are not valid quality predictors.

In summary, this study reveals differences in the behaviour of workers using and not using HIT catchers when it comes to accepting and completing HITs, as well as differences in the quality of results. These findings provide crowdsourcing platforms and job requesters with insights on how to better manage and design HITs to ensure task quality and fair treatment of workers. In addition, our understanding of behaviour-based quality assessment is expanded.

5.5 Discussion

5.5.1 Real-life Evidence of Tragedy of the Commons

This study provides real-life empirical evidence of the phenomenon of the tragedy of the commons found in the simulation study in the previous chapter. Specifically, the original intention of HIT catchers was to improve the crowdworkers' work performance. The aim was to reduce focus transitions caused by workers constantly searching for suitable and available HITs, and to increase the chances of accepting quality HITs. However, as HIT catchers became more popular, the phenomenon of HIT backlogging and abandonment was intensified by the popularity of HIT catchers, leading to the tragedy of the commons effect (Greco & Floridi, 2004). In other words, each crowdworker acts independently according to their own interests and uses automated tools to over-reserve HITs, triggering effects that are detrimental to the common good of all workers. Ultimately, their upfront over-reserving behaviour leads to leaving HITs unopened in queues for longer, and therefore having to continually return them before they are about to expire, or leave them to expire (Section 5.3.2). The abandonment and return of these HITs also deprive the workers themselves of the opportunities to complete. Meanwhile, other workers' job opportunities are compromised. This series of actions leads to a reduction in the overall speed of the completion and quality of results for the HIT group.

5.5.2 Detection and Description of Work Behaviour

In addition, this study designed and implemented strategies to assess crowdworkers' work behaviour more accurately, including real work time while performing specific HITs, attention span, multi-HITs, and malicious account sharing behaviours, based on data provided by scripts from client side and Amazon SQS from server side. By comparing the differences between the two types of workers on the above behaviour indicators, we can have a more comprehensive understanding of their work strategies. This study contributes to the understanding of the reasons for the low quality of microtask outcomes, including providing new explanations for the high number of fraudulent responses (Kennedy et al., 2020). Concerning the behavioural patterns of workers that use these tools, it can be observed that these workers left HITs idle in queues for longer periods of time. spent less time actively working on tasks, spent a greater proportion of attention outside the HIT page, and did more multi-HITing. In some cases, these workers were working in parallel on up to three HITs at the same time, indicating the potential presence of multiple people with the same account. The requesters always require a worker account to be associated with a unique individual, thus reducing data bias due to answer duplications.

Previous studies have attempted to classify the quality of the data provided by crowdworkers by analysing their interaction behaviour with the task interface during the experiment. These include analysis of mouse cursor trajectory, click behaviour, interaction with interface elements and text input (Hirth et al., 2014; Mok et al., 2016). The contribution of this study is not only the detection of multiple behaviours of crowdworkers while performing image annotation tasks (including text editing, focusing, and leaving the task page), but also the extension of the detection of behaviours to the reservation, backlogging and return of HITs. More importantly, this study uncovered specific differences in these detected behaviours between workers using and not using HIT catchers, and the resulting impact on data quality and the overall HIT group completion process. In addition, for sought-after HITs, workers generally face intense competition when trying to accept them. It is far more difficult for them to accept a HIT after previewing it than to accept it directly by skipping the preview step.

Previous studies have shown that the time spent to complete a HIT can be used as a measure of data quality. In turn, under-performing crowdworkers can be identified by setting minimum completion time limits (Difallah et al., 2012; Mason & Suri, 2012; Ribeiro et al., 2011). This study provides a more accurate measure of the real time workers spend completing HITs with the help of HIT page scripts and finds significant differences in the real working time between those with and without detected HIT catchers installed.

5.5.3 The Impact of Work Behaviour on Data Quality

Regarding the quality of the results of the image annotation HITs, non-HIT catcher workers generally have higher Recall and F-score than the HIT catcher workers, indicating higher annotation quality and completeness. Regarding the text responses, non-HIT catcher workers spent more effort on text editing, and had higher text diversity than HIT catcher workers overall. These findings are consistent with the findings in the work behaviour section that HIT catcher workers' time spent on doing HITs was less and they were also less focused when doing HITs.

The lack of quality assurance of data collected from MTurk has always been a problem for requesters (Ahler et al., 2021; Chmielewski & Kucker, 2020; Matherly, 2019). As mentioned above, this study reveals that workers who were detected using HIT catchers had lower levels of work engagement and concentration (Section 5.3.4.4 and 5.3.4.5), as well as less diversity and complexity of textual responses (Section 5.3.6). However, there was no significant difference in image annotation quality between workers using and not using HIT catchers.

In Section 5.3.7, a study on the prediction of pass or fail of annotation quality by linear SVC based on behavioural features revealed that the HIT backlog time and whether the use of HIT catchers was detected were not suitable as metrics for assessing annotation quality. Furthermore, it was also found through Section 5.3.4.1 that a small percentage of participants who used HIT catchers completed many annotation tasks with high quality. In other words, it is more important to assess the quality of the results in combination with metrics such as task focus time and out of focus time, which are directly related to the annotation process. The prediction of data quality based on behavioural features is particularly important for the raw data that are difficult to directly assess quality (Arndt et al., 2022; Bauer et al., 2020). This finding supports the significance of examining participants' level of concentration as highlighted in the study by Aruguete et al. (2019).

Lastly, during the test before publishing HITs, a previously unobserved platform functionality was recognised: when one HIT page is previewed, this HIT would become temporarily unavailable to others, and therefore allow the viewer to stay in the page and accept the HIT without being interrupted. However, this reveals a vulnerability in the platform: Since there is no limit on the number of previews that a worker can make (different from the limitation on the number of HIT reservations), an adversarial attack similar to a digital strike (Checco, Bates, and Demartini 2020) could be devised with minimal coordination between workers, to make a HIT batch indefinitely unavailable by continuously previewing the HITs within a target HIT group.

5.5.4 Complement to Current Reputation System

Current platforms identify quality people when publishing HITs through a reputation system based on HIT approval rate and number of HIT approval, but this is unfair to new workers without sufficient work experience but are dedicated to their work. Moreover, algorithmic control in terms of the number of HIT approval is also negatively affected by this metric. Specifically, extensive task experience would make crowdworkers "professional participants" and thus very familiar with the social research process, including the researcher's techniques of raising questions (Hauser et al., 2018). This familiarity with the research paradigm will in turn bring bias to their responses (Conte et al., 2019). In other words, the experience of answering a large number of questions of a particular type of experiment results in strong predetermined opinions that provide disturbed results. By incorporating the detection of attention switching, multi-HITing, excessive HIT backlogging, etc., it could potentially complement the limitations of the current reputation system to produce a more comprehensive assessment of the result quality.

5.5.5 Countermeasures

Queue size, HIT expiration time, and HIT catching frequency are the main parameters that affect the access and reservation dynamics. By refining the simulator designed in the previous chapter, requesters, platforms and HIT catching script designers could explore the interactions between the parameters mentioned and change them to mitigate their potential negative effects. To summarise, these are some potential generic solutions:

Workers/Script Designers: Reducing the HIT catching frequency can mitigate the tragedy of the commons effect on HIT availability. However, this solution is hard to employ from the workers side because it requires a synchronised and collective effort from multiple actors competing for a resource. Despite that, this is not unimaginable, as workers have often shown willingness to respect self-imposed policies (Bates et al., 2023).

Requesters: The first choice that can mitigate the negative effects described in this work is the manipulation of the HIT expiration time. However, this parameter alone cannot solve these issues, as a too short one would cause an excessive number of expirations, that can reduce the total number of available HITs that are effectively available for workers (since expired HITs cannot be reserved again Other interesting approaches to tackle these issues consist of setting dynamically HIT expiration times and other custom interventions based on worker behaviour detection.

Platform: Crowdsourcing platforms should address the soft-reservation preview vulnerability, review the maximum queue length used, and improve their techniques to detect the multiple accounts and concurrent work.

5.6 From HIT Catchers to Knowledge Sharing

Firstly, similar to the use of HIT catchers, crowd knowledge sharing is also a common collective behaviour in crowdsourcing communities (Gray et al., 2016; LaPlante & Silberman, 2016). Crowdworkers share their experiences on communication tools including forums, thus helping novices to quickly familiarise themselves with crowdsourcing workflows and techniques. When they encounter difficulties, they also seek help in the community and benefit from the crowd collective wisdom.

Second, through the study of El Maarry et al. (2018), it is revealed that the sharing of skill-based knowledge is an important factor driving the popularity of scripting tools among the crowd community. This sharing includes not only work strategies, but also how to effectively utilize tools such as HIT catcher to improve the work efficiency and quality. With support from the community, workers could better master and apply these tools to improve their performance.

However, as Savage et al. (2020) point out, although the sharing of skill-based knowledge facilitates the use of scripting tools, there are still some workers who may not have access to or mastery of these tools due to knowledge gaps, and not everyone benefits equally. This phenomenon of uneven tool use is also revealed by this study. To understand the cause of uneven use of scripting tools, the crowd knowledge sharing behaviour needs to be studied.

In addition, knowledge sharing is recognized as an important factor in driving innovation within an organisation (Nurhidayati & Zaenuri, 2023; Wibowo et al., 2021). Similarly, the generation and improvement of scripting tools including the HIT catchers also benefit from knowledge sharing in the crowd community. Scripting tools such as MTurk Suite, Panda Crazy, while mostly created by individuals or technical teams, rely on the collective efforts of workers to improve the tools through sharing identified issues and suggestions via GitHub[44] and forums[45].

5.7 Chapter Summary

This study quantifies the impact of workers' use of HIT catchers from multiple perspectives using emerging monitoring technology to correlate data from MTurk server and clients. Specifically, by posting image annotation HITs on MTurk, worker behaviour, HIT response, and HIT states related data were collected via client-side web scripts and server-side HIT states notifications. Descriptive statistical analysis was used on worker categorisation, dynamics of HIT state changes, HIT opportunities, HIT results and worker behaviours. The differences between two types of workers regarding their behaviours, opportunities and result qualities were compared within the descriptive analysis. The findings are listed in Section 5.4.

This study provides empirical evidence of the phenomenon of the tragedy of the commons found in the simulation study in the previous chapter. In detail, by monitoring the status of the HITs during the experiment, it was observed that, the sought-after HIT group published for this experiment, was significantly affected by the wide use of the HIT catchers. Many HIT expirations and returns were experienced at the beginning of the experiment, with low HIT availability and completion rates during this phase. In addition, most of the HITs experienced several unsubmitted reservations, which not only slowed down the completion speed of the entire HIT group, but also deprived other workers of work opportunities during the same period.

In addition, this study designed and implemented strategies to assess crowdworkers' work behaviour, including time spent on doing HITs, attention span, multi-HITing, and malicious account sharing behaviours more accurately. It was revealed from this study that HIT catcher workers left HITs idle in queues for longer periods of time, spent less time actively working on HITs, spent a greater proportion of attention outside the HIT page, and did more multi-HITing. In some cases, workers were working in parallel on up to three HITs at the same time, indicating the potential presence of multiple people with the same account.

[44] MTurk Suite related issues raised by crowdworkers: https://github.com/kadauchi/mturk-suite/issues
Panda Crazy related issues raised by crowdworkers: https://github.com/JohnnyRS/PandaCrazy-Max/issues
[45] MTurk Scripts & Resources: https://forum.turkerview.com/forums/mturk-scripts/

Furthermore, the impact of work behaviours between two types of workers on data quality has been studied. Compared to the HIT catcher workers, those not using HIT catchers generally have higher annotation quality and completeness. Regarding the text responses, non-HIT catcher workers spent more effort on text editing and had higher text diversity than HIT catcher workers overall.

Finally, the SVC model trained by machine learning successfully predicts whether the quality of image annotation passes/fails based on behavioural features. Among them, time focusing on HIT and time out-of-focus had the most significant effect on the predicted results.

The next study aims to delve deeper into skill-based knowledge-sharing behaviours in the worker population, particularly from the perspective of the use of communication technologies and social exchanges, and how and to what extent these factors influence the willingness to share and behaviours, and thus the popularity of assistive tools such as the HIT catcher, in the worker population. Through this approach, we can not only understand what factors facilitate skill knowledge sharing, but also explore how the diffusion of scripting tools can be facilitated by optimizing knowledge-sharing mechanisms to promote more equitable use of the tools among the worker population.

Chapter 6 Factors Influencing Knowledge Sharing Behaviour within Crowdworkers

6.1 Introduction

The first two studies in this thesis focus on the phenomenon and impact of using HIT catchers as a crowd collective behaviour. This chapter focuses on the skill-based knowledge sharing that has contributed to the popularity of HIT catchers (Di Gangi et al., 2022; El Maarry et al., 2018; Williams et al., 2019). Specifically, what are the factors that facilitate the crowd collective behaviour of skill-based knowledge sharing, which in turn contributes to a thriving ecosystem of scripting tools, including HIT catchers.

First, based on the theoretical models reviewed in Section 2.6.2 that have been widely used in the study of knowledge sharing behaviours, a factor analysis theoretical framework constructed by UTAUT and SET was constructed from the perspective of the use of communication technologies. This framework examines workers' knowledge sharing behaviour from the standpoint of individual experiences with sharing tools and the exchange of social benefits.

After collecting subjective evaluations of influencing factors from worker groups via a questionnaire, descriptive statistical analysis was implemented to comprehend participants' socio demographic backgrounds, preferences, and knowledge sharing behaviour frequencies. Following this, the impact of each factor on knowledge sharing intention and behaviour was assessed using structural equation modelling.

The results of the study highlight the critical role of the experience of using communication tools in the process of knowledge sharing by contributors and complement the UTAUT model by considering elements of social exchange. It was observed that performance expectancy, effort expectancy, and rewards all significantly influence knowledge sharing behaviour indirectly through intention. Moreover, effort expectancy produces a more pronounced direct than indirect effect on KSB. In addition, this study explores non-technical reasons that may prevent crowdworkers from sharing knowledge. These include personal fears of losing their technical advantages, distrust of unfamiliar members, and doubts about platform policies. Finally, we find that their knowledge sharing contributed to the popularity of scripting tools.

6.2 Theoretical Framework

Crowdworkers, as a typical group of internet users, exchange tips on working on tasks, tips on using tools such as plugins, comments on tasks or requesters, and daily life content through online forums (Turkopticon, Turkerhub, Turkerview), social applications (Facebook, Slack, Discord), plugins (TurkerViewJS), etc. Online knowledge sharing as an internet user behaviour has been studied by researchers using a variety of models: Assegaff et al. explored the perceived benefit of knowledge sharing by extending the Technology Acceptance Model (TAM) to include the perspective of knowledge contributors (Assegaff et al., 2011). The Theory of Planned Behaviour (TPB) has also been widely used in empirical research on knowledge sharing behaviour (Nguyen et al., 2019). TPB focuses on the relationship between behavioural attitudes and intentions. In addition, perceived behavioural control and subjective norms also have an impact on behavioural intentions. through the theory of reasoned action (TRA), Waqar et al. examined the effects of expected extrinsic rewards (AER), sense of self-worth (SSW), organisational-based self-esteem (OBSE) and expected reciprocal relationships (ARR) on the variables including knowledge sharing intention (IKS) and attitudes towards knowledge sharing (ATKS) (Waqar et al., 2018). Hsieh (2021), on the other hand, integrates SCT and IDT as important determinants of willingness to share medical knowledge from a socio-technical perspective and examines the impact of social and technical factors on willingness to share in Shared Decision-Making Platforms.

Based on the theory explained in Section 2.6.2 and their usage scenarios above, this section provides a theoretical basis for the generation of the research framework that follows by understanding the models that have been widely used in the study of knowledge sharing behaviour.

6.2.1 Why Choosing UTAUT and SET

The reason for using UTAUT in this study is that crowdworkers rely on communication technologies to share knowledge. UTAUT is a model widely used to study people's adoption of new technologies. By applying the UTAUT model in this study, factors such as the performance and effort expectations of crowdsourcing workers in adopting communication technologies could be investigated, and thus understand how these factors influence the intention and behaviour. Figure 6.1 reflects the derivation of UTAUT from traditional frameworks such as TAM, TPB, TRA and IDT.

SET is a social psychological theory that emphasizes that people exchange based on costs and benefits in social interactions. In studying the knowledge-sharing behaviour of crowdsourcing workers, SET can complement the UTAUT framework from the perspectives of reciprocal relationships, trust and expectations, and the influence of the social environment on behaviour, so as to explore the behavioural motivations of crowdworkers and the social exchange process behind their decisions. Therefore, UTAUT and SET were applied to understand how the crowdworkers use the communication technology to share knowledge.

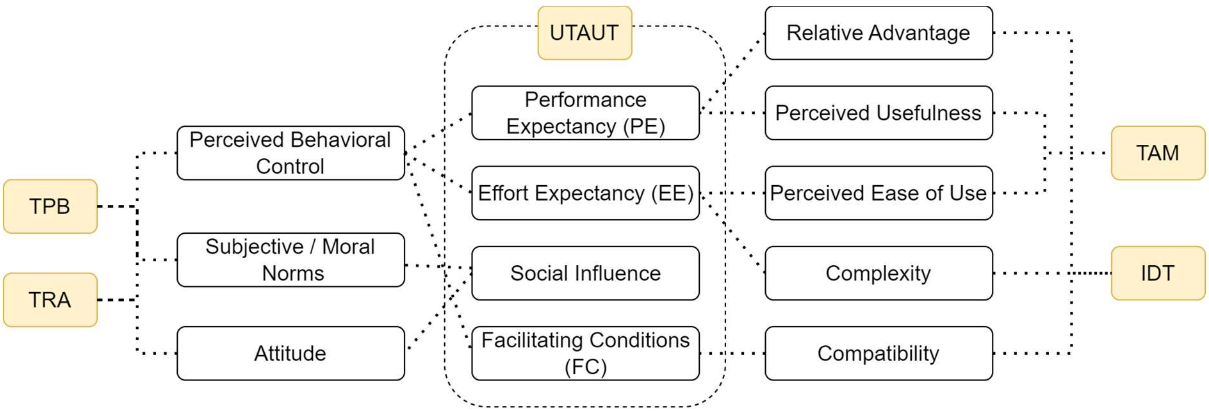

Figure 6.1 An illustration of how each factor within UTAUT is derived from previous theories including TPB, TRA, TAM and IDT.

6.3 Research Model and Hypotheses

In this study, The UTAUT model was used as the theoretical basis and incorporated Social Exchange Theory (SET) to investigate the online knowledge sharing behaviour for crowdworkers. UTAUT is commonly used for examining users' adoption of technology (Marikyan & Papagiannidis, 2021b).

The main changes to the UTAUT model in this study were the addition of intention influencing factors under SET theory and the addition of Trust as an independent influencing factor for behaviour intention. As presented in Section 2.6.1.1, trust is often included as an important factor in studies of knowledge sharing in virtual communities (Chang et al., 2015; Gang & Ravichandran, 2015; Hung et al., 2015).

Finally, as shown in Figure 6.2, the final theoretical model contains four factors in SET, four factors in UTAUT, and the Trust factor, for a total of nine exogenous factors. Then knowledge sharing intention (KSI) and behaviour (KSB) are two endogenous factors. Based on research exploring knowledge sharing within virtual communities, this study developed hypotheses about the factors influencing knowledge sharing intention and behaviour among crowdworkers. Next, the generation of each hypothesis is explained.

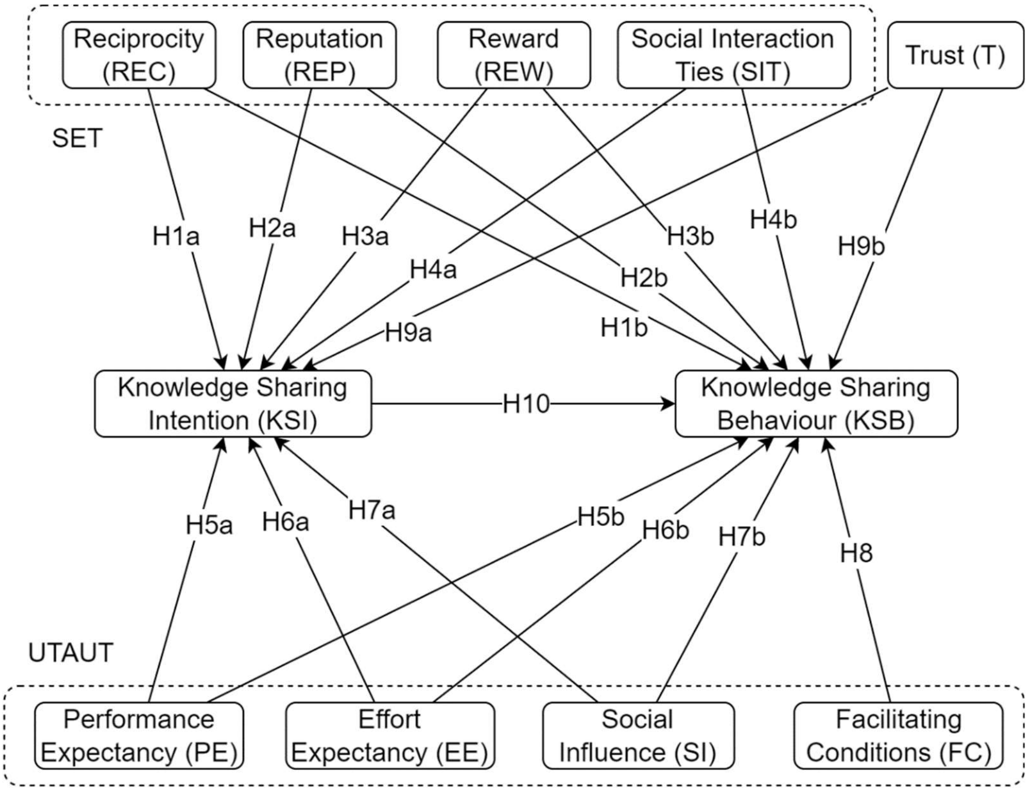

Figure 6.2 Theoretical model for Study 3.

6.3.1 Hypotheses Based on SET

6.3.1.1 Reciprocity

Reciprocity refers to the expectation of individuals to receive rewards for their actions (Nguyen, 2021). Employees who share knowledge online expect that their efforts and valuable content will be rewarded from others. One of the motivations for reciprocity is obligatory: because individuals have collected valuable knowledge from knowledge donors, they are obliged to share their knowledge in return (Feng & Ye, 2016). This means that knowledge donors want the value of their knowledge to be reflected through the mutual giving and acquisition of knowledge, and encourage more members to participate in knowledge sharing (Adamseged & Hong, 2018). Reciprocal knowledge exchange relationships encourage knowledge sharing behaviours, and therefore individuals may be more willing to share their valuable knowledge (Nguyen et al., 2019).

Reciprocity, on the other hand, needs to be achieved with trust, including the belief that other members are knowledgeable and willing to share their knowledge in return (Alwahdani, 2019). Moreover, it is based on the sense of commitment that members develop towards each other when they join a virtual community (González-Anta et al., 2021). However, when the expected feedback based on reciprocity does not occur, individuals' confidence in others to share their knowledge diminishes or disappears and knowledge-sharing actions tend to stop (Jennex, 2019). As one of the most fundamental social norms, reciprocity is characterised by a more equitable exchange of benefits in the expected social interaction (Mustapha & Shamsudin, 2020). Therefore, this study makes the following hypotheses regarding reciprocity:

H1a: Reciprocity has a positive effect on the crowdworkers' intention to share knowledge.

H1b: Reciprocity has a positive effect on the crowdworkers' knowledge sharing behaviour.

In this study, reciprocity was assessed via asking about participants' beliefs on others' KS behaviours and attitudes of sharing.

6.3.1.2 Reputation

Previous research has found that people will share knowledge within a group if they believe that sharing knowledge will enhance their professional reputation (Chang & Chuang, 2011). Specifically, individuals will share knowledge in order to gain the respect of their peers and be treated as an expert in the organisation (Gang & Ravichandran, 2015). When community members perceive that their reputation can continue to be enhanced by sharing their knowledge, they are likely to continue (Jiarui et al., 2022). Therefore, in the context of crowdsourcing, this study makes the following hypotheses with respect to reputation:

H2a: Reputation has a positive effect on the crowdworkers' intention to share knowledge.

H2b: Reputation has a positive effect on the crowdworkers' knowledge sharing behaviour.

In this study, reputation was assessed by measuring the extent to which participants believe that the KS behaviour enhances their community image, recognition, and respect.

6.3.1.3 Reward

Implementing reward systems within virtual communities can be effective in encouraging online users to share knowledge for extrinsic benefits (Wei et al., 2015). Rewards include tangible rewards such as money or voucher redemptions, as well as virtual rewards such as badges, rankings, and special avatars (Anderson et al., 2013; Borst, 2010; Grant & Betts, 2013). Mturk Forum[46], as an example, has tried to select a Turker of the Month based on the results of the members' vote and the number of contributions made during the month. Then the forum moderator would give the Turker of the Month monetary reward via PayPal to motivate the forum members to post helpful content continuously.

Previous studies have also referred to intangible rewards including satisfaction, the pleasure of helping others (Fang & Zhang, 2019; Hung et al., 2015). Perceived self-enjoyment is an intrinsic motivation that makes the individual's perception of sharing knowledge to help others more favourable and leads to sharing behaviour (Cahyaningrum, 2023).

In the context of this study, two intrinsic motivations, satisfaction and enjoyment, were included in the reward factor (Osterbrink & Alpar, 2021). Within the framework of social exchange theory, satisfaction and enjoyment can be considered as rewards because they are both positive outcomes that individuals can experience as a result of their behaviours (Abdou et al., 2022). In addition, knowledge gained from others in the process of knowledge exchange is also included as a type of reward (Ahuja, 2020). Finally, this study makes the following hypotheses based on the group of crowdsourced workers:

H3a: Rewards have a positive effect on the crowdworkers' intention to share knowledge.

H3b: Rewards have a positive effect on the crowdworkers' knowledge sharing behaviour.

In the survey, the reward factor was evaluated by asking the extent to which participants believe that the KS behaviour could benefit them, bring them satisfaction and enjoyment. It should also be noted that reputation in not included in reward in this study. Compared with the reward factor, reputation is generated based on how other crowdworkers evaluate their knowledge sharing behaviour (Cai & Shi, 2022). The worker group may respect an individual because the knowledge they shared is valuable. Although rewards and reputation are both extrinsic motivations, they have different processes of formation.

[46] A forum section showcasing Turker of the Month: http://mturkforum.com/index.php?forums/turker-of-the-month.47/

6.3.1.4 Social Interaction Ties

Previous research has found that if there are close social relationships between members within a virtual community, their online knowledge sharing behaviours will be significantly enhanced (Wang et al., 2022). One explanation is that if community members connect with more people, they can access more relational resources, which can help themselves to get help from others in the future (Nguyen, 2021). This study hypothesizes that in the crowdworkers group:

H4a: Social Interaction Ties (SIT) have a positive effect on the crowdworkers' intention to share knowledge.

H4b: Social Interaction Ties (SIT) have a positive effect on the crowdworkers' knowledge sharing behaviour.

Three indicators have been chosen in the assessment of SIT. In this study, social interaction ties represent the strength of the relationships, the amount of time spent, and communication frequency among members of virtual communities.

6.3.2 Hypotheses Based on UTAUT

6.3.2.1 Performance Expectancy

Performance Expectancy (PE) is the user's perception of the benefits and utility expected from the use of a particular technology (Hassaan et al., 2023). PE as a latent variable contains the observed items of usefulness, effectiveness, perceived speed, and relative advantage (Onaolapo & Oyewole, 2018). Perceived usefulness refers to whether the target technological means are helpful to individuals in their sharing behaviour when using technology for knowledge sharing (Nguyen, 2021). It is also worth noting that effectiveness refers to the expectation that using the technology will help the user realise the purpose effectively. Perceived speed refers to the user's subjective perception of how fast a technology performs a task. Relative advantage refers to the advantages that new technologies offer over alternatives. These advantages can be time efficiency or any other factor that improves task performance. Based on the four observed items about PE, the hypotheses are:

H5a: crowdworkers' Performance Expectation has a positive effect on knowledge sharing intention.

H5b: crowdworkers' Performance Expectation has a positive effect on knowledge sharing behaviour.

6.3.2.2 Effort Expectancy (EE)

EE as a latent variable contains the observed item perceived ease of use (Hung et al., 2019). This study chooses four observed items to measure EE: ease to use technology, ease to access technology, ease to learn technology and technical barriers (Onaolapo & Oyewole, 2018). As a motivation for knowledge sharing behaviour, perceived ease of use emphasises individuals' perceptions of the ease of using technology for knowledge sharing (Lee et al., 2021). It is also important to be clear: the technical barrier is primarily concerned with the technical problems and challenges of using new technologies, such as lack of access to tutorial. The hypotheses regarding EE are:

H6a: crowdworkers' Effort Expectancy has a positive effect on knowledge sharing intention.

H6b: crowdworkers' Effort Expectancy has a positive effect on knowledge sharing behaviour.

6.3.2.3 Social Influence (SI)

As one key construct of SI, subjective norms are external stimuli from the social group that influence individual behaviour (Stok et al., 2015). Social norms arise from the willingness of groups to conform to specific shared expectations (Tesar, 2020). Specifically, official attitudes and policies regarding knowledge sharing create social norms that encourage or discourage this behaviour, which in turn affects employees' motivation to share knowledge. Group behaviour further reinforces this social norm and allows individuals to perceive this social pressure through the workplace climate (Ajzen, 1991; Nguyen, 2021). Previous empirical studies have illustrated that subjective norms are important predictors of behavioural intentions in KS (Dong et al., 2022; Wu et al., 2023). In this study, social norms come from both platforms and the crowd. Therefore, four observed items were used to measure SI: expectations from the platforms, peers, and crowdworkers' attitudes to expectations from both. The hypotheses regarding SI are:

H7a: crowdworkers' Social Influence regarding knowledge sharing has a positive effect on knowledge sharing intention.

H7b: crowdworkers' Social Influence regarding knowledge sharing has a positive effect on knowledge sharing behaviour.

6.3.2.4 Facilitating Conditions (FC)

Facilitating conditions were found to positively impact the adoption of technologies such as digital banking (Nepal & Nepal, 2023). Perceived behaviour control, as one key construct of FC derived from Theory of Planned Behaviour, involves the subjective perceptions of constraints of target behaviour (Liu et al., 2023). In this study, these constraints of target behaviour have been applied and adapted based on the background of crowdwork. As the first observed item, technology integration attempts to evaluate how well the KS tools integrate with other technologies used in crowdwork (Ajzen, 2020). Compatibility, as another observed item of FC, means how the target system fits one's preferred work style based on their own experiences and needs. For the systems for KS, compatibility refers to the ways users communicate and interact via the system. One typical example is: one crowdworker may prefer to chat by voice, but the forum only allows members to type and use images. In summary, participants' scores about technology integration, community and technical support, compatibility and their personal perception regarding the KS tools were used to assess FC.

H8: Facilitating Conditions have a positive effect on the knowledge sharing behaviour of crowdworkers.

6.3.3 Trust

Trust is considered to be an effective factor that facilitates online knowledge sharing (Ismail et al., 2019). As trust is strengthened, individuals perceive less uncertainty and more security, and in turn individuals are more willing to share knowledge (Nguyen, 2021).

In virtual community environments, individuals often do not know enough about other members and therefore lack the most basic trust (Hsu et al., 2007; Mooradian et al., 2006; Wu et al., 2010). In the absence of trust, the initiator of knowledge sharing is forced to contribute without knowing how another actor will respond, which is extremely difficult (Li et al., 2023). Trust between individuals can compensate for the uncertainty caused by this unknown, making knowledge sharing behaviour more likely to occur and helping to build and maintain knowledge exchange relationships.

In other words, sufficient trust between the knowledge provider and the knowledge seeker is a prerequisite for crowdworkers to share knowledge through peer communication. Without peer trust, the knowledge-seeker may question what the other party offers, and the knowledge-provider may not be willing to provide valuable insights to strangers. Crowdworkers may not have a history of working together and may come from different cultural backgrounds and regions. This study therefore hypothesised that:

H9a: Trust has a positive effect on the crowdworkers' intention to share knowledge.

H9b: Trust has a positive effect on the crowdworkers' knowledge sharing behaviour.

The questionnaire for this study contained five questions on trust: the first three focused on participants' trust in each of the three mainstream knowledge sharing tools (forums, social apps, extensions). This was followed by questions on whether participants trusted others to value the knowledge they shared and whether the knowledge could be used unethically.

6.3.4 Knowledge Sharing Intention

Behaviour Intention (BI), according to UTAUT is a factor influencing actual behaviour and has been the focus of much research and predicted through the core structure of UTAUT (Chen et al., 2023; Khalid et al., 2023; Wang et al., 2023). Knowledge sharing intentions indicate the level of effort people are willing to try and how much effort they plan to put into performing the behaviour (Dey & Mukhopadhyay, 2018). This structure is similar to attitudes towards behaviour (TRA, TPB, DTPB), extrinsic and intrinsic motivation (MM) derived from previous models or theories (Lakhal et al., 2013).

In the context of this study, knowledge sharing intention (KSI) indicates the degree to which a crowdworker believes that they will share knowledge with peers. Hypotheses have been developed previously incorporating research on factors affecting KSI. Here we made a further hypothesis around the effect of KSI on behaviour:

H10: Knowledge sharing intentions of crowdworkers have a positive effect on their behaviour to share knowledge.

The latent variable KSI for this study is constructed from four observed variables (KSI1-KSI4). More specifically, KSI1 refers to the willingness to share knowledge in general; KSI2 refers to the degree of tendency of behaviour; KSI3 refers to the importance of behaviour; KSI4 refers to their intentions while answering the question (Bock et al., 2005; Yu et al., 2021).

In summary, based on UTAUT and SET, a theoretical model of the factors influencing the crowdworkers' knowledge sharing intention and behaviour has been proposed as illustrated in Figure 6.3.1.

6.4 Methods

In this chapter, the specific operational steps of the experiment are explained, including how the questionnaire was designed, based on hypotheses as an instrument for data collection, the calculation of sample size, and the subsequent ways ensure data quality was assessed. Finally, this section provides further detail on how SEM was performed as a part of data analysis.

6.4.1 Crowdwork Knowledge Types

This section describes what types of knowledge are included in the knowledge ecosystem formed around crowdwork, the role and significance of each, and most importantly, the type of knowledge to focus on in this research.

6.4.1.1 Skill-related Knowledge

As the focus of this study, this type of knowledge is mainly related to technical issues, working strategies, techniques for finding quality tasks, etc. This knowledge is usually generalised to most HITs. The following (paraphrased to avoid deanonymisation) posts highlight how crowdworkers share skill-related knowledge via the virtual communities:

> *"Perhaps someone can tell me which button on HIT Exporter is best for copying HIT information to the clipboard and pasting it into the forum? Why do we have so many choices?"*
>
> *"Choose XXX. The rest are for different platforms. If you come across one that supports SLK, take advantage of it."*

– Crowdworker shared knowledge on how to export HIT information properly
[Turker Nation: Dec 19, 2022]

“I'm new to this field and would like to know if anyone has made it their primary source of income. If so, could you tell me what kind of earnings I can expect if I work a full 8-10 hour workday five days a week? Your thoughts are greatly appreciated.”

“Some of them are up to you. If you are knowledgeable and resourceful. You need to download some scripts and learn how to do batch work (that's what earns the most). There is an adjustment period. This means that there are now far fewer jobs than there used to be. People usually earned 100 per day in the past. But for today, it is less common or it is just for people with closed qualifications. With a little effort you can probably make 20-30 a day. Read our mturk guide first.”

– Crowdworker shared knowledge on how to earn more as a novice [MTurk Forum: Aug 21, 2021]

What should be mentioned is that there are similarities between this type of knowledge and HIT tutorial knowledge. Specifically, the example above is not only about skills on how to open HIT links safely using a browser, but also a tutorial on how to deal with unexpected situations when doing HITs. However, the definition of skill-based knowledge in this study is more focused on techniques for using / debugging assistive tools and working strategies including selecting specific types of tasks rather than explaining the steps to perform a specific HIT.

6.4.1.2 Opportunity-related Knowledge

This category includes knowledge related to job opportunities, including suggestions of specific HITs, qualifications which are required for completing specific HITs, job opportunities from other platforms, etc. Specifically, regarding the sharing of URLs to HITs, workers could catch HITs automatically through URLs and HIT catchers, thus greatly increasing the efficiency of HIT acceptance. However, it is worth noting that the quality of this type of knowledge is not consistent and depends greatly on the subjective bias of the person posting the information.

“The requester (who posted this qualification) *does have rejections reported by TurkerView, but not for this task (which the qualification relates to), and according to MTurk, this task has a 99% approval rating. The actual HITs will be worth $2 each. I'm hoping for the qualification because I enjoy this type of HIT.”*

"It appears that the requester is rejecting tasks on somewhat subjective reasons and even rejected tasks used for assigning qualification."

"Proceed with caution, I'll notify you once I submit my task and learn whether it was accepted or not."

– Crowdworker shared a qualification opportunity with their reflections [Turker Nation: Nov 29, 2022]

While other job opportunities include those from the platforms similar to MTurk, or those that people can earn rewards without high barriers.

"On this website, https://app.qrowdsy.com/, I recently took part in a survey. In case you're interested, I thought I'd share it with you."

– Crowdworker shared a website with other job opportunities [Turker Nation: Dec 19, 2022]

6.4.1.3 HIT Tutorial Knowledge

Although this knowledge is usually provided by the requester in the HIT description or qualification test, workers also share insights and tips (Amazon Mechanical Turk, 2017). A typical example is workers sharing their methods for completing a particular type of HIT, including task processes, tips, mistakes to avoid, etc.

"I'm a new member of this forum, and I need your assistance to begin working on the request. Hits. I only have a few questions. If "Tesco eggs pack of 6 1 $2.00" were the case. So, should I enter 6 or 1 for the quantity? I would appreciate assistance on this from any of the mentors."

– Crowdworker looking for help on how to do a shopping list HIT properly [MTurk Crowd: Dec 23, 2022]

In the case of this question, when a worker provides an answer regarding the correct input, they are generating and sharing HIT tutorial knowledge in the forum.

6.4.1.4 Evaluation Knowledge

This type of knowledge includes reviews of specific requesters or HITs, and the reviews include ratings and text descriptions. A common form of sharing is for workers to post reviews in a forum with the target task as the subject. Later on, different scripting tools were designed to facilitate the sharing and delivery of this type of knowledge, such as TurkerViewJS[47] and Turkopticon[48].

This type of knowledge serves to help each other avoid the pitfalls of malicious requesters and to help filter out quality HITs. Here is an example of evaluation knowledge:

> *"(Followed by a HIT title) What a scam. I am completely confident in my audio abilities... I can't claim that I wasn't warned! But I figured that if I got rejected, everyone else would as well. Let's just hope they respond and address everyone's rejections."*
>
> *"It's pretty bad what they're doing. I've sent them questions and feedback before, but I haven't received any response. It's hard to think they are legitimate if they are so hard to reach. I suspect they're taking advantage of people by rejecting their work without paying. "*
>
> – Two crowdworker shared their complaints toward the same suspicious HIT [Turker Nation: Jan 31, 2023]

6.4.1.5 Non-job-related Knowledge

This category refers to knowledge that is not related to crowdwork, that comes from the lives of individuals and that has a social element. The aim for sharing this type of knowledge is to satisfy the social needs of individuals, to gain the psychological satisfaction of sharing or even for work-life balance (Shaharuddin et al., 2022).

6.4.2 Data Collection

6.4.2.1 Questionnaire Design

We generated the questionnaire for this study (Appendix B) based on the influencing factors included in the research model generated above and the questionnaire questions used in the related studies. In the subjective attitude assessment section of the questionnaire, participants' subjective attitudes towards skill-related knowledge were first collected based on statements generated from the factors in the UTAUT model. This was followed by a question based on

[47] TurkerViewJS homepage: https://turkerview.com/mturk-scripts/1-turkerviewjs
[48] Turkopticon homepage: https://turkopticon.net/

the statements generated by the SET. Finally, participants' attitudes towards sharing skill-related knowledge were investigated from the perspective of trust. In summary, participants were asked to rate their perceptions of using knowledge sharing tools to share skill-related knowledge based on a 5-point Likert scale ranging from 'strongly disagree' to 'strongly agree' in terms of Performance Expectancy, Effort Expectancy, Social Influences, Facilitating Conditions, etc. In addition, they were asked to explain their choices under different questions, forming complementary feedback.

All statements did not contain negative sentences and therefore do not limit the potential response bias caused by positive wording through negative wording. One reason is that, from a psychological point of view, understanding a negatively worded problem statement requires better linguistic skills and more cognitive load. Specifically, participants tend to develop a mental inertia based on the initial positively worded statement, and the sudden appearance of negative wording may force participants to break this inertia and apply different cognitive processes, thus causing potential comprehension bias (Suárez-Álvarez et al., 2018). Other studies have also found that when using a combination of positive and negative statements, the precision of test and discriminatory power of items could be reduced (Bourque & Shen, 2005; Chiavaroli, 2017; Józsa & Morgan, 2017).

To test the validity of the questionnaire, we first published questionnaire HITs on MTurk and surveyed 20 crowdworkers working on this platform about this study of knowledge sharing behaviour. As existing KS behaviour studies are not constructed based on technology acceptance theory, there are no existing questions related to knowledge sharing behaviour supported by communication technologies. Therefore, in the first phase of the study, the researcher constructed the factors included in the UTAUT based on the context of this study and selected a number of questionnaire questions based on the constructs included in each factor. These questions were then categorised by topics including Socio-demographic Background, Crowdwork Experience, HIT Preference, KS Behaviour, Preferred KS Types, etc. Crowdworkers who had received the questionnaire task were asked to evaluate them, including whether the question belonged to the current category and whether the question needed improvement. The purpose of this phase is to provide an initial check on the construct validity and reliability of the category and item measures. The survey has been piloted through six rounds from the participants of the target population to improve the readability of questions, improve the flow among all questions, and reduce misleading information. The final result was a questionnaire containing 15 sections and 54 questions, which took about 10-15 minutes to complete. Furthermore, this project has received ethical approval from the University of Sheffield on 08/11/2022 (Application Reference Number: 049528).

No personal data like the participants' account names, nor significant fragments from the submission, will be published. Descriptive statistics including counts and percentages will be produced. The analysis here focuses on multiple linear regression. The main independent variables are those related to communication design and knowledge sharing motivation. The data collected will not contain any confidential information and will be impossible to identify the submissions used for the training in our publications.

6.4.2.2 Questionnaire Overview

The beginning of the questionnaire focused on social-demographic background, including age, gender, education and employment status. This was followed by crowdwork experience, including HIT earning to total income, monthly income from MTurk, HIT approval rate. After a brief overview of the participants' backgrounds, they were asked about their experience about knowledge sharing and acquisition, including the type of preferred knowledge, the frequency of sharing/acquisition and the channel through which the knowledge was shared/accessed. It is worth noting that the knowledge acquisition behaviour includes searching and asking questions, and therefore were asked separately.

This was then followed by a session based on the two theoretical models UTAUT, SET and the trust factor. Before designing the questions, it was necessary to define the type of knowledge to focus on in order to make the questions more relevant. If the types of knowledge are defined too broadly, participants may become confused or misunderstand the questions, and the data may lose sufficient reference value.

Based on previous observations, the five types of knowledge (as described in Section 6.4.1) in the MTurk related worker forums include:

1. Skill-related knowledge: solutions for technical issues, working strategies, techniques for finding quality tasks.

2. Opportunity-related knowledge: suggestions of specific HITs, qualifications, or job opportunities from other platforms.

3. HIT tutorial knowledge: tutorials of doing specific HITs.

4. Evaluation knowledge: ratings or comments toward HITs and requesters.

5. Non-job-related knowledge: such as casual conversation.

The skill-related knowledge has been chosen as the main knowledge type of the study. This is because this type of knowledge contains tooling practice knowledge that aligns with the topic of the thesis.

Table 6.1 illustrate the survey questions regarding UTAUT and SET related constructs. In the beginning of the worker perception section, questions on PE, EE, SI and FC involved in UTAUT were asked separately. They were also asked to provide further explanation about the problems they encountered, including why they believed accessing or sharing knowledge was difficult or not effective enough. As can be seen from the table, latent variables such as PE and EE have been decomposed into multiple indicators. In addition, the proper nouns involved in the questionnaire have been annotated. For example, sharing tools are tools or mediums for sharing knowledge, such as forums (MTurk Crowd), Slack channels (Turker Nation), or browser extensions (TurkerViewJS) that you can leave ratings about HITs. Finally, to ensure that each indicator could reflect the participants' attitudes as precisely as possible, and to avoid possible quality problems with the sample data collected, more than three questions have been asked about their attitudes towards each construct. Each question was constructed with reference to previous research and framed based on crowdworkers' sharing of skill-based knowledge through communication tools. Thus, the questions fit the theme of this study.

Table 6.1 Survey questions for UTAUT related constructs.

Performance Expectancy (PE)		
PE1: Usefulness	Sharing tools are useful when I share this type of knowledge. Sharing tools are useful when I get this type of knowledge.	(Chang et al., 2013; Onaolapo & Oyewole, 2018)
PE2: Effectiveness	I can effectively share this type of knowledge using the sharing tools. I can effectively get this type of knowledge using the sharing tools.	
PE3: Perceived Speed	Using the sharing tools makes me share this type of knowledge more quickly. Using the sharing tools makes me get this type of knowledge more quickly.	

	(Optional) If you do not find it effective or useful to share or get this type of knowledge with sharing tools, can you specify why? How do you want to improve it?	
PE4: Relative Advantage	Sharing tools give me relative advantage when I share this type of knowledge. Sharing tools give me relative advantage when I get this type of knowledge.	(Onaolapo & Oyewole, 2018)
Effort Expectancy (EE)		
EE1: Ease of Use	It is easy to use the sharing tools to share this type of knowledge. It is easy to use the sharing tools to get this type of knowledge.	(Chennamaneni et al., 2012; Onaolapo & Oyewole, 2018)
EE2: Ease of Access	I can easily access sharing tools whenever and wherever I want to share or get this type of knowledge.	
EE3: Ease of Learning	Learning to operate the sharing tools is easy for me.	(Chang et al., 2013)
EE4: Technical Barrier	It requires much technical expertise to effectively use sharing tools. (Optional) If you feel it is not easy to share or get this type of knowledge, can you specify why? How do you want to improve it?	(Onaolapo & Oyewole, 2018)
Social Influence (SI)		
SI1: Platforms' Stance	The platform (MTurk, Prolific, Appen, etc.) believes that I should share this type of knowledge with other crowdworkers. (Optional) In your opinion why do they believe so?	(Bock et al., 2005)
SI2: Personal View of Platforms' Stance	I accept and carry out the platform's stance for sharing this type of knowledge even though it is different from mine.	
SI3: Peer Stance	Other crowdworkers believe I should share this type of knowledge with them. (Optional) In your opinion why do they believe so?	
SI4: Personal View of Peer Stance	I respect and put in practice my colleague's stance for sharing this type of knowledge.	
Facilitating Conditions (FC)		
FC1: Technology Integration	The sharing tools integrate well with other technologies I use during crowdwork, such as HIT managers, HIT catchers or visual enhancers. (Optional) If they do not integrate well, can you explain the issues further?	(Ajzen, 2020)
FC2: Community and Technical Support	The sharing tools are well supported by the communities or developers, such as providing guidance and maintenance.	(Hicks, 2020; Moore & Benbasat, 1991)
FC3: Compatibility	The sharing tools fit with my work processes and routines, they also support my work activities and goals	(Kamarozaman & Razak, 2021)
FC4: Personal Perception	Given the resources, opportunities, and knowledge it takes to use such technologies, it is easy for me to use the forums, channels and plugins for sharing knowledge.	(Vanneste et al., 2013)
Reciprocity (REC)		

REC1: Others' Willingness	I believe other crowdworkers actively share this type of knowledge.	(Chang & Chuang, 2011; Perugini et al., 2003)
REC2: Personal Willingness	I want to share tasks tips and insights with others because they will do the same in return.	
REC3: Attitude Towards Mutual Help	It is fair to help each other in forums, channels and platforms.	(Maximiano, 2017)
Reputation (REP)		
REP1: Image	Sharing this type of knowledge improves my image within the community.	(Kankanhalli et al., 2005; Van Den Besselaar et al., 2019)
REP2: Personal Perception	To what extent do you think sharing knowledge could improve your reputation?	
REP3: Respect	When I share this type of knowledge, the people I work with respect me.	
REP4: Recognition	Sharing this type of knowledge improves others recognition of me.	
REP5: General	Have you thought about sharing knowledge due to concerns about how it might affect your reputation?	
Reward (REW)		
REW1: Benefit	I feel that sharing this type of knowledge will benefit me directly.	(Liao et al., 2013)
REW2: Satisfaction	I feel that sharing this type of knowledge will give me satisfaction.	(Fang & Zhang, 2019; Wasko & Faraj, 2005)
REW3: Enjoyment	I feel that sharing this type of knowledge will give me enjoyment.	
REW4: Knowledge	I feel that sharing this type of knowledge will give me valuable information through interaction with peers.	
Social Interaction Ties (SIT)		
SIT1: Importance of maintaining relationship	It is important to maintain close social relationships with other crowdworkers via the sharing tools.	(Wang & Wang, 2013)
SIT2: Support from others	To what extent do your friends or colleagues support or encourage you to use this technology?	
SIT3: Communication frequency	I have frequent communication with other crowdworkers.	
Trust (T)		
T1: Trust via Forums	I trust others when sharing this type of knowledge on forums such as MTurk Crowd.	(Jøsang & Pope, 2005; LaPlante & Silberman, 2016; Mooradian et al., 2006)
T2: Trust via Plugins	I trust others when sharing this type of knowledge on plugins such as TurkerView.	
T3: Trust via Social Apps	I trust others when sharing this type of knowledge on social apps such as Slack, Facebook or Telegram.	
T4: Trust of knowledge being valued	I believe other crowdworkers will value my shared knowledge.	
T5: Trust of knowledge being not misuse	When sharing this type of knowledge with peers, I believe others will not abuse my knowledge or claim it as their own ideas.	
Knowledge Sharing Intention (KSI)		
KSI1: Current Intention	I am willing to share this type of knowledge with other crowdworkers.	(Bock et al., 2005; Yu et al., 2021)
KSI2: Future Intention	To what extent do you plan to share this type of knowledge via the sharing tools in the future?	

KSI3: Importance of KS	From 1 (very unimportant) to 5 (very important), how important is it to you to share this type of knowledge via sharing tools?	
KSI4: Current Intention	How likely are you to share your skill-based knowledge with other members via forums / channels / plugins?	
Knowledge Sharing Behaviour (KSB)		
KSB1: Behaviour Frequency	On average, how often do you post/share knowledge in forums, channels, or platforms about crowdwork?	(Min et al., 2008; Yu et al., 2021)
KSB2: Behaviour upon questions	When I see questions in the sharing tools (such as forums and social apps) that I can answer, I usually share my knowledge with them.	
KSB3: Behaviour after learning	When I have gained a piece of knowledge worth sharing, I share it immediately via the sharing tools.	
KSB4: General	I share skill-based knowledge regularly with peers.	

The validity of web-based experiments has been explored in previous research (Bryant et al., 2004). For example, increased dropout rates, malicious invalid responses and multiple submissions from the same participant can affect the quality of the data results (Hauser et al., 2018). To address the potential threats to data collection in web-based experiments, we describe the intent of the experiment to participants at the beginning of the task. In addition, attention check questions used to check whether participants were serious about answering were added to the questionnaire to further ensure that the data we received was sufficiently credible (Kung et al., 2018; Pei et al., 2020).

Here are the attention check questions applied into the survey, participants need to choose a number from 1 to 5 for each question:

1. Please select the option with the largest number to show you are not responding randomly.
2. This knowledge sharing study is answered carefully. Please choose the option in the middle.
3. I believe the colour of the sky is blue. Make sure to select the option with the smallest number.

6.4.2.3 Participants and Sample Size

While filtering out workers with low HIT approval rates could help to get more high-quality responses, the aim of this study is to include the crowdworkers with low HIT approval rates due to lack of skill-related knowledge. Therefore, the minimum HIT approval rate was not used as a filter for the participants. It is reasonable to assume that: for those respondents with very low HIT approval rates, they could be struggling communicating with peers, resulting in bad performance and therefore low approval rates.

To get the minimum required sample size n, Daniel's equation will be used for the calculation of sample size (Daniel & Cross, 2018):

$$n = \frac{N\left(\frac{za}{2}\right)^2 p(1-p)}{E^2(N-1)+\left(\frac{za}{2}\right)^2 p(1-p)} \tag{6.1}$$

Within this equation, $\left(\frac{za}{2}\right)^2$ is the critical value of the normal distribution at $\frac{a}{2}$. E is the margin of error, and p is the sample population. The confidence level for this study is 95%, and the margin of error is 5%. According to reliable statistics that can be found as of August 2023, it was reported that there were 250,810 MTurk workers worldwide who have completed at least one Human Intelligence Task (HIT) posted through the TurkPrime platform in 2019 (Robinson et al., 2019). Although we don't have access to the most recent statistics, a more reliable sample size can be obtained by considering extreme cases. If we assume that the total number of active workers in 2023 produces a 1000-fold increase over 2019, then the required sample size is 385. In other words, a sample size greater than 385 is required.

6.4.2.4 Method of Sampling

In this study, the surveys were distributed via publishing HIT groups on MTurk on different days during a week. Each HIT group contains 100 HITs, and a total of 6 HIT groups were published. As there was no minimum HIT approval rate to accept this survey task, high approval rate workers were not prioritised.

6.4.2.5 Response Collection

The responses collected through Google Forms were stored in a csv file and used for subsequent quality checks and analysis. To conduct a quality check through the entire set of responses, the raw data was first converted into dataframe format using two Python libraries: Chardet[49] and Pandas[50]. The quality of the participants' responses was assessed through

[49] Project description of Chardet: https://pypi.org/project/chardet/
[50] Pandas homepage: https://pandas.pydata.org/

attention check questions. The responses passing the attention check were determined to be valid for subsequent data analysis.

In order to ensure that participants who provided valid responses were rewarded in full, two strategies were implemented so that a one-to-one match could be made between the collected questionnaires and the identities of the workers who participated in the study. Firstly, participants were given a survey code after submitting their questionnaire responses and were asked to submit the code with their MTurk submission assignment. Second, a Python script was applied to match the Worker ID provided in the survey response from Google Form and the Worker ID from the MTurk server[51]. Surprisingly, as the HIT batches increased, more and more workers started to submit survey codes without completing questionnaires, reaching 46% in one of the batches. While there is no guarantee that every worker account will be restricted to one person in accordance with the platform's policy, The fact that the survey code was spread so quickly does reflect the efficient knowledge sharing about the content of HITs among workers. Another potential reason is that the same worker accepted this HIT using multiple accounts, and submitted the survey code maliciously using other accounts after first completion for more rewards.

Secondly, the high base reward ($0.35) for the HIT led to a large number of randomly completed questionnaires as participants simply wanted to get the base reward by submitting invalid results quickly. To reduce this malicious submission, the base reward was later adjusted downwards to $0.15 and the proportion of the reward for passing quality check was increased, thus encouraging participants to answer with the aim of receiving the full reward.

The amount of payment for micro-tasks was set to minimum UK hourly wage (which is relatively higher than countries crowdworkers are expected to be from US and India). Specifically, each participant automatically received a base reward of $0.15 for submitting a survey response. After passing the quality check, this participant received $2.85 as the bonus of providing a high-quality response. In the end, participants who completed the survey were rewarded $3 in total for the HIT. As the estimated time for completing the survey is about 10 - 15 minutes based on pilot tests, the hourly wage for this survey HIT ($12 ~ $18) is slightly above the minimum UK hourly wage, which is about $11.5 per hour.

[51] Script used to find invalid submissions: https://github.com/howrudoing/Scripts-for-thesis/blob/main/Find%20Malicious%20Turkers_Thesis%20Code.ipynb

Offering financial incentives to participants is a common and effective way of reducing dropout rates in online research and has been shown not to affect participant responses or sample characteristics (O'Neil et al., 2003). To avoid bias caused by multiple answers from the same participant, each crowdworker could only complete one survey for the project. To reduce item non-response, the questionnaire implemented on Google Form was divided into sections, and participants had to answer all questions in the current section before moving on to the next section.

To avoid the bias caused by duplicate answers from the same participant, each crowdworker was allowed to complete the HIT once for this project. As this study had multiple rounds of pilot tests before the final data collection, participants who had completed this survey before the new data collection, regardless of whether their responses were approved or not, were automatically assigned a qualification named "Already Participated in KS Study" via a Python script written for this study. By setting the rule in future questionnaire HITs that workers with this qualification cannot accept this HIT again, the response duplication could be avoided effectively.

To reduce survey abandonment rate and provide timely positive feedback for ongoing participant engagement, the questionnaire was divided into sections with completion progress prompts. Participants had to answer all questions in the current section before moving on to the next section. Finally, questionnaires were published in MTurk in the form of HITs in 6 batches (HIT groups), with 100 HITs for each batch (HIT group).

6.4.3 Data Analysis

This section analyses the quantitative data provided by the participants through three stages. The first stage is descriptive analysis, which focuses on the demographic aspect and is used to present basic information about the questionnaire participants. The second stage is to test the reliability and validity of the measurement model. Several analytical steps were used in this stage, including testing internal consistency of constructs, convergent validity and discriminant validity of the measurement model. Finally, Partial Least Squares Structural Equation Modelling (PLS-SEM) software, Smart PLS 3.0 (Ringle et al., 2022), was applied to test hypotheses arising from the conceptual model, including both direct and indirect effects. For this study, we chose to separate the measurement model from the structural model so that we could refine the measurement model based on reliability and validity tests.

6.4.3.1 Descriptive Analysis

Descriptive analysis summarises the data collected and describes the distribution of gender, age, education, income and HIT approval rates. In addition, participants' frequency of knowledge sharing and acquisition, types of knowledge, and common ways of sharing are described.

6.4.3.2 Internal Consistency of Constructs

Model reliability refers to the consistency of all the questions included in each construct. It also reflects the robustness of the questionnaire (Saunders et al., 2009). In order to assess the overall reliability of the model, the composite reliability (CR) values for each factor included in the model need to be tested. a value of 0.7 or above indicates that the survey questions measure the same construct (Gefen et al., 2000; Tentama & Anindita, 2020; Vinzi et al., 2010). In addition, Cronbach's α has been widely used to assess the construct's internal consistency. However, it is also criticised that it needs to be calculated assuming that all indicators have the same factor loadings (Hair et al., 2014). Therefore, Cronbach's α was used to assist in the evaluation of internal consistency in this study.

6.4.3.3 Factor Reliability

Prior to structural equation modelling analyses, there is a need to ensure that the survey questions can be reasonably constructed for each factor based on the pre-established theories. Previous studies have suggested that factor loadings should be larger than 0.5 or 0.7 (Hair et al., 2014; Hulland, 1999). Otherwise, factors with low factor loadings need to be removed, or the questions used to construct the factors need to be adjusted. The larger factor loading is, the more the observed variable explains each constructed latent variable (Jain & Chetty, 2022).

6.4.3.4 Model Validity

After the reliability test of each factor, the factors of the measurement model need to be further adjusted according to the convergent and discriminant validity of the measurement model.

Convergence validity demonstrates the relationship when two measurements that are meant to measure the same construct. It shows the degree of correlation within factors that measure the same construct (Chin & Yao, 2014). To evaluate the convergence validity, the average variance extracted (AVE) for each construct within the measurement model should be assessed. AVE represents the average variance attributable to the latent construct in the observed variable (dos

Santos & Cirillo, 2023). AVE values higher than 0.5 are considered to have good convergent validity (Bagozzi & Yi, 1988; Fornell & Larcker, 1981; García et al., 2022). In addition, the previously calculated CR can also be used for the convergent validity test.

Discriminant validity is the validity of comparing one construct with another to show the difference between them (Sujati & Gunarhadi, 2020; Taherdoost, 2016). According to the criterion proposed by Fomell & Larcker (1981), the square root of the AVE for each construct needs to be greater than the correlation coefficient between that construct and others.

6.4.3.5 Hypothesis Test

SEM consists of a measurement model and a structural model. Measurement model measure the covariance between latent and observed variables (Hoyle, 2011; Kline, 2011), while structural models test all hypotheses by regressing endogenous latent variables on a number of endogenous and exogenous latent variables (Hoyle, 2011; Kline, 2023). Therefore, after completing the previous tests of reliability and validity of the measurement model, it is necessary to test the proposed hypotheses by selecting the appropriate methods and tools for structural equation modelling analysis based on the sample.

Covariance-based structural equation modelling (CB-SEM) commonly uses Maximum Likelihood Estimation for comparing observed covariance matrices with estimated covariance matrices, in contrast to partial least squares structural equation modelling (PLS-SEM), which is based on Principal Component Analysis and Ordinary Least Squares (Hair et al., 2006). First, in CB-SEM, the model does not always converge (Hair et al., 2017). Moreover, it requires a large sample size, and the data should be normally distributed. In contrast, PLS is less sensitive to sample size and multivariate normal distribution requirements as it uses ordinary least squares (OLS) to explain the total variance (Gefen et al., 2000). Furthermore, PLS-SEM is a non-parametric approach that does not rely on distributional assumptions (Guenther et al., 2023).

6.5 Results

A total of 454 valid samples have been collected after removing 296 invalid responses based on attention check questions. After removing the missing data including participants who claimed not to have shared skill-based knowledge, a total of 413 samples were applied for SEM analysis.

6.5.1 Descriptive Statistical Analysis of the Demographic Variables

6.5.1.1 Social-demographic Background

Table 6.2 is a summary of demographic information of the participants. It is revealed that 57.5% are male and 40.1% are female. The majority of participants are aged from 25 to 44. In terms of education, more than a half of them are Bachelors.

Regarding the income, less than a quarter of the participants earn more than $501 per month, with around 3.4% of workers earning more than $5,000. Possibly benefiting from the increase in overall crowdsourcing industry revenues in recent years, this value has been better than the statistics of El Maarry et al. (2018). However, most of the workers still earn no more than $500 per month from MTurk.

Table 6.2 Sample demographics description.

Gender	Count	Percentage
Female	182	40.1%
Male	261	57.5%
Prefer not to say	11	2.4%
Age		
18-24	24	5.3%
25-34	251	55.4%
35-44	90	19.9%
45-54	54	11.9%
>55	34	7.5%
Education		
High School and below	26	5.7%
Bachelor	297	65.6%
Master or above	130	28.7%
Monthly Income from MTurk		
No more than $100	89	23.4%
$101 - $300	162	42.6%
$301 - $500	52	13.7%
$501 - $1000	36	9.5%
$1001 - $5000	28	7.4%
More than $5000	13	3.4%
HIT Approval Rate		
Less than 90%	33	7.3%
90% - 95%	19	4.2%
95% - 97.5%	70	15.4%
97.5% - 100%	332	73.1%

No minimum HIT approval rate has been used as a restriction for engaging into the study to reduce the response bias caused by allowing only high HIT approval rate crowdworkers. Therefore, those who had previously been frequently rejected were also eligible to complete the questionnaire. While this on the one hand increased the number of malicious responses (including asking for rewards without completing the questionnaire and filling in invalid information), on the other hand it provided a more complete picture of the distribution of the approval rate of workers who are actively looking for HITs.

From Table 6.2 it can be revealed that more than a quarter of the overall participants have an HIT approval rate lower than 97.5%, and more than 10% of the whole sample have a HIT approval rate lower than 95%. However, it is common for requesters to set this approval rate above 95%-98% when posting HITs (Burnette et al., 2022; Hauser & Schwarz, 2016; Kennedy et al., 2020; Saravanos et al., 2021). This means that many crowdworkers who are actively looking for HITs are losing out on stable jobs because of low HIT approval rates.

6.5.1.2 Statistics of Knowledge Sharing and Acquisition

Table 6.3 Distribution of participants' frequencies of knowledge sharing.

Frequency of Knowledge Sharing	Count	Percentage
Never	11	2.4%
Once a week or less	118	26.0%
Once every two/three days	133	29.3%
Once every day	145	31.9%
Multiple times a day	47	10.4%

Table 6.3 shows that only a very small number of participants have not shared knowledge about crowdwork with others, while the majority share knowledge at least weekly. Interestingly, more than 10% of the participants claimed to be sharing knowledge frequently daily. The definition of knowledge sharing was explained together with the questions to ensure that the participants had the correct understanding.

Table 6.4 Number of participants sharing and acquiring knowledge under different types.

Knowledge Type	Knowledge Sharing	Knowledge Acquisition	Both
Skill-related knowledge	262	255	203
Opportunity-related knowledge	305	204	165
HIT tutorial knowledge	211	209	149
Evaluation knowledge	175	196	129
Non-job-related knowledge	34	41	20

It can be revealed from Table 6.4 that participants most frequently share knowledge about job opportunities and skills, and most often searched for or asked for skills-based knowledge. The reasons for this phenomenon were further explored in the participants' feedback:

> *"I need more work opportunities, and if you have any questions you can just say it below. There will be quite a few comments if it's (this job recommendation) great or terrible."*

> *"While sometimes I do want to talk about a script, there really isn't a good place to get started."*

> *"The question about the skills got to be very clear…"*

From the feedback, the difficulties faced by workers in sharing skill-based knowledge are the lack of well-organized topic categorisation and the need for clear questions. In addition, because workers can easily benefit from HIT recommendations without any prior knowledge, they have a higher degree of applicability. This high applicability also makes it easier to receive feedback from others on the sharing of opportunity knowledge, which encourages the continuation of this sharing behaviour.

Another interesting observation is that workers' sharing and access to evaluation knowledge is significantly lower than the previous three types. Evaluation Knowledge includes their reviews and ratings toward HITs and requesters. According to the feedback from workers, it can be speculated that this could be due to lower expected benefits. In other words, workers tend to be most interested in sharing evaluations about the requester or the HIT when they are treated unfairly, in the expectation that their problem could be resolved. In other cases, ratings of the HITs or requesters do not receive a response from either the requester or the worker, and are therefore less motivated for the knowledge contributors.

Table 6.5 Number of participants sharing and acquiring knowledge under different approaches.

Approach	Knowledge Sharing	Knowledge Acquisition	Both
E-mail	229	183	150
Forum	281	283	235
Social Apps	282	291	238
Face-to-face	71	66	35

The numbers of participants using different approaches (Table 6.5) reveal that the most common means of knowledge sharing and acquisition are forums and social apps. It is worth noting that workers also share knowledge with their friends face-to-face. This option has been added into the question after receiving extensive feedback from the pilot tests. Moreover, face-to-face communication among workers has been discussed in a previous study (Gray et al., 2016; Gupta et al., 2014). The advantages of this format over online communication include a higher level of mutual trust, quicker feedback, and richer interactions (Damen et al., 2020).

Besides the ways provided in this survey question, participants have also reported using search engines for knowledge acquisition. This approach is more accessible than a forum channel that requires registration or membership, and in recent years, thanks to deep learning, search engines have become more capable of summarising and organising knowledge on target topics than ever before (Floridi & Chiriatti, 2020).

6.5.2 Preliminary Test of Sample Data

6.5.2.1 Outlier Test

Outliers are observations that are distinct from the majority because they score too high or too low (Hair et al., 2006; Tabachnick et al., 2007). The presence of outliers affects the results of the effect analysis and biases subsequent interpretations. Outliers can be detected via standardised value and boxplot.

The standardised value, standard score, or z-score is obtained by dividing the distance between the current number and the mean by the standard deviation. It can therefore be viewed as the relative position of the current value in the total data. For large samples, standardised values with absolute values greater than 3 are considered outliers (Hair et al., 2006)

In this study, to detect univariate outliers, factors were grouped for each construct and their scores were summed up as the construct score. Using the descriptive statistics function in SPSS, the total score for all constructs was converted to standardised scores for outlier detection. As a result, no outliers were found from the standardised score test.

Boxplot was also applied for outlier detection. Figure 6.3 illustrates the distribution of the total scores under each construct. According to boxplot, only a few observations were found to be slight outliers (interquartile range (IQR) > 1.5) and no extreme outliers were found. Outliers are preserved unless there is evidence that the outliers do indeed deviate from the dataset and do not represent any observations (Hair et al., 2006). Even if the outliers were found to be problematic, they could be treated without seriously biassing the results (Tabachnick et al., 2007). As a result, these mild outliers were ultimately preserved.

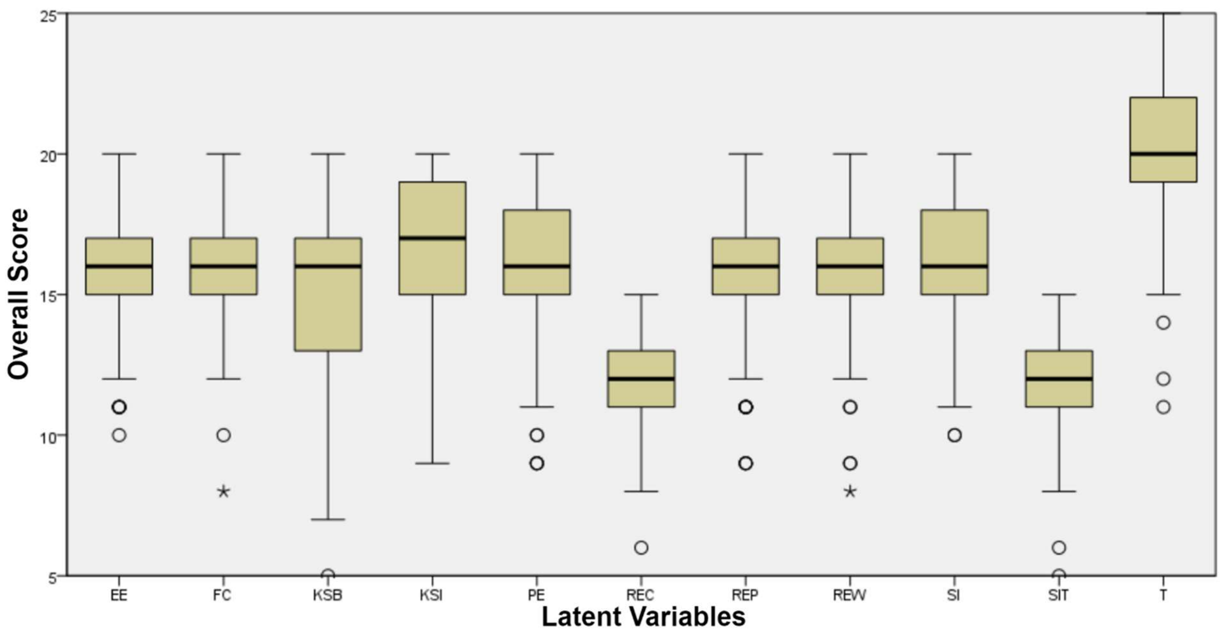

Figure 6.3 Detecting multivariate outliers using boxplot.

6.5.2.2 Normal Distribution Test

The normal distribution test for the sample was performed in this study using skewness and kurtosis (Siraj-Ud-Doulah, 2021). While kurtosis describes the "flatness" of the distribution in comparison to the normal distribution, skewness refers to the distribution's symmetry (Field, 2013; Hair et al., 2006). It has been suggested that absolute values of skewness and kurtosis less than 2.58 indicate that the variable data are normally distributed (Hair et al., 2006). It was also suggested that this absolute value needs to be less than 2 (Bollen & Long, 1993). When the skewness is greater than 3, it is considered to be extremely skewed (Kline, 2011). Extremely non-normally distributed data could potentially make an impact on the significance of parameter estimations (Hair et al., 2011; Henseler et al., 2009). In addition, the larger the sample size, the less the negative impact of non-normality (Hair et al. 2006). In this study, the feasible sample size was 413, thus reducing the negative impact of non-normal distribution of individual factors.

Regarding the total score of all the factors (observed variables), it follows a normal distribution as the absolute values of kurtosis and skewness are both less than 2. However, when testing for independent observed variables (Table 6.6), the data for the three observed variables, PE1, FC4, and KSB2, had moderate deviations from normal distribution because the absolute value of kurtosis was greater than 2 (Curran et al., 1996). Since PLS-SEM uses a nonparametric statistic which makes no distributional assumptions, we do not require that the data necessarily follow a normal distribution.

Table 6.6 Skewness and kurtosis for each observed variable

Factor	Skewness	Kurtosis	Factor	Skewness	Kurtosis	Factor	Skewness	Kurtosis
	Statistic	Statistic		Statistic	Statistic		Statistic	Statistic
PE1	-0.805	2.725	FC3	-0.312	-0.146	SIT2	-0.463	-0.262
PE2	-0.469	0.263	FC4	0.021	2.092	SIT3	-0.649	0.545
PE3	-0.437	-0.345	REC1	-0.434	0.188	T1	-0.587	1.173
PE4	-0.254	0.113	REC2	-0.453	-0.484	T2	-0.585	-0.238
EE1	-0.278	-0.414	REC3	-0.313	-0.328	T3	-0.593	0.969
EE2	-0.345	-0.335	REP1	-0.484	0.161	T4	-0.190	-0.577
EE3	-0.347	-0.321	REP2	-0.422	-0.254	T5	-0.703	1.406
EE4	0.060	0.601	REP3	-0.441	-0.050	KSI1	-0.647	0.767
SI1	-0.649	2.655	REP4	-0.228	0.876	KSI2	-0.450	-0.414
SI2	-0.354	-0.890	REW1	-0.473	-0.106	KSI3	-0.969	1.424
SI3	-0.604	1.231	REW2	-0.521	0.269	KSI4	-0.840	0.377
SI4	-0.355	-0.182	REW3	-0.545	0.014	KSB1	0.015	-0.892
FC1	-0.398	0.462	REW4	-0.200	0.086	KSB2	-0.852	2.252
FC2	-0.432	-0.375	SIT1	-0.556	0.586	KSB3	-0.712	0.408
						KSB4	-0.462	0.881

6.5.2.3 Multicollinearity test

Multicollinearity occurs when there is a strong correlation between independent variables, which can lead to incorrect or unstable results (Byrne, 2016). Multicollinearity can lead to incorrect standardised regression coefficients, which can produce inflated results. Therefore, it is necessary to check the model for multicollinearity. Indicators commonly used to check for covariance include TOL and VIF (Hair et al., 2011). Variance inflation factor (VIF), which is the inverse of TOL, was used to check for covariance in this study. It is generally accepted that if the VIF is greater than 5, the predictor (observed variable) may have strong multicollinearity with another predictor (Myers, 2000). Table 6.7 indicates that the observed variables in the proposed model do not have multicollinearity issue.

Table 6.7 VIF score for each observed variable.

Observed Variable	VIF	Observed Variable	VIF	Observed Variable	VIF	Observed Variable	VIF
EE1	1.655	KSI1	1.143	REP1	1.279	SIT1	1.136
EE2	1.407	KSI2	1.027	REP2	1.066	SIT2	1.037
EE3	1.265	KSI3	1.139	REP3	4.709	SIT3	1.122
EE4	2.367	PE1	1.543	REP4	4.612	T1	1.282
FC1	1.562	PE2	1.407	REW1	1.536	T2	1.202
FC2	1.149	PE3	1.318	REW2	2.871	T3	1.165
FC3	1.526	PE4	1.400	REW3	1.784	T4	1.214
FC4	2.109	REC1	1.088	REW4	4.554	T5	1.320
KSB1	1.080	REC2	1.009	SI1	1.267		
KSB2	1.098	REC3	1.096	SI2	1.073		
KSB3	1.154			SI3	1.179		
				SI4	1.147		

6.5.3 Measurement Model

In this subsection, the internal consistency, factor reliability, convergent validity and discriminant validity of the model were tested based on the previously devised tests for measurement models. All the tests for the proposed measurement model have been assessed via SmartPLS 3 and 4 (Ringle et al., 2022).

6.5.3.1 Internal consistency and factor reliability

The factor loadings for the valid indicators should be greater than 0.70, which represents that the factor extracts sufficient variance from that observed variable (Tavakol & Wetzel, 2020). It was also suggested that the average variance explained (AVE) needs to be at or above 0.50 (Fornell & Larcker, 1981). In other words, all indicators of a factor should explain at least 50% of the variance on average (Ali et al., 2018; Hair et al., 2019). Moreover, it has also been suggested that the factor loadings could be smaller under the sample size of more than 350 (Comrey & Lee, 1992; Tabachnick et al., 2007). For those factors with AVE scores lower than

0.5, composite reliabilities (CR) should be at least higher than the acceptable level of 0.6 (Lam, 2012). However, when the factor loadings are too high, attention also needs to be paid to whether this is because the observed variables are examining the same question repeatedly and may have covariance problems with other observed variables.

Therefore, all the indicators with less than 0.60 factor loadings have been removed since they would significantly lower the AVE of the constructed factors. More specifically, within the UTAUT related factors, EE3 (0.573), SI2 (0.563), SI3 (0.560), FC2 (0.476) and KSB1 (0.596) have been removed from the measurement model. Regarding the SET related factors, REC2 (0.606), REW2 (0.519), REP3 (0.584). Finally, T2 (0.577), T3 (0.454), T5 (0.578) from the Trust factor have been removed. Admittedly there is a risk in doing so, as removing relevant questions to boost the AVE score may result in a loss of content validity. However, these removed questions represent only a specific dimension of measurement under a factor, so the removal of individual questions does not result in the loss of meaningful measurement of the factor. After calculating the AVE and CR score for each factor with the retained indicators, REC, SI and SIT were removed.

Table 6.8 Measurement model confidence and validity analysis.

Construct	Measurement Factor	Factor Loading	Cronbach α	CR	AVE
Performance Expectancy (PE)	PE1	0.799	0.742	0.751	0.563
	PE2	0.709			
	PE3	0.720			
	PE4	0.771			
Effort Expectancy (EE)	EE1	0.796	0.726	0.767	0.651
	EE2	0.697			
	EE4	0.913			
Facilitating Conditions (FC)	FC1	0.807	0.764	0.794	0.679
	FC3	0.766			
	FC4	0.893			
Reward (REW)	REW1	0.744	0.698	0.727	0.624
	REW2	0.742			
	REW3	0.876			
Knowledge Sharing Intention (KSI)	KSI1	0.868	0.719	0.763	0.647
	KSI3	0.634			
	KSI4	0.887			
Knowledge Sharing Behaviour (KSB)	KSB2	0.786	0.754	0.766	0.673
	KSB3	0.757			
	KSB4	0.911			

CR represents the internal consistency of the observed variables in each latent variable. The model would be considered to have good internal consistency when Cronbach α and CR were greater than 0.7 (Fornell and Larcker, 1981). However, these criteria do not necessarily mean that factors that fail to meet the requirements should be removed from the model. Based on the interpretation from George and Mallery (2003), factors with the Cronbach α slightly below 0.7 are still informative if other parameters meet the criteria. Factors with extremely low CR and Cronbach's α have been removed from the final model, including Social Influence (SI), Social Interaction Ties (SIT), Reputation (REP) and Trust (T). As illustrated in Table 6.8, the final selection of constructs and observed variables provides the model with an acceptable reliability.

6.5.3.2 Convergent validity and discriminant validity

Tests of validity include tests of convergent validity and discriminant validity. The purpose of convergent validity is to test whether multiple observed variables belonging to the same construct converge to the same construct. This involves the assessment of CR values, factor loadings and the AVE values.

As indicated before, the factor loadings for all factors exceeded 0.5, and the AVEs of all latent variables all exceeded the threshold of 0.50 (Fornell & Larcker, 1981), suggesting that these factors are empirically distinct. Therefore, the model is considered to have good convergent and discriminant validity.

Cross loading scores were applied to test the discriminant validity of the existing model. From Table 6.9, it can be revealed that the factor loadings of each construct are larger than their cross loadings, indicating good discriminant validity of all the included constructs (Hair et al., 2014; Roubertoux et al., 2020).

Table 6.9 Cross loading matrix for observed variables.

	PE	EE	FC	REW	KSI	KSB
PE1	0.799	0.534	0.553	0.379	0.480	0.301
PE2	0.709	0.507	0.451	0.415	0.353	0.323
PE3	0.720	0.467	0.382	0.399	0.437	0.355
PE4	0.771	0.535	0.517	0.444	0.491	0.474
EE1	0.544	0.796	0.501	0.416	0.420	0.385
EE2	0.451	0.697	0.448	0.334	0.356	0.277
EE4	0.634	0.913	0.611	0.575	0.520	0.505
FC1	0.549	0.509	0.807	0.374	0.446	0.382
FC3	0.440	0.484	0.766	0.384	0.416	0.325
FC4	0.574	0.604	0.893	0.526	0.517	0.469
REW1	0.405	0.444	0.399	0.744	0.418	0.313
REW3	0.381	0.348	0.375	0.742	0.395	0.377
REW4	0.493	0.520	0.468	0.876	0.543	0.411
KSI1	0.573	0.501	0.539	0.483	0.868	0.431
KSI3	0.337	0.326	0.322	0.379	0.634	0.350
KSI4	0.498	0.465	0.472	0.523	0.887	0.570
KSB2	0.461	0.413	0.466	0.433	0.470	0.786
KSB3	0.330	0.354	0.262	0.335	0.445	0.757
KSB4	0.398	0.442	0.438	0.370	0.485	0.911

The correlation coefficient matrix between the variables is shown in Table 6.10 below. The square root of the AVE of each variable is on the diagonal. The square root of the AVE for all latent variables is more significant than their correlation coefficients with other variables, as shown in Table 6.10, indicating that the model has good discriminant validity (Fornell & Larcker, 1981).

Table 6.10 Correlation coefficient matrix and AVE square root values.

	EE	FC	KSB	KSI	PE	REW
EE	0.807					
FC	0.651	0.824				
KSB	0.494	0.483	0.821			
KSI	0.542	0.561	0.569	0.804		
PE	0.680	0.637	0.488	0.593	0.751	
REW	0.561	0.527	0.465	0.579	0.544	0.790

In summary, the measurement model containing PE, EE, FC, REW, KSI, KSB has internal consistency, factor reliability, convergent validity, and discriminant validity. Next, the structural model can be analysed to test the path relationships between the constructs. Figure 6.4 shows the conceptual framework after the measurement model test.

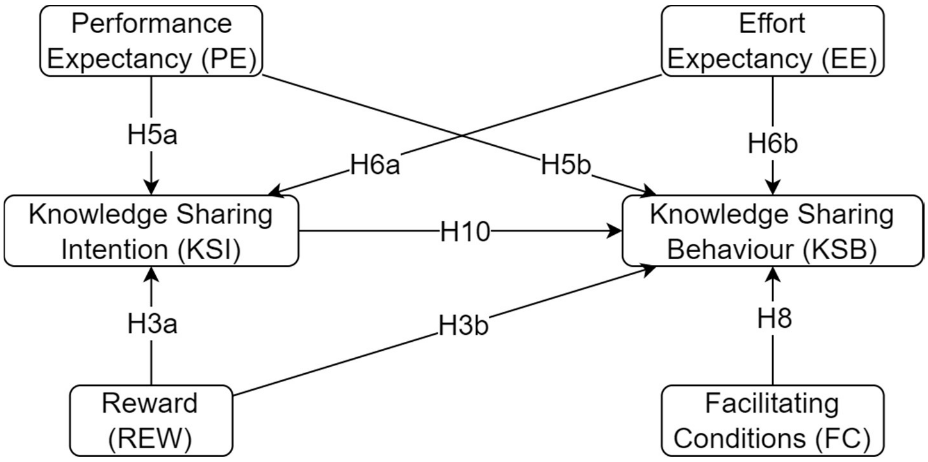

Figure 6.4 Modified conceptual framework.

6.5.4 Structural Model Testing and Results

This section contains significance tests of the hypothesised paths as well as the quality tests of the structural model. The structural model is tested in three aspects: the predictive ability of the model, its explanatory power, and model fitness. In general, this section obeys the following sequence of analyses: firstly, the significance test of the path coefficients, followed by the R^2 and the explanatory effect value f^2. Next step is the assessment of predictive relevance using q^2. Finally, there is an assessment of the model fitness using Goodness of Fit (GoF).

The results of the analysis of all valid paths are shown in Table 6.11 and Figure 6.5. It can be revealed that out of the total 8 testable research hypotheses, 5 hypotheses were supported, and

3 hypotheses were not supported. Due to insufficient evidence, other hypotheses within the initial conceptual model failed to be tested. However, it does not mean the effect did not exist.

Table 6.11 Structural equation model path coefficients.

Hypothesis	Path Relation	Relationship	Path Coefficient (t-value)	Supported or Not
H3a	REW -> KSI	positive	0.323 (6.545) ***	Supported
H5a	PE -> KSI	positive	0.316 (5.835) ***	Supported
H6a	EE -> KSI	positive	0.147 (2.871) **	Supported
H3b	REW -> KSB	positive	0.099 (1.605)	Not Supported
H5b	PE -> KSB	positive	0.084 (1.374)	Not Supported
H6b	EE -> KSB	positive	0.138 (2.182) *	Supported
H8	FC -> KSB	positive	0.101 (1.645)	Not Supported
H10	KSI -> KSB	positive	0.331 (5.191) ***	Supported

Note: *** means $p < 0.001$. It shows very strong evidence against the null hypothesis. For those with a significance value less than 0.01, it shows strong evidence against the null hypothesis (**). When the significance value is between 0.01 and 0.05, it indicates good evidence against the null hypothesis (*).

Specifically, REW (0.323, $p < 0.001$) and PE (0.316, $p < 0.001$) both very significantly affected knowledge sharing intention (KSI) and both had high effects. In addition, EE also significantly influenced KSI (0.147, $p < 0.01$). However, the influence was not as effective as the first two exogenous constructs REW and PE. In contrast, among the constructs directly influencing final behaviour, KSI had a significant effect on KSB (0.331, $p < 0.001$) and had the largest effect. Notably, EE had a relatively significant effect on KSB (0.138, $p < 0.05$). Ultimately, unlike the assumption of the traditional UTAUT model, the hypothesis of the effect of Facilitating Conditions (FC) on KSB was not supported in this study (Attuquayefio & Addo, 2014; Yee & Abdullah, 2021).

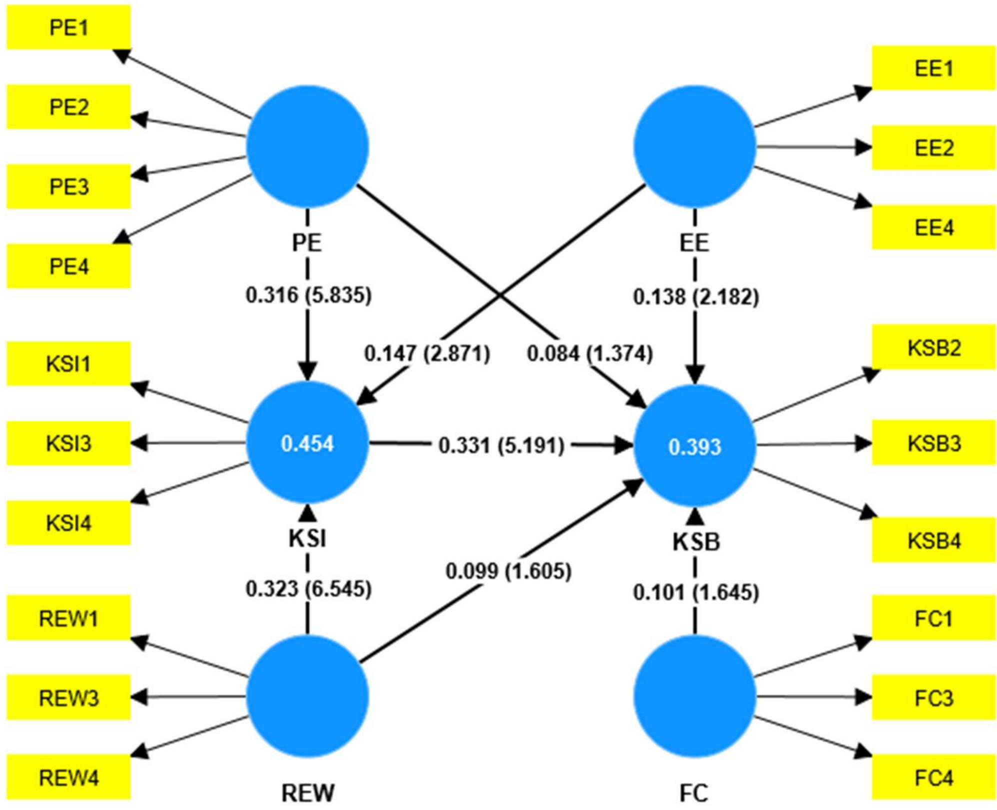

Figure 6.5 Structural model with path coefficients and t value.

6.5.4.1 Explanatory power tests for structural models

After testing the significance of the path coefficients, the quality of the structural model begins to be examined. R^2 is one of the commonly used metrics, which represents the square of the correlation coefficient between the actual and predicted values of a particular dependent construct (KSI, KSB in this model). Thus, it can be used to measure the predictive power of independent constructs in structural models. R^2 ranges between 0 and 1, and higher values means greater explanatory power. R^2 less than 0.25 means weak explanatory power, in the interval [0.25, 0.5) means moderate explanatory power and greater than or equal to 0.5 means strong explanatory power. From Table 6.12, the value of KSI is 0.454, so the model has moderate explanatory power for the construct KSI. Similarly, the value of KSB is 0.393 which also indicates that the structural model has moderate explanatory power for KSB.

Table 6.12 Path coefficient with explanatory power test and predictive ability test scores.

Hypothesis	Path Relation	Path Coefficient (t-value)	Supported or Not	R^2	q^2	95%CI Lower	95%CI Upper	Q^2
H3a	REW -> KSI	0.323 (6.545) ***	Supported		0.161	0.227	0.419	
H5a	PE -> KSI	0.316 (5.835) ***	Supported	0.454	0.108	0.202	0.416	0.441
H6a	EE -> KSI	0.147 (2.871) **	Supported		0.038	0.045	0.246	
H3b	REW -> KSB	0.099 (1.605)	Not Supported		0.165	-0.018	0.219	
H5b	PE -> KSB	0.084 (1.374)	Not Supported		0.152	-0.043	0.201	
H6b	EE -> KSB	0.138 (2.182) *	Supported	0.393	0.140	0.006	0.254	0.314
H8	FC -> KSB	0.101 (1.645)	Not Supported		0.160	-0.016	0.224	
H10	KSI -> KSB	0.331 (5.191) ***	Supported		0.117	0.208	0.455	

Note: *** means p < 0.001. It shows very strong evidence against the null hypothesis. For those with a significance value less than 0.01, it shows strong evidence against the null hypothesis (**). When the significance value is between 0.01 and 0.05, it indicates good evidence against the null hypothesis (*).

6.5.4.2 Predictive ability tests for structural models

Q^2 is the predictive relevance of the structural model in predicting endogenous constructs. This indicator is intended to measure the model predictive relevance. A Q^2 value greater than 0 indicates a good model predictive relevance (Hair et al., 2013), As shown in Table 6.12, the Q^2 for KSI and KSB are both greater than 0, indicating that the model has good predictive relevance.

In contrast, q^2 represents the predictive relevance of specific exogenous constructs to the endogenous constructs. From Table 6.12, it can be observed that the predictive effect values q^2 of REW -> KSI, PE -> KSI, EE -> KSI, EE -> KSB, KSI -> KSB are all between 0.02 and 0.15, which represents the weak predictive relevance of the model to the above relationships. In addition, the q^2 of REW -> KSI, REW -> KSB, FC -> KSB and PE -> KSB were greater than 0.15, indicating that the model has medium predictive relevance for the above relationships. Overall, the exogenous constructs generally have a medium degree of predictive relevance to the endogenous constructs in this structural model.

6.5.4.3 Overall fitness of structural model

Standardised Root Mean Square Residual (SRMR) is defined as the difference between the observed correlation and the model implied correlation matrix. It thus can be used as a metric reference for the model fit. Values less than 0.10 or 0.08 are considered to be a good fit (Hu &

Bentler, 1999). Henseler et al. (2014) used SRMR as a fit metric for PLS-SEM. In this study, the SRMR of the present structural model is 0.072, reflecting a good model fit.

Root mean square residual covariance (RMS-theta) is often used to assess the degree of correlation between measurement residuals, with closer to 0 representing less correlation between measurement residuals. When the RMS-theta is less than 0.12, the model is considered to have a good fit (Henseler et al., 2014). The RMS-theta of the present structural model is 0.176, which is slightly above the optimal interval, but it is still acceptable.

NFI is defined as 1 minus the Chi^2 value of the proposed model divided by the Chi^2 value of the invalid model. Therefore, the NFI value lies between 0 and 1, with the closer the NFI is to 1, the better the fit is, and NFI values above 0.9 usually represent an excellent fit (Bentler & Bonett, 1980). The NFI for our structural model is 0.757, which is slightly less than 0.9. However, the NFI does get biased by model complexity (adding parameters to the model). The more parameters in the model, the greater the NFI result would be (Kenny, 2020). Because of this potential bias, this study does not use this metric as a basis for evaluating the fitness of the present structural model.

Finally, the Goodness of Fit (GoF) has been developed as an overall measure of model fit for PLS-SEM (Tenenhaus et al., 2004). The GoF of this model is 0.373, which is greater than 0.36, indicating the high fitness of our structural model (Table 6.13).

Table 6.13 Metrics to test model fit.

SRMR (<0.08)	GoF (>0.36)	RMS-theta (<0.12)
0.072	0.373	0.176

6.5.4.4 Indirect Effects Test

This study uses bootstrapping to examine the mediating effect within this structural model. The estimates of indirect effects and 95% confidence intervals were derived from 5000 Bootstrap samples. The significance of different effects can be detected not only by p value but also by t value (check if t value > 1.96). From the indirect relationships, it can be seen from Table 6.14 that PE significantly affects KSB indirectly through KSI (t value = 4.462, $p<0.001$), EE significantly affects KSB indirectly through KSI (t value = 2.821, $0.001<p<0.01$), and EE significantly affects KSB indirectly through KSI (t value = 3.787, $p <0.001$). In addition, it is revealed that EE has a relatively significant direct effect on KSB (t value = 2.182, $0.01<p<0.05$), and this effect (path coefficient = 0.138) is greater than the indirect effect on KSB through KSI (path coefficient = 0.049). Combined with the previous path analysis table, it can be found that the construct that causes the largest total effect on KSB is KSI (total effect = 0.331), followed by Reward (total effect = 0.206).

Table 6.14 Illustration of indirect effects.

Independent Variable	Intervening Variable	Dependent Variable	Direct Effect (t-value)	Indirect Effect (t-value)	Total Effect (t-value)
PE			0.084 (1.374)	0.105 (4.462) ***	0.189 (3.096) **
EE	KSI	KSB	0.138 (2.182) *	0.049 (2.821) **	0.187 (2.953) **
REW			0.099 (1.605)	0.107 (3.787) ***	0.206 (3.056) **

Note: *** means $p < 0.001$. It shows very strong evidence against the null hypothesis. For those with a significance value less than 0.01, it shows strong evidence against the null hypothesis (**). When the significance value is between 0.01 and 0.05, it indicates good evidence against the null hypothesis (*).

6.6 Interpretation of the Results

The main objective of this study is to explore the influencing factors of skill-based knowledge sharing among crowdworkers using PLS-SEM model. To explain the relationship between performance expectation, effort expectation, reward, knowledge sharing intention and knowledge sharing behaviour. In the hypothesised model, 4 exogenous variables (PE, EE, REW, FC) and 2 endogenous variables (KSI, KSB) are included.

This section interprets the results of the analyses around the structural model. Specifically, the effects of each exogenous construct on KSI and KSB are interpreted and expanded upon by incorporating qualitative content collected from participant responses regarding the relationship of these effects. Finally, the reasons why workers use HIT catchers, and their use preference are described.

6.6.1 What truly matters, reward or reputation?

In the SET framework, only Reward was included in the final structural model, and Reward showed a significant effect on Knowledge Sharing Intentions (path coefficient = 0.323). The survey questions for the Reward factor include satisfaction, enjoyment and knowledge from peers. The very significant correlation also confirms that participants want to gain enjoyment and satisfaction from sharing skill-based knowledge. In addition, the analysis in this study found that Reward had a significant indirect effect on KSB, and this effect of intrinsic motivation on knowledge sharing behaviour has been widely emphasised in previous studies (Fang & Zhang, 2019; Hung et al., 2015; Maharani, 2017; Osterbrink & Alpar, 2021). In addition, participants' textual responses revealed that non-material rewards also include knowledge shared by others during the communication.

Motivations of KS regarding reward also appeared in other participants' explanations, including *"just interest", "no reason", "it's just important to others", "for fun"*, *"we need to help each other through difficulties",* etc. In summary, an important motivation for workers' willingness to share skill-based knowledge is to exchange knowledge with peers and to gain the pleasure and satisfaction that comes from sharing.

Regarding the value of skill-based knowledge sharing, a number of participants expressed a similar view via textual feedback that what motivated them to try tools including HIT catchers was seeing discussions from others in the forum. In other words, the sharing of skill-based knowledge about the tools significantly contributed to the popularity of the tools including HIT catchers. The responses collected through the questionnaire also support this finding (Table 6.15), with almost three quarters of the participants stating that they started using the scripting tools because they were recommended by others or because they read information shared by others.

Table 6.15 Summary of how participants know about the scripting tools.

	Count	Percentage
Recommended by others (forum members, friends)	157	37.90%
Read information shared by others via forums or channels	142	34.50%
Finding tools online entirely on your own with no recommendations from others	114	27.60%

On the other hand, however, participants expressed concerns in their feedback about losses in opportunity due to knowledge sharing:

> *"Using these tools is very useful when I'm receiving knowledge, but showing my own tricks can potentially be detrimental because then other workers will start taking up HITs and taking work away from me."*

It is reasonable to speculate that this fear of losing future job opportunities simultaneously reduces their subjective expectations of Reward and hence KSI. however, from the answers we collected, about 85.7% of the respondents believe that their technical advantage will be diminished. By summarising the textual responses, it was found that the reasons focused on the fact that it would be easier for others to get and complete HITs, thus leaving themselves with fewer job opportunities. Interestingly, 14.3% of the participants did not believe that their skill advantage would be mitigated. Based on the textual responses provided by the participants, it was found that the reasons included: messages in the channel could be deleted quickly; by sharing knowledge one could become part of new groups and thus get help from others; and the belief that knowledge should be passed on to a wider range of people.

6.6.2 When it's easier to share, I'm more willing to do so.

In the structural model, Effort Expectancy (EE)[52] had a relatively significant positive effect on Knowledge Sharing Intention (KSI) (path coefficient = 0.147), reflecting the fact that the ease of use, access, and learning such knowledge sharing tools were indeed particularly important to participants' willingness to share skill-based knowledge. In addition, in the test of indirect effect, EE was found to have both direct effect (0.138) and indirect effect (0.049) on behaviour

[52] It refers to the ease of use, access, learning and technical barrier of using communication tools.

(KSB). And direct effect is higher. This implies that communication tools that require less efforts allow participants to share more skill-based knowledge. This study's finding that EE directly affects KSB is a complementary improvement to the application of the existing UTAUT model to the knowledge sharing domain. This finding illustrates that the magnitude of efforts using communication technologies directly affects the generation of final sharing behaviours, not just the willingness to share. In other words, after a worker has developed a willingness to share knowledge, they might end up abandoning the sharing of knowledge because of the consideration of the difficulties they may face in the conduct of the behaviour, such as the need to redirect to the appropriate forum thread or the need to register as a channel member. In addition, the lack of studies examining knowledge sharing behaviours from a technological perspective leads to this being a relatively novel finding.

By answering the optional question[53], participants expressed their views on the accessibility of the KS tools and the difficulty of using them:

> *"Sharing can only be truly effective if there's a process everyone is aware of and can contribute to."*
>
> *"Must be familiar with which boards in the forum correspond to which information; otherwise, finding useful ones is difficult."*

It can be observed that the ease of use of tools was effective in increasing their willingness to share. However, the current design of the forum is not friendly enough to both knowledge acquisition and sharing workflows.

In addition, Performance Expectancy (PE)[54] was found to have a significant effect on KSI (path coefficient = 0.316). This implies that the speed of using the knowledge sharing tool and whether it is effective for sharing knowledge significantly affect participants' willingness to share skill-based knowledge. While workers may be unwilling to share knowledge due to concerns about the loss of their technical advantage, the ability of a communication tool to allow the participants to spread their knowledge effectively and quickly can significantly affect their KSI for those who already have an initial willingness.

[53] (Optional) If you feel it is not easy to share or get this type of knowledge, can you specify why? How do you want to improve it?
[54] It refers to the usefulness, effectiveness, perceived speed and relative advantage of using communication tools.

When comparing the total effects of the PE, EE, and REW constructs on final behaviour KSB, it can be found that the total effect from REW (0.206) is slightly higher than that of PE (0.189), and EE (0.187). This implies that participants perceive that the fun, satisfaction, and knowledge from others during communication is more influential on their final sharing behaviour than whether the sharing technique is efficient and effective.

KSI was found to have the highest total effect (path coefficient = 0.331) on KSB, which means participants' willingness to share affect their sharing behaviour the most. This validates the general consensus among studies on the UTAUT framework (Attuquayefio & Addo, 2014; Yee & Abdullah, 2021). Interestingly, in conjunction with the structural model's study of indirect effects, it can be found that PE, EE and REW all further influence the final behaviour via KSI. This echoes our previously mentioned hypothesis about the stage-specific nature of the KSI, whereby a worker's initial willingness to share may influence the role of individual exogenous constructs on their later willingness and eventual sharing behaviour.

In addition, there could be another hypothesis arising from the findings: KSI includes raw behavioural intentions that leave out technological considerations, as well as composite behavioural intentions that include considerations of technological dimensions, and that the KSI should therefore be split into two latent variables when constructing the conceptual model. Workers may only consider possible resistance to behaviour before they develop a willingness to share, and the subjective assessment of resistance may in turn influence their initial willingness.

6.6.3 Insights on the Trust and Social Influence Factor

The trust and social influence factors that were planned to be studied were not included in the SEM analysis due to data quality constraints. However, participants still provided constructive insights into these factors.

6.6.3.1 Lack of Trust and Fear of Being Judged

Although trust did not ultimately result in a reliable impact analysis due to insufficient data quality, participants nevertheless highlighted the importance of trust in knowledge sharing. Here is a response to the question: “(Optional) If you feel it is not easy to share or get this type of knowledge, can you specify why? How do you want to improve it?”

"I think the biggest challenge is that it often requires a great deal of trust, understanding, and collaboration between us. It can be difficult for people to open up and share their knowledge and experience, especially if they feel that it could be used against them or taken advantage of…"

Crowdworkers share skill-based knowledge to gain a mutual competence at work, and therefore this behaviour can be seen as worker collaboration. In the above response, this participant emphasised the importance of trust and mutual understanding for cooperation like knowledge sharing. They would be reluctant to share their knowledge for fear that it would be used by others to take away their job opportunities. This is a reflection of individuals' distrust of others' intentions to use the knowledge they share, as well as their willingness to protect their technical advantage (Xie, Checco, et al., 2023). Here is another response to the question: "(Optional) If you feel it is not easy to share or get this type of knowledge, can you specify why? How do you want to improve it?"

"It is important to create a safe and supportive environment where people feel comfortable sharing their thoughts and ideas without fear of judgement."

In addition, participants mentioned another barrier to sharing skill-based knowledge, which was the fear of judgement from others. This falls under the motivation and attitude of individuals who do not trust judgement from others. Research was conducted to discuss the phenomenon of offensive comments that are prevalent in forums, and which directly increase members' resistance to sharing knowledge (Aroyo et al., 2019; Sood et al., 2012). Therefore, designers of KS tools need to think about how to create a supportive environment in the platform that leads to the release of friendly attitudes between individuals, which in turn reduces participants' distrust of the motives of evaluation from others.

In summary, trust can be divided into trust in others' intentions to use knowledge, and trust in the motivation of others' evaluations. The two trust factors need to be constructed separately in future measurement frameworks to obtain more accurate results. At the same time, the designers of the KS tools have considered how to motivate members to develop supportive evaluation motives, including penalties for malicious comments and rewards for friendly comments.

6.6.3.2 Is such knowledge harmful?

This factor looks at the attitudes of different social groups towards the act of skill-based knowledge sharing. The social groups include platform owners, third-party requesters who post HITs on the platform, crowdworkers who perform HITs on the platform, and even those who make laws related to online work. This study focuses on the attitudes of platform owners, requesters and workers towards skill-based knowledge sharing behaviours from the perspective of questionnaire participants. Firstly, workers interpreted why platforms would support or oppose the sharing of skill-based knowledge between workers. Here is a response to the question: "(Optional) In your opinion why do they believe so?" regarding the statement "The platform (MTurk, Prolific, Appen, etc.) believes that I should share this type of knowledge with other crowdworkers".

> *"Because it helps to reduce the number of complaints and solve technical issues through the platforms, so other workers can work on tasks better."*

Participants claimed that the platform should support the sharing of skill-based knowledge between workers, as this facilitates the successful completion of tasks, as well as the healthy growth of the platform. Here is another response to the same question:

> *"Mturk is complicated with what it allows workers to use in terms of extensions or apps. One has to be careful not to use a tool that violates its terms of service. As for Prolific, there is only one tool I know of, and Prolific itself advertises it as one they allow workers to use."*

Participants also found that the platform had strict terms and conditions restricting the use of third-party tools. In Chapter 4 of the thesis, the platform policies around MTurk have been explored. Specifically, MTurk prohibits the use of scripts that send requests at an excessively high frequency and those that automatically accept HITs (*Acceptable Use Policy*, 2018). This shows that Prolific and MTurk each have different attitudes towards workers' use of assistive tools. There is a large body of research on workers' attitudes towards platforms, but not enough attention has been paid to how the different attitudes of platforms towards assistive tools affect the attitudes of workers towards platforms. Furthermore, the balance between openness and restriction has been one of the challenges the platforms are facing (Wessel, 2017). On the one hand, the platforms need to restrict third-party tools so as to ensure their stable operation; on the other hand, the platforms need these tools to complement their limited functions (including optimisation of the interface, optimisation of search features) so as to improve the user experience.

In addition, participants reflected on the views of other workers in their perspective on sharing skill-based knowledge. Here is a response to the question "(Optional) In your opinion why do they believe so?" regarding the statement "Other crowdworkers believe I should share this type of knowledge with them."

> *"It depends, some (workers) are ok with using these tools, and some, those who don't like having to learn how to use these tools, consider it a way of cheating, but it's not cheating. It's only cheating if it does the work for you, which these tools don't do. they help filter out hits one wouldn't do, doesn't qualify for or doesn't want to do. They help save time by catching work so we can focus on completing the work."*

This response provides an interesting finding: some of the workers who are opposed to sharing skill-based knowledge see assistive tools as more like cheating and therefore resist sharing such knowledge and refuse to learn how to use them. Designers of KS tools should consider the resistance of some workers to knowledge of the tools and try to reduce resistance and misconceptions by showing workers the real purpose of assistive tools and the performance gains they can make, so that they can be more open to knowledge involved in using the tools.

6.7 Chapter Summary and Discussion

This chapter focuses on crowdworkers' behaviour in sharing skill-based knowledge using communication technologies. We examine the latent factors that influence their intentions and behaviours of knowledge sharing using PLS-SEM based on a conceptual model from UTAUT and SET. This leads to a more complete understanding of the cause of HIT catchers being popular among the crowd, which is the skill-based knowledge sharing. The enjoyment, satisfaction, and the information from others through communication promote their willingness to share knowledge. In addition, their effort and performance expectancy of communication tools further facilitate sharing behaviour. Despite the risk of losing their technical advantages and therefore job opportunities, most participants still tend to share knowledge and benefit from the thriving ecosystem of tools, including HIT catchers, that comes with knowledge sharing.

In specific, the theoretical models that have been widely used to study knowledge sharing behaviour was reviewed in the beginning. Then a research framework based on UTAUT and

SET was chosen to construct a research framework to study workers' knowledge sharing behaviour from the perspective of individual experiences of using sharing tools and social benefit exchange. After collecting the subjective ratings of the influencing factors from the worker groups through questionnaire, descriptive statistical analysis was used to understand the participants' social-demographic background, preferences and frequencies of knowledge sharing behaviours. Subsequently, the influence of each factor towards knowledge sharing intention and behaviour was assessed via partial least square structural equation modelling (PLS-SEM). The results showed that PE, EE and REW all significantly influenced workers' willingness and behaviour to share skill-based knowledge, with PE and REW influencing behaviour more indirectly, while EE had a direct impact on sharing behaviour.

It was revealed that a key motivation for workers to share their skill-related knowledge is the pleasure and satisfaction of sharing it with others. They keep a high belief in each other's willingness to share knowledge. Moreover, the effectiveness, promptness, ease of use and accessibility of such knowledge sharing technologies were particularly important for participants' willingness to share knowledge. In addition, whether the sharing tool was well integrated with their work strategies had a positive impact on crowdworkers' eventual knowledge sharing behaviour.

6.7.1 Theoretical Contribution

This study provides a valuable theoretical contribution to the modelling of influencing factors in the context of crowd knowledge sharing. It provides empirical data evidence on the latent factors (PE, EE and REW) influencing willingness and behaviour to share skill-based knowledge. Firstly, this is a preliminary study exploring the influence of factors related to technology use on the sharing intentions of crowdsourced knowledge contributors. The study found a significant indirect effect from performance expectancy (PE), effort expectancy (EE) and reward (REW) to sharing behaviour (KSB) through sharing intention (KSI).

In addition, a relatively significant direct effect of EE on KSB was found, and the direct effect was more than the indirect effect. Subsequently, based on the analysis, a new hypothesis about KSI was proposed: KSI is probably phased and categorical. Specifically, a worker's initial willingness to share may influence the role of exogenous constructs on later willingness to share and sharing behaviour. In addition, a completed KSI probably include an original intention that is unrelated to the use of the technology, as well as an intention to use the technology to achieve the original purpose. On this point, more in-depth research is yet to be carried out by segmenting the different stages of KSI.

This study also explores the non-technical reasons that prevent workers from sharing knowledge. The reasons include individuals' concerns about their loss of technical advantage, mistrust of unfamiliar members, and concerns about platform policies. Finally, we find that the popularity of scripting tools relies on knowledge sharing.

6.7.2 Implication for Practice

By understanding the reasons why crowdworkers adopt KS tools, design principles can be generated to encourage their sharing of skill-related knowledge. It can be revealed from the findings that knowledge exchange platforms such as forums and social apps should promote the ease of use and access (EE) and the benefits to users in their daily crowdwork (FC) to encourage them to participate in the platforms and share their skill knowledge.

For the developers and designers of KS tools, in addition to thinking about how to make the tools less difficult for workers to use and more accessible (e.g. cross-platform compatibility, integrating knowledge sharing features into micro-task platforms or popular plugins), they also need to think about how to increase the enjoyment, satisfaction and other positive feedback for crowdworkers helping others (e.g. the experience value progress bar commonly used in games, or better interaction features with peers). Combined with the feedback from participants in the qualitative analysis, the willingness of workers to share this type of knowledge tends to arise when someone asks a specific technical question. Worse still, this knowledge is often not systematically archived and is scattered throughout the forum in a fragmented form, making it difficult to create a knowledge system like an instruction manual for the group to target and contribute new content. As a result, knowledge is ultimately not shared as effectively as it could be. In the future, tools such as the GPT-3 will be needed to organise knowledge in a systematic way, thus facilitating the exchange of skill-related knowledge among crowdworkers.

In addition, the measurement model generated in this study can also be incorporated into the framework for evaluating the knowledge sharing effectiveness of the knowledge sharing tools, so that usability feedback can be obtained over multiple iterations of the development process. Encouraging users to actively share their knowledge is critical to the popularity of the KS tools, as only sufficient content contributions will continue to attract new members and thus keep the KS tools such as forum or channel maintained and growing.

What should not be overlooked is: the requirements for sharing skill-related knowledge are more demanding than the sharing of opportunities, evaluations and tutorials. Specifically, it requires that the knowledge provider not only has a good grasp of the skills required, but also that they are willing to take the time to post content on a topic that matches the content and accept the loss of technical advantage that would result from sharing it with others. And this loss of technical advantage is likely to diminish their future earnings. The effect of reward on willingness to share found in this study, alongside the qualitative analysis, suggest that workers may be willing to share in order to exchange skills with others, thereby increasing the technical advantage of the whole collective by means of knowledge exchange.

In terms of how to improve knowledge sharing and access, creating a comprehensive knowledge system for skills-related knowledge could be a one-stop solution for workers to easily search, access, and share knowledge. This would link information that is currently scattered across various topics, providing a more efficient and organised way for workers to find the information they need. Such a system would serve as a powerful tool for workers, who could quickly find the information they need to stay up to date on the latest skills and knowledge in their field. Moreover, this system could also be used to share new knowledge and discoveries with other workers, creating a network of knowledge-sharing that would benefit everyone.

6.7.3 Limitations

1. The objectives of this study were limited to skill-based knowledge, so the influences on knowledge sharing intentions and behaviour revealed in the analysis do not necessarily apply to other types of knowledge.

2. The frequency of knowledge sharing behaviour of individual participants in this study was collected through a questionnaire, which may not accurately reflect reality and relied on subjective perceptions. Future research could explore the use of scripts to automatically (or with the assistance of participants) record the number of times they actually shared knowledge over a fixed observation period, in order to obtain more accurate scores for the behavioural factors.

3. There was a lack of reliability and validity in the data for factors including Social Influence, Reciprocity, Reputation, Social Interaction Ties, and Trust. Therefore, these factors were not successfully incorporated into the structural model for impact analysis.

4. With the development of text generation tools such as ChatGPT, short answers from participants have been found to contain responses that are not relevant to the crowdwork. It is worth further exploration to ensure the originality and authenticity of textual responses from participants (Guo et al., 2023).

5. This research primarily focused on English-speaking crowdworkers, which may not be fully representative of the entire crowd community, as different linguistic and cultural backgrounds can potentially influence crowdworking dynamics. Future studies might consider a more diverse participant pool to enhance the generalisability of findings.

6. The data for this study were collected from participants who were available to respond to the published HITs at the time of data collection. This sampling approach may introduce a bias, as those who were available at that specific time might not fully represent the entire population of crowdworkers on MTurk.

Chapter 7 Conclusion

7.1 Introduction

The aim of this thesis was to assess the impact of the use of HIT catchers on job opportunities, work strategies, and quality of outcomes for crowdworker groups, as well as to explore the factors that facilitate the sharing of skill-based knowledge among crowdworkers. This chapter discusses the main findings of the three studies included in this thesis in relation to the literature, aiming to showcase the thesis' contributions.

7.2 Findings

In this thesis, we focus on two types of crowd collective behaviours: the use of HIT catchers and knowledge sharing. Specifically, the behaviour of HIT catcher use and the impact it causes on the microtask ecosystem were examined. Furthermore, the factors influencing the skill-based knowledge sharing behaviours that lead to the popularity of scripting tools were studied. Next, each research question is answered in the context of the findings from each study.

7.2.1 RQ1: What are the impacts of the use of HIT catchers?

The impact of using HIT catchers were found as follows (as is described in Section 4.5):

- The completion of the HIT group was delayed.
- The fairness of the HIT distribution was reduced.
- Long waiting times for other participants, and fewer job opportunities for themselves due to HIT abandonment.
- While it benefits users in the short term, it was observed that a tragedy of the commons might likely occur for all platform members.
- Under the Matthew Effect, short-term unequal job opportunities potentially lead to long-term negative impacts on newcomers and those not using HIT catchers, which in turn undermines the sustainability of the platform. This is further discussed in Section 4.7.
- Subsequently, by including an assessment of data quality, participants using HIT catchers were found to have more similar textual responses, but little difference on the quality of image annotation compared to those not using HIT catchers. This implies that

participants using HIT catchers tended to apply the same textual response to many similar microtasks, thus reducing work time. This is further discussed in Section 5.5.3.

7.2.2 RQ2: What drives the crowd skill-based knowledge sharing within the communities?

- As revealed in Section 6.5.4, the ease of access, use, and learn communication tools, as represented by Effort Expectancy (EE), directly influences participants' skill-based knowledge sharing intention (KSI). This factor also directly affects their knowledge sharing behaviour (KSB).
- The expectation of whether the use of communication tools will enhance knowledge sharing, as represented by Performance Expectancy (PE), affects participants' skill-based KSI directly.
- The pleasure and satisfaction from helping others, as represented by Reward (REW), affects participants' skill-based KSI. Moreover, KSI shows the largest total effect to participants' KSB among all observed factors, followed by REW.
- The sharing of skill-based knowledge has significantly contributed to the popularity of scripting tools including HIT catcher.

7.3 Contributions

Crowdworkers' use of scripting tools and crowd knowledge sharing as two collective behaviours are receiving increasing attention because of their impact on the functioning of microtask platforms and on crowd working conditions. This thesis contributes to three main aspects: (1) the impact of using HIT catchers, (2) crowdwork behaviour and assessment of result quality, and (3) crowd skill-based knowledge sharing:

1. This thesis reveals how the reputation system of microtask platforms contributes to the Matthew effect, whereby those who use HIT catchers can gain at the expense of job opportunities of others not using tools (Section 4.2). The algorithmic control of the current platform, based on reputation system, could lead to the following: as new workers struggle to improve their reputation scores, they may choose to leave the platform due to lack of job opportunities. Long-term reliance on workers with established reputations may lead to a lack of diversity of data source on the platform. Finally, this could trigger a sustainability crisis for the platform (Section 4.5).

2. This thesis provides new perspectives and contributions to the detection methods of worker behaviour. Based on the technique of Application Layer Monitoring (ALM), we incorporate event data of HIT state changes, which makes the exploration of worker behaviour no longer limited to the HIT completion phase (Section 5.2.5). We further investigate worker behaviour in non-completion phases such as HIT acceptance and backlog, which provides a more comprehensive understanding of crowdwork strategies.
3. This thesis explores the quality assessment of image annotation based on behavioural metrics including installation status of HIT catchers and HIT backlog, providing new perspectives and methods for behaviour-based quality assessment (Section 5.3.7). By developing a SVC prediction model, it was found that whether the use of HIT catchers and the backlog time for the task were detected was not a valid predictor for assessing the image annotation quality. In addition, time spent focused on HITs was a significant positive indicator and time spent being distracted was a significant negative indicator.
4. This thesis extends our understanding of the impact of the use of the HIT catchers to microtasks and crowdwork behaviours. This includes a graphical presentation of the HIT state dynamics from publication to completion of the whole HIT group. It is also revealed quantitatively that workers using the HIT catcher were observed to have more attention switching and less focus time (Section 5.3), as well as more HIT backlogging and multi-HITing behaviour. In addition, the wide use of HIT catcher had a direct negative impact on the data collection from requesters. This includes occupying a large number of task opportunities in the early stages of HIT group publication, making the completion time of the entire HIT group significantly longer, and decreasing the data diversity of text generation tasks.
5. This thesis not only confirms the prevalence of multi-device use based on the browser and device information of the participated, but also finds that the behaviour of multi-device use may be potentially associated with multi-HITing and irregular account sharing behaviours (Section 5.3.4.5). We provide a more precise description of such behaviour and its proportion among the participants, which helps to reveal the potential issues from their behavioural patterns and provides useful information for improving the management and monitoring of crowdwork behaviours.
6. This thesis reveals the factors affecting the crowd skill-based knowledge sharing using PLS-SEM (Section 6.5.4). By applying the theoretical framework formed by UTAUT and SET, the study provides a more complete perspective for understanding crowd knowledge sharing behaviour. Specifically, Performance Expectancy (the expectation

of whether the use of communication tools will enhance knowledge sharing), Effort Expectancy (the ease of access, use, and learn communication tools) and Reward (pleasure and satisfaction) are revealed to influences the crowdworkers' Knowledge Sharing Intention (KSI), while EE also directly influences Knowledge Sharing Behaviour (KSB). Among these three exogenous factors, Reward (pleasure and satisfaction of helping) caused the highest total impact on behaviour, reflecting the importance of the community's culture of mutual support and altruism for knowledge sharing. In addition, the study emphasises the critical role of communication tools in the knowledge sharing process. This helps crowdsourcing community administrators and tool developers to better understand how to improve the design and functionality of sharing tools to facilitate knowledge sharing.

7.4 Discussion

Microtask crowdworkers rely on a variety of tools to improve their work experience and engage in community knowledge sharing. However, the widespread use of HIT catchers has created challenges and impacts on platform stability and crowdwork behaviour. This study aims to delve into these influences and the implications of these findings for improving the microtask platform and crowd community.

7.4.1 Research Implication of HIT Catcher Impact

This study begins by revealing the multiple impacts of HIT catchers. Earlier studies explored the causes affecting crowd working conditions mainly in terms of the absence of a regulatory framework (Gerber, 2021; Altenried, 2020) and the unfair distribution of power due to platform design (Fieseler et al., 2019; Irani and Silberman, 2013). This study quantifies the impact of the use of HIT catchers on platform members in short term.

Specifically, while job opportunities for HIT catcher users increase, the HIT-worker diversity and job opportunities for manual workers are significantly limited. Furthermore, current microtask platforms rely heavily on number of HIT completions and HIT approval rate as a reputation system for assessing worker quality. The purpose of this system is to help requesters filter out experienced workers with high quality work. However, this assessment appears unfair to new workers and those who do not use automated scripts. Although they may lack enough HIT completions, this does not mean that their work is of lower quality.

In addition, we explore the reasons for the current high turnover rate on crowdsourcing platforms from the algorithm control perspective in the context of the Matthew Effect. Reputation systems based on HIT completion and approval rates are vulnerable to the Matthew Effect, and workers who do not have the technical skills to use these scripts may find themselves at a competitive disadvantage. This imbalance may not only lead to social polarisation on the platform, but may also increase the gap between experienced and novice workers. Long-term relying on experienced workers with established reputations may lead to a lack of diversity and innovation of the output. Furthermore, this could potentially lead to a sustainability issue for the crowdsourcing platform: on the one hand, there is a continuous loss of new workers (El Maarry et al., 2018). On the other hand, job requesters could gradually abandon the platform due to the result bias caused by excessive number of "professional participants" (Conte et al., 2019; Hauser et al., 2018) (Section 4.5). To create a crowdwork environment with better fairness and sustainability, platforms need to take steps to ensure that new workers unfamiliar with assistive tools have a fair chance to build their reputations. Moreover, they need to consider introducing other quality control mechanisms including behavioural oriented assessment.

We also find that the over-acceptance of HIT catchers led to a large backlog of microtasks during the experiment, which in turn caused the completion process to be slowed down for the entire HIT group. This emphasises the need for the platform to regulate and control scripting tools to ensure the HIT group completions are not slowed down and to maintain the motivation of job requesters.

7.4.2 Behaviour-oriented Quality Assessment

7.4.2.1 Work Behaviour and Data Quality

Current research on crowdwork behaviour focuses on the analysis of microtask completion process, with metrics including mouse behaviour, task completion time, and page switching (Al-Qershi et al., 2021; Mok et al., 2016). This study explored microtask acceptance and backlog behaviour through logs provided by Amazon SQS, extending the scope and approach of crowdwork behaviour research.

This thesis complements the quality assessment metrics for image annotation tasks: in previous research, multidimensional vectors consisting of specific events and timestamps were used as machine learning parameters to predict whether the question about the relevance of the document's theme was correctly answered (Al-Qershi et al., 2021). In addition, task completion and consideration time were also used to predict result quality of text comprehension tasks (Hirth et al., 2014). This thesis reveals the validity of focus time, non-focus time, completion time, and the number of page switches in assessing whether the quality of image annotation is qualified or not by constructing an SVC prediction model.

It was revealed that the HIT backlog time and the installation of HIT were not suitable metrics for quality prediction. In addition, although the use of HIT catchers was not significantly correlated with image annotation quality, it was associated with lower textual response diversity. The study also found that HIT backlog time increased due to workers' use of HIT catchers. These findings help platforms to better understand the crowdwork behaviour under the influence of tools and thus develop more effective quality management strategies.

7.4.2.2 Scripting Tools Influence Crowdwork Strategies

The use of scripting tools is crucial for improving workers' earnings. However, this has also caused an impact on their traditional work strategies, such as the need for an individual to complete many similar microtasks in a limited time after using HIT catchers. Workers are increasingly relying on scripting tools to help them find quality tasks, automate specific tasks, and streamline workflows, therefore to improve efficiency and maximise income (Williams et al., 2019). With the rise of large language models (LLMs), many tasks that require human creativity are being replaced by automated tools (Veselovsky et al., 2023).

Williams et al. (2019) found through interviews that task-switching and multitasking behaviours promoted crowdwork fragmentation. This thesis found that workers' use of HIT catchers led to a rise in the number of attention switches and a fall in concentration time. This demonstrates quantitatively that crowdwork is interrupted more when using HIT catchers (Section 5.3).

Williams et al. (2019) and Gupta et al. (2014) found through interviews in their studies that workers use multiple monitors, or multiple PC devices, with the aid of mobile devices to assist in their crowdwork. This thesis reveals through empirical data the proportion of workers using multi-browsers and mobile devices in the experiment and found that workers using multiple browsers significantly submitted more HITs. We provide insights into how to improve the platform's HIT publishing and completion methods to better meet workers' needs. Researchers could further investigate how to optimise multi-device workflows and whether this has a long-term impact on workers' earnings and satisfaction.

Anti-fraud techniques for crowdwork were mentioned in previous research, including the use of browser fingerprints to detect the same person using multiple accounts, and the use of VPNs to bypass location restrictions (Zhang et al., 2022). However, there is a lack of systematic evaluation of these methods applied on microtasks. This thesis applied these two detection techniques to the microtask domain, thereby quantifying the proportion of workers logging into the same account from multiple addresses in a single experiment, as well as the proportion of workers logging into multiple worker accounts sequentially using the same IP address. These findings could help platforms and requesters to better understand the extent of fraudulent behaviour and how these techniques can be used to identify potential frauds, thus enhancing data credibility.

Prior research began an exploration of the impact of scripting tools, including on HIT market speed and job opportunities (Hanrahan et al., 2018; Williams et al., 2019). This study expands on this theme further: competition for microtasks has grown more intense due to the popularity of automated tools, and this has changed workers' task acceptance strategies: for sought-after HITs in our experiment, workers can skip previewing or use tools to accept the task, which has led to the traditional preview-and-accept-task approach becoming less efficient. In other words, those workers who still preview before accepting tasks may lose job opportunities. It is possible that in the future, workers gradually abandon task previewing to increase the success rate of task acceptance. And this could potentially lead to new problems, such as massive task abandonments due to reserving unsuitable tasks. This challenges the platform designers to consider how to treat the impact of the use of automated tools on the basic functionality of the platforms.

7.4.3 Knowledge Sharing as a Collective Behaviour

7.4.3.1 Knowledge Sharing Leads to Diffusion of Tools

As previous studies have pointed out, the flourishing of the ecosystem of scripting tools is closely linked to knowledge sharing in worker communities (El Maarry et al., 2018; Williams et al., 2019). As another crowd collective behaviour, crowdworkers share job skills, including tools, in their communities (Hanrahan et al., 2015; Irani & Silberman, 2013; Williams et al., 2019). However, not all participants used the HIT catchers, which sparked our interest in knowledge sharing among workers. In addition, higher-income workers are more inclined to use multiple tools and actively participate in the community, further emphasising the link between knowledge exchange and technical advantage (Kaplan et al., 2018).

7.4.3.2 Knowledge Sharing Under Use of Technology and Social Exchange

By combining UTAUT and SET, this thesis provides a comprehensive understanding of skill-based knowledge sharing behaviours from the perspectives of technology use and social exchange. UTAUT provides a theoretical framework on technology adoption, while SET emphasises the role of social exchange and incentives. This promotes theory building in the field of online knowledge sharing. With a more complete factor analysis model, we help online communities or communication tools better design incentives and management strategies to improve the efficiency and quality of online knowledge sharing.

Effort expectancy influences participants' willingness to share skill-based knowledge. This factor also directly affects knowledge sharing behaviour, which is a refinement of the existing UTAUT model in the field of microtask knowledge sharing. This illustrates how the ease of access, use, and learn communication tools can directly affect the final behaviour, not just the willingness to share.

The satisfaction of helping others was found to be a factor that motivate crowdworkers to provide online feedback (Osterbrink & Alpar, 2021). This thesis extends this finding and reveals that the reward factor, which represents satisfaction and happiness, has a greater effect on skill-based knowledge sharing behaviour among workers than effort expectancy and performance expectancy. It implies that workers are more inclined to share their knowledge even if it requires extra effort or the actual sharing is not as effective as it could be. This finding helps platform and tool designers to increase incentives in terms of satisfaction and enjoyment for knowledge contributors. Therefore, to build positive social exchange relationships and promote knowledge exchange in crowdsourcing communities.

7.4.3.3 Impact of Reduced Technical Advantage to Knowledge Sharing

It is revealed that the loss of technical advantage does not significantly affect workers' willingness to contribute knowledge, which is in line with the finding from a previous study (Osterbrink & Alpar, 2021). This reflects the importance of altruism in the worker community in sharing knowledge and promoting community development. Although the balance between protection of technical advantage and altruism is not explicitly mentioned in this study, it can be speculated that workers are willing to share knowledge despite the loss of technical advantage that may affect their income, possibly because they value more on the overall development of the community and altruism.

7.4.4 Summary

This subsection explores the implications of the research findings for improving the functionality of the platform and developing the crowdsourcing ecosystem. We first explore the impact of scripting tools on microtask platforms, revealing the limitations of current reputation systems and the importance of behavioural traces in assessing data quality. Meanwhile, we emphasise the importance of knowledge sharing as a collective behaviour in worker communities. To build a more fair and sustainable crowdsourcing ecosystem, platforms, requesters and workers need to reflect and adapt their strategies of working and quality control. Future explorations are needed on how to better balance the constraints of platforms with the needs of crowdworkers.

7.5 Limitations

Although this study successfully answers the initial research questions, there are limitations to each specific study and the overall thesis research process.

7.5.1 Lack of Diverse Work Strategies in Simulation Model

While our study provides new insights and important contributions with regards to crowdworking, it comes with limitations, primarily due to its exploratory nature. First, we designed our study around a simulation framework, where the data we used derived through the tests on the MTurk Developer Sandbox. This means that, while we are confident with regards to our measurements, it must be made clear that the data are still experimental.

Furthermore, for crowdsourcing platforms like MTurk, the workers are continuously changing. The two main types of workers - those who use scripts and those who do not - employ a variety of strategies in identifying and completing microtasks. Their strategies may evolve over time as they become familiar with the platform and the use of scripts. Therefore, future research should explore the behavioural patterns and strategies of workers in greater depth and systematically, with a view to more accurately assessing ongoing changes in the competence and reputational persistence of both types of workers. This will allow us to improve the current simulation framework but most importantly, it will allow studying the Matthew effects over a longer period.

Finally, to measure for inequalities, we used the Gini coefficient of inequality. The Gini coefficient is widely used in the field of economics (Cowell, 2011) for estimating inequalities in income distribution or better put, in the wealth distribution among a given population. However, it has been criticised as providing potentially misleading results because it is not as sensitive “at the extremes”, i.e., between the very rich and the very poor (Cobham et al., 2016), who, in our study would be those with the highest reputation scores and income and those with the lowest. Therefore, future studies interested in examining inequality and working conditions in crowdsourcing platforms can combine the Gini coefficient with the Palma ratio, which we were unable to do due to the size of our simulation.

7.5.2 Limited Methods of Detecting the Use of Tools

There was a relatively low worker sample size in this study, especially when restricting to Google Chrome workers, the only ones for which the use of HIT catching tools and browser tab switching was detectable. The HIT catching plugin detection ability was incomplete, as workers could use other scripts that cannot be detected. Moreover, the fact of having such a script installed does not mean that the worker will always use it. One possible improvement would be to ask participants in the microtask if they used a specific scripting tool.

While a reduction on the overall quality of work and time spent on a task for workers that use HIT catching plugins was observed, there is still a lack of understanding on potential confounding or mediating variables (such as platform experience) affecting the behavioural difference of workers. In other words, workers with more experience in image annotation tasks may spend less time providing higher quality data regardless of their use of HIT catchers.

7.5.3 Missing Number of HIT Acceptance for Each Participant

As the data collected from SQS on HIT events did not contain *worker_id*, this made it impossible to calculate the number of successful HITs accepted by different workers simply from HIT event records. While it is still possible to study the number of HITs successfully accepted and submitted by workers with and without detected plugins in successfully submitted HITs on the basis of the available data, there seems to be no way to know the number of HITs successfully accepted but not submitted by each worker.

First, this data can make a more convincing demonstration for unequal job opportunities for workers. Second, by comparing the number of HITs abandoned as well as returned by both types of workers, it could be revealed whether workers who use HIT catchers are more likely to abandon their accepted HITs. More importantly, this can help to recognise the extent to which workers' over-acceptance of HITs with HIT catchers has an impact on the completion process of HIT groups.

7.5.4 Lack of Non-survey Methods to Assess Knowledge Sharing Behaviour

The frequency of knowledge sharing behaviour of individual participants in this study was collected through a questionnaire. Like all survey-based methods, we can only measure and assess perceptions, which are by default subjective.

7.5.5 Insufficient Data Reliability and Validity in Some Influencing Factors

While every effort was made to examine the impacts of factors, not all of them were retained in the structural model due to reliability and validity reasons. Specifically, Reciprocity (REC), Social Influence (SI), and Social Interaction Ties (SIT), Reputation (REP) and Trust (T) were removed due to extremely low average variance explained (AVE), composite reliabilities (CR) score or Cronbach's α.

7.5.6 Limited Methods for Quality Check

Ultimately, new mechanisms need to be introduced to ensure the quality of the data, including the use of page script to check the amount of time workers spend focused on the questionnaire pages. Specifically, the total time spent by the worker in completing the questionnaire can be measured by planting a script on the questionnaire page to record the actual time of completion. Based on the estimated completion time, questionnaire results that take insufficient time to complete can be detected as low-quality results. It is also important to introduce the detection of questionnaire auto-completion behaviour. This can be achieved by recording and analysing the cursor trajectory and the text input process of each participant.

In addition, with the development of text generation tools using Large Language Models (LLMs), the short answer questions of the questionnaire have been found to contain responses that are partially irrelevant to the context of the study. How to ensure the originality and authenticity of such textual content is becoming a popular research topic (Guo et al., 2023).

7.6 Future Research

Despite its limitations, this thesis still makes a rich contribution to research in the field of crowdsourcing microtask. Based on the focused objectives as well as the limitations of this study, a set of future research directions has been generated.

7.6.1 Detection and Impact Assessment of Scripting Tools

Future research should continue to expand the exploration of work behaviour and the detection of automation tools including those for automated task completion. This is because such tools have been questioned by requesters due to their potential impact on data quality on the one hand, and are widely used by groups of workers due to the significant increase in worker completion efficiency on the other.

Therefore, a direction worth investigating is the impact that workers' use of auto completion tools has on the quality of HIT results. As an important component of automation tools, auto completion tools have greatly increased the speed of task completion for workers, while also damaging the quality of results. The detection of auto completion tools (monitoring text input behaviour) and the assessment of the validity/quality of the resulting content allow for an understanding of the specific impact that the use of such tools has on workers and on HIT results.

7.6.2 Further Research on Microtask Work Strategies

Future research should further investigate work strategies via empirical studies, so that they can be more realistically reproduced in simulations. Specifically, workers' behaviour towards tool use includes a variety of task acceptance strategies, rest strategies and completion strategies. By studying these strategies, the crowd tool use behaviour can be categorised and the proportion of different types of workers can be assessed. By incorporating more details into the simulation model, simulation results could become more realistic, thus helping the requester and platform to assess the overall impact of work strategies more accurately on the data quality and bias.

Further research is also needed into the causes of changes in time spent for HIT completion and attention switch frequency: factors influencing these changes may come from growth in job skills or changes in work attitudes due to increased fatigue. Research into the dynamics of changes in such behaviour traces will not only help us to understand workers' strategies for using tools, but also provide more perspectives on the causes of low data quality. These could potentially help requesters to obtain more reliable predictions when assessing the data quality. It could also bring insights to help platforms and requesters enhance data quality by improving microtask workflow.

7.6.3 Segmentation of Knowledge Sharing Intention (KSI)

As discussed in Section 6.6.2, our study found that participants' willingness to share knowledge may be multi-stage. In other words, KSI may consist of an initial willingness to share, as well as a later willingness to share that incorporates consideration of technical difficulty. Workers may only consider the technical resistance they may face during their behaviour before developing a later stage of willingness to share, and subjective assessments of resistance may in turn affect initial willingness to share.

Future research could attempt to separate the KSI into initial and later behavioural intentions when constructing measurement models. This can reveal changes in workers' willingness when facing technical difficulties and provide a basis for designing more effective knowledge sharing tools. Meanwhile, this segmentation can also provide a clearer direction for future research, which can help construct a more accurate model for measuring influential factors, thus promoting the development of knowledge sharing research.

7.7 Reflections from PhD

During this journey, I have improved my skills in reflecting on the literature, formulating a research proposal and conducting research, especially in the acquisition, cleaning and organisation of primary data. More importantly, I learned how to deconstruct an unfamiliar and complex concept, which in turn helped to formulate the research steps in an organised manner.

During my PhD, I carried out and participated in research in crowdsourcing collaborations, gamified education, and PhD related topics. Therefore, one of the biggest challenges for me was how to make trade-offs between multiple research topics and personal interests.

In hindsight, when I was in doubt about a research direction, I should be more patient in exploring and listening to perspectives from many people and less limited by my own preconceptions.

In summary, this research journey has been both challenging and rewarding. Beyond academic progress, it has been a journey of personal growth, resilience, and perseverance. More importantly, it makes me start enjoying the process of exploring, making complex problems simple, and reshaping knowledge constantly.

Bibliography

Abbas, T., & Gadiraju, U. (2022). Goal-Setting Behavior of Workers on Crowdsourcing Platforms: An Exploratory Study on MTurk and Prolific. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, *10*(1), 2–13. https://doi.org/10.1609/hcomp.v10i1.21983

Abdou, A. H., Hassan, T. H., Salem, A. E., Albakhit, A. I., Almakhayitah, M. Y., & Salama, W. (2022). The Nexus between Environmentally Sustainable Practices, Green Satisfaction, and Customer Citizenship Behavior in Eco-Friendly Hotels: Social Exchange Theory Perspective. *Sustainability*, *14*(19), 12791. https://doi.org/10.3390/su141912791

Acceptable Use Policy. (2018, May 17). Amazon Mechanical Turk. https://www.mturk.com/worker/acceptable-use-policy

Adamseged, H., & Hong, J. (2018). Knowledge sharing among university faculty members. *Journal of Education and Practice*, *9*(24), 1–10.

Adnan, M. M., Rahim, M. S. M., Rehman, A., Mehmood, Z., Saba, T., & Naqvi, R. A. (2021). Automatic image annotation based on deep learning models: A systematic review and future challenges. *IEEE Access*, *9*, 50253–50264.

Ahuja, A. (2020). A study on the impact of rewards on knowledge sharing in education sector. *Studies in Indian Place Names, UGC Care Journal*, *40*(27).

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211.

Ajzen, I. (2020). The theory of planned behavior: Frequently asked questions. *Human Behavior and Emerging Technologies*, *2*(4), 314–324.

Alalawneh, A. A., & Alkhatib, S. F. (2021). The barriers to big data adoption in developing economies. *The Electronic Journal of Information Systems in Developing Countries*, *87*(1), e12151.

Alamir, P., & Navimipour, N. J. (2016). Trust evaluation between users of social networks using the quality of service requirements and call log histories. *Kybernetes*.

Al-Husseini, S., & Elbeltagi, I. (2018). The role of knowledge sharing in enhancing innovation: A comparative study of public and private higher education institutions in Iraq. *Innovations in Education and Teaching International*, *55*(1), 23–33.

Ali, F., Rasoolimanesh, S. M., Sarstedt, M., Ringle, C. M., & Ryu, K. (2018). An assessment of the use of partial least squares structural equation modeling (PLS-SEM) in hospitality research. *International Journal of Contemporary Hospitality Management*.

Alibrahim, H., & Ludwig, S. A. (2021). Hyperparameter Optimization: Comparing Genetic Algorithm against Grid Search and Bayesian Optimization. *2021 IEEE Congress on Evolutionary Computation (CEC)*, 1551–1559. https://doi.org/10.1109/CEC45853.2021.9504761

Almuqrin, A. H. (2022). Social exchange theory and theory of reasoned action affecting knowledge sharing: A case from Saudi Arabia. *Journal of Information Studies & Technology (JIS&T)*, *2022*(1), 5. https://doi.org/10.5339/jist.2022.5

Al-Qershi, F., Al-Qurishi, M., Aksoy, M., Faisal, M., & Algabri, M. (2021). A Time-Series-Based New Behavior Trace Model for Crowd Workers That Ensures Quality Annotation. *Sensors*, *21*(15), 5007. https://doi.org/10.3390/s21155007

Altenried, M. (2020). The platform as factory: Crowdwork and the hidden labour behind artificial intelligence. *Capital & Class*, *44*(2), 145–158.

Alwahdani, A. (2019). The Impact of Trust and Reciprocity on Knowledge Exchange: A Case Study in IT Outsourcing. *Journal of Information Systems Engineering & Management*, *4*(1). https://doi.org/10.29333/jisem/5738

Ambati, V., Vogel, S., & Carbonell, J. (2011). *Towards Task Recommendation in Micro-Task Markets*. 4.

Anderson, B. A., Laurent, P. A., & Yantis, S. (2013). Reward predictions bias attentional selection. *Frontiers in Human Neuroscience*, *7*, 262.

Antonelli, C., & Crespi, F. (2013). Technological Forecasting & Social Change. *Technological Forecasting*, 12.

Araman, V. F., Calmon, A., & Fridgeirsdottir, K. (2019). *Pricing and Job Allocation in Online Labor Platforms*.

Aringhieri, R. (2010). An integrated DE and AB simulation model for EMS management. *2010 IEEE Workshop on Health Care Management (WHCM)*, 1–6. https://doi.org/10.1109/WHCM.2010.5441260

Arndt, A. D., Ford, J. B., Babin, B. J., & Luong, V. (2022). Collecting samples from online services: How to use screeners to improve data quality. *International Journal of Research in Marketing*, *39*(1), 117–133. https://doi.org/10.1016/j.ijresmar.2021.05.001

Aroyo, L., Dixon, L., Thain, N., Redfield, O., & Rosen, R. (2019). Crowdsourcing Subjective Tasks: The Case Study of Understanding Toxicity in Online Discussions. *Companion Proceedings of The 2019 World Wide Web Conference*, 1100–1105. https://doi.org/10.1145/3308560.3317083

Aruguete, M. S., Huynh, H., Browne, B. L., Jurs, B., Flint, E., & McCutcheon, L. E. (2019). How serious is the 'carelessness' problem on Mechanical Turk? *International Journal of Social Research Methodology*, *22*(5), 441–449. https://doi.org/10.1080/13645579.2018.1563966

Asratian, A. S., Denley, T. M., & Häggkvist, R. (1998). *Bipartite graphs and their applications* (Vol. 131). Cambridge university press.

Assegaff, S., Hussin, A. R. C., & Dahlan, H. M. (2011). Perceived benefit of knowledge sharing: Adapting TAM model. *2011 International Conference on Research and Innovation in Information Systems*, 1–6. https://doi.org/10.1109/ICRIIS.2011.6125744

Attuquayefio, S., & Addo, H. (2014). Review of studies with UTAUT as conceptual framework. *European Scientific Journal*, *10*(8).

Axtell, R. (2000). *Why agents?: On the varied motivations for agent computing in the social sciences*.

Azzeh, M., Elsheikh, Y., Nassif, A. B., & Angelis, L. (2023). Examining the performance of kernel methods for software defect prediction based on support vector machine. *Science of Computer Programming*, *226*, 102916. https://doi.org/10.1016/j.scico.2022.102916

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *16*, 74–94.

Banks, J., CARSON II, J. S., Barry, L., & others. (2005). *Discrete-event system simulationfourth edition*. Pearson.

Baptista, R., & Neves-Silva, R. (2021). Agent-based simulation of consumer occupancy distribution in shopping centers. *2021 International Young Engineers Forum (YEF-ECE)*, 92–97. https://doi.org/10.1109/YEF-ECE52297.2021.9505136

Barbosa, C., & Azevedo, A. (2017). Hybrid Simulation for Complex Manufacturing Value-chain Environments. *Procedia Manufacturing*, *11*, 1404–1412. https://doi.org/10.1016/j.promfg.2017.07.270

Bard, A., Söderqvist, J., & Larsson, A. (2019). Behind the history of labor: Technology as the driving force 1. In *The Digital Transformation of Labor (Open Access)* (pp. 15–30). Routledge.

Bates, J., Gerakopoulou, E., & Checco, A. (2023). Addressing labour exploitation in the data science pipeline: Views of precarious US-based crowdworkers on adversarial and co-operative interventions. *Journal of Information, Communication and Ethics in Society*.
Bauer, B., Larsen, K. L., Caulfield, N., Elder, D., Jordan, S., & Capron, D. (2020). *Review of best practice recommendations for ensuring high quality data with amazon's mechanical turk*.
Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588–606. https://doi.org/10.1037/0033-2909.88.3.588
Bergin, D. L. (2015). Cyber-attack and defense simulation framework. *The Journal of Defense Modeling and Simulation*, *12*(4), 383–392. https://doi.org/10.1177/1548512915593528
Bergvall-Kåreborn, B., & Howcroft, D. (2014). Amazon Mechanical Turk and the commodification of labour: Amazon Mechanical Turk. *New Technology, Work and Employment*, *29*(3), 213–223. https://doi.org/10.1111/ntwe.12038
Bernstein, M. S., Little, G., Miller, R. C., Hartmann, B., Ackerman, M. S., Karger, D. R., Crowell, D., & Panovich, K. (2015). Soylent: A word processor with a crowd inside. *Communications of the ACM*, *58*(8), 85–94. https://doi.org/10.1145/2791285
Blaising, A., & Dabbish, L. (2022). Managing the Transition to Online Freelance Platforms: Self-Directed Socialization. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CSCW2), 1–26. https://doi.org/10.1145/3555201
Bock, G.-W., Zmud, R. W., Kim, Y.-G., & Lee, J.-N. (2005). Behavioral intention formation in knowledge sharing: Examining the roles of extrinsic motivators, social-psychological forces, and organizational climate. *MIS Quarterly*, 87–111.
Bohannon, J. (2016). Mechanical Turk upends social sciences. *Science*, *352*(6291), 1263–1264. https://doi.org/10.1126/science.352.6291.1263
Bollen, K. A., & Long, J. S. (Eds.). (1993). *Testing structural equation models*. Sage Publications.
Bornschein, J., Visin, F., & Osindero, S. (2020). Small data, big decisions: Model selection in the small-data regime. *International Conference on Machine Learning*, 1035–1044.
Borromeo, R. M., Laurent, T., Toyama, M., & Amer-Yahia, S. (2017). *Fairness and Transparency in Crowdsourcing* [dataset]. OpenProceedings.org. https://doi.org/10.5441/002/EDBT.2017.46
Borshchev, A. (2013). *The Big Book of Simulation Modeling: Multimethod Modeling with AnyLogic 6*. AnyLogic North America. https://books.google.co.uk/books?id=2c5FnwEACAAJ
Borst, I. (2010). *Understanding Crowdsourcing: Effects of motivation and rewards on participation and performance in voluntary online activities*.
Bosnjak, M., Ajzen, I., & Schmidt, P. (2020). The theory of planned behavior: Selected recent advances and applications. *Europe's Journal of Psychology*, *16*(3), 352–356. https://doi.org/10.5964/ejop.v16i3.3107
Bouarfa, S., Blom, H. A., Curran, R., & Everdij, M. H. (2013). Agent-based modeling and simulation of emergent behavior in air transportation. *Complex Adaptive Systems Modeling*, *1*, 1–26.
Boudreau, K. (2010). Open platform strategies and innovation: Granting access vs. Devolving control. *Management Science*, *56*(10), 1849–1872.
Bourque, L. B., & Shen, H. (2005). Psychometric characteristics of Spanish and English versions of the Civilian Mississippi Scale. *Journal of Traumatic Stress: Official Publication of The International Society for Traumatic Stress Studies*, *18*(6), 719–728.
Brailsford, S. C., & Hilton, N. A. (2001). *A comparison of discrete event simulation and system dynamics for modelling health care systems*.

Brailsford, S. C., Sykes, J., & Harper, P. R. (2006). Incorporating human behavior in healthcare simulation models. *Proceedings of the 2006 Winter Simulation Conference*, 466–472.

Brandon, D. M., Long, J. H., Loraas, T. M., Mueller-Phillips, J., & Vansant, B. (2014). Online instrument delivery and participant recruitment services: Emerging opportunities for behavioral accounting research. *Behavioral Research in Accounting*, *26*(1), 1–23.

Brawley, A. M., & Pury, C. L. S. (2016). Work experiences on MTurk: Job satisfaction, turnover, and information sharing. *Computers in Human Behavior*, *54*, 531–546. https://doi.org/10.1016/j.chb.2015.08.031

Bryman, A. (2012). *Social research methods* (4th ed). Oxford University Press.

Buchan, I. (2022). *Gini Coefficient of Inequality*. StatsDirect. https://www.statsdirect.com/help/default.htm#nonparametric_methods/gini.htm

Burmania, A., Parthasarathy, S., & Busso, C. (2016). Increasing the Reliability of Crowdsourcing Evaluations Using Online Quality Assessment. *IEEE Transactions on Affective Computing*, *7*(4), 374–388. https://doi.org/10.1109/TAFFC.2015.2493525

Burnette, C. B., Luzier, J. L., Bennett, B. L., Weisenmuller, C. M., Kerr, P., Martin, S., Keener, J., & Calderwood, L. (2022). Concerns and recommendations for using Amazon MTurk for eating disorder research. *International Journal of Eating Disorders*, *55*(2), 263–272.

Byrne, B. M. (2016). *Structural Equation Modeling With AMOS: Basic Concepts, Applications, and Programming, Third Edition* (0 ed.). Routledge. https://doi.org/10.4324/9781315757421

Cahyaningrum, A. O. (2023). Improving knowledge sharing through intrinsic motivation in the integration of self determination theory and theory of reason for action. *Enrichment : Journal of Management*, *13*(1), 370–375. https://doi.org/10.35335/enrichment.v13i1.1219

Cai, Y., & Shi, W. (2022). The influence of the community climate on users' knowledge-sharing intention: The social cognitive theory perspective. *Behaviour & Information Technology*, *41*(2), 307–323. https://doi.org/10.1080/0144929X.2020.1808704

Callison-Burch, C. (2014). *Crowd-Workers: Aggregating Information Across Turkers To Help Them Find Higher Paying Work*. 2.

Carmen, R., Defraeye, M., & Nieuwenhuyse, I. (2015). A Decision Support System for Capacity Planning in Emergency Departments. *International Journal of Simulation Modelling*, *14*, 299–312. https://doi.org/10.2507/IJSIMM14(2)10.308

Castille, C. M., Avet, S., & Daigle, K. (2019). *Comparing MTurk and the US population's occupational diversity*.

Chang, C.-C., Liang, C., Yan, C.-F., & Tseng, J.-S. (2013). The Impact of College Students' Intrinsic and Extrinsic Motivation on Continuance Intention to Use English Mobile Learning Systems. *The Asia-Pacific Education Researcher*, *22*(2), 181–192. https://doi.org/10.1007/s40299-012-0011-7

Chang, C.-M., Hsu, M.-H., & Lee, Y.-J. (2015). Factors Influencing Knowledge-Sharing Behavior in Virtual Communities: A Longitudinal Investigation. *Information Systems Management*, *32*(4), 331–340. https://doi.org/10.1080/10580530.2015.1080002

Chang, H. H., & Chuang, S.-S. (2011). Social capital and individual motivations on knowledge sharing: Participant involvement as a moderator. *Information & Management*, *48*(1), 9–18. https://doi.org/10.1016/j.im.2010.11.001

Chang, J. C., Amershi, S., & Kamar, E. (2017). Revolt: Collaborative Crowdsourcing for Labeling Machine Learning Datasets. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2334–2346. https://doi.org/10.1145/3025453.3026044

Chauhan, S., & Jaiswal, M. (2016). Determinants of acceptance of ERP software training in business schools: Empirical investigation using UTAUT model. *The International Journal of Management Education*, *14*(3), 248–262.

Checco, A., Bates, J., & Demartini, G. (2018). All that glitters is gold—An attack scheme on gold questions in crowdsourcing. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, *6*, 2–11.

Chen, L., Jia, J., & Wu, C. (2023). Factors influencing the behavioral intention to use contactless financial services in the banking industry: An application and extension of UTAUT model. *Frontiers in Psychology*, *14*, 1096709. https://doi.org/10.3389/fpsyg.2023.1096709

Chen, S.-C., Shing-Han, L., & Chien-Yi, L. (2011). Recent related research in technology acceptance model: A literature review. *Australian Journal of Business and Management Research*, *1*(9), 124.

Chennamaneni, A., Teng, J. T. C., & Raja, M. K. (2012). A unified model of knowledge sharing behaviours: Theoretical development and empirical test. *Behaviour & Information Technology*, *31*(11), 1097–1115. https://doi.org/10.1080/0144929X.2011.624637

Chiavaroli, N. (2017). Negatively-worded multiple choice questions: An avoidable threat to validity. *Practical Assessment, Research, and Evaluation*, *22*(1), 3.

Chien, T.-W., Chow, J. C., Chang, Y., & Chou, W. (2018). Applying Gini coefficient to evaluate the author research domains associated with the ordering of author names. *Medicine*, *97*(39). https://doi.org/10.1097/MD.0000000000012418

Chilton, L. B., Horton, J. J., Miller, R. C., & Azenkot, S. (2010). Task search in a human computation market. *Proceedings of the ACM SIGKDD Workshop on Human Computation - HCOMP '10*, 1. https://doi.org/10.1145/1837885.1837889

Chin, C.-L., & Yao, G. (2014). Convergent Validity. In A. C. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 1275–1276). Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_573

Chmielewski, M., & Kucker, S. C. (2020). An MTurk Crisis? Shifts in Data Quality and the Impact on Study Results. *Social Psychological and Personality Science*, *11*(4), 464–473. https://doi.org/10.1177/1948550619875149

Cho, E., Xie, H., Lalor, J. P., Kumar, V., & Campbell, W. M. (2019). Efficient Semi-Supervised Learning for Natural Language Understanding by Optimizing Diversity. *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 1077–1084. https://doi.org/10.1109/ASRU46091.2019.9003747

ChrisTurk. (2017, January 17). *Human Intelligence Tasks (HITs)*. TurkerView mTurk Forum | Scripts, HITs, & mTurk Community. https://forum.turkerview.com/wiki/Human-Intelligence-Tasks-HITs

ChrisTurk. (2022). *TurkerViewJS* [Computer software]. https://turkerview.com/mturk-scripts/1-turkerviewjs

Chu, J., & Manchanda, P. (2016). Quantifying cross and direct network effects in online consumer-to-consumer platforms. *Marketing Science*, *35*(6), 870–893.

Cobham, A., Schlögl, L., & Sumner, A. (2016). Inequality and the tails: The Palma proposition and ratio. *Global Policy*, *7*(1), 25–36.

Cohen, E., Venkateswaran, M., Sankar, N., & Callison-Burch, C. (2020). *Turkish Judge: A Peer Evaluation Framework for Crowd Work Appeals*.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research Method in education 6th ed London: Routledge Falmer*.

Comrey, A., & Lee, H. (1992). A first course in factor analysis (2nd edn.) Lawrence Earlbaum associates. *Publishers: Hillsdale, New Jersey*.

Connelly, C. E., Fieseler, C., Černe, M., Giessner, S. R., & Wong, S. I. (2021). Working in the digitized economy: HRM theory & practice. *Human Resource Management Review*, *31*(1), 100762. https://doi.org/10.1016/j.hrmr.2020.100762

Conte, A., Levati, M. V., & Montinari, N. (2019). Experience in public goods experiments. *Theory and Decision*, *86*(1), 65–93.

Cook, K. S. (2015). Exchange: Social. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 482–488). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.32056-6

Coppock, A. (2019). Generalizing from Survey Experiments Conducted on Mechanical Turk: A Replication Approach. *Political Science Research and Methods*, *7*(3), 613–628. https://doi.org/10.1017/psrm.2018.10

Cordier, M., Uehara, T., Weih, J., & Hamaide, B. (2017). An Input-output Economic Model Integrated Within a System Dynamics Ecological Model: Feedback Loop Methodology Applied to Fish Nursery Restoration. *Ecological Economics*, *140*, 46–57. https://doi.org/10.1016/j.ecolecon.2017.04.005

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., & Schiele, B. (2016). *The Cityscapes Dataset for Semantic Urban Scene Understanding*. https://doi.org/10.48550/ARXIV.1604.01685

Cowell, F. (2011). *Measuring inequality* (3rd ed.). Oxford University Press.

Crotty, M. J. (1998). The foundations of social research: Meaning and perspective in the research process. *The Foundations of Social Research*, 1–256.

Crowston, K., Mitchell, E., & Østerlund, C. (2018). *Coordinating Advanced Crowd Work: Extending Citizen Science*. Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2018.212

Cui, L., Chen, J., He, W., Li, H., Guo, W., & Su, Z. (2021). Achieving Approximate Global Optimization of Truth Inference for Crowdsourcing Microtasks. *Data Science and Engineering*, *6*(3), 294–309. https://doi.org/10.1007/s41019-021-00164-2

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*(1), 16–29. https://doi.org/10.1037/1082-989X.1.1.16

Damen, D., van der Wijst, P., van Amelsvoort, M., & Krahmer, E. (2020). The Effect of Perspective-Taking on Trust and Understanding in Online and Face-to-Face Mediations. *Group Decision and Negotiation*, *29*(6), 1121–1156. https://doi.org/10.1007/s10726-020-09698-8

Damodaran, S., & Couretas, J. M. (2015). Cyber modeling & simulation for cyber-range events. *Summer Simulation Multiconference*.

Daniel, W. W., & Cross, C. L. (2018). *Biostatistics: A foundation for analysis in the health sciences*. Wiley.

Dannefer, D. (1987). Aging as intracohort differentiation: Accentuation, the Matthew effect, and the life course. *Sociological Forum*, *2*(2), 211–236.

Davahli, M. R., Karwowski, W., & Taiar, R. (2020). A System Dynamics Simulation Applied to Healthcare: A Systematic Review. *International Journal of Environmental Research and Public Health*, *17*(16), 5741. https://doi.org/10.3390/ijerph17165741

Davis, E., & Marcus, G. (2016). The scope and limits of simulation in automated reasoning. *Artificial Intelligence*, *233*, 60–72. https://doi.org/10.1016/j.artint.2015.12.003

D'Cruz, P., & Noronha, E. (2016). Positives outweighing negatives: The experiences of Indian crowdsourced workers. *Work Organisation, Labour & Globalisation*, *10*(1), 44–63. JSTOR. https://doi.org/10.13169/workorgalaboglob.10.1.0044

Deng, X., & Guo, K. (2018). Understanding knowledge sharing in virtual communities: A network perspective. *Library Hi Tech*, *39*(4), 1174–1186. https://doi.org/10.1108/LHT-09-2018-0119

Deng, X., Joshi, K. D., & Galliers, R. D. (2016). The duality of empowerment and marginalization in microtask crowdsourcing: Giving voice to the less powerful through value sensitive design. *Mis Quarterly*, *40*(2), 279–302.

Dennis, S. A., Goodson, B. M., & Pearson, C. (2019). *Online Worker Fraud and Evolving Threats to the Integrity of MTurk Data: A Discussion of Virtual Private Servers and the Limitations of IP-Based Screening Procedures* (SSRN Scholarly Paper ID 3233954). Social Science Research Network. https://doi.org/10.2139/ssrn.3233954

Developer Sandbox. (2021). Amazon Mturk Requester. https://requester.mturk.com/developer/sandbox

Dey, T., & Mukhopadhyay, S. (2018). Influence of Behavioral Intentions, Affective Trust and Affective Commitment on Knowledge Sharing Behavior: *International Journal of Knowledge Management*, *14*(2), 37–51. https://doi.org/10.4018/IJKM.2018040103

Di Gangi, P. M., McAllister, C. P., Howard, J. L., Thatcher, J. B., & Ferris, G. R. (2022). Can you see opportunity knocking? An examination of technology-based political skill on opportunity recognition in online communities for MTurk workers. *Internet Research*, *32*(4), 1041–1075. https://doi.org/10.1108/INTR-03-2021-0175

Difallah, D. E., Catasta, M., Demartini, G., Ipeirotis, P. G., & Cudré-Mauroux, P. (2015). The Dynamics of Micro-Task Crowdsourcing: The Case of Amazon MTurk. *Proceedings of the 24th International Conference on World Wide Web - WWW '15*, 238–247. https://doi.org/10.1145/2736277.2741685

Difallah, D. E., Demartini, G., & Cudré-Mauroux, P. (2012). Mechanical cheat: Spamming schemes and adversarial techniques on crowdsourcing platforms. *CrowdSearch*, 26–30.

Djanatliev, A., & German, R. (2013). Prospective healthcare decision-making by combined system dynamics, discrete-event and agent-based simulation. *2013 Winter Simulations Conference (WSC)*, 270–281. https://doi.org/10.1109/WSC.2013.6721426

Dong, L., Ji, T., Zhou, G., & Zhang, J. (2022). Subjective Norms or Psychological Empowerment? Moderation Effect of Power Distance on Knowledge Sharing. *Sustainability*, *14*(21), 14407. https://doi.org/10.3390/su142114407

DonovanM. (2018). *Turkmaster (Mturk)* (1.5.11) [Computer software]. https://greasyfork.org/scripts/4771-turkmaster-mturk

dos Santos, P. M., & Cirillo, M. Â. (2023). Construction of the average variance extracted index for construct validation in structural equation models with adaptive regressions. *Communications in Statistics - Simulation and Computation*, *52*(4), 1639–1650. https://doi.org/10.1080/03610918.2021.1888122

Drapeau, R., Chilton, L., Bragg, J., & Weld, D. (2016). MicroTalk: Using Argumentation to Improve Crowdsourcing Accuracy. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, *4*, 32–41. https://doi.org/10.1609/hcomp.v4i1.13270

Dror, G., Koren, Y., Maarek, Y., & Szpektor, I. (2011). I want to answer; who has a question? Yahoo! answers recommender system. *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1109–1117. https://doi.org/10.1145/2020408.2020582

Duffy, M., & Chenail, R. J. (2009). Values in Qualitative and Quantitative Research. *Counseling and Values*, *53*(1), 22–38. https://doi.org/10.1002/j.2161-007X.2009.tb00111.x

Duggan, J. (2016). An Introduction to System Dynamics. In J. Duggan (Ed.), *System Dynamics Modeling with R* (pp. 1–24). Springer International Publishing. https://doi.org/10.1007/978-3-319-34043-2_1

Dunn, M. (2020). Making gigs work: Digital platforms, job quality and worker motivations. *New Technology, Work and Employment*, *35*(2), 232–249.
Dunn, S., & Hedges, M. (2014). How the crowd can surprise us: Humanities crowdsourcing and the creation of knowledge. *Crowdsourcing Our Cultural Heritage*, 231–246.
Düntsch, I., & Gediga, G. (2019). Confusion Matrices and Rough Set Data Analysis. *Journal of Physics: Conference Series*, *1229*(1), 012055. https://doi.org/10.1088/1742-6596/1229/1/012055
Easterby-Smith, M., Thorpe, R., & Jackson, P. R. (2012). *Management research*. Sage.
El Maarry, K., Milland, K., & Balke, W.-T. (2018). A Fair Share of the Work?: The Evolving Ecosystem of Crowd Workers. *Proceedings of the 10th ACM Conference on Web Science*, 145–152. https://doi.org/10.1145/3201064.3201074
Ergün, E., & Avcı, Ü. (2018). Knowledge sharing self-efficacy, motivation and sense of community as predictors of knowledge receiving and giving behaviors. *Journal of Educational Technology & Society*, *21*(3), 60–73.
Fait, M., & Sakka, G. (2021). Knowledge sharing: An innovative organizational approach to engage volunteers. *EuroMed Journal of Business*, *16*(3), 290–305. https://doi.org/10.1108/EMJB-10-2019-0131
Fan, S., Gadiraju, U., Checco, A., & Demartini, G. (2020). CrowdCO-OP: Sharing Risks and Rewards in Crowdsourcing. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW2), 1–24.
Fang, C., & Zhang, J. (2019). Users' continued participation behavior in social Q&A communities: A motivation perspective. *Computers in Human Behavior*, *92*, 87–109. https://doi.org/10.1016/j.chb.2018.10.036
Farhi, F., Jeljeli, R., Zahra, A., Saidani, S., & Feguiri, L. (2023). Factors Behind Virtual Assistance Usage Among iPhone Users: Theory of Reasoned Action. *International Journal of Interactive Mobile Technologies (iJIM)*, *17*(02), 42–61. https://doi.org/10.3991/ijim.v17i02.36021
Feng, Y., & Ye, H. (Jonathan). (2016). Why do you return the favor in online knowledge communities? A study of the motivations of reciprocity. *Computers in Human Behavior*, *63*, 342–349. https://doi.org/10.1016/j.chb.2016.05.007
Fernández-Macías, E., & Bisello, M. (2020). *A taxonomy of tasks for assessing the impact of new technologies on work*. JRC Working Papers Series on Labour, Education and Technology.
Ferran-Ferrer, N. (2015). Volunteer participation in citizen science projects. *Profesional de La Información*, *24*(6), 827–837.
Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. sage.
Fieseler, C., Bucher, E., & Hoffmann, C. P. (2019). Unfairness by Design? The Perceived Fairness of Digital Labor on Crowdworking Platforms. *Journal of Business Ethics*, *156*(4), 987–1005. https://doi.org/10.1007/s10551-017-3607-2
Floridi, L., & Chiriatti, M. (2020). GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds and Machines*, *30*(4), 681–694. https://doi.org/10.1007/s11023-020-09548-1
Fofana, F., Bazeley, P., & Regnault, A. (2020). Applying a mixed methods design to test saturation for qualitative data in health outcomes research. *PloS One*, *15*(6), e0234898.
Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, *18*(1), 39. https://doi.org/10.2307/3151312
Fralich, R., & Bitektine, A. (2020). "Invincibles" and "invisibles": CEO status and the 'Matthew effect' in strategic decision-making. *Long Range Planning*, *53*(3), 101887.
Fruggeri, L. (2021). Social research as a process of interaction. *Murmurations: Journal of Transformative Systemic Practice*, *3*(2), 81–99. https://doi.org/10.28963/3.2.6

Gadiraju, U., Checco, A., Gupta, N., & Demartini, G. (2017). Modus Operandi of Crowd Workers: The Invisible Role of Microtask Work Environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(3), 1–29. https://doi.org/10.1145/3130914

Gadiraju, U., Kawase, R., & Dietze, S. (2014). A taxonomy of microtasks on the web. *Proceedings of the 25th ACM Conference on Hypertext and Social Media - HT '14*, 218–223. https://doi.org/10.1145/2631775.2631819

Gadiraju, U., Kawase, R., Dietze, S., & Demartini, G. (2015). Understanding Malicious Behavior in Crowdsourcing Platforms: The Case of Online Surveys. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 1631–1640. https://doi.org/10.1145/2702123.2702443

Gadiraju, U., Yang, J., & Bozzon, A. (2017). Clarity is a Worthwhile Quality: On the Role of Task Clarity in Microtask Crowdsourcing. *Proceedings of the 28th ACM Conference on Hypertext and Social Media*, 5–14. https://doi.org/10.1145/3078714.3078715

Gang, K., & Ravichandran, T. (2015). Exploring the Determinants of Knowledge Exchange in Virtual Communities. *IEEE Transactions on Engineering Management*, *62*(1), 89–99. https://doi.org/10.1109/TEM.2014.2376521

Gao, R., & Shah, C. (2020). Toward creating a fairer ranking in search engine results. *Information Processing & Management*, *57*(1), 102138.

García, L. F., Aluja, A., Urieta, P., & Gutierrez, F. (2022). High convergent validity among the five-factor model, PID-5-SF, and PiCD. *Personality Disorders: Theory, Research, and Treatment*, *13*(2), 119–132. https://doi.org/10.1037/per0000486

Garcia Martinez, M. (2017). Inspiring crowdsourcing communities to create novel solutions: Competition design and the mediating role of trust. *Technological Forecasting and Social Change*, *117*, 296–304. https://doi.org/10.1016/j.techfore.2016.11.015

Gefen, D., Straub, D., & Boudreau, M.-C. (2000). Structural Equation Modeling and Regression: Guidelines for Research Practice. *Communications of the Association for Information Systems*, *4*. https://doi.org/10.17705/1CAIS.00407

Geiger, D., & Schader, M. (2014). Personalized task recommendation in crowdsourcing information systems—Current state of the art. *Decision Support Systems*, *65*, 3–16. https://doi.org/10.1016/j.dss.2014.05.007

George, D., & Mallery, M. (2003). *Using SPSS for Windows step by step: A simple guide and reference*.

Gerber, C. (2021). Community building on crowdwork platforms: Autonomy and control of online workers? *Competition & Change*, *25*(2), 190–211.

Ghojogh, B., & Crowley, M. (2019). The theory behind overfitting, cross validation, regularization, bagging, and boosting: Tutorial. *arXiv Preprint arXiv:1905.12787*.

Ghosh, S., Sperling, R., & Hooper, S. (2019). Using Amazon MTurk for research in academia: A beginner's guide for using Qualtrics, detecting VPN/proxy, limiting countries using geolocation & other tips. *Detecting VPN/Proxy, Limiting Countries Using Geolocation & Other Tips (September 18, 2019)*.

Glassman, M., Kuznetcova, I., Peri, J., & Kim, Y. (2021). Cohesion, collaboration and the struggle of creating online learning communities: Development and validation of an online collective efficacy scale. *Computers and Education Open*, *2*, 100031. https://doi.org/10.1016/j.caeo.2021.100031

Gol, E. S., Stein, M.-K., & Avital, M. (2019). Crowdwork platform governance toward organizational value creation. *The Journal of Strategic Information Systems*, *28*(2), 175–195.

González Pinto, J. M., El-Maary, K., & Balke, W.-T. (2019). *Fine-tuning Gold Questions in Crowdsourcing Tasks using Probabilistic and Siamese Neural Network Models*. https://doi.org/10.34962/JWS-68

González-Anta, B., Orengo, V., Zornoza, A., Peñarroja, V., & Martínez-Tur, V. (2021). Understanding the Sense of Community and Continuance Intention in Virtual Communities: The Role of Commitment and Type of Community. *Social Science Computer Review*, *39*(3), 335–352. https://doi.org/10.1177/0894439319859590

Göritz, A. S., Borchert, K., & Hirth, M. (2021). Using Attention Testing to Select Crowdsourced Workers and Research Participants. *Social Science Computer Review*, *39*(1), 84–104. https://doi.org/10.1177/0894439319848726

Gosling, S. D., & Mason, W. (2015). Internet research in psychology. *Annual Review of Psychology*, *66*, 877–902.

Goyal, T., McDonnell, T., Kutlu, M., Elsayed, T., & Lease, M. (2018). Your Behavior Signals Your Reliability: Modeling Crowd Behavioral Traces to Ensure Quality Relevance Annotations. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, *6*, 41–49. https://doi.org/10.1609/hcomp.v6i1.13331

Graham, M., Hjorth, I., & Lehdonvirta, V. (2017). Digital labour and development: Impacts of global digital labour platforms and the gig economy on worker livelihoods. *Transfer: European Review of Labour and Research*, *23*(2), 135–162. https://doi.org/10.1177/1024258916687250

Grant, S., & Betts, B. (2013). Encouraging user behaviour with achievements: An empirical study. *2013 10th Working Conference on Mining Software Repositories (MSR)*, 65–68.

Gray, D. E. (2021). *Doing research in the real world*. sage.

Gray, M. L., Suri, S., Ali, S. S., & Kulkarni, D. (2016). The Crowd is a Collaborative Network. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 134–147. https://doi.org/10.1145/2818048.2819942

Greco, G. M., & Floridi, L. (2004). The tragedy of the digital commons. *Ethics and Information Technology*, *6*(2), 73–81.

Guenther, P., Guenther, M., Ringle, C. M., Zaefarian, G., & Cartwright, S. (2023). Improving PLS-SEM use for business marketing research. *Industrial Marketing Management*, *111*, 127–142. https://doi.org/10.1016/j.indmarman.2023.03.010

Gumata, N., & Ndou, E. (2017). The Efficiency Wage Hypothesis, Labour Productivity and the Minimum Wage. In N. Gumata & E. Ndou (Eds.), *Labour Market and Fiscal Policy Adjustments to Shocks: The Role and Implications for Price and Financial Stability in South Africa* (pp. 127–143). Springer International Publishing. https://doi.org/10.1007/978-3-319-66520-7_8

Guo, B., Zhang, X., Wang, Z., Jiang, M., Nie, J., Ding, Y., Yue, J., & Wu, Y. (2023). *How Close is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection* (arXiv:2301.07597). arXiv. https://doi.org/10.48550/arXiv.2301.07597

Gupta, B., Dasgupta, S., & Gupta, A. (2008). Adoption of ICT in a government organization in a developing country: An empirical study. *The Journal of Strategic Information Systems*, *17*(2), 140–154.

Gupta, N., Martin, D., Hanrahan, B. V., & O'Neill, J. (2014). Turk-Life in India. *Proceedings of the 18th International Conference on Supporting Group Work*, 1–11. https://doi.org/10.1145/2660398.2660403

Hagger, M. S., & Hamilton, K. (2023). Longitudinal tests of the theory of planned behaviour: A meta-analysis. *European Review of Social Psychology*, 1–57. https://doi.org/10.1080/10463283.2023.2225897

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., Tatham, R. L., & others. (2006). *Multivariate data analysis (Vol. 6)*.

Hair, J. F., Matthews, L. M., Matthews, R. L., & Sarstedt, M. (2017). PLS-SEM or CB-SEM: updated guidelines on which method to use. *International Journal of Multivariate Data Analysis*, *1*(2), 107–123. https://doi.org/10.1504/IJMDA.2017.087624

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a Silver Bullet. *Journal of Marketing Theory and Practice*, *19*(2), 139–152. https://doi.org/10.2753/MTP1069-6679190202

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2013). Partial least squares structural equation modeling: Rigorous applications, better results and higher acceptance. *Long Range Planning*, *46*(1–2), 1–12.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, *31*(1), 2–24.

Hair, J., Sarstedt, M., Hopkins, L., & G. Kuppelwieser, V. (2014). Partial least squares structural equation modeling (PLS-SEM): An emerging tool in business research. *European Business Review*, *26*(2), 106–121. https://doi.org/10.1108/EBR-10-2013-0128

Halder, A. (2019). An Adaptive Non-linear Statistical Salt-and-Pepper Noise Removal Algorithm using Interquartile Range. *ELCVIA Electronic Letters on Computer Vision and Image Analysis*, *18*(1), 76. https://doi.org/10.5565/rev/elcvia.1145

Han, L., Roitero, K., Gadiraju, U., Sarasua, C., Checco, A., Maddalena, E., & Demartini, G. (2019). All Those Wasted Hours: On Task Abandonment in Crowdsourcing. *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, 321–329. https://doi.org/10.1145/3289600.3291035

Hanrahan, B. V., Chen, A., Ma, J., Ma, N. F., Squicciarini, A., & Savage, S. (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), 128:1-128:23. https://doi.org/10.1145/3449202

Hanrahan, B. V., Martin, D., Willamowski, J., & Carroll, J. M. (2018). Investigating the Amazon Mechanical Turk Market Through Tool Design. *Computer Supported Cooperative Work (CSCW)*, *27*(3–6), 1255–1274. https://doi.org/10.1007/s10606-018-9312-6

Hanrahan, B. V., Willamowski, J. K., Swaminathan, S., & Martin, D. B. (2015). TurkBench: Rendering the Market for Turkers. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, 1613–1616. https://doi.org/10.1145/2702123.2702279

Hanum, L., Ari, D. P. S., & Susilo, H. (2020). Determinant of Tax Compliance Intention in Java-Indonesia: Theory of reasoned action approach. *Proceedings of the 2nd Annual International Conference on Business and Public Administration (AICoBPA 2019)*. 2nd Annual International Conference on Business and Public Administration (AICoBPA 2019), Malang, Indonesia. https://doi.org/10.2991/aebmr.k.201116.036

Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., & Bigham, J. P. (2018). A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–14. https://doi.org/10.1145/3173574.3174023

Hartanti, B., Romadon, A., Anisah, N., & Utomo, L. P. (2021). *Information Systems Behavior on System Security in the Perspective of "Theory of Reasoned Action":* 2nd Annual Conference on Social Science and Humanities (ANCOSH 2020), Malang, Indonesia. https://doi.org/10.2991/assehr.k.210413.001

Hasan, S. (2018). *Mturk-engine* (1.8.3) [TypeScript]. https://github.com/Anveio/mturk-engine

Hassaan, M., Li, G., & Akhtar, R. (2023). Mobile Payment Adoption in Pakistan: Extending UTAUT with Islamic Religiosity. *Journal of Social Sciences Review*, *3*(2), 219–231. https://doi.org/10.54183/jssr.v3i2.224

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*, 400–407.

Hauser, D., Paolacci, G., & Chandler, J. J. (2018). *Common Concerns with MTurk as a Participant Pool: Evidence and Solutions* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/uq45c

Heer, J., & Bostock, M. (2010). Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 203–212. https://doi.org/10.1145/1753326.1753357

Hein, A., Schreieck, M., Riasanow, T., Setzke, D. S., Wiesche, M., Böhm, M., & Krcmar, H. (2020). Digital platform ecosystems. *Electronic Markets*, *30*(1), 87–98. https://doi.org/10.1007/s12525-019-00377-4

Hellman, B. (2023). *MTurk Suite* [JavaScript]. https://github.com/Kadauchi/mturk-suite

Henseler, J., Dijkstra, T. K., Sarstedt, M., Ringle, C. M., Diamantopoulos, A., Straub, D. W., Ketchen, D. J., Hair, J. F., Hult, G. T. M., & Calantone, R. J. (2014). Common Beliefs and Reality About PLS: Comments on Rönkkö and Evermann (2013). *Organizational Research Methods*, *17*(2), 182–209. https://doi.org/10.1177/1094428114526928

Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. In R. R. Sinkovics & P. N. Ghauri (Eds.), *Advances in International Marketing* (Vol. 20, pp. 277–319). Emerald Group Publishing Limited. https://doi.org/10.1108/S1474-7979(2009)0000020014

Herrera, M., Pérez-Hernández, M., Kumar Parlikad, A., & Izquierdo, J. (2020). Multi-agent systems and complex networks: Review and applications in systems engineering. *Processes*, *8*(3), 312.

Hettiachchi, D., Schaekermann, M., McKinney, T. J., & Lease, M. (2021). The challenge of variable effort crowdsourcing and how visible gold can help. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 1–26.

Hicks, J. (2020). Generating conditions of strong social support for wind power: Insights from community-owned wind energy projects. *Australasian Journal of Environmental Management*, *27*(2), 137–155. https://doi.org/10.1080/14486563.2020.1758807

Hirth, M., Scheuring, S., Hossfeld, T., Schwartz, C., & Tran-Gia, P. (2014). Predicting result quality in Crowdsourcing using application layer monitoring. *2014 IEEE Fifth International Conference on Communications and Electronics (ICCE)*, 510–515. https://doi.org/10.1109/CCE.2014.6916756

Holden, M. T., & Lynch, P. (2004). Choosing the Appropriate Methodology: Understanding Research Philosophy. *The Marketing Review*, *4*(4), 397–409. https://doi.org/10.1362/1469347042772428

Hoyle, R. K. (2011). Structural equation modeling for social and personality psychology. *Structural Equation Modeling for Social and Personality Psychology*, 1–120.

Hsieh, P.-J. (2021). Determinants of Knowledge-Sharing Intentions for Shared Decision-Making Platforms. *Journal of Computer Information Systems*, 1–12. https://doi.org/10.1080/08874417.2021.1980747

Hsu, M.-H., Ju, T., Yen, C.-H., & Chang, C.-M. (2007). Knowledge Sharing Behavior in Virtual Communities: The Relationship between Trust, Self-Efficacy, and Outcome Expectations. *International Journal of Human-Computer Studies*, *65*, 153–169. https://doi.org/10.1016/j.ijhcs.2006.09.003

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.

Hulland, J. (1999). Use of partial least squares (PLS) in strategic management research: A review of four recent studies. *Strategic Management Journal*, *20*(2), 195–204.

Hung, D. N., Tham, J., Azam, S. M. F., & Khatibi, A. A. (2019). An Empirical Analysis of Perceived Transaction Convenience, Performance Expectancy, Effort Expectancy and Behavior Intention to Mobile Payment of Cambodian Users. *International Journal of Marketing Studies*, *11*(4), 77. https://doi.org/10.5539/ijms.v11n4p77

Hung, S.-Y., Lai, H.-M., & Chou, Y.-C. (2015). Knowledge-sharing intention in professional virtual communities: A comparison between posters and lurkers. *Journal of the Association for Information Science and Technology*, *66*(12), 2494–2510. https://doi.org/10.1002/asi.23339

Hunter, D., McCallum, J., & Howes, D. (2019). Defining exploratory-descriptive qualitative (EDQ) research and considering its application to healthcare. *Journal of Nursing and Health Care*, *4*(1).

Ibrahim, A., & Shiring, E. (2022). The Relationship between Educators' Attitudes, Perceived Usefulness, and Perceived Ease of Use of Instructional and Web-Based Technologies: Implications from Technology Acceptance Model (TAM). *International Journal of Technology in Education*, *5*(4), 535–551. https://doi.org/10.46328/ijte.285

Ihl, A., Strunk, K. S., & Fiedler, M. (2020). The mediated effects of social support in professional online communities on crowdworker engagement in micro-task crowdworking. *Computers in Human Behavior*, *113*, 106482. https://doi.org/10.1016/j.chb.2020.106482

Imamov, M., & Semenikhina, N. (2021). The impact of the digital revolution on the global economy. *Linguistics and Culture Review*, *5*(S4), 968–987.

Irani, L. (2015). The cultural work of microwork. *New Media & Society*, *17*(5), 720–739. https://doi.org/10.1177/1461444813511926

Irani, L., & Silberman, M. S. (2013). Turkopticon: Interrupting worker invisibility in amazon mechanical turk. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, 611. https://doi.org/10.1145/2470654.2470742

Ismail, I. R., Tajuddin, N., & Yunus, N. K. M. (2019). Trust and Intention to Share as Predictors of Online Knowledge-Sharing Behavior. In M. Y. Mohamad Noor, B. E. Ahmad, M. R. Ismail, H. Hashim, & M. A. Abdullah Baharum (Eds.), *Proceedings of the Regional Conference on Science, Technology and Social Sciences (RCSTSS 2016)* (pp. 57–67). Springer Singapore. https://doi.org/10.1007/978-981-13-0203-9_6

Jacques, J. T., & Kristensson, P. O. (2019). Crowdworker Economics in the Gig Economy. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–10. https://doi.org/10.1145/3290605.3300621

Jahan, N., & Kim, S. W. (2021). Understanding online community participation behavior and perceived benefits: A social exchange theory perspective. *PSU Research Review*, *5*(2), 85–100. https://doi.org/10.1108/PRR-12-2019-0036

Jain, R., & Chetty, P. (2022, March 28). Confirmatory Factor Analysis (CFA) in SEM using SPSS Amos. *Project Guru*.

Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education*, *38*(12), 1217–1218. https://doi.org/10.1111/j.1365-2929.2004.02012.x

Jennex, M. E. (Ed.). (2019). *Effective Knowledge Management Systems in Modern Society:* IGI Global. https://doi.org/10.4018/978-1-5225-5427-1

Jennings, N. R. (2000). On agent-based software engineering. *Artificial Intelligence*, *117*(2), 277–296.

Jiarui, W., Xiaoli, Z., & Jiafu, S. (2022). Interpersonal Relationship, Knowledge Characteristic, and Knowledge Sharing Behavior of Online Community Members: A TAM Perspective. *Computational Intelligence and Neuroscience*, *2022*, 1–11. https://doi.org/10.1155/2022/4188480

Johnson, P., & Duberley, J. (2000). *Understanding management research: An introduction to epistemology*. Sage.

Jonassen, D. H. (1991). Objectivism versus constructivism: Do we need a new philosophical paradigm? *Educational Technology Research and Development*, *39*, 5–14.

Jøsang, A., & Pope, S. (2005). Semantic constraints for trust transitivity. *Proceedings of the 2nd Asia-Pacific Conference on Conceptual Modelling-Volume 43*, 59–68.

Józsa, K., & Morgan, G. A. (2017). Reversed items in Likert scales: Filtering out invalid responders. *Journal of Psychological and Educational Research*, *25*(1), 7–25.

Juba, B., & Le, H. S. (2019). Precision-recall versus accuracy and the role of large data sets. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(01), 4039–4048.

Kamarozaman, Z., & Razak, F. Z. A. (2021). The role of facilitating condition in enhancing user's continuance intention. *Journal of Physics: Conference Series*, *1793*(1), 012022. https://doi.org/10.1088/1742-6596/1793/1/012022

Kang, M. (2022). Motivational affordances and survival of new askers on social Q&A sites: The case of Stack Exchange network. *Journal of the Association for Information Science and Technology*, *73*(1), 90–103. https://doi.org/10.1002/asi.24548

Kankanhalli, A., Tan, B. C., & Wei, K.-K. (2005). Contributing knowledge to electronic knowledge repositories: An empirical investigation. *MIS Quarterly*, 113–143.

Kaplan, T., Saito, S., Hara, K., & Bigham, J. (2018). Striving to Earn More: A Survey of Work Strategies and Tool Use Among Crowd Workers. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, *6*, 70–78. https://doi.org/10.1609/hcomp.v6i1.13327

Kässi, O., & Lehdonvirta, V. (2018). Online labour index: Measuring the online gig economy for policy and research. *Technological Forecasting and Social Change*, *137*, 241–248.

Katz, L. F. (1986). Efficiency Wage Theories: A Partial Evaluation. *NBER Macroeconomics Annual*, *1*, 235–276. https://doi.org/10.1086/654025

Kavak, H., Padilla, J. J., Vernon-Bido, D., Diallo, S. Y., Gore, R., & Shetty, S. (2021). Simulation for cybersecurity: State of the art and future directions. *Journal of Cybersecurity*, *7*(1), tyab005. https://doi.org/10.1093/cybsec/tyab005

Ke, T. T., & Zhu, Y. (2021). Cheap Talk on Freelance Platforms. *Management Science*, *67*(9), 5901–5920. https://doi.org/10.1287/mnsc.2020.3782

Kennedy, R., Clifford, S., Burleigh, T., Waggoner, P. D., Jewell, R., & Winter, N. J. G. (2020). The shape of and solutions to the MTurk quality crisis. *Political Science Research and Methods*, *8*(4), 614–629. https://doi.org/10.1017/psrm.2020.6

Kenny, D. (2020, June 5). *SEM: Measuring Model Fit*. https://davidakenny.net/cm/fit.htm

Khalid, K., Ismail, R., & Mat-Jizat, J. E. (2023). Teachers and Crowdsourcing Platform: Using UTAUT to Examine Behaviour Intention and Use Behaviour. *International Journal of Academic Research in Progressive Education and Development*, *12*(1), Pages 1930-1946. https://doi.org/10.6007/IJARPED/v12-i1/15737

Kim, W., & Park, J. (2017). Examining structural relationships between work engagement, organizational procedural justice, knowledge sharing, and innovative work behavior for sustainable organizations. *Sustainability*, *9*(2), 205.

Kirsch, L. S. (1997). Portfolios of control modes and IS project management. *Information Systems Research*, *8*(3), 215–239.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed). Guilford Press.

Kulkarni, A., Can, M., & Hartmann, B. (2012). Collaboratively crowdsourcing workflows with turkomatic. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work - CSCW '12*, 1003. https://doi.org/10.1145/2145204.2145354

Kung, F. Y. H., Kwok, N., & Brown, D. J. (2018). Are Attention Check Questions a Threat to Scale Validity? *Applied Psychology*, *67*(2), 264–283. https://doi.org/10.1111/apps.12108

Kurup, A. R., & Sajeev, G. P. (2020). A task recommendation scheme for crowdsourcing based on expertise estimation. *Electronic Commerce Research and Applications*, *41*, 100946. https://doi.org/10.1016/j.elerap.2020.100946

Kwek, A. (2020). Crowdsourced Research: Vulnerability, Autonomy, and Exploitation. *Ethics & Human Research*, *42*(1), 22–35.

Lai, H.-M., & Chen, T. T. (2014). Knowledge sharing in interest online communities: A comparison of posters and lurkers. *Computers in Human Behavior*, *35*, 295–306. https://doi.org/10.1016/j.chb.2014.02.004

Lai, H.-M., Huang, Y.-W., & Hung, S.-Y. (2018). A Meta-analysis of Knowledge Sharing in Virtual Communities: The Moderating Effect of Membership Types. In L. Uden, B. Hadzima, & I.-H. Ting (Eds.), *Knowledge Management in Organizations* (pp. 83–93). Springer International Publishing. https://doi.org/10.1007/978-3-319-95204-8_8

Lakhal, S., Khechine, H., & Pascot, D. (2013). Student behavioural intentions to use desktop video conferencing in a distance course: Integration of autonomy to the UTAUT model. *Journal of Computing in Higher Education*, *25*(2), 93–121. https://doi.org/10.1007/s12528-013-9069-3

Lam, L. W. (2012). Impact of competitiveness on salespeople's commitment and performance. *Journal of Business Research*, *65*(9), 1328–1334.

LaPlante, R., & Silberman, M. S. (2016). Building trust in crowd worker forums: Worker ownership, governance, and work outcomes. *Proceedings of WebSci16. ACM*, 35–63.

Lascau, L., Gould, S. J. J., Brumby, D. P., & Cox, A. L. (2022). Crowdworkers' Temporal Flexibility is Being Traded for the Convenience of Requesters Through 19 'Invisible Mechanisms' Employed by Crowdworking Platforms: A Comparative Analysis Study of Nine Platforms. *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 1–8. https://doi.org/10.1145/3491101.3519629

LaToza, T. D., Di Lecce, A., Ricci, F., Towne, W. B., & Van Der Hoek, A. (2015). Ask the crowd: Scaffolding coordination and knowledge sharing in microtask programming. *2015 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, 23–27.

Lee, A. S.-H., Wang, S., Yeoh, W., & Ikasari, N. (2021). Understanding the Use of Knowledge Sharing Tools. *Journal of Computer Information Systems*, *61*(5), 458–470. https://doi.org/10.1080/08874417.2020.1752850

Lee, J. (2018). The effects of knowledge sharing on individual creativity in higher education institutions: Socio-technical view. *Administrative Sciences*, *8*(2), 21.

Leemis, L. M., & Park, S. K. (2006). *Discrete-Event Simulation: A First Course* (1st edition). Pearson.

Lenart-Gansiniec, R. (2017). Virtual Knowledge Sharing in Crowdsourcing: Measurement Dilemmas. *Journal of Entrepreneurship, Management and Innovation*, *13*(3), 95–123.

Leung, G. S.-K., Cho, V., & Wu, C. H. (2021). Crowd Workers' Continued Participation Intention in Crowdsourcing Platforms: An Empirical Study in Compensation-Based Micro-Task Crowdsourcing. *Journal of Global Information Management*, *29*(6), 1–28. https://doi.org/10.4018/JGIM.20211101.oa13

Li, J. (2015). Knowledge sharing in virtual communities: A social exchange theory perspective. *Journal of Industrial Engineering and Management*, *8*(1), 170–183. https://doi.org/10.3926/jiem.1389

Li, L. P., Frethey-Bentham, C., Juric, B., & Brodie, R. J. (2023). A Negative Actor Engagement Scale for Online Knowledge-Sharing Platforms. *Australasian Marketing Journal*, *31*(1), 36–48. https://doi.org/10.1177/18393349211022044

Liang, Y., Li, Y., & Shin, B.-S. (2022). Decentralized crowdsourcing for human intelligence tasks with efficient on-chain cost. *Proceedings of the VLDB Endowment*, *15*(9), 1875–1888. https://doi.org/10.14778/3538598.3538609

Liao, C., To, P.-L., & Hsu, F.-C. (2013). Exploring knowledge sharing in virtual communities. *Online Information Review*.

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442. https://doi.org/10.3758/s13428-016-0727-z

Litman, L., Robinson, J., & Rosenzweig, C. (2015). The relationship between motivation, monetary compensation, and data quality among US-and India-based workers on Mechanical Turk. *Behavior Research Methods*, *47*(2), 519–528.

Liu, K., Liu, Y., Kou, Y., Yang, X., & Hu, G. (2023). Formation mechanism for collaborative behaviour among stakeholders in megaprojects based on the theory of planned behaviour. *Building Research & Information*, *51*(6), 667–681. https://doi.org/10.1080/09613218.2023.2188444

Loepp, E., & Kelly, J. T. (2020). Distinction without a difference? An assessment of MTurk Worker types. *Research & Politics*, *7*(1), 2053168019901185. https://doi.org/10.1177/2053168019901185

Luo, C., Lan, Y., (Robert) Luo, X., & Li, H. (2021). The effect of commitment on knowledge sharing: An empirical study of virtual communities. *Technological Forecasting and Social Change*, *163*, 120438. https://doi.org/10.1016/j.techfore.2020.120438

Maaß, L., Zhang, X., & Klinger, J. (2020). Discrete Event Simulation in einer Notaufnahme. In W. Rogowski (Ed.), *Management im Gesundheitswesen* (pp. 171–184). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-26982-1_13

Macal, C. M., & North, M. J. (2009). Agent-based modeling and simulation. *Proceedings of the 2009 Winter Simulation Conference (WSC)*, 86–98. https://doi.org/10.1109/WSC.2009.5429318

Maffie, M. D. (2020). Are we 'sharing' or 'gig-ing'? A classification system for online platforms. *Industrial Relations Journal*, *51*(6), 536–555. https://doi.org/10.1111/irj.12312

Maharani, R. S. (2017). The Influence of Motivation and Social Capital to Knowledge Sharing in Online Communities: Study on Female Daily Online Community. *1st Indonesia International Graduate Conference on Communication (IGCC) 2017*.

Majid, M. A. (2011). *Human behaviour modelling: An investigation using traditional discrete event and combined discrete event and agent-based simulation*.

Margaryan, A. (2016). Understanding crowdworkers' learning practices. *The Internet, Policy and Politics Biannual Conference*.

Margaryan, A. (2019). Workplace learning in crowdwork: Comparing microworkers' and online freelancers' practices. *Journal of Workplace Learning*, *31*(4), 250–273. https://doi.org/10.1108/JWL-10-2018-0126

Marikyan, D., & Papagiannidis, S. (2021a). *Technology Acceptance Model: A review*. https://open.ncl.ac.uk/theories/1/technology-acceptance-model/

Marikyan, D., & Papagiannidis, S. (2021b). Unified Theory of Acceptance and Use of Technology. *TheoryHub Book*, 16.

Marouf, L. (2016). The role of knowledge sharing culture in business performance. *VINE Journal of Information and Knowledge Management Systems*.
Marshall, C. C., Goguladinne, P. S. R., Maheshwari, M., Sathe, A., & Shipman, F. M. (2023). Who Broke Amazon Mechanical Turk?: An Analysis of Crowdsourcing Data Quality over Time. *Proceedings of the 15th ACM Web Science Conference 2023*, 335–345. https://doi.org/10.1145/3578503.3583622
Martin, D., Hanrahan, B. V., O'Neill, J., & Gupta, N. (2014). Being A Turker. *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 224–235. https://doi.org/10.1145/2531602.2531663
Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*(1), 1–23.
Matherly, T. (2019). A panel for lemons? Positivity bias, reputation systems and data quality on MTurk. *European Journal of Marketing*, *53*(2), 195–223. https://doi.org/10.1108/EJM-07-2017-0491
Matli, W., & Wamba, S. F. (2023). Work from anywhere: Inequalities in technology infrastructure distribution for digit workers. *Digital Transformation and Society*, *2*(2), 149–162. https://doi.org/10.1108/DTS-08-2022-0042
Maximiano, S. (2017). Measuring Reciprocity: Do Survey and Experimental Data Correlate? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3589577
McInnis, B., Cosley, D., Nam, C., & Leshed, G. (2016). Taking a HIT: Designing around Rejection, Mistrust, Risk, and Workers' Experiences in Amazon Mechanical Turk. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2271–2282. https://doi.org/10.1145/2858036.2858539
Merton, R. K. (1968). The Matthew effect in science: The reward and communication systems of science are considered. *Science*, *159*(3810), 56–63.
Min, Q., Ji, S., & Qu, G. (2008). Mobile commerce user acceptance study in China: A revised UTAUT model. *Tsinghua Science & Technology*, *13*, 257–264. https://doi.org/10.1016/S1007-0214(08)70042-7
Mingers, J. (2003). A classification of the philosophical assumptions of management science methods. *Journal of the Operational Research Society*, *54*(6), 559–570. https://doi.org/10.1057/palgrave.jors.2601436
Mohammadi, A., & Hashemi Golpayegani, S. A. (2021). A Financial Incentive Mechanism for Truthful Reporting Assurance in Online Crowdsourcing Platforms. *Journal of Theoretical and Applied Electronic Commerce Research*, *16*(6), 2014–2030. https://doi.org/10.3390/jtaer16060113
Mok, R., Chang, R., & Li, W. (2016). Detecting Low-Quality Workers in QoE Crowdtesting: A Worker Behavior-Based Approach. *IEEE Transactions on Multimedia*, *PP*, 1–1. https://doi.org/10.1109/TMM.2016.2619901
Mooradian, T., Renzl, B., & Matzler, K. (2006). Who trusts? Personality, trust and knowledge sharing. *Management Learning*, *37*(4), 523–540.
Moore, G. C., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research*, *2*(3), 192–222.
Moss, A. J., Rosenzweig, C., Robinson, J., & Litman, L. (2020). Demographic Stability on Mechanical Turk Despite COVID-19. *Trends in Cognitive Sciences*, *24*(9), 678–680. https://doi.org/10.1016/j.tics.2020.05.014
Müller, K., Vignaux, T., Lünsdorf, O., & Scherfke, S. (2021). *simpy: Event discrete, process based simulation for Python.* (4.0.1) [Python; OS Independent]. https://simpy.readthedocs.io

Mustafee, N., Katsaliaki, K., & Taylor, S. J. (2010). Profiling literature in healthcare simulation. *Simulation*, *86*(8–9), 543–558.
Mustapha, N. N. S. N., & Shamsudin, M. F. (2020). The power of reciprocity theory in marketing. *Journal of Postgraduate Current Business Research*, *5*(1).
Myers, R. H. (2000). *Classical and modern regression with applications* (2. ed). Duxbury.
Naderi, B., Zequeira Jiménez, R., Hirth, M., Möller, S., Metzger, F., & Hoßfeld, T. (2021). Towards speech quality assessment using a crowdsourcing approach: Evaluation of standardized methods. *Quality and User Experience*, *6*(1), 2. https://doi.org/10.1007/s41233-020-00042-1
Nepal, S., & Nepal, B. (2023). Adoption of Digital Banking: Insights from a UTAUT Model. *Journal of Business and Social Sciences Research*, *8*(1), 17–34. https://doi.org/10.3126/jbssr.v8i1.56580
Nguyen, M., Fry, M.-L., & Arli, D. (2023). Share or Not Share Knowledge: Posters Versus Lurkers in Organisational Online Knowledge Sharing and Internal Marketing. *Australasian Marketing Journal*, 144135822311670. https://doi.org/10.1177/14413582231167035
Nguyen, T.-M. (2021). Four-dimensional model: A literature review in online organisational knowledge sharing. *VINE Journal of Information and Knowledge Management Systems*, *51*(1), 109–138. https://doi.org/10.1108/VJIKMS-05-2019-0077
Nguyen, T.-M., Ngo, L. V., & Gregory, G. (2022). Motivation in organisational online knowledge sharing. *Journal of Knowledge Management*, *26*(1), 102–125. https://doi.org/10.1108/JKM-09-2020-0664
Nguyen, T.-M., Nham, T. P., Froese, F. J., & Malik, A. (2019). Motivation and knowledge sharing: A meta-analysis of main and moderating effects. *Journal of Knowledge Management*, *23*(5), 998–1016. https://doi.org/10.1108/JKM-01-2019-0029
Nurhidayati, N. A., & Zaenuri, N. A. (2023). Innovation behaviour improvement strategy through knowledge-sharing behaviour based on knowledge-oriented leadership and knowledge-sharing climate. *International Journal of Knowledge Management Studies*, *14*(3), 283–306. https://doi.org/10.1504/IJKMS.2023.132043
Oanţă, R. M. (2020). Knowledge Sharing Through Platform Sharing. *Scientific Bulletin*, *25*(1), 41–49. https://doi.org/10.2478/bsaft-2020-0006
Oelen, A. (2022). *Leveraging human-computer interaction and crowdsourcing for scholarly knowledge graph creation* [DoctoralThesis, Hannover : Institutionelles Repositorium der Leibniz Universität Hannover]. https://doi.org/10.15488/13066
Onaolapo, S., & Oyewole, O. (2018). Performance Expectancy, Effort Expectancy, and Facilitating Conditions as Factors Influencing Smart Phones Use for Mobile Learning by Postgraduate Students of the University of Ibadan, Nigeria. *Interdisciplinary Journal of E-Skills and Lifelong Learning*, *14*, 095–115. https://doi.org/10.28945/4085
O'Neil, K. M., Penrod, S. D., & Bornstein, B. H. (2003). Web-based research: Methodological variables' effects on dropout and sample characteristicsWeb-based research: Methodological variables' effects on dropout and sample characteristics. *Faculty Publications, Department of Psychology*, 324.
Osterbrink, L., & Alpar, P. (2021). Silence of crowdworkers—Reasons and implications for work conditions and quality. *International Studies of Management & Organization*, *51*(2), 136–161. https://doi.org/10.1080/00208825.2021.1927311
Oyong, S., & Ekong, V. (2019). An explorative survey of formal and agile software development methods. *Global Journal of Pure and Applied Sciences*, *25*(1), 71–79.
Özgün, O., & Barlas, Y. (2009). *Discrete vs. Continuous Simulation: When Does It Matter?*

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Park, S., & Kim, E.-J. (2018). Fostering organizational learning through leadership and knowledge sharing. *Journal of Knowledge Management*, *22*(6), 1408–1423.

Park, Y. S., Konge, L., & Artino, A. R. (2020). The Positivism Paradigm of Research: *Academic Medicine*, *95*(5), 690–694. https://doi.org/10.1097/ACM.0000000000003093

Participation Agreement. (2020, March 25). https://www.mturk.com/participation-agreement

Patel, A., Berdunov, V., Quayyum, Z., King, D., Knapp, M., & Wittenberg, R. (2020). Estimated societal costs of stroke in the UK based on a discrete event simulation. *Age and Ageing*, *49*(2), 270–276. https://doi.org/10.1093/ageing/afz162

Paulino, D., Correia, A., Reis, A., Guimarães, D., Rudenko, R., Nunes, C., Silva, T., Barroso, J., & Paredes, H. (2022). Cognitive Personalization in Microtask Design. In M. Antona & C. Stephanidis (Eds.), *Universal Access in Human-Computer Interaction. Novel Design Approaches and Technologies* (pp. 79–96). Springer International Publishing. https://doi.org/10.1007/978-3-031-05028-2_5

Paulino, D., Guimarães, D., Correia, A., Ribeiro, J., Barroso, J., & Paredes, H. (2023). A Model for Cognitive Personalization of Microtask Design. *Sensors*, *23*(7), Article 7. https://doi.org/10.3390/s23073571

Pei, W., Mayer, A., Tu, K., & Yue, C. (2020). Attention Please: Your Attention Check Questions in Survey Studies Can Be Automatically Answered. *Proceedings of The Web Conference 2020*, 1182–1193. https://doi.org/10.1145/3366423.3380195

Perugini, M., Gallucci, M., Presaghi, F., & Ercolani, A. P. (2003). The personal norm of reciprocity. *European Journal of Personality*, *17*(4), 251–283. https://doi.org/10.1002/per.474

Petrović, N., Moyà-Alcover, G., Varona, J., & Jaume-i-Capó, A. (2020). Crowdsourcing human-based computation for medical image analysis: A systematic literature review. *Health Informatics Journal*, *26*(4), 2446–2469. https://doi.org/10.1177/1460458220907435

Pi, S.-M., Chou, C.-H., & Liao, H.-L. (2013). A study of Facebook Groups members' knowledge sharing. *Computers in Human Behavior*, *29*(5), 1971–1979. https://doi.org/10.1016/j.chb.2013.04.019

Piasna, A., & Myant, M. (2017). *Myths of employment deregulation: How it neither creates jobs nor reduces labour market segmentation—*. 255.

Poncet, A., Courvoisier, D. S., Combescure, C., & Perneger, T. V. (2016). Normality and Sample Size Do Not Matter for the Selection of an Appropriate Statistical Test for Two-Group Comparisons. *Methodology*, *12*(2), 61–71. https://doi.org/10.1027/1614-2241/a000110

Ponciano, L., & Brasileiro, F. (2015). Finding volunteers' engagement profiles in human computation for citizen science projects. *arXiv Preprint arXiv:1501.02134*.

Ponis, S. T., Delis, A., Gayialis, S. P., Kasimatis, P., & Tan, J. (2013). Applying discrete event simulation (DES) in healthcare: The case for outpatient facility capacity planning. *International Journal of Healthcare Information Systems and Informatics (IJHISI)*, *8*(3), 58–79.

Popiel, P. (2017). "Boundaryless" in the creative economy: Assessing freelancing on Upwork. *Critical Studies in Media Communication*, *34*(3), 220–233. https://doi.org/10.1080/15295036.2017.1282618

Posch, L., Bleier, A., Flöck, F., Lechner, C. M., Kinder-Kurlanda, K., Helic, D., & Strohmaier, M. (2022). Characterizing the Global Crowd Workforce: A Cross-Country Comparison of Crowdworker Demographics. *Human Computation*, *9*(1). https://doi.org/10.15346/hc.v9i1.106

Powers, D. (2011). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies*, *2*(1), 37–63.

Pring, R. (2004). *The philosophy of education*. Bloomsbury Publishing.

Procter, L., Angus, D. J., Blaszczynski, A., & Gainsbury, S. M. (2019). Understanding use of consumer protection tools among Internet gambling customers: Utility of the Theory of Planned Behavior and Theory of Reasoned Action. *Addictive Behaviors*, *99*, 106050. https://doi.org/10.1016/j.addbeh.2019.106050

Qiu, S., Bozzon, A., Birk, M. V., & Gadiraju, U. (2021). Using Worker Avatars to Improve Microtask Crowdsourcing. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 322:1-322:28. https://doi.org/10.1145/3476063

Ramadhiani, F. R., & Adnan, R. S. (2023). The Role of Social Capital and Knowledge Capital for Freelance Creative Workers in Audiovisual Industry:-. *The Journal of Society and Media*, *7*(1), 19–48.

Ramirez, J. (2023). *PandaCrazy Max* (0.9.31) [Computer software]. https://github.com/JohnnyRS/PandaCrazy-Max

Rani, U., & Furrer, M. (2019). On-Demand Digital Economy: Can Experience Ensure Work and Income Security for Microtask Workers? *Jahrbücher Für Nationalökonomie Und Statistik*, *239*(3), 565–597. https://doi.org/10.1515/jbnst-2018-0019

Reschke, B. P., Azoulay, P., & Stuart, T. E. (2018). Status Spillovers: The Effect of Status-conferring Prizes on the Allocation of Attention. *Administrative Science Quarterly*, *63*(4), 819–847. https://doi.org/10.1177/0001839217731997

Reynolds, J., & Kincaid, R. (2023). Gig Work and the Pandemic: Looking for Good Pay from Bad Jobs During the COVID-19 Crisis. *Work and Occupations*, *50*(1), 60–96. https://doi.org/10.1177/07308884221128511

Ribeiro, F., Florêncio, D., Zhang, C., & Seltzer, M. (2011). Crowdmos: An approach for crowdsourcing mean opinion score studies. *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2416–2419.

Ringle, C. M., Wende, S., & Becker, J.-M. (2022). *SmartPLS 4*. SmartPLS. https://www.smartpls.com/

Robinson, J., Rosenzweig, C., Moss, A. J., & Litman, L. (2019). Tapped out or barely tapped? Recommendations for how to harness the vast and largely unused potential of the Mechanical Turk participant pool. *PLOS ONE*, *14*(12), e0226394. https://doi.org/10.1371/journal.pone.0226394

Rothwell, S., Carter, S., Elshenawy, A., & Braga, D. (2016). Job Complexity and User Attention in Crowdsourcing Microtasks. *AAAI*, 6.

Roubertoux, P. L., Tordjman, S., Caubit, X., Di Cristopharo, J., Ghata, A., Fasano, L., Kerkerian-Le Goff, L., Gubellini, P., & Carlier, M. (2020). Construct Validity and Cross Validity of a Test Battery Modeling Autism Spectrum Disorder (ASD) in Mice. *Behavior Genetics*, *50*(1), 26–40. https://doi.org/10.1007/s10519-019-09970-x

Rybina, L., & Garkavenko, V. (2020). Theory of reasoned action and antismoking social marketing. *The Journal of Economic Research & Business Administration*, *132*(2). https://doi.org/10.26577/be.2020.v132.i2.10

Rzeszotarski, J., & Kittur, A. (2012). CrowdScape: Interactively visualizing user behavior and output. *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, 55–62. https://doi.org/10.1145/2380116.2380125

Rzeszotarski, J. M., & Kittur, A. (2011). Instrumenting the crowd: Using implicit behavioral measures to predict task performance. *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, 13–22.

Safran, M., & Che, D. (2018). Efficient Learning-Based Recommendation Algorithms for Top-*N* Tasks and Top-*N* Workers in Large-Scale Crowdsourcing Systems. *ACM Transactions on Information Systems*, *37*(1), 2:1-2:46. https://doi.org/10.1145/3231934

Saito, S., Chiang, C.-W., Savage, S., Nakano, T., Kobayashi, T., & Bigham, J. P. (2019). *Predicting the Working Time of Microtasks Based on Workers' Perception of Prediction Errors*. 28.

Sajjadi, M. S., Bachem, O., Lucic, M., Bousquet, O., & Gelly, S. (2018). Assessing generative models via precision and recall. *Advances in Neural Information Processing Systems*, *31*.

Sannon, S., & Cosley, D. (2019). Privacy, Power, and Invisible Labor on Amazon Mechanical Turk. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3290605.3300512

Saravanos, A., Zervoudakis, S., Zheng, D., Stott, N., Hawryluk, B., & Delfino, D. (2021). The hidden cost of using Amazon Mechanical Turk for research. *HCI International 2021-Late Breaking Papers: Design and User Experience: 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings 23*, 147–164.

Saremi, R., Yang, Y., Vesonder, G., Ruhe, G., & Zhang, H. (2021). *CrowdSim: A Hybrid Simulation Model for Failure Prediction in Crowdsourced Software Development* (arXiv:2103.09856). arXiv. http://arxiv.org/abs/2103.09856

Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students*. Pearson education.

Savage, S., Chiang, C., Saito, S., Toxtli, C., & Bigham, J. (2020). *Becoming the Super Turker:Increasing Wages via a Strategy from High Earning Workers*. 12.

Scholz, T. (2016). Platform Cooperativism: Challenging the Corporate Sharing Economy. *Rosa Luxemburg Stiftung*.

Schulte, J., Schlicher, K. D., & Maier, G. W. (2020). Working everywhere and every time?—Chances and risks in crowdworking and crowdsourcing work design. *Gruppe. Interaktion. Organisation. Zeitschrift Für Angewandte Organisationspsychologie (GIO)*, *51*(1), 59–69. https://doi.org/10.1007/s11612-020-00503-3

Schultz, R. (2020). *Stax* [Computer software]. https://gitlab.com/StructByLightning/stax

Sedighi, M., Van Splunter, S., Brazier, F., Van Beers, C., & Lukosch, S. (2016). Exploration of multi-layered knowledge sharing participation: The roles of perceived benefits and costs. *Journal of Knowledge Management*, *20*(6), 1247–1267. https://doi.org/10.1108/JKM-01-2016-0044

Seneta, E. (2013). A tricentenary history of the law of large numbers. *Bernoulli*, *19*(4), 1088–1121.

Shaharuddin, W. Y. W., Ahmad, M., Omar, S., & See, Y. (2022). Gossip as a medium of informal communication: A measurement model of gossip engagement at the workplace. *SEARCH Journal of Media and Communication Research*.

Shateri, K., & Hayat, A. (2020). Investigating the mediating role of organizational trust in the relationship between perceived organizational support and Knowledge sharing. *Knowledge Management & E-Learning: An International Journal*, 298–314. https://doi.org/10.34105/j.kmel.2020.12.016

Shen, H., Jin, H., Cabrera, Á. A., Perer, A., Zhu, H., & Hong, J. I. (2020). Designing Alternative Representations of Confusion Matrices to Support Non-Expert Public Understanding of Algorithm Performance. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW2), 1–22. https://doi.org/10.1145/3415224

Shen, W., Hu, Y. J., & Ulmer, J. R. (2015). Competing for Attention: An empirical study of online reviewers' strategic behavior. *Mis Quarterly*, *39*(3), 683–696.

Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Homayouni, S. (2020). Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 6308–6325. https://doi.org/10.1109/JSTARS.2020.3026724

Siebers, P.-O., Aickelin, U., Celia, H., & Clegg, C. W. (2014). Towards the development of a simulator for investigating the impact of people management practices on retail performance. *Springer*.

Siebers, P.-O., Macal, C. M., Garnett, J., Buxton, D., & Pidd, M. (2010). Discrete-event simulation is dead, long live agent-based simulation! *Journal of Simulation*, *4*(3), 204–210.

Silberman, M. S. (2015). Operating an employer reputation system: Lessons from Turkopticon, 2008-2015. *Comp. Lab. L. & Pol'y J.*, *37*, 505.

Silberman, M. S., Tomlinson, B., LaPlante, R., Ross, J., Irani, L., & Zaldivar, A. (2018). Responsible research with crowds: Pay crowdworkers at least minimum wage. *Communications of the ACM*, *61*(3), 39–41. https://doi.org/10.1145/3180492

Silva, A. A., Tavares, M. W., Carrasquilla, A., Misságia, R., & Ceia, M. (2020). Petrofacies classification using machine learning algorithms. *Geophysics*, *85*(4), WA101–WA113.

Siraj-Ud-Doulah, Md. (2021). An Alternative Measures of Moments Skewness Kurtosis and JB Test of Normality: *Journal of Statistical Theory and Applications*, *20*(2), 219. https://doi.org/10.2991/jsta.d.210525.002

Smith, D., & Srinivas, S. (2019). A simulation-based evaluation of warehouse check-in strategies for improving inbound logistics operations. *Simulation Modelling Practice and Theory*, *94*, 303–320.

Sodré, I., & Brasileiro, F. (2017). An Analysis of the Use of Qualifications on the Amazon Mechanical Turk Online Labor Market. *Computer Supported Cooperative Work (CSCW)*, *26*(4–6), 837–872. https://doi.org/10.1007/s10606-017-9283-z

Sood, S., Antin, J., & Churchill, E. (2012). Profanity use in online communities. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1481–1490. https://doi.org/10.1145/2207676.2208610

Spurk, D., & Straub, C. (2020). Flexible employment relationships and careers in times of the COVID-19 pandemic. *Journal of Vocational Behavior*, *119*, 103435. https://doi.org/10.1016/j.jvb.2020.103435

Stafford, L. (2017). Social Exchange Theory. In D. O. Braithwaite, E. A. Suter, & K. Floyd (Eds.), *Engaging Theories in Family Communication* (2nd ed., pp. 279–289). Routledge. https://doi.org/10.4324/9781315204321-25

Stok, F. M., De Vet, E., De Wit, J. B., Luszczynska, A., Safron, M., & De Ridder, D. T. (2015). The proof is in the eating: Subjective peer norms are associated with adolescents' eating behaviour. *Public Health Nutrition*, *18*(6), 1044–1051. https://doi.org/10.1017/S1368980014001268

Suárez-Álvarez, J., Pedrosa, I., & Lozano, L. M. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema*, *30.2*, 149–158. https://doi.org/10.7334/psicothema2018.33

Sujati, H., & Gunarhadi, S. (2020). Testing the construct validity and reliability of curiosity scale using confirmatory factor analysis. *Journal of Educational and Social Research*, *20*(4).

Sumari, S., Ibrahim, R., Hawaniah Zakaria, N., & Hamijah Ab Hamid, A. (2013). Comparing Three Simulation Model Using Taxonomy: System Dynamic Simulation, Discrete Event Simulation and Agent Based Simulation. *International Journal of Management Excellence*, *1*(3), 54–59. https://doi.org/10.17722/ijme.v1i3.17

Suprunenko, O. (2021). Combined approach architecture development to simulation modeling of systems with parallelism. *Eastern-European Journal of Enterprise Technologies*, *4*(4(112)), 74–82. https://doi.org/10.15587/1729-4061.2021.239212

Sveen, A. F., Erichsen, A. S. S., & Midtbø, T. (2020). Micro-tasking as a method for human assessment and quality control in a geospatial data import. *Cartography and Geographic Information Science*, *47*(2), 141–152. https://doi.org/10.1080/15230406.2019.1659187

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics (Vol. 5)*. Boston, MA: pearson.

Taherdoost, H. (2016). Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3205040

Tam, K.-P., Leung, A. K. -y, & Clayton, S. (2021). Research on climate change in social psychology publications: A systematic review. *Asian Journal of Social Psychology*, *24*(2), 117–143. https://doi.org/10.1111/ajsp.12477

Tang, J.-H., & Yang, H.-L. (2005). User role and perception of requirements in a web-based community of practice. *Online Information Review*.

Tang, W., Yin, M., & Ho, C.-J. (2019). Leveraging Peer Communication to Enhance Crowdsourcing. *The World Wide Web Conference*, 1794–1805. https://doi.org/10.1145/3308558.3313554

Tavakol, M., & Wetzel, A. (2020). Factor Analysis: A means for theory and instrument development in support of construct validity. *International Journal of Medical Education*, *11*, 245–247. https://doi.org/10.5116/ijme.5f96.0f4a

Teevan, J., Iqbal, S. T., & von Veh, C. (2016). Supporting Collaborative Writing with Microtasks. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2657–2668. https://doi.org/10.1145/2858036.2858108

Tenenhaus, M., Amato, S., & Vinzi, V. E. (2004). A global Goodness–of–Fit index for PLS structural equation modelling. *Proceedings of the XLII SIS Scientific Meeting*, 739–742.

Tentama, F., & Anindita, W. D. (2020). Employability scale: Construct validity and reliability. *International Journal of Scientific & Technology Research*, *9*(4), 3166–3170.

Tesar, J. (2020). How Do Social Norms and Expectations About Others Influence Individual Behavior?: A Quantum Model of Self/Other-Perspective Interaction in Strategic Decision-Making. *Foundations of Science*, *25*(1), 135–150. https://doi.org/10.1007/s10699-019-09582-y

Tønnessen, Ø., Dhir, A., & Flåten, B.-T. (2021). Digital knowledge sharing and creative performance: Work from home during the COVID-19 pandemic. *Technological Forecasting and Social Change*, *170*, 120866. https://doi.org/10.1016/j.techfore.2021.120866

Toxtli, C., Suri, S., & Savage, S. (2021). Quantifying the Invisible Labor in Crowd Work. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 319:1-319:26. https://doi.org/10.1145/3476060

Treviso, F., Trinchero, R., & Canavero, F. G. (2021). Bayesian Optimization of Hyperparameters in Kernel-Based Delay Rational Models. *IEEE Electromagnetic Compatibility Magazine*, *10*(2), 90–93. https://doi.org/10.1109/MEMC.2021.9477255

Tseng, H., Yeh, H.-T., & Tang, Y. (2019). A Close Look at Trust Among Team Members in Online Learning Communities: *International Journal of Distance Education Technologies*, *17*(1), 52–65. https://doi.org/10.4018/IJDET.2019010104

Turnitsa, C., Padilla, J. J., & Tolk, A. (2010). Ontology for Modeling and Simulation. *Proceedings of the 2010 Winter Simulation Conference*, 643–651. https://doi.org/10.1109/WSC.2010.5679124

Uchiyama, Y., Furuoka, F., & Akhir, M. N. M. (2022). Gig Workers, Social Protection and Labour Market Inequality: Lessons from Malaysia. *Jurnal Ekonomi Malaysia*, *56*(3), 165–184.

Ulfha, S., Rosmadi, M., Widiastuti, E., & Raspati, M. I. (2019). Role of Knowledge Sharing and Organizational Commitment to Development Micro Small and Medium Enterprises (MSMEs). *Budapest International Research and Critics Institute-Journal (BIRCI-Journal)*, 307–312.

Using the sandbox. (2021). Amazon Mechanical Turk Developer Guide. https://docs.aws.amazon.com/AWSMechTurk/latest/AWSMechanicalTurkRequester/mturk-use-sandbox.html

Uzor, S., Jacques, J. T., Dudley, J. J., & Kristensson, P. O. (2021). Investigating the Accessibility of Crowdwork Tasks on Mechanical Turk. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3411764.3445291

Vahdat, S., Rajabion, L., Naeini, S., Hassani, A., & Charband, Y. (2020). A new model for investigation the impact of social media and hospital climate on the intention toward medical knowledge sharing: The mediating role of cultural willingness. *VINE Journal of Information and Knowledge Management Systems*.

Van Den Besselaar, P., Sandström, U., & Mom, C. (2019). Recognition through performance and reputation. *ISSI*, 2065–2069.

Van Lier, T., Caris, A., & Macharis, C. (2016). Sustainability SI: Bundling of outbound freight flows: Analyzing the potential of internal horizontal collaboration to improve sustainability. *Networks and Spatial Economics*, *16*, 277–302.

Van Looy, B., Ranga, M., Callaert, J., Debackere, K., & Zimmermann, E. (2004). Combining entrepreneurial and scientific performance in academia: Towards a compounded and reciprocal Matthew-effect? *Research Policy*, *33*(3), 425–441.

Vancea, D. C. P., Nemirschi, A., & others. (2020). Challenges in Using AI in Online Commerce. *Ovidius University Annals, Economic Sciences Series*, *20*(2), 559–563.

Vanneste, D., Vermeulen, B., & Declercq, A. (2013). Healthcare professionals' acceptance of BelRAI, a web-based system enabling person-centred recording and data sharing across care settings with interRAI instruments: A UTAUT analysis. *BMC Medical Informatics and Decision Making*, *13*(1), 129. https://doi.org/10.1186/1472-6947-13-129

Varshney, M., Pickett, K., & Bagrodia, R. (2011). A Live-Virtual-Constructive (LVC) framework for cyber operations test, evaluation and training. *2011 - MILCOM 2011 Military Communications Conference*, 1387–1392. https://doi.org/10.1109/MILCOM.2011.6127499

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 425–478.

Vinzi, V. E., Chin, W. W., Henseler, J., Wang, H., & others. (2010). *Handbook of partial least squares* (Vol. 201). Springer.

Wagner, N., & Agrawal, V. (2014). An agent-based simulation system for concert venue crowd evacuation modeling in the presence of a fire disaster. *Expert Systems with Applications*, *41*(6), 2807–2815.

Wakeland, W. W., Gallaher, E. J., Macovsky, L. M., & Aktipis, C. A. (2004). A comparison of system dynamics and agent-based simulation applied to the study of cellular receptor dynamics. *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of The*, 10 pp.-. https://doi.org/10.1109/HICSS.2004.1265242

Waldkirch, M., Bucher, E., Schou, P. K., & Grünwald, E. (2021). Controlled by the algorithm, coached by the crowd–how HRM activities take shape on digital work platforms in the gig economy. *The International Journal of Human Resource Management*, 1–36.

Wallace, L. G., & Sheetz, S. D. (2014). The adoption of software measures: A technology acceptance model (TAM) perspective. *Information & Management*, *51*(2), 249–259. https://doi.org/10.1016/j.im.2013.12.003

Walsham, G. (2006). Doing interpretive research. *European Journal of Information Systems*, *15*(3), 320–330.

Walters, K., Christakis, D. A., & Wright, D. R. (2018). Are Mechanical Turk worker samples representative of health status and health behaviors in the US? *PloS One*, *13*(6), e0198835.

Wan, J., Lu, Y., Wang, B., & Zhao, L. (2017). How attachment influences users' willingness to donate to content creators in social media: A socio-technical systems perspective. *Information & Management*, *54*(7), 837–850. https://doi.org/10.1016/j.im.2016.12.007

Wang, E. S.-T., & Wang, M. C.-H. (2013). Social Support and Social Interaction Ties on Internet Addiction: Integrating Online and Offline Contexts. *Cyberpsychology, Behavior, and Social Networking*, *16*(11), 843–849. https://doi.org/10.1089/cyber.2012.0557

Wang, G., Xu, J., & Lin, S. (2023). Influencing factors of using behaviour for computational advertising under the theoretical model of UTAUT. *Journal of Control and Decision*, *10*(1), 19–25. https://doi.org/10.1080/23307706.2022.2082568

Wang, H.-K., Yen, Y.-F., & Tseng, J.-F. (2015). Knowledge sharing in knowledge workers: The roles of social exchange theory and the theory of planned behavior. *Innovation*, *17*(4), 450–465. https://doi.org/10.1080/14479338.2015.1129283

Wang, N., Wang, L., Ma, Z., & Wang, S. (2022). From knowledge seeking to knowledge contribution: A social capital perspective on knowledge sharing behaviors in online Q&A communities. *Technological Forecasting and Social Change*, *182*, 121864. https://doi.org/10.1016/j.techfore.2022.121864

Wang, Y., Wang, K., & Miao, C. (2020). Truth Discovery against Strategic Sybil Attack in Crowdsourcing. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 95–104. https://doi.org/10.1145/3394486.3403052

Waqar, M., ul Haq, M. A., & Usman, M. (2018). Investigating Factors Affecting Knowledge Sharing Intentions among Islamic Banking Employees: A TRA Perspective. *Journal of Islamic Business and Management*, *8*(2).

Wasko, M. M., & Faraj, S. (2005). Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly*, 35–57.

Watwani, K. (2023). *Turk Guru* (1.9.3) [Computer software]. https://www.turk-guru.com

Wei, X., Chen, W., & Zhu, K. (2015). Motivating User Contributions in Online Knowledge Communities: Virtual Rewards and Reputation. *2015 48th Hawaii International Conference on System Sciences*, 3760–3769. https://doi.org/10.1109/HICSS.2015.452

Wessel, M., Thies, F., & Benlian, A. (2017). Opening the Floodgates: The Implications of Increasing Platform Openness in Crowdfunding. *Journal of Information Technology*, *32*(4), 344–360. https://doi.org/10.1057/s41265-017-0040-z

Wibowo, M. A., Widodo, Fachrunnisa, O., Adhiatma, A., Nugroho, M., & Prabowo, Y. (2021). Knowledge Sharing, Innovation Strategy and Innovation Capability: A Systematic Literature Review. *Complex, Intelligent and Software Intensive Systems: Proceedings of the 15th International Conference on Complex, Intelligent and Software Intensive Systems (CISIS-2021)*, 473–483.

Wicaksono, A., & Maharani, A. (2020). The Effect of Perceived Usefulness and Perceived Ease of Use on the Technology Acceptance Model to Use Online Travel Agency. *Journal of Business Management Review*, *1*(5), 313–328. https://doi.org/10.47153/jbmr15.502020

Wiesenberg, M., Godulla, A., Tengler, K., Noelle, I.-M., Kloss, J., Klein, N., & Eeckhout, D. (2020). Key challenges in strategic start-up communication: A systematic literature review and an explorative study. *Journal of Communication Management*.

Williams, A. C., Mark, G., Milland, K., Lank, E., & Law, E. (2019). The Perpetual Work Life of Crowdworkers: How Tooling Practices Increase Fragmentation in Crowdwork. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–28. https://doi.org/10.1145/3359126

Willis, J. W., Jost, M., & Nilakanta, R. (2007). *Foundations of qualitative research: Interpretive and critical approaches*. sage.

Wojtusiak, J., Warden, T., & Herzog, O. (2012). Machine learning in agent-based stochastic simulation: Inferential theory and evaluation in transportation logistics. *Computers & Mathematics with Applications*, *64*(12), 3658–3665.

Wood, A. J., Graham, M., Lehdonvirta, V., & Hjorth, I. (2019). Good Gig, Bad Gig: Autonomy and Algorithmic Control in the Global Gig Economy. *Work, Employment and Society*, *33*(1), 56–75. https://doi.org/10.1177/0950017018785616

Woodcock, J., & Graham, M. (2019). The gig economy. *A Critical Introduction. Cambridge: Polity*.

Wu, W., & Gong, X. (2020). Motivation and sustained participation in the online crowdsourcing community: The moderating role of community commitment. *Internet Research*, *31*(1), 287–314. https://doi.org/10.1108/INTR-01-2020-0008

Wu, Y., Hu, X., Wei, J., & Marinova, D. (2023). The effects of attitudes toward knowledge sharing, perceived social norms and job autonomy on employees' knowledge-sharing intentions. *Journal of Knowledge Management*, *27*(7), 1889–1903. https://doi.org/10.1108/JKM-06-2022-0468

Xia, S., Xiong, Z., Luo, Y., & Dong, L. (2015). A method to improve support vector machine based on distance to hyperplane. *Optik*, *126*(20), 2405–2410. https://doi.org/10.1016/j.ijleo.2015.06.010

Xiao, Y., Zhang, X., & Ordóñez De Pablos, P. (2017). How does individuals' exchange orientation moderate the relationship between transformational leadership and knowledge sharing? *Journal of Knowledge Management*, *21*(6), 1622–1639. https://doi.org/10.1108/JKM-03-2017-0120

Xie, H., Checco, A., & Zamani, E. D. (2023). The Unintended Consequences of Automated Scripts in Crowdwork Platforms: A Simulation Study in MTurk. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-023-10373-x

Xie, H., Maddalena, E., Qarout, R., & Checco, A. (2023). The Dark Side of Recruitment in Crowdsourcing: Ethics and Transparency in Micro-Task Marketplaces. *Computer Supported Cooperative Work (CSCW)*. https://doi.org/10.1007/s10606-023-09464-9

Yan, B., Jian, L., Ren, R., Fulk, J., Sidnam-Mauch, E., & Monge, P. (2021). The Paradox of Interaction: Communication Network Centralization, Shared Task Experience, and the Wisdom of Crowds in Online Crowdsourcing Communities. *Communication Research*, *48*(6), 796–818. https://doi.org/10.1177/0093650220915033

Yan, J., Yu, W., & Zhao, J. L. (2015). How signaling and search costs affect information asymmetry in P2P lending: The economics of big data. *Financial Innovation*, *1*(1), 19. https://doi.org/10.1186/s40854-015-0018-1

Yanagisawa, R., Saito, S., Nakano, T., Kobayashi, T., & Ogawa, T. (2022). PostMe: Unsupervised Dynamic Microtask Posting For Efficient and Reliable Crowdsourcing. *2022 IEEE International Conference on Big Data (Big Data)*, 4049–4054. https://doi.org/10.1109/BigData55660.2022.10020590

Yang, J., & Choi, Y. (2021). LQR-Based Sparsification Algorithms of Consensus Networks. *Electronics*, *10*(9), 1082. https://doi.org/10.3390/electronics10091082

Yang, Z. (2021). Deconstruction of the discourse authority of scientists in Chinese online science communication: Investigation of citizen science communicators on Chinese knowledge sharing networks. *Public Understanding of Science*, *30*(8), 993–1007. https://doi.org/10.1177/09636625211005106

Ye, H. J., & Kankanhalli, A. (2017). Solvers' participation in crowdsourcing platforms: Examining the impacts of trust, and benefit and cost factors. *J. Strateg. Inf. Syst.*, *26*, 101–117.

Yee, M. L. S., & Abdullah, M. S. (2021). A Review of UTAUT and Extended Model as a Conceptual Framework in Education Research. *Jurnal Pendidikan Sains Dan Matematik Malaysia*, *11*, 1–20. https://doi.org/10.37134/jpsmm.vol11.sp.1.2021

Yin, M., Gray, M. L., Suri, S., & Vaughan, J. W. (2016). The Communication Network Within the Crowd. *Proceedings of the 25th International Conference on World Wide Web*, 1293–1303. https://doi.org/10.1145/2872427.2883036

Yoshikawa, K., Wu, C.-H., & Lee, H.-J. (2018). Generalized Social Exchange and Its Relevance to New Era Workplace Relationships. *Industrial and Organizational Psychology*, *11*(3), 486–492. https://doi.org/10.1017/iop.2018.100

Yu, L., Chen, Z., Yao, P., & Liu, H. (2021). A Study on the Factors Influencing Users' Online Knowledge Paying-Behavior Based on the UTAUT Model. *Journal of Theoretical and Applied Electronic Commerce Research*, *16*(5), Article 5. https://doi.org/10.3390/jtaer16050099

Zhang, J., Qi, L., Shi, Y., & Gao, Y. (2022). Generalizable model-agnostic semantic segmentation via target-specific normalization. *Pattern Recognition*, *122*, 108292. https://doi.org/10.1016/j.patcog.2021.108292

Zhang, X. (2018). Application of discrete event simulation in health care: A systematic review. *BMC Health Services Research*, *18*(1), 687. https://doi.org/10.1186/s12913-018-3456-4

Zhang, X., Liu, S., Deng, Z., & Chen, X. (2017). Knowledge sharing motivations in online health communities: A comparative study of health professionals and normal users. *Computers in Human Behavior*, *75*, 797–810. https://doi.org/10.1016/j.chb.2017.06.028

Zhang, Z., Zhu, S., Mink, J., Xiong, A., Song, L., & Wang, G. (2022). Beyond Bot Detection: Combating Fraudulent Online Survey Takers. *Proceedings of the ACM Web Conference 2022*, 699–709. https://doi.org/10.1145/3485447.3512230

Zhao, Y., Gong, X., Lin, F., & Chen, X. (2023). Data Poisoning Attacks and Defenses in Dynamic Crowdsourcing With Online Data Quality Learning. *IEEE Transactions on Mobile Computing*, *22*(5), 2569–2581. https://doi.org/10.1109/TMC.2021.3133365

Zhu, D., & Carterette, B. (2010). An analysis of assessor behavior in crowdsourced preference judgments. *SIGIR 2010 Workshop on Crowdsourcing for Search Evaluation*, 17–20.

Zyskowski, K., & Milland, K. (2018). A Crowded Future: Working against Abstraction on Turker Nation. *Catalyst: Feminism, Theory, Technoscience*, *4*. https://doi.org/10.28968/cftt.v4i2.29581

Appendix A Ethical Considerations

Part 1: Ethics Approval for Study in Chapter 5

The University Of Sheffield.

Downloaded: 31/07/2023
Approved: 27/06/2021

Alessandro Checco
Information School

Dear Alessandro

PROJECT TITLE: Understanding crowd work platform work dynamics
APPLICATION: Reference Number 041062

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 27/06/2021 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 041062 (form submission date: 07/06/2021); (expected project end date: 07/06/2022).
- Participant information sheet 1092918 version 1 (07/06/2021).
- Participant consent form 1092919 version 1 (07/06/2021).

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Your responsibilities in delivering this research project are set out at the end of this letter.

Yours sincerely

Paul Reilly
Ethics Administrator
Information School

Please note the following responsibilities of the researcher in delivering the research project:

- The project must abide by the University's Research Ethics Policy: https://www.sheffield.ac.uk/research-services/ethics-integrity/policy
- The project must abide by the University's Good Research & Innovation Practices Policy: https://www.sheffield.ac.uk/polopoly_fs/1.671066!/file/GRIPPolicy.pdf
- The researcher must inform their supervisor (in the case of a student) or Ethics Administrator (in the case of a member of staff) of any significant changes to the project or the approved documentation.
- The researcher must comply with the requirements of the law and relevant guidelines relating to security and confidentiality of personal data.
- The researcher is responsible for effectively managing the data collected both during and after the end of the project in line with best practice, and any relevant legislative, regulatory or contractual requirements.

Participant Information Sheet

The University Of Sheffield.

You are being invited to take part in a research project. Before you decide whether or not to participate, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. Thank you. The University of Sheffield has approved this project following the GRIP policy, and obtain ethics approval following the Research Ethics procedure. Feel free to contact the research ethics coordinator Dr Paul Reilly for more information, p.j.reilly@sheffield.ac.uk.

Research Project Title: Interaction with Mturk platform

1. **What is the project's purpose?**
   The purpose of this study is to understand the interaction between crowd workers and the crowdsourcing platform.

2. **Why have I been chosen?**
   You have been chosen to participate because you are a worker in the MTurk platform.

3. **Do I have to take part?**
   It is up to you to decide whether or not to take part. If you do decide to take part, you can keep this information sheet (and be asked to click agree on a electronic consent form). You can withdraw at any time without any negative consequences (please inform the researchers, see ne. 13). Please note that dependent on the time of your withdrawal, it might not be possible to remove from the study any input you have provided.

4. **What will happen to me if I take part? What do I have to do?**
   You will be asked to complete a crowdsourcing job, for which you will be compensated.

5. **What are the possible disadvantages and risks of taking part?**
   The risks of taking part in the work is no more than those experienced normal life. You can take breaks and interrupt completely the work at any moment.

6. **What are the possible benefits of taking part?**
   There are no immediate benefits, but there is a monetary compensation.

7. **Will my taking part in this project be kept confidential?**
   All the information that we collect about you during the study will be kept strictly confidential and will only be accessible to members of the research team. You will not be able to be identified in any reports or publications.

8. **What is the legal basis for processing my personal data?**
   We will collect some personal data (such as worker ID, connections web logs). The legal basis for this collection is "a task in the public interest".

9. **What will happen to the data collected, and the results of the research project?**
   The data collected will be used by the project researchers to understand to understand how workers get allocated and complete their work in this crowdsourcing platform.

   All the data collected in this study will only be accessible to the researchers in the project. The results may be published in project reports, journal articles or conference papers. Participants will not be able to be identified in any report or publication.

10. **Who is organising and funding the research?**
    This project is funded by the University of Sheffield.

11. **Who is the Data Controller?**
    The University of Sheffield will act as the Data Controller for this study. This means that the University is responsible for looking after your information and using it properly.

12. **Who has ethically reviewed the project?**
    This project has been ethically approved via the University of Sheffield's Ethics Review Procedure, as administered by the Information School. The University's Research Ethics Committee monitors the application and delivery of the University's Ethics Review Procedure across the University.

13. **Contact for further information**
    If you wish to obtain further information about this study, please do not hesitate to contact us:

    - Dr Alessandro Checco, University of Sheffield (a.checco@sheffield.ac.uk)

Thank you in advance for taking part in the project.

Participant Consent Form

Risks and opportunities of Big Data Project

Taking part in the project I confirm that:

- I have read and understood the project information sheet or the project has been fully explained to me. (If you will answer No to this question please do not proceed with this consent form until you are fully aware of what your participation in the project will mean.)
- I have been given the opportunity to ask questions about the project (by contacting a.checco@sheffield.ac.uk).
- I agree to take part in the project. I understand that taking part in the project will include completing a crowdsourcing job in the Amazon Mechanical Turk platform.
- I understand that my taking part is voluntary and that I can withdraw from the study at any time; I do not have to give any reasons for why I no longer want to take part and there will be no adverse consequences if I choose to withdraw. (Please note that dependent of the time of your withdrawal, it might not be possible to remove from the study any input you have provided, such as contributions to group discussions.)

How my information will be used during and after the project, I confirm that:

- I understand my personal details such as worker ID, IP etc. will not be revealed to people outside the project.
- I understand and agree that my words may be quoted in publications, reports, web pages, and other research outputs. I understand that I will not be named in these outputs unless I specifically request this.
- I understand and agree that other authorised researchers will have access to this data only if they agree to preserve the confidentiality of the information as requested in this form.
- I understand and agree that other authorised researchers may use my data in publications, reports, web pages, and other research outputs, only if they agree to preserve the confidentiality of the information as requested in this form.
- I give permission for the data that I provide to be deposited in the University of Sheffield project repository in Google Drive and it can be used for future research and learning.

So that the information you provide can be used legally by the researchers, I confirm that:

- I agree to assign the copyright I hold in any materials generated as part of this project to The University of Sheffield.

Research Ethics Coordinator: Dr Paul Reilly (p.j.reilly@sheffield.ac.uk), Information School, University of Sheffield, Sheffield, United Kingdom Project contact details for further information: Dr Alessandro Checco (a.checco@sheffield.ac.uk), Information School, University of Sheffield, Sheffield, United Kingdom The template of this consent form has been approved by the University of Sheffield Research Ethics Committee and is available to view here: https://www.sheffield.ac.uk/rs/ethicsandintegrity/ethicspolicy/further-guidance/homepage

Part 2: Ethics Approval for Study in Chapter 6

The University Of Sheffield.

Downloaded: 27/02/2023
Approved: 08/11/2022

Haoyu Xie
Registration number: 190269621
Information School
Programme: Social Science

Dear Haoyu

PROJECT TITLE: Factors influencing the knowledge sharing behaviour among crowdworkers
APPLICATION: Reference Number 049528

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 08/11/2022 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 049528 (form submission date: 04/11/2022); (expected project end date: 28/11/2022).
- Participant information sheet 1111417 version 3 (26/10/2022).
- Participant consent form 1111227 version 2 (29/09/2022).

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Your responsibilities in delivering this research project are set out at the end of this letter.

Yours sincerely

Peter Bath
Ethics Administrator
Information School

Please note the following responsibilities of the researcher in delivering the research project:

- The project must abide by the University's Research Ethics Policy:https://www.sheffield.ac.uk/research-services/ethics-integrity/policy
- The project must abide by the University's Good Research & Innovation Practices Policy: https://www.sheffield.ac.uk/polopoly_fs/1.671066!/file/GRIPPolicy.pdf
- The researcher must inform their supervisor (in the case of a student) or Ethics Administrator (in the case of a member of staff) of any significant changes to the project or the approved documentation.
- The researcher must comply with the requirements of the law and relevant guidelines relating to security and confidentiality of personal data.
- The researcher is responsible for effectively managing the data collected both during and after the end of the project in line with best practice, and any relevant legislative, regulatory or contractual requirements.

Factors influencing the knowledge sharing behaviour on crowdsourcing platforms Consent Form

Please tick the appropriate boxes	**Yes**	**No**
Taking Part in the Project		
I have read and understood the project information sheet dated 10/08/2022 or the project has been fully explained to me. (If you will answer No to this question please do not proceed with this consent form until you are fully aware of what your participation in the project will mean.)	☐	☐
I have been given the opportunity to ask questions about the project.	☐	☐
I agree to take part in the project. I understand that taking part in the project will include a walkthrough of a selected region and answered several questions where some were just selections, while others were text entry questions. The questions are about safety, sense of belonging, community spirit, etc.	☐	☐
I understand that by choosing to participate as a volunteer in this research, this does not create a legally binding agreement nor is it intended to create an employment relationship with the University of Sheffield.	☐	☐
I understand that my taking part is voluntary and that I can withdraw from the study at any time; I do not have to give any reasons for why I no longer want to take part and there will be no adverse consequences if I choose to withdraw.	☐	☐
How my information will be used during and after the project		
I understand my personal details such as name, phone number, address and email address etc. will not be revealed to people outside the project.	☐	☐
I understand and agree that my words may be quoted in publications, reports, web pages, and other research outputs. I understand that I will not be named in these outputs unless I specifically request this.	☐	☐
I understand and agree that other authorised researchers will have access to this data only if they agree to preserve the confidentiality of the information as requested in this form.	☐	☐
I understand and agree that other authorised researchers may use my data in publications, reports, web pages, and other research outputs, only if they agree to preserve the confidentiality of the information as requested in this form.	☐	☐
I give permission for the questionnaire response that I provide to be deposited in the University of Sheffield project repository in Google Drive so it can be used for future research and learning	☐	☐
So that the information you provide can be used legally by the researchers		
I agree to assign the copyright I hold in any materials generated as part of this project to The University of Sheffield.	☐	☐

Name of participant Signature Date

Name of Researcher Haoyu Signature Date

Project contact details for further information:

Lead researcher: Haoyu Xie (hxie5@sheffield.ac.uk)

Supervisor: Dr Suvodeep Mazumdar (s.mazumdar@sheffield.ac.uk); Dr Efpraxia Zamani (e.zamani@sheffield.ac.uk)

Participant Information Sheet

You are being invited to take part in a research project. Before you decide whether or not to participate, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. Thank you.
The University of Sheffield has approved this project following the GRIP policy, and obtain ethics approval following the Research Ethics procedure.

1. Research Project Title:

Factors influencing the knowledge sharing behaviour among crowdworkers

2. Invitation paragraph

The aim of this research is to explore the phenomenon of knowledge being shared among crowdworkers, to understand the crowdworkers' knowledge sharing behaviours, including their perceptions and factors influencing their knowledge sharing practice.
You are being invited to take part in a research project. Before you decide whether or not to participate, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following Information carefully and please get in touch with us if you would like more Information. Take time to decide whether or not you wish to take part.

3. What is the project's purpose?

The purpose is to identify existing practice of knowledge sharing behaviour among crowdworkers, and to understand what are the facilitators and barriers that influence how crowdworkers share knowledge with other crowdworkers. Therefore, your personal background, task preferences, knowledge preferences, and perceptions toward numerous factors knowledge sharing behaviour are needed for a relatively comprehensive understanding and analysis of the research topic.

4. Why have I been chosen?

You have been chosen to participate because you are a crowdworker in this crowdsourcing platform

5. Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part, you can keep this information sheet (and be asked to click agree on an electronic consent form). You can withdraw at any time without any negative consequences. Please note that dependent on the time of your withdrawal, it might not be possible to remove from the study any input you have provided. You will be allowed to withdraw anytime before clicking the final submission button for your survey response, and all the input you have provided will be removed at the time of your withdrawal.

6. What will happen to me if I take part? What do I have to do?

You will be asked to complete a crowdsourcing job, for which you will be compensated.

7. What are the possible disadvantages and risks of taking part?

The risks of taking part in the work Is no more than those experienced in normal life. You can take breaks and Interrupt completely the work at any moment.

8. What are the possible benefits of taking part?

You will be paid for filling in the survey based on UK minimum wage standard.

9. Will my taking part in this project be kept confidential?

All of the information that we collect about you during the study will be kept strictly confidential and will only be accessible to members of the research team. You will not be able to be identified in any reports or publications.

10. What is the legal basis for processing my personal data?

We will collect some personal data (such as worker ID, age, gender). The legal basis for this collection is "a task in the public interest".

11. What will happen to the data collected, and the results of the research project?

The data collected will be used by the project researchers to understand how workers communicate within peers to share knowledge and how the factors influence their behaviours. All the data collected in the study will only be accessible to the researchers in the project. The results may be published in project reports, journal articles or conference papers. Participants will not be able to be identified in any report or publication.

12. Who is organising and funding the research?

The project is funded by the FashionBrain (R/147955)

13. Who is the Data Controller?

The University of Sheffield will act as the Data Controller for the study. This means that the University is responsible for looking after your information and using it properly.

14. Who has ethically reviewed the project?

This project has been ethically approved via the University of Sheffield's Ethics Review Procedure, as administered by the Information School. The University's Research Ethics Committee monitors the application and delivery of the University's Ethics Review procedure across the University.

15. What if something goes wrong and I wish to complain about the research or report a concern or incident?

If you are dissatisfied with any aspect of the research and wish to make a complaint, please contact Haoyu Xie (hxie5@sheffield.ac.uk) in the first instance. If you feel your complaint has not been handled in a satisfactory way you can contact the Head of the Department of the Information School, Prof. Val Gillet (v.gillet@sheffield.ac.uk). If the complaint relates to how your personal data has been handled, you can find information about how to raise a complaint in the University's Privacy Notice: https://www.sheffield.ac.uk/govern/data-protection/privacy/general.

16. Contact for further information

If you wish to obtain further information about this study, please do not hesitate to contact us:

- Haoyu Xie, University of Sheffield (hxie5@sheffield.ac.uk)
- Dr. Suvodeep Mazumdar, University of Sheffield (s.mazumdar@sheffield.ac.uk)
- Dr. Efpraxia Zamani, University of Sheffield (e.zamani@sheffield.ac.uk)

Thank you in advance for taking part in the project.

Appendix B Questionnaire for Knowledge Sharing Study

Participation Instructions

1: Thank you for agreeing to take part in this survey. Please take some time to read through the Information Sheet before you start with this survey.

Please read the following consent form and select 'yes' if you agree to all of the following statements. If you do not agree to all the statements, please select 'no' to exit the survey:

Taking part in the project

I have read and understood the project information sheet dated 08.11.2022 or the project has been fully explained to me.

I have been given the opportunity to ask questions about the project by contacting hxie5@sheffield.ac.uk I agree to take part in the project.

I understand that taking part in the project will include completing a crowdsourcing job in the crowdsourcing platform.

I understand that by choosing to participate as a volunteer in this research, this does not create a legally binding agreement nor is it intended to create an employment relationship with the University of Sheffield.

I understand that my taking part is voluntary and that I can withdraw from the study at any time;

I do not have to give any reasons for why I no longer want to take part and there will be no adverse consequences if I choose to withdraw. (Please note that dependent of the time of your withdrawal, it might not be possible to remove from the study any input you have provided, for example, when all inputs have been aggregated.)

How my information will be used during and after the project

I understand my personal details such as worker ID, IP etc. will not be revealed to people outside the project.

I understand and agree that my words may be quoted in publications, reports, web pages, and other research outputs.

I understand that I will not be named in these outputs unless I specifically request this.

I understand and agree that other authorised researchers will have access to this data only if they agree to preserve the confidentiality of the information as requested in this form.

I understand and agree that other authorised researchers may use my data in publications, reports, web pages, and other research outputs, only if they agree to preserve the confidentiality of the information as requested in this form.

I give permission for the data that I provide to be deposited in the University of Sheffield project repository in Google Drive so it can be used for future research and learning

So that the information you provide can be used legally by the researchers

I agree to assign the copyright I hold in any materials generated as part of this project to The University of Sheffield.

▢ Yes

▢ No

Social-demographic Background

2: Age: What is your age?

▢ 18-24 years old

▢ 25-34 years old

▢ 35-44 years old

▢ 45-54 years old

▢ 55-64 years old

▢ 65-74 years old

▢ 75 years or older

3: Gender: To which gender identity do you most identify?

If you prefer to self identify, please choose "other" and input your answer.

▢ Woman

▢ Man

▢ I prefer not to say

4: Education: What is the highest degree or level of school you have completed? If currently enrolled, highest degree received.

- ▢ Primary school
- ▢ Secondary school
- ▢ Higher or secondary or further education (A-levels, BTEC, etc.)
- ▢ College or university
- ▢ Post-graduate degree
- ▢ I prefer not to say

5: Would you define yourself as a freelancer or full-time employee?

- ▢ Freelancer
- ▢ Part-time employee
- ▢ Full-time employee

6: How much percentage does the rewards from doing HITs (via MTurk, Appen, Prolific, etc.) contribute to your overall income?

No more than 10% ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 ▢ 6 ▢ 7 ▢ 8 ▢ 9 ▢ 10 100%

Crowdwork Experience

7: What is your HIT approval rate (0~100)?

____

8: What is your average monthly income in USD$ from MTurk?

____

Ways and Frequency of Knowledge Sharing - 1

9: On average, how often do you post/share knowledge in forums, channels, or platforms about crowdwork?

(The knowledge could be a Solution for technical issues, ratings or comments toward HITs and requesters, tutorials of doing specific HITs, etc.)

- ▢ Never
- ▢ Once a week or less
- ▢ Once every two/three days
- ▢ Once every day

▢ Multiple times a day

10: What types of knowledge do you normally share?

Select all that apply

▢ Techniques for finding good HITs, solutions for technical issues, general working strategies

▢ Suggestions of specific HITs, qualifications, job opportunities from other platforms.

▢ Tutorials of doing specific HITs

▢ Ratings or comments toward HITs and requesters

▢ Non-job-related such as casual conversation

11: Where do you normally share knowledge about crowdwork?

Select all that apply

▢ E-mail

▢ Forums (Reddit, Turker Nation, mturk forum, MTurk Crowd, turkopticon, etc.)

▢ Social messaging apps (Facebook messenger, Slack, WhatsApp, Discord, etc.)

▢ Face-to-face

Ways and Frequency of Knowledge Acquisition - 1

12: On average, how often do you seek information / knowledge in forums, channels, or platforms about crowdwork?

(The information could be a Solution for technical issues, ratings or comments toward HITs and requesters, tutorials of doing specific HITs, etc.)

▢ Never

▢ Once a week or less

▢ Once every two/three days

▢ Once every day

▢ Multiple times a day

13: What types of information / knowledge do you normally look for?

Select all that apply

▢ Techniques for finding good HITs, solutions for technical issues, general working strategies

▢ Suggestions of specific HITs, qualifications, job opportunities from other platforms

▢ Tutorials of doing specific HITs

▢ Ratings or comments toward HITs and requesters

▢ Non-job-related such as casual conversation

14: Where do you normally seek those information / knowledge?

Select all that apply

▢ E-mail

▢ Forums (Reddit, Turker Nation, mturk forum, MTurk Crowd, turkopticon, etc.)

▢ Social messaging apps (Facebook messenger, Slack, WhatsApp, Discord, etc.)

▢ Face-to-face

15: On average, how often do you ask questions in forums, channels, or platforms about crowdwork?

▢ Never

▢ Once a week or less

▢ Once every two/three days

▢ Once every day

▢ Multiple times a day

16: How long do you spend browsing the forums and social messaging apps about crowdwork each day? For example, Turker Nation, mturk forum, etc.

▢ Never

▢ No more than 10 minutes

▢ More than 10 minutes but less than 1 hour

▢ More than 1 hour but less than 3 hours

▢ More than 3 hours

Ways and Frequency of Knowledge Sharing - 2

17: If there are other ways of sharing or searching for knowledge, please specify here:

____

UTAUT related questions on your perceptions of knowledge sharing - 1

Knowledge sharing tools: A tool or medium for sharing knowledge, such as Turker forums, Slack channels, browser extensions that you can leave ratings about HITs.

18: Do you use knowledge sharing tools such as forums (MTurk Crowd), Slack channels (Turker Nation), or browser extensions (TurkerViewJS) that you can leave ratings about HITs?

▢ Yes

▢ No

Regarding the knowledge about techniques for finding good HITs (e.g. using HIT catchers), solutions for technical issues (e.g. cannot install script), general working strategies (e.g. ways to access more HITs):

Sharing tools: A tool or medium for sharing knowledge, such as forums (MTurk Crowd), Slack channels (Turker Nation), or browser extensions (TurkerViewJS) that you can leave ratings about HITs.

19: Sharing tools are useful when I share this type of knowledge.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

20: Sharing tools are useful when I get this type of knowledge.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

21: I can effectively share this type of knowledge using the sharing tools

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

22: I can effectively get this type of knowledge using the sharing tools

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

23: Using the sharing tools makes me share this type of knowledge more quickly.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

24: Using the sharing tools makes me get this type of knowledge more quickly.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

25: (Optional) If you do not find it effective or useful to share or get this type of knowledge with sharing tools, can you specify why? How do you want to improve it?

____

____

26: Sharing tools give me relative advantage when I share this type of knowledge.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

27: Sharing tools give me relative advantage when I get this type of knowledge.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

28: It is easy to use the sharing tools to share this type of knowledge.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

29: It is easy to use the sharing tools to get this type of knowledge.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

30: I can easily access sharing tools whenever and wherever I want to share or get this type of knowledge.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

31: Learning to operate the sharing tools is easy for me.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

32: It requires much technical expertise to effectively use sharing tools.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

33: (Optional) If you feel it is not easy to share or get this type of knowledge, can you specify why? How do you want to improve it?

____

____

Regarding the knowledge about techniques for finding good HITs (e.g. using HIT catchers), solutions for technical issues (e.g. cannot install script), general working strategies (e.g. ways to access more HITs):

Sharing tools: A tool or medium for sharing knowledge, such as forums (MTurk Crowd), Slack channels (Turker Nation), or browser extensions (TurkerViewJS) that you can leave ratings about HITs.

34: The platform (MTurk, Prolific, Appen, etc.) believes that I should share this type of knowledge with other crowdworkers.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

35: I accept and carry out the platform's stance for sharing this type of knowledge even though it is different from mine.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

36: Other crowdworkers believe I should share this type of knowledge with them.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

37: I respect and put in practice my colleague's stance for sharing this type of knowledge.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

38: Please select the option with the largest number to show you are not responding randomly.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Aagree

Facilitating Conditions

Sharing tools: A tool or medium for sharing knowledge, such as forums (MTurk Crowd), Slack channels (Turker Nation), or browser extensions (TurkerViewJS) that you can leave ratings about HITs.

39: The sharing tools integrate well with other technologies I use during crowdwork, such as HIT managers, HIT catchers or visual enhancers.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

40: The sharing tools are well supported by the communities or developer, such as providing guidance and maintenance.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

41: The sharing tools fit with my work processes and routines, they also support my work activities and goals.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

42: Given the resources, opportunities and knowledge it takes to use such technologies, it is easy for me to use the forums, channels and plugins for sharing knowledge.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

43: This knowledge sharing study is answered carefully. Please choose the option in the middle.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

44: If you do not use the knowledge sharing tools like TurkView and MTurk forums, can you specify why?

____

____

Reciprocity:

45: I believe other crowdworkers actively share this type of knowledge.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

46: I want to share this type of knowledge with others because they will do the same in return.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

47: It is fair to help each other in forums, channels and platforms.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

48: Sharing this type of knowledge improves my image / reputation within the community.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

49: To what extent do you think sharing knowledge could improve your reputation?

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

Reputation:

50: Sharing this type of knowledge improves others recognition of me.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

51: When I share this type of knowledge, the people I work with respect me.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

52: Have you thought about sharing knowledge due to concerns about how it might affect your reputation?

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

Reward:

53: I feel that sharing this type of knowledge will benefit me directly.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

54: I feel that sharing this type of knowledge will give me satisfaction.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

55: I feel that sharing this type of knowledge will give me enjoyment.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

56: I feel that sharing this type of knowledge will give me valuable information through interaction with peers.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

Social Interaction Ties:

57: It is important to maintain close social relationships with other crowdworkers via the sharing tools.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

58: To what extent do your friends or colleagues support or encourage you to use this technology?

Very unsupportive ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Very supportive

59: I have frequent communication with other crowdworkers.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

Trust:

60: I trust others when sharing this type of knowledge on forums such as MTurk Crowd.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

61: I trust others when sharing this type of knowledge on browser extensions such as TurkerView.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

62: I trust others when sharing this type of knowledge on social apps such as Slack, Facebook or Telegram.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

63: I believe other crowdworkers will value my shared knowledge.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

64: When sharing this type of knowledge with peers, I believe others will not abuse my knowledge or claim it as their own ideas.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

Knowledge Sharing Intention:

65: I am willing to share this type of knowledge with other crowdworkers.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

66: To what extent do you plan to share this type of knowledge via the sharing tools in the future?

Very unlikely ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Very likely

67: From 1 (very unimportant) to 5 (very important), how important is it to you to share this type of knowledge via sharing tools?

Very unimportant ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Very important

68: How likely are you to share your skill-based knowledge with other members via forums / channels / plugins?

Very unimportant ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Very important

69: I believe the colour of the sky is blue. Make sure to select the option with the smallest number.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

Knowledge Sharing Behaviour:

70: When I see questions in the sharing tools (such as forums and social apps) that I can answer, I usually share my knowledge with them.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

71: When I have gained a piece of knowledge worth sharing, I share it immediately via the sharing tools.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

72: I share skill-based knowledge regularly with peers.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

HIT catchers

73: Do you use HIT catching tools (like Panda Crazy Max, Turkmaster, MTurk Suite)?

▢ Yes

▢ No

If you use HIT catching tools

74: Please tick all HIT catching tools you are using recently.

Select all that apply

▢ MTurk Suite

▢ Panda Crazy Max

▢ Turkmaster

▢ Mturk Engine

▢ Stax

▢ Turk Guru

75: When do you use them?

▢ Every time I work on MTurk

▢ Only when I cannot find the HITs I want

▢ Always keep them running whether I am working or not

76: How do you customise settings in your HIT catching tools?

____

____

77: Do you only run them when searching for specific types of HITs? If so, what are these types?

____

____

78: How do you use HIT catchers with other tools (e.g. HIT Tracker, HIT Finder, Queue helper)?

79: If you do not use HIT catching tools, can you explain why?

▢ I haven't heard about them at all

▢ I have heard about them but do not know how to use them

▢ I know how to use them but they are not useful

Ease of getting HITs

80: It is easy to get desired HITs without HIT catching tools.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

81: It is easy to get desired HITs with HIT catching tools.

Strongly Disagree ▢ 1 ▢ 2 ▢ 3 ▢ 4 ▢ 5 Strongly Agree

82: (Optional) If you do not find it easy to get HITs, can you explain why?

Sharing knowledge about tools

83: Do you think sharing scripting related knowledge with others reduce your technical advantages compared to others?

scripting related knowledge could be how to install Panda Crazy, MTurk Suite.

▢ Yes

▢ No

84: Can you specify why?

85: What made you start knowing and using scripting tools?

▢ Recommended by others (forum members, friends)

▢ Read information shared by others via forums or channels

▢ Finding tools online entirely on your own with no recommendations from others

86: What are the most important factors that make you share information about scripting tools?

Select all that apply

▢ Performance expectancy of the forums/channels/tools you use to share information, such as how useful / effective / fast to share information

▢ Effort Expectancy of the forums/channels/tools you use to share information, such as ease of use / access / learning

▢ Platform (MTurk)'s and other Turkers' opinions on whether or not you should share such information

▢ How well the forums/channels/tools you use to share information integrate with your work style, also the community sopport

▢ Reciprocity: belief that everyone would share such information with each other

▢ Reputation: belief that it could improve your reputation

▢ Reward: belief that it could benefit you with money, satisfaction, enjoyment, etc.

▢ Social Interaction Ties: believe the Importance of maintaining relationship via sharing information

▢ Trust others when sharing information via forums/channels/tools

▢ The intention of sharing such information

Before Submission

87: Worker ID: Please provide your MTurk Worker ID for authentication purpose

Worker ID can be found on your MTurk Dashboard or in the upper left corner of the HITs website. Responses without a correct Worker ID will not receive a reward.