A Culturally-Aware AI Tool for Crowdworkers: Leveraging Chronemics to Support Diverse Work Styles

CARLOS TOXTLI, Clemson University, USA
CHRISTOPHER CURTIS, Northeastern University, USA
SAIPH SAVAGE, Northeastern University & Universidad Nacional Autonoma de Mexico (UNAM), USA & Mexico

Fig. 1. Overview of CultureFit's Functionality.

Crowdsourcing markets are expanding worldwide, but often feature standardized interfaces that ignore the cultural diversity of their workers, negatively impacting their well-being and productivity. To transform these workplace dynamics, this paper proposes creating culturally-aware workplace tools, specifically designed to adapt to the cultural dimensions of monochronic and polychronic work styles. We illustrate this approach with "CultureFit," a tool that we engineered based on extensive research in Chronemics and culture theories. To study and evaluate our tool in the real world, we conducted a field experiment with 55 workers from 24 different countries. Our field experiment revealed that CultureFit significantly improved the earnings of workers from cultural backgrounds often overlooked in design. Our study is among the pioneering efforts to examine culturally aware digital labor interventions. It also provides access to a dataset with over two million data points on culture and digital work, which can be leveraged for future research in this emerging field. The paper concludes by discussing the importance and future possibilities of incorporating cultural insights into the design of tools for digital labor.

CCS Concepts: • **Human-centered computing** → **Field studies**; **User studies**; **Open source software**.

Additional Key Words and Phrases: AI, Crowdsourcing, Culture, System, Chronemics

ACM Reference Format:
Carlos Toxtli, Christopher Curtis, and Saiph Savage. 2024. A Culturally-Aware AI Tool for Crowdworkers: Leveraging Chronemics to Support Diverse Work Styles. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW2, Article 360 (November 2024), 34 pages. https://doi.org/10.1145/3686899

1 Introduction

Crowdworkers significantly enhance AI services [53, 89, 104], yet often face poor working conditions [57], particularly non-US/European workers whose cultural backgrounds differ from the intended

Authors' Contact Information: Carlos Toxtli, Clemson University, Clemson, USA; Christopher Curtis, Northeastern University, Boston, USA; Saiph Savage, Northeastern University & Universidad Nacional Autonoma de Mexico (UNAM), Boston, USA & Mexico.

ACM 2573-0142/2024/11-ART360
https://doi.org/10.1145/3686899

design of crowdsourcing platforms [27, 139, 180]. This issue usually stems from the assumption that crowdworkers are a homogeneous group [56], neglecting their diverse cultural backgrounds [90]. Moreover, a notable trend in design has emerged advocating for minimizing cultural impact in work interfaces, aiming for global uniformity in their design rather than customizing these systems to accommodate cultural nuances [133, 134, 193]. Consequently, many work interfaces have strived for uniform standards, and have ignored worker diversity [76, 84, 88].

However, interfaces often reflect the cultural biases of their designers [18], inadvertently embedding their cultural norms [146, 150, 177]. This can lead to designs that unintentionally require "outside workers" to adapt or modify their behaviors [126, 177], potentially hindering their success and effectiveness in their jobs [24, 60, 64, 85]. A solution can be to create culturally aware tools for crowdworkers, yet research into integrating culture theory into such designs remains limited [108, 118, 163]. Further research is crucial to assess these systems' effectiveness and their potential benefits for crowdworkers from varied cultural backgrounds.

To address this knowledge gap, we focus on designing a tool that aims to enhance crowdworkers' experiences by incorporating cultural considerations. Drawing from culture theory, we apply Chronemics—a discipline used in various social sciences like Organizational Psychology and Anthropology—to inform our interface design [6, 12, 64, 155]. Chronemics helps us distinguish between "monochronic" and "polychronic" work practices, which influence how different cultures manage time and tasks [60, 130]. For instance, monochronic cultures (usually involving people from the United States, Germany, Scandinavia) focus on sequential task handling, whereas polychronic cultures (usually involving people from Latin America, Africa, South Asia) prioritize multitasking and social interactions [36, 182].

For this purpose, we implemented these concepts into CultureFit[1], a tool that dynamically adjusts its interface to notify workers of potential tasks according to "Monochronic" or "Polychronic" settings, thereby catering to the diverse cultural preferences of crowdworkers. Figures 1 and 2 illustrate CultureFit and show its interface in use. CultureFit equips polychronic workers with a notification system that extends beyond the crowdsourcing platform, enabling them to consider work opportunities while involved in various other activities, such as using desktop applications or browsing social media. This flexibility, supported by culture theory [11], leverages the multitasking preferences of polychronic workers. Conversely, the system notifies monochronic workers of available tasks only when they are on the crowdsourcing platform and have completed their current tasks, minimizing distractions and enhancing focus. This approach, which helps maintain their preferred work schedules, aligns with the preferences culture theory identifies for monochronic workers [50, 60]. By acknowledging workers' cultural differences, our tool proposes a novel design space that shifts the adaptation burden from workers to the interface itself, enabling digital workers from diverse backgrounds to potentially thrive in a space that respects and integrates their cultural work patterns [3, 57].

To evaluate the effectiveness of our novel tool, we conducted an IRB-approved field study to investigate its potential to improve the experiences of crowdworkers. We employed a between-subjects design for our study, where workers who identified with either polychronic or monochronic cultural traits used CultureFit, while polychronic and monochronic workers in the control groups did not. Fig. 3 presents an overview of our field experiment. Together, we were able to recruit crowdworkers spanning 24 countries across regions including the United States, Africa, Latin America, Europe, and South and Southeast Asia. Workers in our study completed over 2,300 tasks for 158 requesters on the Toloka crowdsourcing platform, generating more than two million anonymized data points with information about workers' digital labor and cultural traits. We plan

[1]Plugin link:https://github.com/ai4he/toloka-cultural-assistant

to publicly release this dataset upon the paper's publication. Notably, through our results we found that polychronic workers—often disadvantaged in crowd work [26, 59, 108, 180]—saw a 258% wage increase when using CultureFit.

In this paper, we contribuite: 1) a culturally aware crowd work notification tool; 2) a case study on applying culture theories to crowd work tools; 3) a two-week field experiment showcasing how CultureFit enhances wages for polychronic crowdworkers; 4) design recommendations for incorporating culture theories into crowd work tools; 5) over two million anonymized data points on digital labor and culture, to be publicly available post-publication, for future research studying cultural impacts on computer-supported collaborative work systems.

Positionality Statement. As researchers in the HCI community, our commitment lies in designing and evaluating technology that embraces diverse perspectives and experiences. We acknowledge the problematic nature of crowd work, recognizing the unfavorable conditions it imposes on workers, which pose significant challenges to their well-being [121]. Extensive research has shown that crowd work disproportionately impacts workers outside the United States and Europe, resulting in significantly lower earnings for these individuals [59]. Consequently, our research aims to foster inclusive technologies that foster positive outcomes for all. Our diverse team includes authors from Latin America and the United States, featuring individuals with indigenous heritage and a leader in diversity, equity, and inclusion at their institution. Among us, two authors practice "Polychronic" work styles, while one adheres to a "Monochronic" approach.

2 Related Work

Our research connects with the following key areas of prior literature:

2.1 Universal Design.

Vast work in the field of universal design has questioned the usefulness of culturally-informed interfaces [134, 136, 174]. While universal design is aware of the clashes that exist due to the lack of cultural awareness in design, it also argues that people will eventually adjust to the non-culturally-aware interfaces [101, 134]. This view-point has allowed universal design to make products and provide "equivalent experiences" for a wider range of possible users than addressing the specific accessibility or inclusive goals of smaller targeted populations [75, 174, 176]. Universal design has thus positioned itself as a "culturally-neutral" design framework [71, 72]. Norman and the researchers who follow his work have argued that designs become more approachable to a wider global audience upon the removal of individualistic and cultural human elements [135]. Norman argues that people will adapt to activities in ways not necessarily innate, supporting the prevailing importance of designing for "activity over culture" [134].

Our research considers that while it might be true that people will eventually adapt to interfaces not tailored to their culture [134], the adaptation period can generate harm [148, 169], e.g., cost people their livelihood [57]. Prior work has also established that there is a link between cross- cultural differences and outcomes like mental health and job performance [50]. Current research on homogeneous designs also indicates that prevailing design norms may indeed lead to reduced success for some crowdworkers [76, 84, 88]. Additionally, there is concern that crowdworkers from cultures underrepresented in design frequently have to engage in higher rates of unpaid labor, leading to poorer work experiences [26, 50, 59, 94, 132, 180]. Consequently, we advocate for culturally-aware design.

2.2 Chronemics and Different Cultural Dimensions within Time Management.

A key goal of designing culturally aware interfaces is determining what long-term aspects of a culture are important to consider [119, 144]. Hall and Hall [62, 63], argued that key cultural variables to consider are people's concepts of time. This falls under the study of Chronemics, which is the study of time perceptions, and "*... includes time orientation, understanding and organisation, the use of and reaction to time pressures, the innate and learned awareness of time...*" [151]. This theory extends to the social and cultural level, and can be used describe work patterns across cultures and societies [63]. Hall and Hall identified two main cultural concepts of time: (1) monochronic, and (2) polychronic [63]. They established that cultures with a monochronic time use, view time as linear [63]. This helps these individuals focus on doing one thing at a time, as time becomes something that can be scheduled and compartmentalized [63, 125]. According to the theory, Monochronic workers typically prioritize schedules highly [182], often valuing them above building social relationships [13]. They also prefer to limit multi-tasking and minimize distractions [108, 182].

By contrast, polychronic cultures view time as something occurring within the context of multiple events and the constant involvement of other people [62, 63]. In these cultures, it is more important to complete human transactions than to adhere to schedules [182]. Polychronic workers thus tend to value receiving unpredictable alerts, particularly those related to relationship building [182], as well as favor flexible, spontaneous, and concurrent work schedules [61, 62, 109], and prefer to work while engaging in other activities [50, 94]. We use insights from this culture theory to design better tools for global crowdworkers.

Our research is also inspired by previous studies on designing culturally aware systems [149, 177], as well as recent HCI research that has analyzed the monochronic and polychronic characteristics of crowdworkers, providing targeted design recommendations [107, 108]. Overall, we utilize this prior work to guide our tool design.

2.3 Culture and Crowd Work.

Crowdsourcing platforms allow non-experts to find new jobs [57]. These platforms are usually also not tied to a particular geographical region [47, 73]. The results are that more people from different parts of the world are exploring crowd work as a viable job option [47, 95]. The growing global nature of crowd work has led HCI researchers to study the demographics of crowdworkers [69, 157], helping to shed light on the different cultures present in crowd work and the problems they face [47, 66, 121, 129, 132]. On the other hand, recent research has identified that many crowdworkers have a tendency to follow polychronic work patterns [107]. Workers with this type of cultural background have also tended to experience more hardships [26, 55, 81, 117, 121], and have been joining crowdsourcing platforms in increasing numbers [47, 129, 132]. This prior work serves as motivation for our research, as it highlights the need to create tools that can support workers with diverse cultural backgrounds.

2.4 The Tooling Add-on Movement.

Crowdsourcing platforms have recently been augmented with a suite of tools to improve them for both workers and requesters [39, 42, 67, 161, 162]. Such a tooling movement has focused on creating change for workers and requesters without needing buy-in from the crowdsourcing platforms themselves [100, 162]. Most of these tools have taken the form of "web add-ons" (plugins) that provide additional functionalities to the platforms [153, 159, 191]. For instance, Turkopticon enhances the functionality of the crowdsourcing platform of Amazon Mechanical Turk by enabling workers to rate requesters [86, 170]. This "add-on" has helped workers fight power imbalances and information asymmetries [86, 99, 159]. We take inspiration from this to now augment crowdsourcing platforms with culturally aware interfaces.

2.5 Worker Community System Designs.

A number of data-driven tools have facilitated collaboration and community-building among workers [17, 44, 74, 93, 141, 143, 145], including crowdworkers. These tools have improved how workers collaborate and help foster a sense of unity. For example, Coworker.org, an NGO dedicated to empowering workers and developing power-building strategies in the modern economy [46], has developed a calculator tool that skillfully analyzes the wage data of platform workers [37]. This tool clarifies payment practices and encourages mutual support among workers, enabling them to challenge unfair pay practices from requesters.

Other tools have been designed to help workers collaborate with each other to identify potential wage theft and unpaid work [21, 54, 137, 180], as well as flag safety risks [2, 33]. Tools like Turkopticon and Turkerview are specif- ically designed to facilitate crowdworker collaboration by allowing them to share information about requesters, ultimately aiding in the identification of more favorable job opportunities [86, 160]. Additionally, online forums like Turker Nation have played a crucial role in cultivat- ing a community and promoting collaborative learning among crowdworkers, enhancing their collective knowl- edge and support network [121, 189, 197]. These plat- forms and tools have begun to foster mutual support for workplace challenges. Our paper explores how to enhance support for crowdworkers by recognizing their cultural identities.

3 CultureFit

CultureFit is a tool designed to improve the experiences of crowdworkers by tailoring job notifications to their cultural backgrounds. Drawing from the "Tooling Add- on Movement" in crowd work [86, 153, 159, 170, 191], CultureFit functions autonomously as a Chrome plugin, operating independently without requiring official sup- port from crowdsourcing platforms such as Toloka. Fig. 1 presents an overview of CultureFit, and Fig. 2 presents screenshots of it. Grounded in culture theory [64, 65], our tool is designed to accommodate cultural variations in time perception, specifically monochronic and poly- chronic cultural orientations. CultureFit leverages these distinctions to tailor notifications about available tasks to workers [6, 82, 111, 122].

It can be important to observe that by notifying workers about tasks on Toloka, CultureFit can help to streamline the process of finding work [57]. This can help to minimize the unpaid labor of task searching—one of the most burdensome types of unpaid labor in crowd work

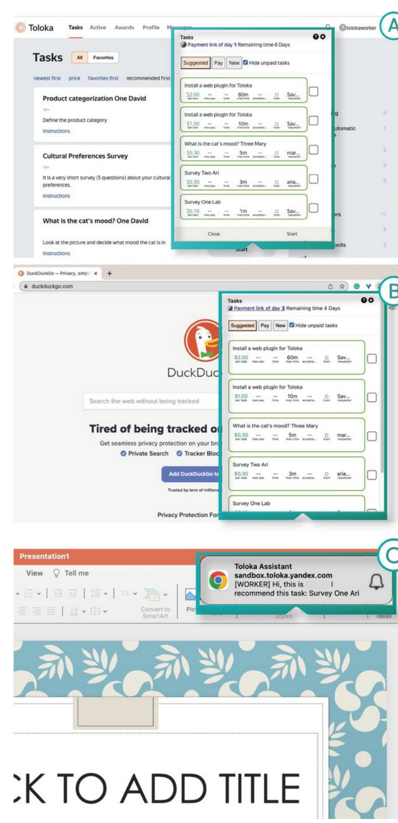

Fig. 2. Screenshots of CultureFit adapting its notification interface to workers' culture. For monochronic workers: (a) CultureFit no- tifies about tasks on Toloka. For polychronic workers: CultureFit notifies about tasks on Toloka while: (b) browsing other sites; or (c) engaged in other computer activities.

[180]. Overall, CultureFit is a crowd work tool that streamlines the discovery of relevant tasks through culturally aware notifications, thereby reducing workers' need to navigate extensive lists. Unlike other tools aimed at aiding workers in finding tasks [38, 160, 190, 192], CultureFit distinctively integrates cultural sensitivity into its notification design.

Another thing to observe is that CultureFit has a task recommendation algorithm, which determines which tasks workers are notified about. For more details on how this recommendation algorithm works, please see our Appendix. Note that while interesting, this recommendation algorithm is not our main research contribution and is hence not detailed here.

CultureFit enhances cultural awareness in crowd work through two culturally sensitive notifications:

·

3.1 Polychronic Work Notification Interface.

CultureFit caters to workers with polychronic cultural traits, who, according to the theory, favor flexible, spontaneous, and concurrent work schedules [61, 62, 109]. Thus, our tool opportunistically notifies polycronic workers about job opportunities across the worker's computer operating system, web browser, or crowdsourcing marketplace. In this way, the tool is able to reach workers while they are doing different activities. Fig. 2 presents how CultureFit can notify workers about labor opportunities within different contexts, and also in a peripheral manner, to facilitate opportunistic work schedules and multi-tasking (something related work recommends for polychronic workers [108]). CultureFit's Polychronic Interface also alerts workers to social media updates, emails, and messages from requesters and fellow crowdworkers. This feature, informed by culture theory [62, 109], aims to support the building of social connections, which is important to polychronic cultures, who prioritize meaningful human interactions and relationships.

·

3.2 Monochronic Notification Interface.

In the case of workers who follow Monochronic labor patterns [65], CultureFit notifies about tasks in a way that will allow them to manage their work days in a focused and efficient manner. For this purpose, CultureFit provides task notifications only when workers are on the Toloka platform (Fig. 2.a), considering that these workers, according to the theory, prefer to limit multitasking [13]. CultureFit also makes sure to notify workers about tasks only after the workers have finished the labor they are currently doing, and only notifies them about tasks that will fit within the worker's established schedule. For this purpose, the tool favors notifying about tasks that can be completed within the worker's schedule over tasks that might require working over time, even if the tasks match the worker's preferred type or come from their favorite requesters (to predict the amount of time a task will take, we use techniques from prior work [158, 160]). For monochronic workers, CultureFit also strategically limits notifications from social media, emails, and messages from requesters on the crowdsourcing platform. This approach aligns with research findings that emphasize monochronic workers' preferences for fewer interruptions, and a more focused approach to social connections and work tasks [13, 108].

4 User Scenarios

We present user scenarios on how monochronic and polychronic workers would use CultureFit. We aim to enhance understanding of the context in which CultureFit is used, especially within the Toloka crowdsourcong platform.

4.1 User Scenario for Monochronic Crowdworker: Bob

Bob is a dedicated monochronic crowdworker who prefers to work in a structured manner. He likes to focus on one task at a time and finds it distracting to sift through multiple tasks on traditional crowdsourcing platforms. Bob's typical workday is well-planned, with specific times allocated for different activities, including work, meals, and leisure.

Before CultureFit, Bob would log into Toloka every morning spending a significant amount of time searching through tasks to find the ones that fit his schedule and expertise [98]. This process was time-consuming and led to frustration, as Bob felt that he could have used this time to actually work on tasks. He also found it challenging to resist the temptation of non-work-related notifications (e.g., activities on social media), which occasionally disrupted his focus.

After adopting CultureFit, Bob now receives notifications about tasks on Toloka that are aligned with his schedule and work preferences. This change has improved his work efficiency and satisfaction. He no longer needs to manually search for tasks each morning. Instead, Bob receives a curated list of tasks right before his designated work time, allowing him to dive straight into focused work. CultureFit also blocks non-work-related notifications during his work hours, helping him maintain his concentration. Bob appreciates how CultureFit understands his monochronic work style and tailors the notification system to enhance his focus. This allows him to be more productive and achieve a better work-life balance, as he can now dedicate his planned work time more effectively and enjoy his leisure time without worrying about missing out on tasks.

4.2 User Scenario for Polychronic Crowdworker: Alejandro

Alejandro is a dynamic polychronic crowdworker who thrives on multitasking. He enjoys working on multiple tasks simultaneously and often juggles work with other activities like socializing or hobbies. Alejandro prefers a flexible work environment where he can switch between tasks as his interest and energy dictate.

Before CultureFit, Alejandro found traditional crowdsourcing platforms somewhat restrictive and inefficient for his working style. He would often miss out on social happenings or interesting tasks because he was too engrossed in another activity. Alejandro wanted a system that would automatically notify him of new opportunities, eliminating the need to continuously monitor the crowdsourcing platform or sift through lengthy lists of available tasks.

After adopting CultureFit, Alejandro's work experience has been transformed as the tool provides notifications that support his multitasking nature. Now, while engaged in one task on Toloka, he receives alerts for other tasks that match his interests, allowing him to seamlessly transition between jobs without losing momentum. Additionally, CultureFit alerts him to social happenings and networking opportunities, ensuring he does not miss out on valuable human-to-human experiences. Alejandro appreciates the tool's flexibility, particularly its ability to recognize his polychronic tendencies and offer a work experience tailored to his dynamic style. With CultureFit, he feels more connected to work opportunities and social engagements, enhancing both his professional and personal life.

5 Methods

Our IRB-approved field experiment focused on comparing the experiences of crowdworkers who utilized CultureFit and those who did not. This helped us better grasp the impact of culturally aware tools on crowdworkers. We thus implemented a between-subjects design, which included control groups that did not use CultureFit and intervention groups that did. These groups were further divided into sub-groups based on the cultural orientations of the workers, categorizing them as either monochronic or polychronic. We had overall the following groups:

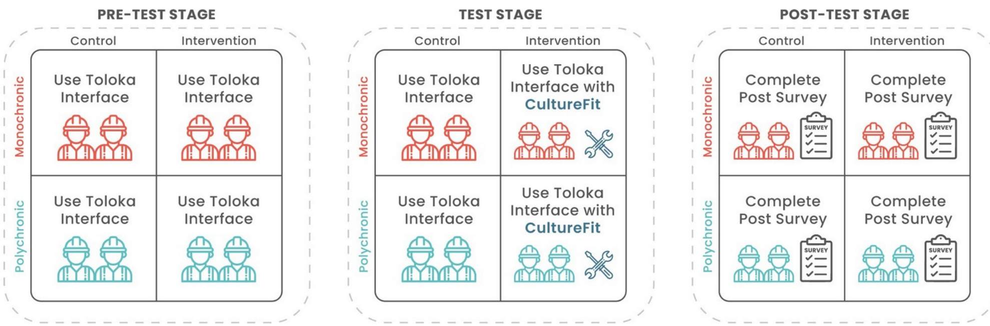

Fig. 3. Overview of our study that involved a 2x2 between subject study with three different stages.

(1) Monochronic workers using CultureFit on Toloka (intervention group).
(2) Polychronic workers using CultureFit on Toloka (intervention group).
(3) Monochronic workers following their usual work routine on Toloka WITHOUT CultureFit (control group).
(4) Polychronic workers following their usual work routine on Toloka WITHOUT CultureFit (control group).

Having these control and intervention groups, laid the groundwork for a comparative study of workers' experiences. For example, it facilitates an investigation into the experiences of monochronic workers who engaged with CultureFit as opposed to monochronic workers in the control group without access. It also allowed us to compare the experiences of monochronic and polychronic workers with access to CultureFit. Note that both control and intervention groups worked on the existing Toloka interface, but only the intervention groups accessed also the CultureFit web-plugin. Note that for our study, we also implemented a "randomized control-group pre-test/post-test" design where we divided our study into three phases: pre-test, test, and post-test [31]. The test phase signifies the period when workers in the intervention groups begin utilizing CultureFit. The pre-test and post-test phases are designed to capture the state of affairs before and after the application of CultureFit, respectively. Fig. 3 presents an overview of this experiment design. It is also important to highlight that given the potential market fluctuations in crowdsourcing markets [10, 29, 160], merely comparing data from before and after our tool was used may not accurately reflect our tool's effectiveness. This is due to the possibility that any observed improvements in workers' wages might be attributed to general market trends rather than our tool specifically. This is also why we implemented both control and intervention conditions alongside the "before-and-after" stages. By comparing worker groups with and without access to our tool (intervention and control conditions), and by conducting before and after analyses of each condition, we aimed to more effectively isolate the actual impact of CultureFit. Overall, this methodological approach helps us discern our tool's real world contribution to workers' experiences amidst the dynamic nature of the Toloka crowdsourcing market.

5.1 Pre-Test Stage.

We initiated our study with a 7-day "Pre-Test Stage" applicable to all workers. As mentioned above, this pre-test phase helped us establish a baseline understanding of the wages, perceptions, and digital behaviors among crowdworkers across our four conditions. The duration and structure of this stage, as well as subsequent stages in our study, were informed by prior research [29, 160, 180]. To establish our baseline, we guided all workers through a structured process:

(1) **Pre-Survey Completion**: All workers completed a pre-survey that we crafted to gather information about their demographics, crowd work experiences, and to classify them as either polychronic or monochronic workers. Our Appendix presents our pre-survey.
(2) **Web-plugin Installation**: All workers installed a web-plugin equipped with telemetry tracking, which we designed in accordance with prior work [158, 180]. The web-plugin helps us to monitor and quantify workers' digital behavior (e.g., hourly wages and task completion rates). We allowed the uninstallation of our web-plugin at any time and offered financial rewards for study participation. After installing our web-plugin and finishing the pre-survey, all workers received $2 USD, an amount set to surpass the US federal minimum wage of
$7.25/hour for a task under 10 minutes [171]. Workers also earned an extra $0.5 daily for simply keeping our web-plugin installed, totaling $3.5 for the 7-day Pre-Test Stage.
(3) **Do Crowd Work**: Participants did crowd work as usual for 7 days. (e.g., they engaged in activities such as data labeling, completing surveys, communicating with requesters, or searching for tasks [180]). This allowed us to observe their baseline.

5.2 Test Stage.

During our 7-day Test Stage, workers in the intervention group gained access to CultureFit. CultureFit was activated via the web-plugin previously installed. Crowdworkers using CultureFit began receiving culturally-aware notifications about potential tasks. These workers were encouraged to use CultureFit at their discretion. Meanwhile, the control groups continued their usual crowd work activities, mirroring the Pre-Test Stage. Note that having these control and intervention groups, as well as the different stages, helped us to identify market fluctuations that could influence the results we observed with CultureFit.

5.3 Post-Test Stage.

At the experiment's end, workers in all conditions completed a post-survey about their experiences during the Test Stage, earning $1.0 for participation. Overall, participants received $10 for the entire study. They were instructed to uninstall CultureFit, with the plugin automatically deactivating if not manually removed.

5.4 Data Analysis of Workers' Survey Data.

We conducted both quantitative and qualitative analyses of participant survey responses. We analyzed their Likert scale answers from surveys, calculating the median of the five-level questions for each condition. We also used affinity diagramming to study open-ended responses from pre- and post-surveys, identifying common themes [70, 115]. This involved color-coding responses for our four experiment conditions and organizing them by question. We further categorized responses using intra-question affinity diagramming, with the authors independently coding data and collaboratively developing three thematic codes.

5.5 Participants.

Similar to prior work [108], we used the Toloka crowdsourcing platform to recruit participants. We posted a detailed task description outlining all study requirements and steps, including completion of surveys, web-plugin installation, and data collection. Workers were also informed that they could drop out of the study at any time and would be compensated for the duration of their participation. Workers interested in participating were required to complete a pre-survey questionnaire upon accepting the study terms. This pre-survey used the Multitasking Preference Inventory (MPI) to evaluate workers' polychronic or monochronic tendencies [138]. The MPI, a 14-item self-assessment

Question	Monochronic	Polychronic
How often do you use your strengths when working on Toloka?	4 ($\mu = 4.0, \sigma = 1.0$)	4 ($\mu = 3.8, \sigma = 1.1$)
How fast can you find tasks to work on in Toloka?	5 ($\mu = 4.4, \sigma = 0.6$)	4 ($\mu = 3.6, \sigma = 1.2$)
How much do you enjoy planning your tasks on Toloka?	4 ($\mu = 3.8, \sigma = 1.0$)	3 ($\mu = 3.2, \sigma = 0.9$)
How much do you like having someone oversee the work you do?	4 ($\mu = 3.6, \sigma = 1.2$)	2 ($\mu = 2.0, \sigma = 1.1$)

Table 1. Overview of Pre-survey responses.

tool, is more effective than earlier scales like the Inventory of Polychronic Values (IPV) in distinguishing individual and cultural work styles [12]. It includes statements rated on a 5-point scale, assessing multitasking behavior and task-switching preferences [138]. Scores indicate a preference for either monochronic or polychronic work habits. We categorized workers into monochronic or polychronic groups based on their stated preferences. We recruited 55 participants (23 monochronic, 32 polychronic) from a diverse pool across 24 countries, including the US, Middle-East, Central and South America, various African countries, South and Southeast Asia, and Eastern and Western Europe. Note that a power analysis for a dichotomous endpoint-one-sample study indicated the need for at least 9 participants per condition, assuming a α (Type I error) of 0.2, β (Type II error) of 0.05, an anticipated incidence of 0.5, and a power of 0.8 ($1-\beta$).

We then grouped participants into four conditions:

- **Monochronic/CultureFit (M/CF).** 12 workers from monochronic cultures used CultureFit.
- **Polychronic/CultureFit (P/CF).** 17 workers from polychronic cultures used CultureFit.
- **Monochronic/Control (M/S).** 11 workers from monochronic cultures, who conducted their work as (s)tandard.
- **Polychronic/Control (P/S).** 15 workers from polychronic cultures, who conducted their work as (s)tandard.

6 Results

We originally recruited 126 Toloka workers, of whom 55 remained throughout our two-week field experiment. This retention rate is typical for real-world experiments conducted on crowdsourcing platforms [29, 68, 83, 160, 180]. This paper focuses on the 55 workers who completed our study throughout its duration. However, section 6.7 compares these 55 participants with those who dropped out to identify any tendencies or traits possibly influencing dropout rates in real-world longitudinal studies.

6.1 Results: Overview

Our study included 55 workers—23 monochronic and 32 polychronic, with a median age of 30—which is a typical sample size for real world deployments [29, 83, 160, 180]. We collected over two million telemetry logs on Toloka, tracking workers' mouse clicks, scrolls, keyboard activity, and page transitions, along with wage and task data, and interactions with our tool's notifications. Over two weeks, workers engaged with 2,303 tasks, with monochronic workers interacting with a median of 89 tasks ($\mu = 156, \sigma = 193$), and polychronic workers with a median of 83 tasks ($\mu = 255, \sigma = 587$).

6.2 Pre-Test Stage: Survey Results.

Fig. 4 and Table 1 present our pre-test survey results. We investigated differences in views between monochronic and polychronic workers by computing median Likert scale values and applying

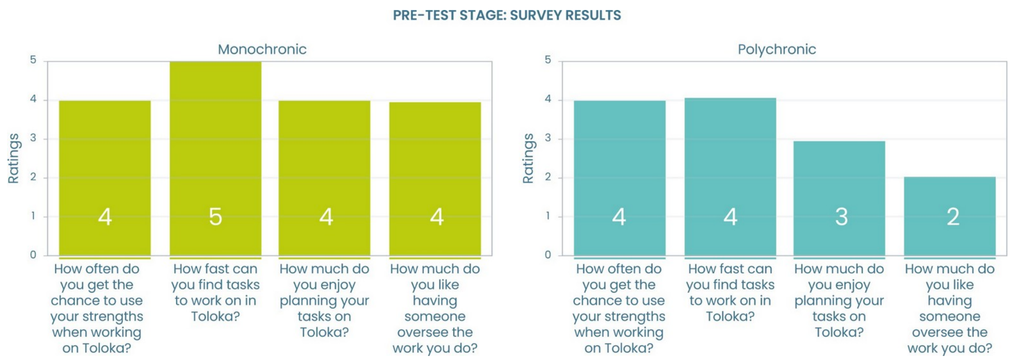

Fig. 4. Pre-survey Overview: Monochronic vs. Polychronic Workers

the Mann-Whitney U Test. Results revealed no significant differences in workers' perceived use of their strengths in their jobs ($H = 233$, $p = 0.32$). However, significant cultural differences emerged in workers' preferences for planning. In specific, monochronic workers showed a stronger preference than polychronic workers for planning their work on Toloka ($U = 394$, $p = 0.07$). Monochronic workers also preferred closer supervision, similarly marking a significant difference in supervision preferences ($U = 170$, $p = 0.001$). These findings are consistent with the known cultural preferences of monochronic individuals, who typically favor structured schedules and more formal work relationships [61, 181]. Conversely, under half of the monochronic workers (10 out of 23) reported using web forums (e.g., Reddit, Quora, Facebook Groups) for work assistance, without favoring any specific platform. Meanwhile, two-thirds of polychronic participants used web forums, notably favoring Facebook. Only 10% of monochronic workers used job assistance tools, with no polychronic workers doing so.

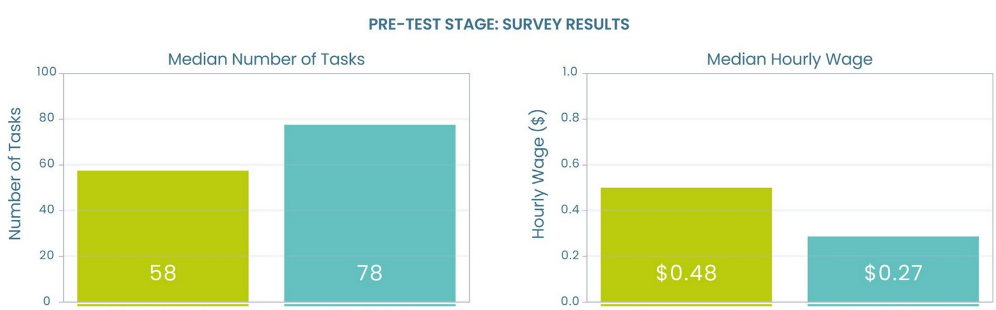

Fig. 5. Pre-Test Telemetry Log Summary for workers of different cultural groups.

6.3 Pre-Test Stage: Telemetry Logs Results.

We also collected data on the wages and number of tasks that workers in our pre-test stage completed using our web plugins with Telemetry tracking, following the same methodologies from prior work [160, 180]. Table 2 and Fig. 5 presents a summary of these results. Monochronic workers finished fewer tasks but earned slightly higher hourly wages than polychronic workers. Next, we

Variable	Monochronic	Polychronic
Median Number of Tasks	58 tasks ($\mu = 85, \sigma = 84$)	78 tasks ($\mu = 89, \sigma = 76$)
Median Hourly Wage	\$0.48 ($\mu =$ \$1.05, $\sigma =$ \$1.40)	\$0.27 ($\mu =$ \$1.51, $\sigma =$ \$2.46)

Table 2. Telemetry Log Results for the Pre-Test Stage.

wanted to study if these differences were significant. We thus first computed for each cultural group the Shapiro-Wilk test on their distribution of tasks and their distribution of wages. Across groups, this test indicated that we were working with non-normal distributions (p-value $< .05$). Based on this, to evaluate potential significant differences in the task and wage distributions between monochronic and polychronic workers, we decided to employ the Kruskal-Wallis Omnibus test. This non-parametric test, designed for comparing median values across independent distributions, allowed us to first analyze the disparities in task distributions between monochronic and polychronic workers. Subsequently, we applied the same technique to assess differences in the wage distributions of polychronic and monochronic workers. Our findings showed no significant differences in neither the number of tasks that monochronic and polychronic workers completed ($H = 7, p = 0.06$) nor in their wages ($H = 4, p = 0.25$).

6.4 Test Stage: Telemetry Logs Results.

Our goal in the Test Stage was to investigate the effects of embedding culture theory within crowd work tools. This includes examining shifts in workers' wages and digital behaviors as indicators of change.

Condition	Pre-Test	Test
Monochronic/ CultureFit	\$0.34 ($\mu =$ \$0.60, $\sigma =$ \$0.92)	\$0.39 ($\mu =$ \$1.65, $\sigma =$ \$1.70)
Polychronic/CultureFit	\$0.16 ($\mu =$ \$1.27, $\sigma =$ \$1.78)	\$0.68 ($\mu =$ \$1.31, $\sigma =$ \$1.63)
Monochronic/Control	\$0.62 ($\mu =$ \$1.50, $\sigma =$ \$1.89)	\$0.69 ($\mu =$ \$1.65, $\sigma =$ \$1.70)
Polychronic/Control	\$0.39 ($\mu =$ \$1.76, $\sigma =$ \$3.14)	\$1.04 ($\mu =$ \$1.89, $\sigma =$ \$2.04)

Table 3. Workers' median earnings during the Pre-Test and Test stages across the different conditions.

6.4.1 Hourly Wages. To assess the potential impact of CultureFit on workers' wages, we first established the baseline median hourly wage for each condition (control and CultureFit) during the Pre-Test phase. We then compared this baseline to workers' median hourly wage during the Test Stage. It is important to note that we calculated workers' hourly wages using the same methodologies as previous studies [160, 180]. The results, detailed in Table 3, show workers' median hourly wages across the four different conditions and stages. Figure 6 illustrates the percentage increase in the median hourly wages of workers between the Pre-Test and Test Stages. To calculate this percentage change, we subtracted the median Pre-Test wage from the median Test wage and then divided the result by the median Pre-Test wage. This highlights how median hourly wages evolved across the different conditions during our study.

Our results uncovered that the wages for workers increased during the Test stage across all conditions. Next, we studied whether this increase was significant. For this purpose, we first conducted the Shapiro-Wilk test on the wage distributions across conditions during the Pre-

Test and during the Test stages, finding non-normal distributions (p-value < .05). Based on this, we

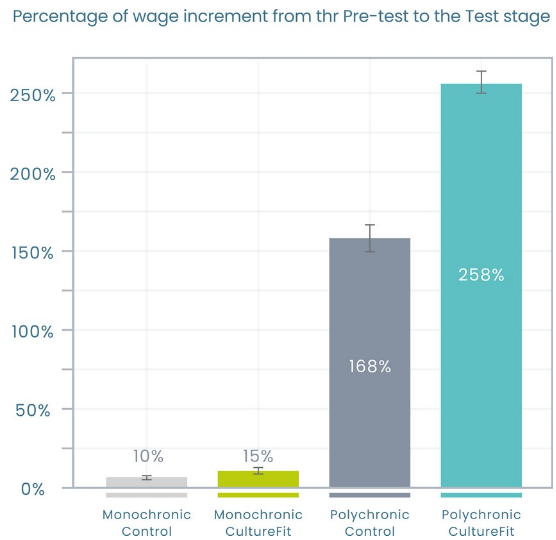

Fig. 6. Overview of workers' increase in wages from the Pre-Test Stage to the Test Stage across conditions. Error bars represent 95% confidence intervals. CultureFit significantly increased the wages of polychronic workers.

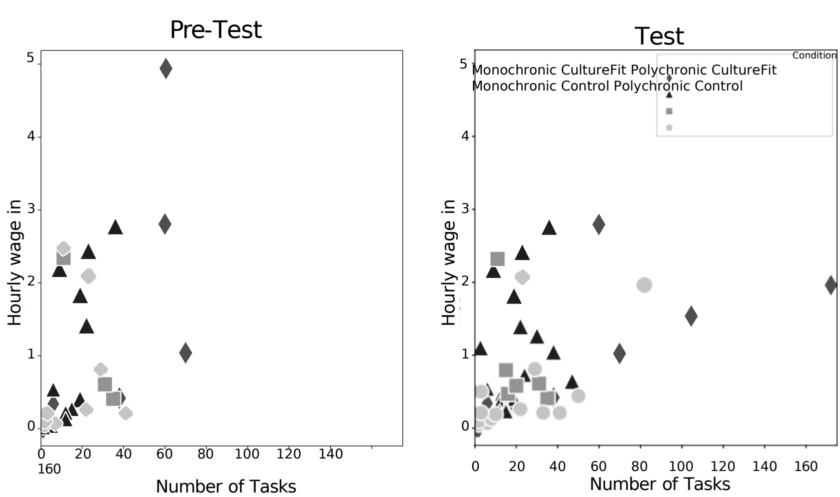

Fig. 7. Overview per condition of the number of tasks each worker completed (X-axis) and the total wages they received (Y-axis).

decided to use the Kruskal-Wallis Omnibus test to assess if the wage changes observed in each condition were significant. We found that workers in the polychronic CultureFit condition did change significantly their wages ($Z = 11, p = 0.01$). The median wages of these workers increased 258%, going from \$0.19 to \$0.68 USD. For the other conditions, we did not observe any significant wage increases. For instance, polychronic workers in the control condition saw median wages

rise from \$0.39 to \$1.04 USD, indicating an increase of 168%. But, the change was not significant ($Z = 9, p = 0.48$). This lack of significance is likely due to high wage variability among polychronic workers in the control group ($\sigma = 2.04$), indicating inconsistent wage increases across workers. Similarly, monochronic workers using CultureFit experienced a 15% wage increase, from \$0.62 to
\$0.69 USD per hour, but this was also not significant ($Z = 5, p = 0.31$). Monochronic workers in the control condition had a 10% wage increase, which was again not statistically significant ($Z = 4$,
$p = 0.43$). These non-significant variations in wages between the pre-test and test stages could just reflect natural fluctuations in the crowdsourcing market [160].

6.4.2 Analyzing Shifts in Digital Work PaFerns. To further study the changes our tool might have created in workers' digital traces, we created scatter plots from the Pre-Test and Test stages (Fig. 7), plotting each worker's task count (X-axis) against their earnings (Y-axis). In these scatter plots, each point represents a worker that is color and figure coded according to their condition. Comparing Fig. 7's Pre-Test scatter plot to its Test one, we see two key trends: an upward shift indicating higher hourly wages and a rightward shift reflecting increased task completion. The upward trend is most notable in the Polychronic CultureFit group where workers experienced significant wage increases.

6.5 Post-Test Stage: Post-Survey Quantitative Results.

Fig. 8. Median Post-Survey Response Summary.

After the Test Stage, our tool ceased operations, halting notifications and behavior tracking, and redirected workers to a post-survey. Fig 8 and Table 4 shows the median perceptions of workers from the post-survey under different conditions. We studied whether CultureFit usage led to significantly different perceptions compared to non-users, i.e., we analyzed the responses between workers utilizing CultureFit (Figs. 8.A and 8.C) and those who did not (Figs. 8.B and 8.D). For this purpose, we first computed the Shapiro-Wilk test, which indicated non-normal distributions for both groups across survey questions ($p < .05$). As a result, we utilized the Kruskal-Wallis test to determine whether significant differences existed among the responses from these groups. Our analysis uncovered significant differences between CultureFit users (Figs. 8.A and 8.C) and

Question	Monochronic/CultureFit	Polychronic/CultureFit	Monochronic/Control	Polychronic/Control
How often did you get to use your strengths while working on Toloka this week?	4 ($\mu = 4.0, \sigma = 0.8$)	5 ($\mu = 4.5, \sigma = 0.8$)	4 ($\mu = 4.0, \sigma = 0.8$)	4 ($\mu = 4.1, \sigma = 0.8$)
How different did your experience feel while working on Toloka this week?	4 ($\mu = 3.9, \sigma = 0.9$)	4 ($\mu = 3.6, \sigma = 1.3$)	2 ($\mu = 3.0, \sigma = 1.5$)	3 ($\mu = 2.9, \sigma = 1.3$)
How much has your work schedule on Toloka changed this week compared to previous weeks?	4 ($\mu = 3.4, \sigma = 1.2$)	2 ($\mu = 2.4, \sigma = 1.5$)	4 ($\mu = 3.2, \sigma = 1.5$)	1 ($\mu = 2.3, \sigma = 1.8$)
How much did you get to try out new tasks on Toloka this week?	4 ($\mu = 3.7, \sigma = 0.7$)	4 ($\mu = 3.2, \sigma = 1.4$)	3 ($\mu = 3.1, \sigma = 1.6$)	2 ($\mu = 2.6, \sigma = 1.2$)

Table 4. Overview of the median Post-survey responses.

non-users (Figs. 8.B and 8.D) in reporting changes to their experiences ($U = 210$, $p = 0.03$) and schedules ($U = 209$, $p = 0.03$) during the Test-Stage.

We also investigated CultureFit's impact on enhancing workers' perceptions of utilizing their strengths. To quantify this change, we utilized the Wilcoxon signed-rank Z test, comparing percep- tions from the Pre-Test to the Test stages. This test evaluates the impact of an intervention on a participant group by comparing the variations in their responses before and after the intervention. It does not require the differences between paired observations to adhere to a normal distribution. Polychronic workers in the CultureFit condition had a significant shift in how much they felt that they utilized their strengths (Fig 8.C), moving from a median of 4 ($\mu = 3.8$, $\sigma = 1.1$) in the Pre-Test to 5 ($\mu = 4.5$, $\sigma = 0.8$) ($Z = 3$, $p = 0.01$) in the Test stage. We did not observe significant changes in other conditions (Fig 8.A, B, and D).

6.6 Post-Test Stage: Post-Survey Qualitative Results.

We analyzed workers' open-ended survey responses, identifying common themes:

6.6.1 Culturally-Aware Interfaces and Improved Work Practices. Workers using CultureFit reported that the tool brought about positive changes in their work behavior: *"Despite it [CultureFit] changing my behavior, it was for the best " [P/CF 10].* Workers using CultureFit felt that it transformed their work habits for the better by providing a better understanding of time dynamics in the crowdsourcing market, which enabled them to manage their work schedule more effectively: *"It [the tool] gave me a better sense of how much time I should work. The tool allowed me to stay on top of what jobs were in Toloka and what was new. So I was able to schedule my time more effectively." [M/CF 7].* Further, it was interesting to observe that workers considered that despite the tools changing their work practices, they did not feel that the tool interfered with their work: *"I love how the plugin doesn't really interfere and affect your work" [P/CF 4].* It is important to note that the belief of CultureFit leading to better work practices occurred for both monochronic and polychronic workers. Overall, workers from both cultures perceived a positive change in their work processes when using the tool. Polychronic workers were pleased with the positive change that CultureFit had on their work, and did not express that the tool negatively interfered with their work. Meanwhile, participants from monochronic cultures also noticed a positive change. The tool helped them to be more mindful of the time spent using the platform.

6.6.2 Culture and Social Features in Work Tools. Polychronic workers differed from monochronic workers in the features they desired in future tools, regardless of the conditions they were. Poly- chronic workers valued having tools with "social features". In specific, polychronic workers men- tioned that they preferred applications that facilitated communication with the crowdsourcing platform, requesters, and other workers. Also, polychronic workers with access to CultureFit expressed that they valued that the interface helped them to connect properly with requesters and other workers: *"The tool provides a friendly link between the client [requesters] and the service provider [workers], that interests me [P/CF 8]."* Note that the "friendly link" mentioned by workers primarily refers to notifications about messages from requesters on Toloka—notifications that Toloka's interface design did not prioritize or effectively alert workers to. Further, polychronic workers in the control mentioned that in future tools they would like to see the integration of social features: *"Plugins should also be tools that increase communication between clients [requesters] and providers [workers]. The ability to communicate more [with requesters] and hear back about whether or not my efforts are appreciated would be helpful [P/C 9]."* Finally, monochronic workers did not mention wanting social features. This might be because culture theory argues that such features are not as important in monochronic cultures [15, 184].

6.6.3 Positive Experiences with Culturally-Aware Notification Tools. Participants in the CultureFit conditions reported that our tool led them to have more positive and engaging labor experiences. This result resembles what prior research has found about crowdworkers and their positive expe- riences with tooling [91, 191]. The positive sentiment with our tool led some to engage in more crowd work. As one participant expressed, *"Since installing the plugin, I felt more inclined to check out Toloka on a daily basis, and by doing that, I found a few more tasks that I usually wouldn't have seen, this meant I worked on a few more tasks, and I enjoyed these [M/C 5]."*

6.7 Analyzing the Crowdworker Dropouts

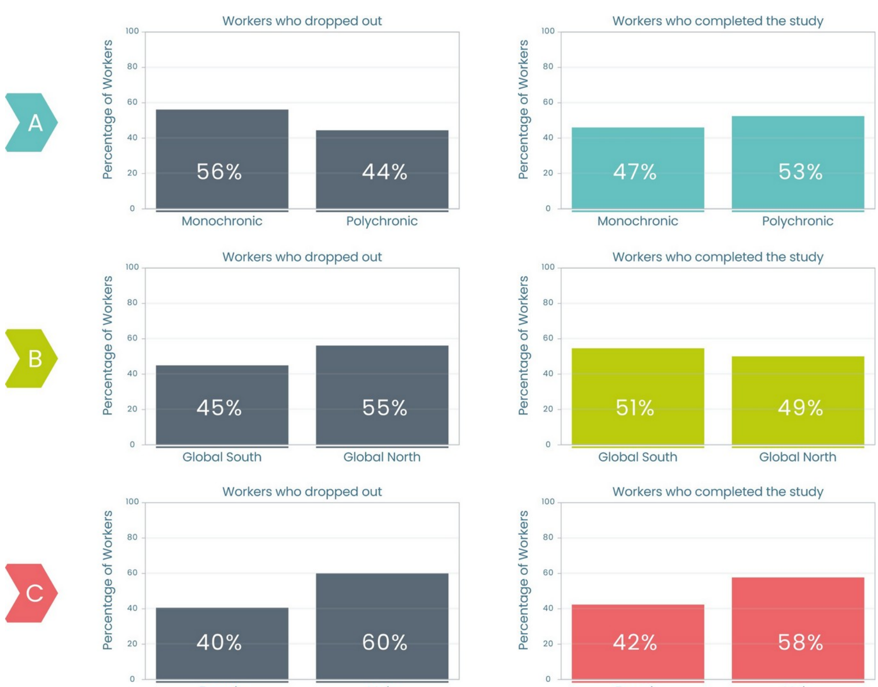

Fig. 9. Characteristics of "persisting" crowdworkers and those who "dropped out".

This paper focuses on the 55 crowdworkers who completed our study. However, we also examined the characteristics of the "dropouts" to gain insights into the dynamics of conducting real-world experiments with new crowd work tool designs. Figure 9 presents different traits contrasting "dropouts" with those who stayed.

Fig. 9.A shows the completion and dropout rates among workers with different cultural traits, with nearly equal numbers of monochronic (56%) and polychronic (44%) workers dropping out.

Next, we investigated if there were significant differences in cultural traits between workers who dropped out and those who remained in the study. A Chi-square test of independence revealed that there was no significant association between the cultural traits of workers who dropped out or those who continued participating in the study, X^2 (1, N=77)=1.91, p=0.16.

In Figure 9.B, we categorized workers' countries of origin into global south or north based on existing literature [1, 175, 187], and displayed the distribution through bar plots showing the percentage of workers from each region. A Chi-square test of independence indicated no significant link between workers' geographic regions of origin and their choices to either discontinue or persist in the study, X^2 (34, N=77)=32.66, p=0.53.

In Fig. 9.C, we display the gender distributions for workers who dropped out versus those who stayed in the study. A Chi-square test of independence found no significant relationship between the gender of workers and those who dropped out or remained in the study, X^2 (1, N=77)=0.005, p=0.94.

Overall, these findings from the Chi-square tests can help to dispel the notion of survivorship bias linked to the demographics of workers. Next, we examined if reductions in wages or the number of tasks completed might have prompted workers to exit our study, questioning if our tool could have negatively affected workers' productivity and led to their departure. However, our analysis showed that all dropouts occurred within the pre-test stage (usually also within the first day), before workers received any notifications from our tool. Therefore, dropouts were not due to a decline in productivity from using CultureFit. Future studies could benefit from interviewing the dropouts to understand why some crowdworkers persist in longitudinal studies while others choose to leave.

7 Discussion

Our research pioneers ways to start integrating cultural dimensions into the design of tools for crowdworkers . Such direction is vital to the field of CSCW, as it opens up new opportunities for creating computational artifacts that can support the global workforce that exists on crowdsourcing platforms [40, 41, 48, 140, 185]. Through our real-world field experiments, we started to see the potential benefits of having a culturally-sensitive tool, particularly for workers with polychronic traits [41, 108, 110, 128]. This can be important , as previous studies had shown that these workers are often disproportionately affected by the challenges of crowd work [26, 57, 108, 180].

7.1 Understanding the Integration of Cultural Insights in Crowd Work Tool Design

Inspired by culture theory [6, 12, 64, 155], we created a culturally aware tool tailored to both polychronic and monochronic crowdworkers. Based on the theory, we anticipated that our tool would enhance productivity and satisfaction for both groups. We expected polychronic workers would find value in our tool's notifications designed to enhance multitasking and adapt to flexible schedules [64]. Conversely, monochronic workers would appreciate our tool's notifications that promote structured and sequential task management, supporting their preference for orderly progress [12, 60, 64, 130]. In summary, we expected that our tool's culturally aware notifications would help both groups thrive within crowd work.

However, our results were unexpected. Initially, in the Pre-Test phase of our field experiment, both polychronic and monochronic workers demonstrated similar performance in terms of tasks completed and hourly wages. Yet, when they began using our tool during the Test phase, a no- table divergence occurred: polychronic workers saw a significant increase in their wages, while monochronic workers experienced no such improvement (see Fig. 6). Consequently, it became clear that our culturally aware tool did not universally enhance labor outcomes.

For example, our scatter plot, illustrated in Figure 7, reveals that most monochronic workers in the Test stage are clustered at the lower regions of both the X and Y axes, mirroring their positioning in the control condition and in the Pre-Test phase. This similarity further showcases that our tool did not impact the economic outcomes or the number of tasks monochronic workers completed. In contrast, Figure 7 in the Test stage shows a noticeable shift for polychronic workers. Initially concentrated around the origin point (0,0) in the Pre-Test phase, these workers now appear more prominently distributed across higher values of both the X and Y axes in the Test visualization. This distribution shift indicates that they completed more tasks and received higher wages, suggesting that our tool effectively aligned with their multitasking abilities and preferences for engaging in diverse and simultaneous activities.

7.1.1 Why might we see these results? The differential impact that CultureFit had to monochornic and polychronic workers could be due to the design bias in workplace technologies [65, 120, 172], which are often unconsciously tailored to monochronic preferences — organized, sequential, and linear task presentations [8**?**]. This inherent design focus may mean that monochronic workers do not experience as significant a change with the introduction of a culturally aware tool because the tool aligns closely with the existing cultural bias in tool design, which already favors monochronic workers. Conversely, polychronic workers, who are less catered to by standard designs [65, 134, 172], likely experience more noticeable benefits when a tool is finally adapted to fit their cultural work style. Essentially, while existing tools support monochronic workers well [65], our culturally aware tool likely fills an important gap for polychronic workers whose natural work tendencies are often forgotten in design [79, 147].

We believe that the disparity in tool effectiveness revealed by our results underscores the need for designing culturally aware tools, especially for populations traditionally overlooked in the design process. Culturally aware tools are likely to have greater impact on them. We could thus envision future research focusing on designing tools that cater specifically to the culture and needs of rural crowdworkers in the United States, a group often forgotten in mainstream tool design [22, 47, 96]. Creating tools that resonate with the culture of rural areas could be more impactful than continuing to design primarily for urban workers, who might already have access to a wide array of specialized tools [25, 43]. Similarly, it may be more beneficial to design tools that are adapted to the cultures of the Global South rather than continuing to focus predominantly on the Global North [34, 142], where there is already an abundance of systems tailored to local needs. This approach can not only promote inclusivity, but can also maximize the potential impact of the computational artifacts by being tailored to cultures that are traditionally underrepresented in design.

7.2 Understanding the Feasibility and Challenges of Culturally-Aware Tools in Crowd Work

Next, we discuss the challenges and feasibility of our proposed system design.

Feasibility. Prior work has provided important design recommendations for crowdworkers with polychronic or monochronic traits [106, 108]. However, several of these recommendations focus on completely re-designing crowdsourcing platforms [51], or forcing workers and requesters to change their behaviors [16, 52, 179]. However, pressing people and platforms to change is not always feasible [28]. In our design of CultureFit, we focused on designing a tool that could co-exist within existing crowdsourcing platforms, and could automatically adapt to the cultural background of the workers (without neither forcing workers, requesters, nor platforms to have to make significant changes). Our approach can thus make it more feasible to start to create culturally-aware crowd work tools. Overall, in system design, we acknowledge the advantages of implementing a "tool add-on" architecture that complements existing work environments without major disruption

[19, 131]. Future tools for digital labor platforms should embrace this approach to enhance their feasibility. Note that feasibility also entails ensuring equal access to tools, which is important for preventing economic disparities among crowdworkers. [57, 191]. To foster equitable tool access, we aim to opensource CultureFit, engage with worker collectives, and employ proven global tool access and adoption strategies [4, 102].

Challenges. There are the challenges that can arise with our culturally-aware tool: Requesters from monochronic cultures may be unhappy with polychronic workers who utilize CultureFit for multitasking. This dissatisfaction can stem from cultural differences, as monochronic requesters could perceive multitasking negatively [13, 109]. To tackle this issue, we propose keeping workers' practices hidden from requesters and only showing them the final outcomes of their work. However, a complication arises from the fact that certain digital labor platforms now employ "surveillance mechanisms" allowing requesters and platforms to monitor workers' actions and determine payment based on their work practices [183]. This surveillance assumes that only one work practice is correct, e.g., one that excludes multitasking. To address this challenge, we envision implementing CultureFit on the requesters' side as well. This would provide guidance to requesters on embracing the cultural diversity of their workforce [97].

7.3 Designing the Future of Culturally-Aware Tools for Crowd Work

Our paper presents findings that we hope will inspire the development of novel crowd work tools. Next, we outline new design directions inspired by our findings and previous research:

/.3.1 Designing Crowd Work Tools for Play and Socialization. Crowd work often isolates workers [57], limiting their social connections [194]. Our post-study revealed that some workers from poly- chronic cultures appreciated CultureFit's socialization notifications, emphasizing the significance of enabling social interactions at work for them. Future crowd work tools could improve social- ization by integrating CSCW research on friendship interfaces [45, 164], or research on designing work interfaces for "play for play's sake [14, 92, 127, 196]," akin to having ping pong tables in the workplace [7, 49]. These tools could better cater to diverse cultural preferences, thereby enhancing worker engagement.

/.3.2 Designing Crowd Work Tools for Multiple Goals. Our scatter plot (Fig. 7) showed that some of the monochronic workers who used our tool increased their task completion, as indicated by more advanced points along the X-axis. It is likely that our culturally aware tool helped these workers concentrate better and complete tasks without interruption [87], which explains the uptick in activity. However, the plot also showed no corresponding rise along the Y-axis, which measured hourly wages. Overall, our findings indicate that although CultureFit is culturally aware, it did not significantly increase earnings for monochronic workers. In the future, research could delve into creating tools that not only promote cultural awareness but also directly focus on boosting earnings or assist workers in achieving their different goals. Incorporating insights from previous studies focused on diverse objectives—such as wage increases, skill development, and creative pursuits [29, 160, 168, 178] —can help researchers develop a more comprehensive approach to designing tools for crowd work. This approach would enhance task efficiency and support workers' broader career development, offering more inclusive support systems [105].

/.3.3 Designing Tools for Cultural Understanding in Crowd Work. A novel direction for culturally aware crowd work tools could involve integrating a real-time cultural exchange platform into the workflow. This feature would allow workers from diverse backgrounds to interactively share cultural insights and practices, thereby enriching the work environment and fostering mutual understanding. For example, workers could engage in brief, timed exchanges sharing cultural

greetings, traditions, or personal experiences before starting collaborative tasks. This would deepen understanding and appreciation among workers. Additionally, integrating AI could help overcome language barriers and merge cultural connectivity with productivity, enhancing worker engagement and broadening global perspectives. This future research could benefit from connections with prior CSCW work in "Cross-Cultural Communication Studies" [9, 112, 113, 167, 195], "Educational
Technology Research" [35, 77, 154, 173], and "Social Computing" [30, 32, 123]. These fields could guide the tool's design to handle diverse communication styles and cultural norms [80], make the platform more engaging and educational with immersive technologies [58], facilitate social interactions after or during work [93, 124], and even create a supportive worker community with effective moderation tools [103, 165, 166].

Limitations and Future Work.

The study's findings were limited by the methodology, population, and platform used. Future research should consider field experiments which focus on more granular comparisons - such as comparing polychronic workers in Latin America and India. This would compliment recent research which includes investigations into individual regions and niches, such as a new study profiling the working conditions of data anotators in India [188]. Furthermore, the study showed that workers changed their digital behaviors when using the tool, but it remains unclear how they would adapt and use it over the long term. However, there are known challenges regarding the difficulties of performing long term studies with crowdworkers [83], additionally, given the novel nature of the work there is little to no established digital infrastructure to support a longitudinal study at this time. This leaves long-term studies and development in the means to do so as valuable future work. Within our limitations we also have to recognize the challenge of "demand characteristics" — subtle signals in experiments that can shape participants' behavior to match perceived expectations. Demand characteristics might have affected participants' reported satisfaction with our tool. However, we took several steps to minimize their influence. We carefully used neutral language in our instructions and avoided leading questions to try to ensure that responses were genuine. To reduce the sense of being studied, we also designed our tool as a lightweight web plugin integrated seamlessly into workers' usual digital labor platform. Additionally, we established a control group to help discern the true impact of these characteristics. Moreover, we guaranteed anonymity and confidentiality for all responses, encouraging participants to share their honest feedback without reservation. While our study demonstrates the potential benefits of incorporating cultural awareness into crowd work tools, particularly concerning cultural dimensions of time management styles, we acknowledge that this represents only one facet of cultural diversity. Future work could explore how other cultural dimensions, such as power distance or individualism vs. collectivism [79, 80], might be integrated into culturally-aware crowd work tool design. Despite its limitations, we hope to inspire researchers to incorporate cultural dimensions into their crowd work designs.

8 Conclusion

Crowdsourcing markets often feature standardized interfaces that do not accommodate the cultural diversity of workers, which can adversely affect their well-being and productivity [80, 119, 126, 139, 152, 156, 177]. Our research studies how considering the cultural dimensions of monochronic and polychronic work styles can positively transform crowdworkers' experiences. Our paper proposes the creation of culturally customized workplace systems, exemplified by our tool "CultureFit," designed based on Chronemics and culture theories. Through a field experiment involving 55 workers from 24 countries, we found that CultureFit significantly enhanced earnings for culturally diverse workers frequently overlooked in design [65, 78, 139]. Moreover, we will introduce a novel dataset on culture and digital work, laying the groundwork for further research in the area. Overall, our findings underline the significance of incorporating cultural insights into digital labor tool design, and providing insights for future research directions.

Acknowledgments. This research was supported by an NSF CAREER grant (2339443) and an NSF Medium grant (2403252). Special thanks to the workers who participated in this research, as well as Dmitry Ustalov, Anastasia Lucas, Eeshani Mondal for their invaluable feedback.

References

[1] 2023. Global North and Global South. https://en.wikipedia.org/wiki/Global_North_and_Global_South. Accessed: 2023-04-27.

[2] Jake Alimahomed-Wilson and Ellen Reese. 2021. Surveilling Amazon's warehouse workers: racism, retaliation, and worker resistance amid the pandemic. *Work in the Global Economy* 1, 1-2 (2021), 55–73.

[3] Adriana Alvarado Garcia, Juan F Maestre, Manuhuia Barcham, Marilyn Iriarte, Marisol Wong-Villacres, Oscar A Lemus, Palak Dudani, Pedro Reynolds-Cuéllar, Ruotong Wang, and Teresa Cerratto Pargman. 2021. Decolonial pathways: Our manifesto for a decolonizing agenda in hci research and design. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–9.

[4] Walter Ángel, Saiph Savage, and Nataly Moreno. 2015. Participatory stoves: Designing renewable energy technologies for the rural sector. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*. 259–262.

[5] Ricardo Baeza-Yates, Berthier Ribeiro-Neto, et al. 1999. *Modern information retrieval*. Vol. 463. ACM press New York.

[6] Dawna I Ballard and David R Seibold. 2004. Communication-related organizational structures and work group temporal experiences: the effects of coordination method, technology type, and feedback cycle on members' construals and enactments of time. *Communication Monographs* 71, 1 (2004), 1–27.

[7] Franklin D Becker. 1990. The total workplace: Facilities management and the elastic organization. *(No Title)* (1990).

[8] H Bigelow. 1998. Cultural Differences in the Workplace. *Unpublished manuscript, RMIT University, Victoria, Australia* (1998).

[9] Pernille Bjørn, Luigina Ciolfi, Mark Ackerman, Geraldine Fitzpatrick, and Volker Wulf. 2016. Practice-based CSCW Research: ECSCW bridging across the Atlantic. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*. 210–220.

[10] Jacqueline Bloomfield and Murray J Fisher. 2019. Quantitative research design. *Journal of the Australasian Rehabilita- tion Nurses Association* 22, 2 (2019), 27–30.

[11] Allen C Bluedorn. 2002. *The human organization of time: Temporal realities and experience*. Stanford University Press.

[12] Allen C Bluedorn, Thomas J Kalliath, Michael J Strube, and Gregg D Martin. 1999. Polychronicity and the Inventory of Polychronic Values (IPV): The development of an instrument to measure a fundamental dimension of organizational culture. *Journal of managerial psychology* 14, 3/4 (1999), 205–231.

[13] Allen C Bluedorn, Carol Felker Kaufman, and Paul M Lane. 1992. How many things do you like to do at once? An introduction to monochronic and polychronic time. *Academy of Management Perspectives* 6, 4 (1992), 17–26.

[14] Mark Blythe and Andrew Monk. 2004. Funology 2. (2004).

[15] Ricarda B Bouncken. 2004. Cultural diversity in entrepreneurial teams: Findings of new ventures in Germany. *Creativity and Innovation Management* 13, 4 (2004), 240–253.

[16] Daren C Brabham. 2012. Crowdsourcing: A model for leveraging online communities. In *The participatory cultures handbook*. Routledge, 120–129.

[17] Alexander Braylan and Matthew Lease. 2021. Aggregating Complex Annotations via Merging and Matching. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 86–94.

[18] Jan Brejcha. 2015. *Cross-cultural human-computer interaction and user experience design: a semiotic perspective*. CRC Press.

[19] Duncan P. Brumby, Anna L. Cox, Jonathan Back, and Sandy JJ Gould. 2013. Recovering from an interruption: Investigating speed accuracy trade offs in task resumption behavior. *Journal of Experimental Psychology: Applied* 19, 2 (2013), 95–107.

[20] Stefan Buttcher, Charles LA Clarke, and Gordon V Cormack. 2016. *Information retrieval: Implementing and evaluating search engines*. Mit Press.

[21] Dan Calacci and Alex Pentland. 2022. Bargaining with the Black-Box: Designing and Deploying Worker-Centric Tools to Audit Algorithmic Management. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–24.

[22] Patrick J Carr and Maria J Kefalas. 2009. *Hollowing out the middle: The rural brain drain and what it means for America*. Beacon press.

[23] Ben Carterette and Rosie Jones. 2007. Evaluating search engines by modeling the relationship between relevance and clicks. *Advances in neural information processing systems* 20 (2007).

[24] Antonio A Casilli. 2017. Global digital culture| Digital labor studies go global: Toward a digital decolonial turn. *International Journal of Communication* 11 (2017), 21.

[25] Robert Chambers. 2014. *Rural development: Putting the last first*. Routledge.

[26] Angela Chen. 2020. Desperate Venezuelans are making money by training AI for self-driving cars. https://www. technologyreview.com/2019/08/22/65375/venezuela-crisis-platform-work-trains-self-driving-car-ai-data/

[27] Angela Chen. 2021. Desperate Venezuelans are making money by training AI for self-driving cars.

[28] EunJeong Cheon. 2024. Examining Algorithmic Metrics and their Effects through the Lens of Reactivity. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. 3179–3192.

[29] Chun-Wei Chiang, Anna Kasunic, and Saiph Savage. 2018. Crowd coach: Peer coaching for crowd workers' skill growth. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–17.

[30] Nicholas A Christakis and James H Fowler. 2009. *Connected: The surprising power of our social networks and how they shape our lives*. Little, Brown Spark.

[31] Larry B Christensen, Burke Johnson, Lisa Anne Turner, and Larry B Christensen. 2011. *Research methods, design, and analysis*. Pearson Boston.

[32] Claudio U Ciborra. 1997. *Groupware and teamwork: Invisible aid or technical hindrance?* John Wiley & Sons, Inc.

[33] Shroma Worker Collective. 2023. Monitoring - Shroma. https://shroma.ge/monitor/?lang=en. Accessed: 2023-12-10.

[34] Paul Collier. 2008. *The bottom billion: Why the poorest countries are failing and what can be done about it*. Oxford University Press, USA.

[35] Gráinne Conole. 2012. *Designing for learning in an open world*. Vol. 4. Springer Science & Business Media.

[36] June Cotte and Srinivasan Ratneshwar. 1999. Juggling and hopping: what does it mean to work polychronically? *Journal of Managerial Psychology* (1999).

[37] Shipt Creators. 2023. Shipt Transparency Calculator. https://home.coworker.org/shiptcalc/. Accessed: 2023-12-10.

[38] Maya De Los Santos, Norma Elva Chávez, Alberto Navarrete, Cristina Martínez Pinto, Luz Elena González, Tatiana Telles-Calderon, and Saiph Savage. 2023. La Independiente: Designing Ubiquitous Systems for Latin American and Caribbean Women Crowdworkers. In *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing*. 404–406.

[39] Maya De los Santos, José Emilio García, Norma Elva Chávez, Alberto Navarrete, Cristina Martínez, Luz Elena González, Tatiana Telles, Paola Ricaurte, and Saiph Savage. 2023. La Independiente: an AI-enhanced Platform Co-Designed with Latin-American Crowd-Workers. *Avances en Interacción Humano-Computadora* 8, 1 (2023), 6–10.

[40] Valerio De Stefano. 2015. The rise of the just-in-time workforce: On-demand work, crowdwork, and labor protection in the gig-economy. *Comp. Lab. L. & Pol'y J.* 37 (2015), 471.

[41] Djellel Difallah, Elena Filatova, and Panos Ipeirotis. 2018. Demographics and dynamics of mechanical turk workers. In *Proceedings of the eleventh ACM international conference on web search and data mining*. 135–143.

[42] Kimberly Do, Maya De Los Santos, Michael Muller, and Saiph Savage. 2024. GigSousveillance: Designing Gig Worker Centric Sousveillance Tools. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*.

[43] Cynthia M Duncan. 2015. *Worlds apart: Poverty and politics in rural America*. Yale University Press.

[44] Tom Edixhoven, Sihang Qiu, Lucie Kuiper, Olivier Dikken, Gwennan Smitskamp, and Ujwal Gadiraju. 2021. Improving Reactions to Rejection in Crowdsourcing Through Self-Reflection. In *13th ACM Web Science Conference 2021*. 74–83.

[45] Thomas Erickson and Wendy A Kellogg. 2000. Social translucence: an approach to designing systems that support social processes. *ACM transactions on computer-human interaction (TOCHI)* 7, 1 (2000), 59–83.

[46] Susan R Fiorentino and Sandra M Tomkowicz. 2019. E-Petitions and Protected Concerted Activity: The Millennial Response to Organized Labor? *ABAJ Lab. & Emp. L.* 34 (2019), 71.

[47] Claudia Flores-Saviaga, Yuwen Li, Benjamin Hanrahan, Jeffrey Bigham, and Saiph Savage. 2020. The Challenges of Crowd Workers in Rural and Urban America. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 8, 1 (2020), 159–162.

[48] Sarah E Fox, Vera Khovanskaya, Clara Crivellaro, Niloufar Salehi, Lynn Dombrowski, Chinmay Kulkarni, Lilly Irani, and Jodi Forlizzi. 2020. Worker-centered design: Expanding HCI methods for supporting labor. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems*. 1–8.

[49] Ron Friedman. 2014. *The best place to work: The art and science of creating an extraordinary workplace*. TarcherPerigee.

[50] Ashley Fulmer, Brandon Crosby, and Michele Gelfand. 2014. *Cross-cultural Perspectives on Time*. https://doi.org/10. 4324/9781315798370

[51] Snehal Gaikwad, Durim Morina, Rohit Nistala, Megha Agarwal, Alison Cossette, Radhika Bhanu, Saiph Savage, Vishwajeet Narwal, Karan Rajpal, Jeff Regino, et al. 2015. Daemo: A self-governed crowdsourcing marketplace. In *Adjunct proceedings of the 28th annual ACM symposium on user interface software & technology*. 101–102.

[52] Snehalkumar (Neil) S Gaikwad, Durim Morina, Adam Ginzberg, Catherine Mullings, Shirish Goyal, Dilrukshi Gamage, Christopher Diemert, Mathias Burton, Sharon Zhou, Mark Whiting, et al. 2016. Boomerang: Rebounding the consequences of reputation feedback on crowdsourcing platforms. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. 625–637.

[53] Yashesh Gaur, Walter S Lasecki, Florian Metze, and Jeffrey P Bigham. 2016. The effects of automatic speech recognition quality on human transcription latency. In *Proceedings of the 13th International Web for All Conference*. 1–8.

[54] Shromis Gemo. 2024. Introduction to Wage Theft. https://shroma.ge/en/wagetheft-intro-en/. Accessed: 2023-12-10.

[55] Wen Gong. 2009. National culture and global diffusion of business-to-consumer e-commerce. *Cross cultural manage- ment: an international journal* (2009).

[56] Mitchell L Gordon, Michelle S Lam, Joon Sung Park, Kayur Patel, Jeff Hancock, Tatsunori Hashimoto, and Michael S Bernstein. 2022. Jury learning: Integrating dissenting voices into machine learning models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–19.

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

[58] Sue Gregory, Mark JW Lee, Barney Dalgarno, and Belinda Tynan. 2016. *Learning in virtual worlds: Research and applications*. Athabasca University Press.

[59] Neha Gupta, David Martin, Benjamin V Hanrahan, and Jacki O'Neill. 2014. Turk-life in India. In *Proceedings of the 18th International Conference on Supporting Group Work*. 1–11.

[60] Edward Hall. 1971. The paradox of culture. *B. Landis and ES Tauber (Eds.): In the Name of Life. Essays in Honor of Erich Fromm, New York (Holt, Rinehart and Winston) 1970, pp. 218-235.* (1971).

[61] Edward Twitchell Hall. 1989. *Beyond culture*. Anchor.

[62] Edward T Hall and Edward Twitchell Hall. 1989. *The dance of life: The other dimension of time*. Anchor.

[63] Edward Twitchell Hall and Mildred Reed Hall. 1990. *Understanding cultural differences:[keys to success in West Germany, France, and the United States]*. Intercultural Press.

[64] Edward Twitchell Hall, Mildred Reed Hall, et al. 1989. *Understanding cultural differences*. Intercultural press.

[65] Edward Twitchell Hall and T Hall. 1959. *The silent language*. Vol. 948. Anchor books.

[66] Benjamin V Hanrahan, Ning F Ma, Eber Betanzos, and Saiph Savage. 2020. Reciprocal research: Providing value in design research from the outset in the rural united states. In *Proceedings of the 2020 International Conference on Information and Communication Technologies and Development*. 1–5.

[67] B. V. Hanrahan, J. K. Willamowski, S. Swaminathan, and D. B. Martin. 2015. TurkBench: Rendering the market for Turkers. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 1613–1616.

[68] Kotaro Hara, Abigail Adams, Kristy Milland, Saiph Savage, Chris Callison-Burch, and Jeffrey P Bigham. 2018. A data-driven analysis of workers' earnings on Amazon Mechanical Turk. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–14.

[69] Kotaro Hara, Abigail Adams, Kristy Milland, Saiph Savage, Benjamin V Hanrahan, Jeffrey P Bigham, and Chris Callison-Burch. 2019. Worker demographics and earnings on amazon mechanical turk: An exploratory analysis. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6.

[70] Gunnar Harboe and Elaine M Huang. 2015. Real-world affinity diagramming practices: Bridging the paper-digital gap. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 95–104.

[71] Jean Hardy. 2019. How the design of social technology fails rural America. In *Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion*. 189–193.

[72] Jean Hardy, Dharma Dailey, Susan Wyche, and Norman Makoto Su. 2018. Rural computing: Beyond access and infrastructure. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 463–470.

[73] Jean Hardy, Chanda Phelan, Morgan Vigil-Hayes, Norman Makoto Su, Susan Wyche, and Phoebe Sengers. 2019. Designing from the rural. *Interactions* 26, 4 (2019), 37–41.

[74] Danula Hettiachchi, Mike Schaekermann, Tristan J McKinney, and Matthew Lease. 2021. The Challenge of Variable Effort Crowdsourcing and How Visible Gold Can Help. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–26.

[75] Ann Heylighen. 2014. About the nature of design in universal design. *Disability and rehabilitation* 36, 16 (2014), 1360–1368.

[76] E Jeffrey Hill, Chongming Yang, Alan J Hawkins, and Maria Ferris. 2004. A cross-cultural test of the work-family interface in 48 countries. *Journal of Marriage and the Family* (2004), 1300–1316.

[77] Cindy E Hmelo-Silver. 2013. The international handbook of collaborative learning. (2013).

[78] Geert Hofstede. 1984. Cultural dimensions in management and planning. *Asia Pacific journal of management* 1 (1984), 81–99.

[79] Geert Hofstede. 2001. *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Sage publications.

[80] Geert Hofstede, Gert Jan Hofstede, and Michael Minkov. 2005. *Cultures and organizations: Software of the mind*. Vol. 2. Mcgraw-hill New York.

[81] John Hooker. 2003. *Working across cultures*. Stanford University Press.

[82] Ding-Long Huang, Pei-Luen Patrick Rau, Hui Su, Nan Tu, and Chen Zhao. 2010. Effects of communication style and time orientation on notification systems and anti-virus software. *Behaviour & Information Technology* 29, 5 (2010), 483–495.

[83] Ting-Hao Kenneth Huang, Walter S Lasecki, Amos Azaria, and Jeffrey P Bigham. 2016. " Is There Anything Else I Can Help You With?" Challenges in Deploying an On-Demand Crowd-Powered Conversational Agent. In *Fourth AAAI Conference on Human Computation and Crowdsourcing*.

[84] Lilly Irani. 2015. The cultural work of microwork. *New media & society* 17, 5 (2015), 720–739.

[85] Lilly Irani, Janet Vertesi, Paul Dourish, Kavita Philip, and Rebecca E Grinter. 2010. Postcolonial computing: a lens on design and development. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1311–1320.

[86] Lilly C Irani and M Six Silberman. 2013. Turkopticon: Interrupting worker invisibility in amazon mechanical turk. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 611–620.

[87] Christian P. Janssen, Shamsi T. Iqbal, Andrew L. Kun, and Stella F. Donker. 2019. Interrupted by my car? Implications of interruption and interleaving research for automated vehicles. *International Journal of Human-Computer Studies* 130 (2019), 221–233.

[88] Gareth R Jones. 1983. Transaction costs, property rights, and organizational culture: An exchange perspective. *Administrative Science Quarterly* (1983), 454–467.

[89] Hyun Joon Jung and Matthew Lease. 2012. Improving quality of crowdsourced labels via probabilistic matrix factorization. In *Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence*.

[90] Shivani Kapania, Alex S Taylor, and Ding Wang. 2023. A hunt for the Snark: Annotator Diversity in Data Practices. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.

[91] Toni Kaplan, Susumu Saito, Kotaro Hara, and Jeffrey P Bigham. 2018. Striving to earn more: a survey of work strategies and tool use among crowd workers. In *Sixth AAAI Conference on Human Computation and Crowdsourcing*.

[92] Anna Kasunic, Chun-Wei Chiang, Geoff Kaufman, and Saiph Savage. 2019. Crowd Work on a CV? Understanding How AMT Fits into Turkers' Career Goals and Professional Profiles. *arXiv preprint arXiv:1902.05361* (2019).

[93] Anna Kasunic, Chun-Wei Chiang, Geoff Kaufman, and Saiph Savage. 2019. Turker Tales: Integrating Tangential Play into Crowd Work. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. 21–34.

[94] Carol Kaufman-Scarborough and Jay Lindquist. 1999. Time Management and Polychronicity: Comparisons, Contrasts, and Insights for the Workplace. *Journal of Managerial Psychology* 14 (06 1999), 288–312. https://doi.org/10.1108/ 02683949910263819

[95] Elizabeth Kaziunas, Michael S Klinkman, and Mark S Ackerman. 2019. Precarious interventions: Designing for ecologies of care. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–27.

[96] Vera Khovanskaya, Phoebe Sengers, Melissa Mazmanian, and Charles Darrah. 2017. Reworking the gaps between design and ethnography. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 5373–5385.

[97] P. Kim, E. Cheon, and S. Sawyer. 2023. Online Freelancing on Digital Labor Platforms: A Scoping Review. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*. 259–266.

[98] S. Kingsley, M. S. Silberman, C. Wang, R. Lambeth, J. Zhi, M. Eslami, B. Li, and J. Bigham. 2024. 'Your Duties Are To Sweep A Floor Remotely': Low Information Quality in Job Advertisements is a Barrier to Low-Income Job-Seekers' Successful Use of Digital Platforms. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work*. 1–20.

[99] Sara Constance Kingsley, Mary L. Gray, and Siddharth Suri. 2015. Accounting for market frictions and power asymmetries in online labor markets. *Policy & Internet* 7, 4 (2015), 383–400.

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

[101] Satoshi Kose. 1998. From barrier-free to universal design: an international perspective. *Assistive technology* 10, 1 (1998), 44–50.

[102] Kenneth L Kraemer, Jason Dedrick, and Prakul Sharma. 2009. One laptop per child: vision vs. reality. *Commun. ACM* 52, 6 (2009), 66–73.

[103] Robert E Kraut and Paul Resnick. 2012. *Building successful online communities: Evidence-based social design*. Mit Press.

[104] Ranjay Krishna, Yuke Zhu, Oliver Groth, Justin Johnson, Kenji Hata, Joshua Kravitz, Stephanie Chen, Yannis Kalantidis, Li-Jia Li, David A Shamma, et al. 2017. Visual genome: Connecting language and vision using crowdsourced dense image annotations. *International journal of computer vision* 123 (2017), 32–73.

[105] Andrew L. Kun and Orit Shaer. 2023. Designing an Inclusive and Engaging Hybrid Event: Experiences From CHIWORK. *IEEE Pervasive Computing* 22, 3 (2023), 52–57.
[106] L. Lascau, D. P. Brumby, S. J. Gould, and A. L. Cox. 2024. "Sometimes It's Like Putting the Track in Front of the Rushing Train": Having to Be 'On Call' for Work Limits the Temporal Flexibility of Crowdworkers. *ACM Transactions on Computer-Human Interaction* 31, 2 (29 Jan 2024), 1–45.
[107] Laura Lascau, Sandy JJ Gould, Duncan P Brumby, and Anna L Cox. 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. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–8.
[108] Laura Lascau, Sandy JJ Gould, Anna L Cox, Elizaveta Karmannaya, and Duncan P Brumby. 2019. Monotasking or multitasking: Designing for crowdworkers' preferences. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14.
[109] Heejin Lee. 1999. Time and information technology: monochronicity, polychronicity and temporal symmetry. *European Journal of Information Systems* 8, 1 (1999), 16–26.
[110] Kiefer Lee and Steve Carter. 2011. Global marketing management. *Strategic Direction* 27, 1 (2011).
[111] Richard Lewis. 2010. *When cultures collide: Leading across cultures*. Nicholas Brealey International.
[112] Xiaoyan Li, Naomi Yamashita, Wen Duan, Yoshinari Shirai, and Susan R Fussell. 2023. Improving Non-Native Speakers' Participation with an Automatic Agent in Multilingual Groups. *Proceedings of the ACM on Human-Computer Interaction* 7, GROUP (2023), 1–28.
[113] Silvia Lindtner, Ken Anderson, and Paul Dourish. 2012. Cultural appropriation: information technologies as sites of transnational imagination. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*. 77–86.
[114] Pasquale Lops, Dietmar Jannach, Cataldo Musto, Toine Bogers, and Marijn Koolen. 2019. Trends in content-based recommendation. *User Modeling and User-Adapted Interaction* 29, 2 (2019), 239–249.
[115] Andrés Lucero. 2015. Using affinity diagrams to evaluate interactive prototypes. In *IFIP conference on human-computer interaction*. Springer, 231–248.
[116] Christopher D Manning, Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to information retrieval*. Cambridge university press.
[117] Lalita A Manrai and Ajay K Manrai. 1995. Effects of cultural-context, gender, and acculturation on perceptions of work versus social/leisure time usage. *Journal of Business Research* 32, 2 (1995), 115–128.
[118] Masoumeh Mansouri and Henry Taylor. 2024. Does Cultural Robotics Need Culture? Conceptual Fragmentation and the Problems of Merging Culture with Robot Design. *International Journal of Social Robotics* 16, 2 (2024), 385–401.
[119] Aaron Marcus. 2006. Cross-cultural user-experience design. In *International Conference on Theory and Application of Diagrams*. Springer, 16–24.
[120] Aaron Marcus and Emilie West Gould. 2000. Crosscurrents: cultural dimensions and global Web user-interface design. *interactions* 7, 4 (2000), 32–46.
[121] David Martin, Benjamin V Hanrahan, Jacki O'Neill, and Neha Gupta. 2014. Being a turker. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. 224–235.
[122] DS McCrickard. 2003. Chewar. CM, Somervell, JP, & Ndiwalana. *A." A Model for Notification Systems Evaluation– Assessing User Goals for Multitasking Activity." ACM Transactions on Computer-Human Interaction* (2003), 312–338.
[123] Gabriele Meiselwitz. 2020. *Social Computing and Social Media. Design, Ethics, User Behavior, and Social Network Analysis: 12th International Conference, SCSM 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part I*. Vol. 12194. Springer Nature.
[124] Jonathan Mella, Ioanna Iacovides, and Anna Cox. 2024. 'Jumping Out from the Pressure of Work and into the Game': Curating Immersive Digital Game Experiences for Post-Work Recovery. *ACM Games: Research and Practice* (2024).
[125] Jack R Meredith, Scott M Shafer, and Samuel J Mantel Jr. 2017. *Project management: a strategic managerial approach*. John Wiley & Sons.
[126] Erin Meyer. 2014. *The culture map: Breaking through the invisible boundaries of global business*. Public Affairs.
[127] Andrew Monk, Marc Hassenzahl, Mark Blythe, and Darren Reed. 2002. Funology: designing enjoyment. In *CHI'02 extended abstracts on human factors in computing systems*. 924–925.
[128] Tony Morden. 1999. Models of national culture–a management review. *Cross Cultural Management: An International Journal* 6, 1 (1999), 19–44.
[129] Bruno Moreschi, Gabriel Pereira, and Fabio G Cozman. 2020. The Brazilian Workers in Amazon Mechanical Turk: Dreams and realities of ghost workers. *Contracampo* 39, 1 (2020).
[130] Luciara Nardon and Richard M Steers. 2009. The culture theory jungle: Divergence and convergence in models of national culture. *Cambridge handbook of culture, organizations, and work* (2009), 3–22.
[131] Joseph W. Newbold, Anna Rudnicka, David Cook, Marta E. Cecchinato, Sandy JJ Gould, and Anna L. Cox. 2022. The new normals of work: a framework for understanding responses to disruptions created by new futures of work.

Human–Computer Interaction 37, 6 (2022), 508–531.
[132] Gemma Newlands and Christoph Lutz. 2021. Crowdwork and the mobile underclass: Barriers to participation in India and the United States. *new media & society* 23, 6 (2021), 1341–1361.
[133] Don Norman. 2012. Does culture matter for product design. *Core 77 Design magazine &* (2012).
[134] Don Norman. 2013. *The design of everyday things: Revised and expanded edition*. Basic books.
[135] Donald A Norman. 1991. Cognitive artifacts. *Designing interaction: Psychology at the human-computer interface* 1, 1 (1991), 17–38.
[136] Donald A Norman. 2016. *Living with complexity*. MIT press.
[137] Georgia Fair Labor Platform. 2020. Wage Theft Calculator. Georgia Fair Labor Platform.
[138] Elizabeth M Poposki and Frederick L Oswald. 2010. The multitasking preference inventory: Toward an improved measure of individual differences in polychronicity. *Human Performance* 23, 3 (2010), 247–264.
[139] Julian Posada. 2022. *The Coloniality of Data Work: Power and Inequality in Outsourced Data Production for Machine Learning*. Ph. D. Dissertation. University of Toronto (Canada).
[140] Nastazja Potocka-Sionek. 2022. Crowdwork and global supply chains: Regulating digital piecework. In *A Research Agenda for the Gig Economy and Society*. Edward Elgar Publishing, 215–234.
[141] Vivek Krishna Pradhan, Mike Schaekermann, and Matthew Lease. 2022. In Search of Ambiguity: A Three-Stage Workflow Design to Clarify Annotation Guidelines for Crowd Workers. *Frontiers in Artificial Intelligence* 5 (2022).
[142] CK Prahalad. 2008. The fortune at the bottom of the pyramid: Eradicating poverty through profits. *McKinsey briefing notes series* 36, 3 (2008), 52–74.
[143] Sihang Qiu, Alessandro Bozzon, Max V Birk, and Ujwal Gadiraju. 2021. Using worker avatars to improve microtask crowdsourcing. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–28.
[144] Whitney Quesenbery and Daniel Szuc. 2011. *Global UX: design and research in a connected world*. Elsevier.
[145] Md Mustafizur Rahman, Mucahid Kutlu, and Matthew Lease. 2019. Constructing test collections using multi-armed bandits and active learning. In *The World Wide Web Conference*. 3158–3164.
[146] Arun Rai, Andrew Burton-Jones, Hsinchun Chen, Alok Gupta, Alan R Hevner, Wolfgang Ketter, Jeffrey Parsons, H Raghav Rao, Sumit Sarkar, and Youngjin Yoo. 2017. Editor's comments: Diversity of design science research. (2017).
[147] Pei-Luen Patrick Rau, Tom Plocher, and Yee-Yin Choong. 2012. *Cross-cultural design for IT products and services*. CRC Press.
[148] PL Patrick Rau. 2015. *Cross-Cultural Design Methods, Practice and Impact: 7th International Conference, CCD 2015, Held as Part of HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015, Proceedings, Part I*. Vol. 9180. Springer.
[149] Katharina Reinecke. 2010. *Culturally adaptive user interfaces*. Ph. D. Dissertation. University of Zurich.
[150] Katharina Reinecke and Abraham Bernstein. 2011. Improving performance, perceived usability, and aesthetics with culturally adaptive user interfaces. *ACM Transactions on Computer-Human Interaction (TOCHI)* 18, 2 (2011), 1–29.
[151] Cecil R Reynolds and Elaine Fletcher-Janzen. 2007. *Encyclopedia of Special Education: A Reference for the Education of Children, Adolescents, and Adults with Disabilities and Other Exceptional Individuals, Volume 3*. Vol. 3. John Wiley & Sons.
[152] Paola Ricaurte. 2019. Data epistemologies, the coloniality of power, and resistance. *Television & New Media* 20, 4 (2019), 350–365.
[153] Veronica A Rivera and David T Lee. 2021. I Want to, but First I Need to: Understanding Crowdworkers' Career Goals, Challenges, and Tensions. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–22.
[154] Tim S Roberts. 2005. Computer-supported collaborative learning in higher education. In *Computer-supported collaborative learning in higher education*. IGI global, 1–18.
[155] William Peter Robinson and Howard Giles. 1990. *Handbook of language and social psychology*. Wiley Chichester, UK.
[156] A. M. Roffarello, L. De Russis, D. Lottridge, and M. E. Cecchinato. 2023. Understanding digital wellbeing within complex technological contexts. *International Journal of Human-Computer Studies* 175 (2023), 103034.
[157] Joel Ross, Lilly Irani, M Six Silberman, Andrew Zaldivar, and Bill Tomlinson. 2010. Who are the crowdworkers? Shifting demographics in Mechanical Turk. In *CHI'10 extended abstracts on Human factors in computing systems*. 2863–2872.
[158] Susumu Saito, Chun-Wei Chiang, Saiph Savage, Teppei Nakano, Tetsunori Kobayashi, and Jeffrey Bigham. 2019. Turkscanner: Predicting the hourly wage of microtasks. In *The World Wide Web Conference*. 3187–3193.
[159] Shruti Sannon and Dan Cosley. 2019. Privacy, power, and invisible labor on Amazon Mechanical Turk. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–12.
[160] Saiph Savage, Chun Wei Chiang, Susumu Saito, Carlos Toxtli, and Jeffrey Bigham. 2020. Becoming the super turker: Increasing wages via a strategy from high earning workers. In *Proceedings of The Web Conference 2020*. 1241–1252.
[161] Saiph Savage and Martha Garcia-Murillo. 2024. Tools for crowdworkers coding data for AI. In *Handbook of Artificial Intelligence at Work*, Joy Xiang and Peter Lee (Eds.). Edward Elgar Publishing, 76–94.

[162] S Savage, C Toxtli, and E Betanzos. 2021. Research Methods to Study & Empower Crowd Workers. *Oxford University Press* (2021).

[163] Sergio Sayago. 2023. *Cultures in human-computer interaction*. Springer Nature.

[164] Victoria Schwanda Sosik, Xuan Zhao, and Dan Cosley. 2012. See friendship, sort of: How conversation and digital traces might support reflection on friendships. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*. 1145–1154.

[165] Joseph Seering, Juan Pablo Flores, Saiph Savage, and Jessica Hammer. 2018. The social roles of bots: evaluating impact of bots on discussions in online communities. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–29.

[166] Joseph Seering, Ole J. Mengshoel, Quinn Liao, and Geoffrey Kaufman. 2019. Moderator engagement and community development in the age of algorithms. *New Media & Society* 21, 7 (2019), 1417–1443.

[167] Leslie D Setlock, Susan R Fussell, and Christine Neuwirth. 2004. Taking it out of context: collaborating within and across cultures in face-to-face settings and via instant messaging. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*. 604–613.

[168] Orit Shaer, Angelora Cooper, Osnat Mokryn, Andrew L. Kun, and Hagit Ben Shoshan. 2024. AI-Augmented Brainwriting: Investigating the use of LLMs in group ideation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–17.

[169] Mildred Sikkema and Agnes Niyekawa. 1987. *Design for Cross-Cultural Learning.* ERIC.

[170] M Silberman. 2015. Operating an employer reputation system: Lessons from Turkopticon, 2008-2015. *Comp. Lab. L. & Pol'y J.* 37 (2015), 505.

[171] M Six Silberman, Bill Tomlinson, Rochelle LaPlante, Joel Ross, Lilly Irani, and Andrew Zaldivar. 2018. Responsible research with crowds: pay crowdworkers at least minimum wage. *Commun. ACM* 61, 3 (2018), 39–41.

[172] Jennifer Daryl Slack and John Macgregor Wise. 2005. *Culture+ technology: A primer*. Peter Lang.

[173] Gerry Stahl. 2006. *Group cognition: Computer support for building collaborative knowledge (acting with technology)*. The MIT Press.

[174] Edward Steinfeld and Jordana Maisel. 2012. *Universal design: Creating inclusive environments*. John Wiley & Sons.

[175] Joseph Stiglitz and Robert M Pike. 2004. Globalization and its Discontents. (2004).

[176] Molly Follette Story. 2001. Principles of universal design. *Universal design handbook* (2001).

[177] Huatong Sun. 2012. *Cross-cultural technology design: Creating culture-sensitive technology for local users*. OUP USA.

[178] Ryo Suzuki, Niloufar Salehi, Michelle S Lam, Juan C Marroquin, and Michael S Bernstein. 2016. Atelier: Repurposing expert crowdsourcing tasks as micro-internships. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 2645–2656.

[179] Carlos Toxtli, Angela Richmond-Fuller, and Saiph Savage. 2020. Reputation agent: Prompting fair reviews in gig markets. In *Proceedings of The Web Conference 2020*. 1228–1240.

[180] Carlos Toxtli, Siddharth Suri, and Saiph Savage. 2021. Quantifying the invisible labor in crowd work. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–26.

[181] Fons Trompenaars and Charles Hampden-Turner. 2004. *Managing people across cultures*. Capstone Chichester.

[182] Fons Trompenaars and Charles Hampden-Turner. 2011. *Riding the waves of culture: Understanding diversity in global business*. Nicholas Brealey International.

[183] Upwork. 2022. Use your work diary - upwork customer service & support. https://support.upwork.com/hc/en- us/articles/211068518-Use-Your-Work-Diary

[184] Oana Ursu and Elena Ciortescu. 2021. Exploring Cultural Patterns in Business Communication. Insights from Europe and Asia. *CES Working Papers* 13, 2 (2021), 149–158.

[185] Rama Adithya Varanasi, Divya Siddarth, Vivek Seshadri, Kalika Bali, and Aditya Vashistha. 2022. Feeling Proud, Feeling Embarrassed: Experiences of Low-income Women with Crowd Work. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–18.

[186] Ellen M Voorhees, Donna K Harman, et al. 2005. *TREC: Experiment and evaluation in information retrieval*. Vol. 63. Citeseer.

[187] Immanuel Wallerstein. 2020. *World-systems analysis: An introduction*. duke university Press.

[188] Ding Wang, Shantanu Prabhat, and Nithya Sambasivan. 2022. Whose AI Dream? In Search of the Aspiration in Data Annotation.. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 582, 16 pages. https://doi.org/10.1145/3491102.3502121

[189] Xinyi Wang, Haiyi Zhu, Yangyun Li, Yu Cui, and Joseph Konstan. 2017. A community rather than a union: Understanding self-organization phenomenon on Mturk and how it impacts Turkers and requesters. In *Proceedings of the 2017 CHI conference extended abstracts on human factors in computing systems*. 2210–2216.

[190] Senuri Wijenayake, Danula Hettiachchi, and Jorge Goncalves. 2023. Combining Worker Factors for Heterogeneous Crowd Task Assignment. In *Proceedings of the ACM Web Conference 2023*. 3794–3805.

[191] Alex C Williams, Gloria Mark, Kristy Milland, Edward Lank, and Edith Law. 2019. The Perpetual Work Life of Crowdworkers: How Tooling Practices Increase Fragmentation in Crowdwork. *Proceedings of the ACM on Human- Computer Interaction* 3, CSCW (2019), 1–28.

[192] Haoyu Xie. 2023. *Understanding the Use of HIT Catchers and Crowd Knowledge Sharing: A Case Study of Crowdworkers on Amazon Mechanical Turk*. Ph. D. Dissertation. University of Sheffield.

[193] Judith Yaaqoubi. 2020. *Practitioners' Views on Cultural Adaptation of Web-based Products*. University of Washington.

[194] Zheng Yao, Silas Weden, Lea Emerlyn, Haiyi Zhu, and Robert E Kraut. 2021. Together but alone: Atomization and peer support among gig workers. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–29.

[195] Chien Wen Yuan, Leslie D Setlock, Dan Cosley, and Susan R Fussell. 2013. Understanding informal communication in multilingual contexts. In *Proceedings of the 2013 conference on Computer supported cooperative work*. 909–922.

[196] G Zichermann. 2020. Gamification at Work: Designing Engaging Business Software. *Retrieved from The Interaction Design Foundation: https://www. interactiondesign. org/literature/book/gamification-at-work-designing-engaging- businesssoftware* (2020).

[197] Kathryn Zyskowski and Kristy Milland. 2018. A crowded future: Working against abstraction on Turker Nation. *Catalyst: Feminism, Theory, Technoscience* 4, 2 (2018), 1–30.

Appendix A

Our Appendix provides information on: 1) CultureFit's recommender system that powers its notification system's task feeds; 2) the surveys conducted before and after the study.

A.1 CultureFit's Recommendation Algorithm

CultureFit utilizes a content-based recommendation algorithm framed as a regression model, which operates through a feed-forward back-propagation neural network implemented via TensorFlow.js[2] [114]. This model is designed to predict the likelihood that a task will be initiated or completed by a crowdworker, taking into account the specific characteristics of individual tasks or task batches. The input features for the model include payment per task, payment per batch of tasks, the number of tasks previously completed by the worker from the requester who is posting the task, task acceptance rate (a metric indicating the percentage of tasks accepted by workers, which can reflect task attractiveness or suitability), task duration, one-hot-encoded task category, and the task type (regular, training, or exam).

Given the relatively sparse data available per worker, we implemented a continuous learning approach, enabling dynamic updates to the model as tasks are completed. The model is re-trained locally each time a worker completes a new type of task, allowing it to adapt in real-time to the evolving actions and preferences of the worker. This strategy enhances the model's accuracy over time and tailors task recommendations more effectively to individual workers.

Our continuous learning framework employs regression techniques to consistently update and refine predictions based on incoming data. Regression is particularly well-suited for generating continuous outputs, such as the probability of task completion in this context. This output precision facilitates meticulous computation of ranking metrics, ensuring that the tasks most likely to be completed by the worker are prioritized. The model uses the Adam optimizer and Mean Squared Error (MSE) as the loss function, with a learning rate of 0.001 and a training period spanning 1,000 epochs. CultureFit maintains a consistent application of this algorithm across various cul- tural contexts—both monochronic and polychronic—ensuring uniformity in task recommendation practices.

A.1.1 General Evaluation of CultureFit's Recommendation Algorithm. We evaluated the perfor- mance of the recommendation algorithm integrated into our tool. For this purpose, we computed

[2] https://www.tensorflow.org/js

the mean average precision (MAP) at k (mAP@k) per participant to understand the relevance of the recommendations. CultureFit's recommendation algorithm achieved overall MAP scores of
0.68 at mAP@3 and 0.52 at mAP@5. For monochronic workers, the mAP@3 score was 0.67 and the mAP@5 score was 0.52. For polychronic workers, the mAP@3 score reached 0.69, while the mAP@5 was 0.51. Ultimately, this means that the first two task recommendation items were likely to be completed by both polychronic and monochronic workers.

Note that we decided to use MAP to evaluate the recommender system we integrated into CultureFit due to:

(a) **MAP Evaluates Precision and Ranking of Recommendations**: MAP is commonly used for evaluating the results of search engines [20, 116, 186], particularly because it can measure the precision of top-ranked items and assesses how well these rankings align with user preferences [5, 23]. For our tool, designed to minimize workers' search time, accurately measuring the relevance of top tasks is important—if these tasks are not relevant, we do not effectively reduce workers' search time. Therefore, MAP's sensitivity to the order of task presentation was a key reason for selecting it to evaluate the recommendation component of CultureFit. Additionally, since CultureFit employs regression techniques to refine predictions and produce a ranked output of task completion probabilities, MAP is particularly suitable. It inherently assesses the quality of these rankings by evaluating the precision at various list depths [5, 20].

(b) **Monitors Adaptation:** For CultureFit's recommendation algorithm, we integrated a continuous learning model with regression techniques, which evolves by learning from new data. On the other hand, MAP can be recalculated periodically to monitor any improvements or declines in the precision of recommendations over time. Therefore, in our context, MAP proved to be a useful tool for assessing the effectiveness of the evolving recommendation algorithm integrated into CultureFit.

(c) **Feedback Integration:** MAP can also help to quantify how well a recommendation system integrates user feedback to refine its recommendations by measuring the precision of the recommendation list at various cutoff points. This approach can provide clear insights into how a recommender system adapts to user feedback. For CultureFit, MAP is particularly useful in evaluating how the recommendation system incorporates feedback from both completed and uncompleted tasks into its model updates. This capability is important for ensuring that the recommendations accurately reflect observed worker behaviors and preferences.

Overall, MAP effectively measures whether the most relevant recommendations are presented first, factoring in user satisfaction and engagement within dynamic learning environments, such as those where CultureFit's algorithm operates, making it a suitable method for assessing our tool's success.

A.1.2 Comparing CultureFit's Recommendation Algorithm to Baselines. To further validate the effectiveness of CultureFit's recommendation algorithm, we compared its predictive performance against two baselines: a simpler, less resource-intensive algorithm, and a version of our model that is updated less frequently.

The less resource-intensive baseline, which uses a heuristic approach of recommending tasks based on the most frequently completed tasks across all workers, achieved a mean average precision at k (mAP@k) of 0.64 at mAP@3 and 0.49 at mAP@5. This indicates that although simpler, this model underperforms in predicting task relevance compared to CultureFit's more sophisticated model, which integrates continuous learning. Recall that CultureFit's recommendation algorithm achieved overall MAP scores of 0.68 at mAP@3 and 0.52 at mAP@5.

We also evaluated a version of our recommendation algorithm that is retrained only once every three days, as opposed to immnediate retraining for the main model. This less frequently updated model scored 0.66 at mAP@3 and 0.47 at mAP@5, indicating a decrease in predictive accuracy over time, likely due to stale data. However it is important to note that this model did outperform the heuristic baseline. These comparisons underscore that although CultureFit's more advanced model demands additional resources and frequent updates, it does deliver more relevant task recommendations. There is always a trade-off between resource consumption and predictive performance. Nonetheless, we hope that these results will guide researchers in future decisions regarding the optimization of update frequencies and algorithm complexity when deploying AI- enhanced tools for workers across various scenarios.

A.1.3 Evaluating the Benefits of Continuous Learning in CultureFit's Recommendation Algorithm: An Ablation Study. We also aimed to evaluate the effectiveness of the continuous learning component within CultureFit's recommendation module. For this purpose, we conducted an ablation study on CultureFit's recommendation algorithm, comparing its performance with and without retraining (i.e., the continuous learning component). We assessed the MAP at various levels of k for both configurations. The results showed that the model with continuous retraining—CultureFit's standard model—achieved MAP scores of 0.68 at mAP@3 and 0.52 at mAP@5, as previously mentioned. This outperformed the non-retraining version, which scored 0.49 at mAP@3 and 0.36 at mAP@5. This highlights the role of the continuous learning feature in adapting to new data and enhancing task prediction accuracy. Furthermore, these findings underscore the benefits of dynamic model updates for optimizing performance in real-time task prediction scenarios within crowd work tools.

A.2 Surveys

In our study, crowdworkers were required to complete a pre-survey before gaining access to our tool and a post-survey upon concluding the study. Both surveys were administered to participants in both the control group and those using the CultureFit system.

A.2.1 Pre-survey. In the following we present the pre-survey we gave to participants.

Question	Options
EXPERIENCES WITH CROWD WORK	
How long have you been working on Toloka?	Less than a month; Between 1 and 3 months; Between 3 and 6 months; Between 6 months and 1 year; Between 1 and 2 years; More than 2 and 3 years; More than 3 years
How often do you work in Toloka?	Everyday; From five to six days a week; From three to four days a week; Once or twice a week; Less than once a week
How fast can you find tasks to work on in Toloka?	1) Very Slowly - It takes me a long time to find tasks I can work on; 2) Slowly - It often takes some time to find tasks; 3) Moderately - I can find tasks in a reasonable amount of time; 4) Quickly - I usually find tasks to work on quickly; 5)Very Quickly - I can find tasks to work on immediately.

How often do you get the chance to use your strengths when working on Toloka?	1. No Extent - I do not use my strengths at all when working on Toloka; 2. Slight Extent - I use my strengths to a slight extent when work- ing on Toloka; 3. Moderate Extent - I use my strengths to a moderate extent when working on Toloka; 4. Great Extent - I use my strengths to a great extent when working on Toloka; 5. Very Great Extent - I use my strengths to a very great extent when working on Toloka.
How much do you enjoy planning your tasks on Toloka?	1) Not at all - I do not enjoy planning my tasks at all; 2) Slightly - I enjoy planning my tasks a little; 3) Moderately - I somewhat enjoy plan- ning my tasks; 4) Quite a bit - I enjoy planning my tasks quite a lot; 5) Extremely - I enjoy planning my tasks immensely.
How much do you like having someone over-see the work you do?	1) Not at all - I do not like having someone oversee my work at all; 2) Slightly - I slightly dislike having someone oversee my work; 3) Neutral - I am neutral about having someone oversee my work; 4) Somewhat - I somewhat like having someone oversee my work; 5) Very much - I very much like having someone over- see my work.
How often do you use Toloka on your Desktop?	1) Never - I never use Toloka on a desktop; 2) Rarely - I rarely use Toloka on a desktop; 3) Sometimes - I sometimes use Toloka on a desk- top. 4) Often - I often use Toloka on a desktop; 5) Always - I always use Toloka on a desktop.
How often do you use Toloka on your Smart-phone?	1) Never - I never use Toloka on a smartphone; 2) Rarely - I rarely use Toloka on a smartphone; 3) Sometimes - I sometimes use Toloka on a smartphone; 4) Often - I often use Toloka on a smartphone; 4) Always - I always use Toloka on a smartphone.
What web browsers do you use? Check all that apply.	Chrome; Firefox; Opera; Yandex; Brave; Other
Name a few of the tools you use to help you do work on Toloka?	Open-ended
Which forums do you visit to discuss topics re-lated to Toloka, including tasks and requesters?	Open-ended
What other crowdsourcing or Gig markets have you worked on previously? Check all that apply.	Amazon Mechanical Turk; LiveOps; Sama-source; Galaxy Zoo; Prolific; Upwork; Other

Since when have you been working on other Crowdsourcing or Gig Markets?	Never; Less than a month; Between 1 and 3 months; Between 3 and 6 months; Between 6 months and 1 year; Between 1 and 2 years; More than 2 and 3 years; More than 3 years
MPI QUESTIONS	
I prefer to work on several projects in a day, rather than completing one project and then switching to another.	1) Not at all; 2) Slightly; 3) Neutral; 4) Some- what; 5) Very much.
I would like to work in a job where I was con- stantly shifting from one task to another, like a receptionist or an air traffic controller.	1) Not at all; 2) Slightly; 3) Neutral; 4) Some- what; 5) Very much.
I lose interest in what I am doing if I have to focus on the same task for long periods of time, without thinking about or doing something else.	1) Not at all; 2) Slightly; 3) Neutral; 4) Some- what; 5) Very much.
When doing a number of assignments, I like to switch back and forth between them rather than do one at a time.	1) Not at all; 2) Slightly; 3) Neutral; 4) Some- what; 5) Very much.
To see if you're still paying attention, please select the choice that says neutral.	Strongly Disagree 1 2 3 4 5 Strongly Agree
I like to finish one task completely before fo- cusing on anything else.	1) Not at all; 2) Slightly; 3) Neutral; 4) Some- what; 5) Very much.
It makes me uncomfortable when I am not able to finish one task completely before focusing on another task.	1) Not at all; 2) Slightly; 3) Neutral; 4) Some- what; 5) Very much.
I am much more engaged in what I am doing if I am able to switch between several different tasks.	1) Not at all; 2) Slightly; 3) Neutral; 4) Some- what; 5) Very much.
I do not like having to shift my attention be- tween multiple tasks.	1) Not at all; 2) Slightly; 3) Neutral; 4) Some- what; 5) Very much.
I would rather switch back and forth between several projects than concentrate my efforts on just one.	1) Not at all; 2) Slightly; 3) Neutral; 4) Some- what; 5) Very much.
I would prefer to work in an environment where I can finish one task before starting the next.	1) Not at all; 2) Slightly; 3) Neutral; 4) Some- what; 5) Very much.
I do not like when I have to stop in the middle of a task to work on something else.	1) Not at all; 2) Slightly; 3) Neutral; 4) Some- what; 5) Very much.
When I have a task to complete, I like to break it up by switching to other tasks intermittently.	1) Not at all; 2) Slightly; 3) Neutral; 4) Some- what; 5) Very much.
I have a "one-track" mind.	1) Not at all; 2) Slightly; 3) Neutral; 4)

	Some- what; 5) Very much.
I prefer not to be interrupted when working on a task.	1) Not at all; 2) Slightly; 3) Neutral; 4) Some- what; 5) Very much.
BACKGROUND AND DEMOGRAPHICS	

Please select your current country of residence from the list below.	List of Countries
Which country have you lived in for the ma- jority of your life? Please specify below.	List of Countries
Have you lived in countries other than your current residence? If so, please list them:	Open-ended
Please state what is your educational back- ground?	No schooling completed; Elementary school; Some high school, no diploma; High school graduate, diploma or the equivalent (for ex- ample: GED); Some college credit, no degree; Trade/technical/vocational training; Associate degree; Bachelor's degree; Master's degree; Professional degree; Doctorate degree
Please state what is your gender:	Male; Female; Non-binary; Prefer not to say
Please state 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

A.2.2 *Post-survey.* In the following we present the post-survey we gave to participants.

Question	Options
How much has your work schedule on Toloka changed this week compared to previous weeks?	Not at all 1 2 3 4 5 Very Much
How much did you get to try out new tasks on Toloka this week?	Not at all 1 2 3 4 5 Very Much
How often did you get to use your strengths while working on Toloka this week?	1. No Extent - I did not use my strengths at all when working on Toloka this week; 2. Slight Extent - I used my strengths to a slight ex- tent when working on Toloka this week; 3. Moderate Extent - I used my strengths to a moderate extent when working on Toloka this week; 4. Great Extent - I used my strengths to a great extent when working on Toloka this week; 5. Very Great Extent - I used my strengths to a very great extent when working on Toloka this week.
How different did your experience feel while working on Toloka this week?	Not different at all 1 2 3 4 5 Very different
What did you like most about the plugin, and why?	Open-ended
What aspect of the plugin did you like the least, and why?	Open-ended
If you could magically change one thing about the plugin by adding or removing something to it, what would it be and why?	Open-ended
To see if you're still paying attention, please select the choice that says strongly disagree.	Strongly disagree 1 2 3 4 5 Strongly agree
Do you have any final thoughts or comments? Feel free to share also any questions you might have for us, or anything else you'd like to discuss.	Open-ended

Received July 2023; revised April 2024; accepted July 2024