Conversational Crowdsourcing

Sihang Qiu, Ujwal Gadiraju, Alessandro Bozzon, Geert-Jan Houben
Web Information Systems
Delft University of Technology
{s.qiu-1, u.k.gadiraju, a.bozzon, g.j.p.m.houben}@tudelft.nl

Abstract

The trend of remote work leads to the prosperity of crowdsourcing marketplaces. In crowdsourcing marketplaces, online workers can select their preferable tasks and then complete them to get paid, while requesters design and publish tasks to acquire their desirable data. The standard user interface of the crowdsourcing task is the web page, where users provide answers using HTML-based web elements, and the task-related information (including instructions and questions) is displayed on a single web page. Although the traditional way of presenting tasks is straightforward, it could negatively affect workers' satisfaction and performance by causing problems such as boredom and fatigue. To address this challenge, we proposed a novel concept — conversational crowdsourcing, which employs conversational interfaces to facilitate crowdsourcing task execution. With conversational crowdsourcing, workers receive task information as messages from a conversational agent, and provide answers by sending messages back to the agent. In this vision paper, we introduce our recent work in terms of using conversational crowdsourcing to improve worker performance and experience by employing novel human-computer interaction affordances. Our findings reveal that conversational crowdsourcing has important implications in improving the worker satisfaction and requester-worker relationship in crowdsourcing marketplaces.

1 Introduction

The world is now experiencing an incredible development of artificial intelligence, machine learning, and robotics. The importance of human input for such novel techniques has been widely acknowledged for building training datasets, evaluating AI systems, carrying out human-related experiments, etc [6, 9]. Crowdsourcing has become a primary means to effectively collect human from anonymous users of the Internet, which leads to the prosperity of crowdsourcing marketplaces, such as Amazon's Mechanical Turk[1], Yandex Toloka[2], and Prolific[3]. The prosperity of crowdsourcing markets has attracted an increasing number of people to work full-time online. In a crowdsourcing marketplace, crowd workers can select and complete tasks, offered by requesters who demands the data, to get paid. Traditionally, crowdsourcing tasks are firstly designed by requesters, and then executed by crowd workers, both based on the web page. Current crowdsourcing-related studies have taken great strides in improving the worker performance and output quality [5], however, the importance of user experience and satisfaction has been underestimated. Considering the great potential of crowdsourcing marketplaces, researchers have identified that the future of crowdsourcing will depend on both organizational performance and worker satisfaction [16]. However, recent studies have

[1] https://www.mturk.com/
[2] https://toloka.yandex.com/
[3] https://www.prolific.co/

NeurIPS 2020 Crowd Science Workshop: Remoteness, Fairness, and Mechanisms as Challenges of Data Supply by Humans for Automation. Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR-WS.org/Vol-2736/paper1.pdf

revealed that crowdsourcing in such a monotonous way can lead to problems such as boredom, fatigue, and high drop-out rates [10, 22]. Such problems can negatively affect worker satisfaction and engagement.

Researchers have attempted to design tasks to engage and motivate workers [19, 29]. However, current motivation designs depend on the task type and context. We still look forward to a general solution that can effectively better engage workers. We have noticed that there has been a rise in the use of conversational interfaces, which can provide a human-like means of interaction between the user and virtual assistants, chatbots, or messaging services. We are also witnessing a rapid proliferation of messaging services such as WhatsApp, Telegram, and Messenger, as smartphones have been extensively used worldwide. The growing familiarity of people with such messaging services enables users to master conversational interfaces with a lower barrier. Our previous work [20] has explored whether conversational interfaces can be alternatives to standard Web interfaces for microtask crowdsourcing, by carrying out experiments in a variety of popularly task types. We found that conversational interfaces could positively affect workers' satisfaction and an intention for future use of similar interfaces, and meanwhile achieve comparable worker performance in terms of both execution time and output accuracy.

However, previous work has only shown that conversational interfaces could be an equivalent platform to the traditional web interfaces. The advantages of conversational interfaces in microtask crowdsourcing remain unexplored. Therefore, in this vision paper, we propose a novel concept — conversational crowdsourcing. To this end, we proposed the workflow of conversational crowdsourcing and developed a tool to deploy web-based conversational interfaces on popular crowdsourcing platforms. Based on this, we researched conversational crowdsourcing from two perspectives — conversation design and novel UI affordances. In terms of conversation design, we specifically investigated approaches for estimating conversational styles and the effects of conversational styles on worker performance. As for novel UI affordances, we combined web search with conversational interfaces to study the effects on human memorability during search sessions. Furthermore, we also considered gamification elements, and implemented the avatar customization function on conversational interfaces to understand how avatars could affect worker satisfaction. In the following section, we will delve into several research questions to fill the knowledge gap.

2 Research Questions

Web-based Interface for Conversational Crowdsourcing. Online microtask crowdsourcing enables the possibility of accomplishing tasks requiring a large number of people. Tasks such as image annotation, sentiment analysis, and speech transcription can be easily accomplished on the online crowdsourcing marketplaces. During this process, the crowdsourcing platform is responsible for worker selection, microtask generation, microtask assignment and answer aggregation, while online workers interact with a crowdsourcing system to accept and execute a microtask using a worker interface.

Traditional web-based user interfaces are widely used for the interaction between crowdsourcing platforms and workers in the majority of prior work, to communicate with workers, transmit instructions and gather responses thereafter. In our introduced concept of *conversational crowdsourcing*, a conversational agent is able to interface online workers with the crowdsourcing platform, facilitating task execution and task completion [20, 23]. To this end, we attempt to address the following research question:

RQ1: How can the logic and workflow of the conversational crowdsourcing be designed to support the task execution?

Improving Worker Engagement. Our previous findings have suggested the use of conversational interfaces as a viable alternative to the existing standard web interfaces. However, little is known about the impact of conversational microtasking on the engagement of workers. Previous works have studied the nature of tasks that are popularly crowdsourced on Amazon's Mechanical Turk, showing that tasks are often deployed in large batches consisting of similar HITs [1, 7]. Long and monotonous batches of HITs pose challenges with regards to engaging workers, potentially leading to sloppy work due to boredom and fatigue [4]. There is a lack of understanding of whether conversational microtasking would either alleviate or amplify the concerns surrounding worker engagement. Therefore, we aim to address the following research question.

RQ2: To what extent can conversational interfaces improve the worker engagement in microtask crowdsourcing?

Conversational Style Analysis. The design of the conversation can affect the crowdsourcing outcome. Previous works in the field of psychology have shown the important role that *conversational styles* have on inter-human communication [17, 26, 27]. Having been developed in the context of human conversations, the insights and conclusions of these works are not directly applicable to conversational microtasking, since the contrasting goal of workers is to optimally allocate their effort rather than being immersed in conversations. To the best of our knowledge, the conversational style of neither the conversational agent (particularly for crowdsourcing) nor the online users (particularly for workers in the context of microtask crowdsourcing) have been ever studied. Understanding the role of conversational styles in human computation can help us better adapt strategies to improve output quality and worker engagement, or better assist and guide workers in the training process. To this end, there is a need for novel methods for the estimation of conversational styles in the context of microtask crowdsourcing. Therefore, we will delve into the following research questions:

RQ3: How do conversational agents with different conversational styles affect the performance of workers and their cognitive load while completing tasks?

RQ4: How can the conversational style of a crowd worker be reliably estimated?

RQ5: To what extent does the conversational style of crowd workers relate to their work outcomes, perceived engagement, and cognitive task load in different types of tasks?

Enhancing Long-term Memorability. Information finding tasks are rather popular and more acceptable in online crowdsourcing platforms since it combines with the learning process. Prior studies in online learning have revealed that conversational systems can significantly improve learning outcomes [11, 18, 25]. As the goal of learning is to develop a deep understanding of some information, memorization is an important element [15, 2]. Although conversation can produce unique context linked with information, the effect of conversational systems on human memorability needs further exploration. We investigated the role of text-based conversational interfaces in online information finding tasks [23]. We found that a conversational interface could better engage online users. However, the question of whether improved user engagement through conversational interfaces leads to better memorability of information remains unanswered.

To this end, we aim to fill this knowledge gap by proposing novel approaches to improve human memorability during information search. We specifically focus on information retrieval activities carried out through the Web search using desktop browsers. Through rigorous experiments, we seek to address the following research questions.

RQ6: How can human memorability of information consumed in informational web search sessions be improved?

RQ7: How does the use of conversational interfaces affect the search behavior of users?

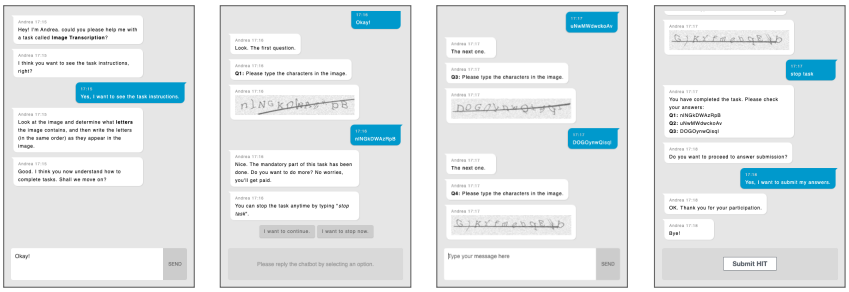
Improving Worker Experience. To increase participant engagement and satisfaction, the use of gamification has received attention in recent crowdsourcing-related works. While most gaming elements need to be designed based on specific task types, avatar customization is directly applicable in most contexts. Relevant work in the field of games research has shown that identifying with avatars can be effective in improving players' enjoyment and satisfaction [28, 3]. The contexts of games and crowd work are underlined by the need to motivate and engage participants, yet the potential of using worker avatars to promote identification and improve worker satisfaction in microtask crowdsourcing has remained unexplored. This is important to investigate, since using worker avatars and assigning avatars characteristics or personality traits can increase identification [28, 21]. Avatar identification has been studied from three perspectives — similarity identification, embodied identification, and wishful identification. Prior works have shown that avatar appearance and characteristics can affect similarity and wishful identification respectively [12, 13], whereas embodied identification demands more avatar operations and interactions, which is very common in video games but not essential in crowdsourcing. Since the influence of worker avatars in crowd work has remained unexplored, we know little about their impact on both conventional task interfaces as well as novel conversational interfaces. We thereby delve into this comparison through our work, to address the following research questions:

RQ8: How do worker avatars affect worker experience and quality-related outcomes in conventional web and novel conversational interfaces?

RQ9: How does avatar customization and characterization selection affect worker performance and satisfaction?

3 Insights

To answer **RQ1**, we developed a tool for quickly deploying crowdsourcing tasks in a customizable conversational interface, named TickTalkTurk [24]. We designed the logic and workflow of the conversational agent. Our conversational crowdsourcing tasks are mostly performed based on TickTalkTurk, as shown in Figure 1.



(a) Greetings and Task Instructions. (b) Interacting with the chatbot using buttons. (c) Interacting with the chatbot using free text. (d) submitting HIT using a customized HTML component.

Figure 1: The user interfaces of conversational crowdsourcing.

To answer **RQ2**, we conducted online experiments on AMT. We found that conversational interfaces have positive effects on worker engagement, as well as the perceived cognitive load in comparison to traditional web interfaces in general. As to **RQ3**, we found that a suitable conversational style has the potential to engage workers further (in specific task types), although our results were inconclusive in this regard. Our work takes crucial strides towards furthering the understanding of conversational interfaces for microtasking, revealing insights into the role of conversational styles across a variety of tasks [23].

To answer **RQ4** and **RQ5**, we conducted experiments to investigate the feasibility of conversational style estimation for online crowdsourcing. Our results revealed that workers with an *Involvement* conversational style have significantly higher output quality, higher user engagement and less cognitive task load while they are completing a high-difficulty task, and have less task execution time in general. The findings have important implications on worker performance prediction, task scheduling and assignment in microtask crowdsourcing.

To answer **RQ6** and **RQ7**, we conducted an online crowdsourcing experiment in a classical information retrieval setup. Results revealed that conversational interfaces have the potential to augment long-term memorability (7.5% lower long-term information loss). Furthermore, we found that users leveraging conversational interfaces showed a completely different behavior pattern (such behaviors have been proved to be beneficial for human memorability by previous studies), compared to traditional web users. Our findings suggest that the conversational interface is a promising tool for augmenting human memorability, particularly in information finding tasks.

To answer **RQ8** and **RQ9**, we support workers in building their own representations by customizing the appearance of their avatars. We also ask workers to characterize their avatars before they begin task execution – by selecting one out of three worker characterizations drawn from related literature (*diligent worker, competent worker, balanced worker*) [14, 8]. We designed worker avatars and studied the influence of avatar customization. Experiments have shown that using avatar customization has significantly impacts on fostering the sense of success in performance and lowering the perceived task complexity. The analysis of workers' behaviors and performances shows the existence of similarity and wishful avatar identification. Our findings have important implications in terms of reducing perceived workload and improving sense of success in crowdsourcing task design, which is crucial to the sustainability of the online freelancing marketplace.

References

[1] Alan Aipe and Ujwal Gadiraju. Similarhits: Revealing the role of task similarity in microtask crowdsourcing. In *Proceedings of the 29th on Hypertext and Social Media*, pages 115–122. ACM, 2018.

[2] John B Biggs. *Student Approaches to Learning and Studying. Research Monograph.* ERIC, 1987.

[3] Max V Birk, Cheralyn Atkins, Jason T Bowey, and Regan L Mandryk. Fostering intrinsic motivation through avatar identification in digital games. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 2982–2995, 2016.

[4] Peng Dai, Jeffrey M Rzeszotarski, Praveen Paritosh, and Ed H Chi. And now for something completely different: Improving crowdsourcing workflows with micro-diversions. In *Proceeding of The 18th ACM Conference on Computer-Supported Cooperative Work and Social Computing*, pages 628–638. ACM, 2015.

[5] Florian Daniel, Pavel Kucherbaev, Cinzia Cappiello, Boualem Benatallah, and Mohammad Allahbakhsh. Quality control in crowdsourcing: A survey of quality attributes, assessment techniques, and assurance actions. *ACM Computing Surveys (CSUR)*, 51(1):1–40, 2018.

[6] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[7] Djellel Eddine Difallah, Michele Catasta, Gianluca Demartini, Panagiotis G Ipeirotis, and Philippe Cudré-Mauroux. The dynamics of micro-task crowdsourcing: The case of amazon mturk. In *Proceedings of the 24th international conference on world wide web*, pages 238–247, 2015.

[8] Ujwal Gadiraju, Gianluca Demartini, Ricardo Kawase, and Stefan Dietze. Crowd anatomy beyond the good and bad: Behavioral traces for crowd worker modeling and pre-selection. *Computer Supported Cooperative Work (CSCW)*, 28(5):815–841, 2019.

[9] Mary L Gray and Siddharth Suri. *Ghost work: how to stop Silicon Valley from building a new global underclass*. Eamon Dolan Books, 2019.

[10] Lei Han, Kevin Roitero, Ujwal Gadiraju, Cristina Sarasua, Alessandro Checco, Eddy Maddalena, and Gianluca Demartini. The impact of task abandonment in crowdsourcing. *IEEE Transactions on Knowledge and Data Engineering*, 2019.

[11] Bob Heller, Mike Proctor, Dean Mah, Lisa Jewell, and Bill Cheung. Freudbot: An investigation of chatbot technology in distance education. In *EdMedia+ Innovate Learning*, pages 3913–3918. Association for the Advancement of Computing in Education (AACE), 2005.

[12] Cynthia Hoffner. Children's wishful identification and parasocial interaction with favorite television characters. *Journal of Broadcasting & Electronic Media*, 40(3):389–402, 1996.

[13] Cynthia Hoffner and Martha Buchanan. Young adults' wishful identification with television characters: The role of perceived similarity and character attributes. *Media psychology*, 7(4):325–351, 2005.

[14] Gabriella Kazai, Jaap Kamps, and Natasa Milic-Frayling. Worker types and personality traits in crowdsourcing relevance labels. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 1941–1944, 2011.

[15] David Kember. The intention to both memorise and understand: Another approach to learning? *Higher Education*, 31(3):341–354, 1996.

[16] Aniket Kittur, Jeffrey V Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. The future of crowd work. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 1301–1318, 2013.

[17] Robin Tolmach Lakoff. Stylistic strategies within a grammar of style. *Annals of the New York Academy of Sciences*, 327(1):53–78, 1979.

[18] Annabel Latham, Keeley Crockett, David McLean, and Bruce Edmonds. A conversational intelligent tutoring system to automatically predict learning styles. *Computers & Education*, 59(1):95–109, 2012.

[19] Andrew Mao, Ece Kamar, and Eric Horvitz. Why stop now? predicting worker engagement in online crowdsourcing. In *Proceedings of the First AAAI Conference on Human Computation and Crowdsourcing*, pages 103–111. AAAI, 2013.

[20] Panagiotis Mavridis, Owen Huang, Sihang Qiu, Ujwal Gadiraju, and Alessandro Bozzon. Chatterbox: Conversational interfaces for microtask crowdsourcing. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*, pages 243–251. ACM, 2019.

[21] Michael P McCreery, S Kathleen Krach, Peter G Schrader, and Randy Boone. Defining the virtual self: Personality, behavior, and the psychology of embodiment. *Computers in Human Behavior*, 28(3):976–983, 2012.

[22] Brian McInnis, Dan Cosley, Chaebong Nam, and Gilly Leshed. Taking a hit: Designing around rejection, mistrust, risk, and workers' experiences in amazon mechanical turk. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 2271–2282, 2016.

[23] Sihang Qiu, Ujwal Gadiraju, and Alessandro Bozzon. Improving worker engagement through conversational microtask crowdsourcing. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2020.

[24] Sihang Qiu, Ujwal Gadiraju, and Alessandro Bozzon. Ticktalkturk: Conversational crowdsourcing made easy. In *Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing*, pages 1–5, 2020.

[25] Donggil Song, Eun Young Oh, and Marilyn Rice. Interacting with a conversational agent system for educational purposes in online courses. In *2017 10th international conference on human system interactions (HSI)*, pages 78–82. IEEE, 2017.

[26] Deborah Tannen. Conversational style. *Psycholinguistic models of production*, pages 251–267, 1987.

[27] Deborah Tannen. *Conversational style: Analyzing talk among friends*. Oxford University Press, 2005.

[28] Sabine Trepte and Leonard Reinecke. Avatar creation and video game enjoyment. *Journal of Media Psychology*, 2010.

[29] Mengdie Zhuang and Ujwal Gadiraju. In what mood are you today? an analysis of crowd workers' mood, performance and engagement. In *Proceedings of the 10th ACM Conference on Web Science*, pages 373–382, 2019.