

Subcontracting Microwork

Meredith Ringel Morris¹, Jeffrey P. Bigham², Robin Brewer³, Jonathan Bragg⁴,
Anand Kulkarni⁵, Jessie Li², and Saiph Savage⁶

¹Microsoft Research, Redmond, WA, USA

²Carnegie Mellon University, Pittsburgh, PA, USA

³Northwestern University, Evanston, IL, USA

⁴University of Washington, Seattle, WA, USA

⁵University of California, Berkeley, CA, USA

⁶West Virginia University, Morgantown, WV, USA

merrie@microsoft.com, {bigham, jessieli}@cs.cmu.edu, rnbrewer@u.northwestern.edu,
jbragg@cs.washington.edu, anandk@berkeley.edu, savage@mail.wvu.edu

ABSTRACT

Mainstream crowdwork platforms treat microtasks as indivisible units; however, in this article, we propose that there is value in re-examining this assumption. We argue that crowdwork platforms can improve their value proposition for all stakeholders by supporting *subcontracting* within microtasks. After describing the value proposition of subcontracting, we then define three models for microtask subcontracting: *real-time assistance*, *task management*, and *task improvement*, and reflect on potential use cases and implementation considerations associated with each. Finally, we describe the outcome of two tasks on Mechanical Turk meant to simulate aspects of subcontracting. We reflect on the implications of these findings for the design of future crowd work platforms that effectively harness the potential of subcontracting workflows.

Author Keywords

Crowdsourcing; microwork; subcontracting; human computation; task selection; task design.

ACM Classification Keywords

H.5.3. Group and Organization Interfaces.

INTRODUCTION

Microwork, small tasks performed on crowd work platforms (e.g., Amazon’s Mechanical Turk, Upwork, LeadGenius, SamaSource, and many others [37]), is an increasingly important form of digital labor, representing both a valued, flexible job opportunity for workers [39] and a viable means of accomplishing work at scale for businesses [17, 31],

scientific enterprises [10, 11], and community or volunteer ventures [5]. Mainstream crowdwork platforms treat microtasks (also called human intelligence tasks or HITs) as indivisible units; however, we propose that there is value in re-examining this assumption. We argue that crowdwork platforms can improve their value proposition for all stakeholders by supporting *subcontracting* within microtasks, i.e. outsourcing one or more aspects of a microtask to additional workers.

After describing the value proposition of subcontracting, we then define three models for subcontracting microtasks: *real-time assistance*, *task management*, and *task improvement*, and reflect on potential use cases and implementation considerations associated with each. Finally, we describe the outcome of two tasks on Mechanical Turk meant to simulate aspects of subcontracting; these simple HITs demonstrate the potential of subcontracting for improving aspects of crowdwork such as making workers and/or requesters aware of accessibility bugs, and for better matching task components to workers’ skills and interests. We reflect on the implications for the design of future crowd work platforms that effectively harness the potential of subcontracting workflows.

Value Proposition

We propose that enabling subcontracting of microwork, if properly implemented, can enhance crowdwork processes and outcomes. Subcontracting can provide value from the perspective of requesters, workers, and platform owners.

Value for Workers

From the perspective of a subcontracting-enabled crowd platform, there are two classes of workers, whom we will refer to as the *primary worker* (the original worker accepting a task, who may initiate subcontracting), and *secondary workers* (the workers who perform the subcontracted microwork).

Microtaskers performing primary worker roles in a subcontracting-enabled system would have added responsibilities that bring opportunities to begin to learn new

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
CHI 2017, May 06 - 11, 2017, Denver, CO, USA
Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM 978-1-4503-4655-9/17/05...\$15.00

DOI: <http://dx.doi.org/10.1145/3025453.3025687>

skills, such as management and task design, providing scaffolding for career growth, a dynamic currently lacking from most mainstream crowd platforms [20]. The added responsibility associated with the role of primary workers should result in increased wages for these roles, in an economically fair system [1]. Subcontracting also fundamentally enables new work types within platforms, such as task labeling, headhunting, management, etc.; these roles might encourage more diverse audiences to engage with crowd work by offering tasks that better match their skills and/or interests.

Secondary (subcontracted) work can offer low-reputation workers (e.g., platform newcomers, workers lacking skills with tasks of certain categories, etc.) a path to building skills and reputation. Secondary work can also lower the barrier for novice or casual workers by allowing them to invest less time and effort in task selection, by instead following a trusted primary worker or by benefiting from task decomposition and metadata added by other secondary workers.

Value for Requesters

One potential benefit of a subcontracting-based system for task requesters is that it may allow them to reduce the time and effort necessary for optimizing the deconstruction of tasks into microtasks, or of adding metadata or instructions to tasks, as such tasks may be delegated to workers themselves through the subcontracting structure (see the “Models of Subcontracting” section for more in-depth explanations of these processes).

Additionally, for requesters the addition of a layer of indirection through subcontracting mitigates the risk and workload involved in working with a full crowd of potentially unreliable or untrusted workers, as some of this risk is now delegated to primary workers. This benefit is analogous to the benefit of using a contracting or temp agency in the traditional economy – the temp agency manager is analogous to the primary worker, while individual temps map to secondary workers.

Value for Platforms

Subcontracting can benefit platform operators by enabling a more efficient and productive work system. Via subcontracting’s potential to enhance task design, better match workers with tasks, and reduce rates of partially-completed tasks, platforms can expect to see better throughput and higher-quality work outcomes. By empowering workers and offering them more flexibility and initiative within their work, platforms may also see enhanced employee productivity and reduced turnover.

MODELS OF SUBCONTRACTING

In this section, we unpack the concept of subcontracting as it might apply to microtasking platforms, identifying three primary models of subcontracting work in this context: *real-time assistance*, *task management*, and *task improvement*.

Note that it is not necessarily the case that subcontracting crowd work results in workers interacting with sub-components of an original microtask/HIT (e.g., working on a component that has been doled out by a primary worker in the task management model). For instance, in some cases secondary workers might simply support a complete, original HIT (e.g., in the real-time assistance model). Subcontracting could even enlarge the original HIT to be more than it first embodied (e.g., the addition of metadata in the task improvement model).

While not explicitly focused on the need for subcontracting, several prior systems and studies illustrate scenarios in which one or more of our three subcontracting models may be useful and/or have implemented a specific workflow that exemplifies these different types of subcontracting. We discuss this prior work in the context of each of the three subcontracting models.

Real-Time Assistance

Real-time assistance encompasses a model of subcontracting in which the primary worker engages one or more secondary workers to provide real-time advice, assistance, or support during a task. Such interactions might occur with the integration of widgets for real-time assistance, and might include easy integration of video-conferencing, screen sharing, VOIP calls, live-streaming, instant-messaging, or other types of real-time communications for task sharing.

For instance, Kobayashi et al. [21] and Brewer et al. [6] postulate IM or video chat interactions with experienced crowd workers during HIT execution may be necessary for supporting older adults (or other novice or less tech-savvy workers) in understanding complex task interfaces or workflows. Such chat support would constitute a type of real-time assistance subcontracting, since the support-giver is not the original worker who accepted the HIT, but is facilitating its completion. Similarly, real-time assistance could help facilitate task completion (or better task quality) by connecting workers with people with whom they can discuss and clarify interpretations of task instructions, or with people who may have domain expertise or other skills relevant to completion of a task.

Zyskowski et al. [39] noted that some crowd workers with disabilities have to leave tasks incomplete due to difficulties encountered mid-way through (e.g., a blind participant discovering that one sub-component of a task requires examining an uncaptioned image). Real-time assistance could allow a worker with disabilities to connect with someone who has the capabilities to assist with a specific component of a task, as an alternative to the status quo practice of task abandonment in such situations.

Prior work on seeking assistance online has shown how crowd workers form informal or formal communities to offer and receive help from others [15, 18, 27]. A community assistance model, such as the real-time assistance

subcontracting we propose, could be a compelling avenue to explore with subcontracting, but asking for and receiving help remains a challenging task. Prior work shows how people do not know how to ask for help and may prefer automated recommendations, particularly if they develop trust with the system [2, 9, 33]. Further, trust plays a role in less automated forms of help, as prior work suggests microworkers may ask for and receive assistance but may be more trusting of responses they get from people they know [29]. This suggests that an important component of implementing real time assistance within a platform is to provide context about the helpers (e.g. their helper rating, their profession, skills, or other aspects of demographics or reputation, etc.); the capability to request assistance from a specific, known worker may also be valued.

Task Management

Task management subcontracting applies to situations in which a primary worker takes on a meta-work role for a complex task, delegating components to secondary workers and taking responsibility for integrating and/or approving the products of the secondary workers' labor. Headhunting work (i.e., identifying workers with specific talents or skills appropriate for aspects of a complex task) is another component of task management subcontracting.

Task management work may offer desired career growth opportunities for experienced workers [20]. Headhunting may be particularly valued by both requesters and workers, as finding tasks that are a good match for a worker's skills is an unresolved problem in mainstream crowd work platforms, currently addressed somewhat inefficiently through informal backchannels [15, 18].

Leadership via task management by the primary worker may be helpful for efficient task completion, but online leadership is challenging, particularly for leaders of multiple projects who can feel overburdened, resulting in failed or unfinished projects [25]. Using the theory of distributed leadership, tools have been designed for leadership redistribution [25], a concept that could be used within a task management subcontracting model. Pipeline is one such tool, that focuses on decentralization for sharing leadership responsibilities and automation through a trusted member system to improve task completion speed and efficiency [25]. Pipeline allowed workers to self-select microtasks to complete, yet for larger and varied projects, it may be beneficial for leaders to assign tasks similar to [35].

The Atelier system has explored a mentorship/internship crowdwork model where novice workers are assigned more experienced mentors to help them complete tasks [35]. These mentors not only provide feedback to the interns, but also help to structure the tasks into manageable milestones and answer questions, thus illustrating examples of both real-time assistance and task management subcontracting styles; the Atelier model also illustrates how different subcontracting styles may be used in an integrative fashion.

Task Improvement

Task improvement subcontracting entails allowing a primary worker to edit task structure, including clarifying instructions, fixing user interface components, changing the task workflow, and adding, removing, or merging sub-tasks. Clarifying instructions and fixing user interface components in particular would help to address the first two of seven main risk factors for workers identified by McInnis et al. [28]. Task improvement may also involve adding metadata to tasks (or creating meta-tasks that instruct secondary workers to add metadata), such as metadata that might help workers better identify the skills or abilities required to complete a task, its difficulty level, compliance with accessibility guidelines [39], etc. These improved tasks and sub-tasks can then be assigned to or discovered by secondary workers. Task improvement can make a task more accessible to a larger pool of workers, less risky, faster to complete, and/or more likely to generate quality output.

Current metadata about tasks is limited to subjective ratings [18] or implicit feedback [7, 16] related to pay and interactions with the requester. Information on other task aspects, such as usability or accessibility, may be available in an unstructured form through forum discussions or reviews [15, 18]. However, workers who are unable to find this information may choose to assume unnecessary risk by attempting the task or avoiding the task altogether, resulting in market inefficiency.

Some task improvements can be handed off entirely to workers, but others may require interaction with the requester. Kulkarni et al. [22] had workers decompose tasks into subtasks, enabling a kind of distributed workflow specification. One could also imagine assigning this managerial role to a smaller set of skilled workers, who could perhaps make use of tools originally designed for requesters [19, 32]. On the other hand, subcontracting tasks like clarifying instructions may require occasional feedback from the requester to ensure that the instructions align with the requester's desired work product and evaluation criteria.

Task improvement roles address the fact that many requesters, due to inexperience, time pressure, disinterest, or other reasons, do a poor job in decomposing and specifying their tasks [30]. Allowing workers to iteratively improve task instructions, decompositions, and workflows (rather than merely giving such requesters and tasks low ratings on external forums like Turkopticon [18]) may be a valued service to requesters. Experimental platforms like Stanford's Daemo [34] enable requesters to create "prototype tasks" in order to iterate on and improve task design – task improvement subcontracting takes task prototyping one step further, by placing the power to iterate and make changes in the hands of primary workers, rather than with requesters alone.

IMPLEMENTATION CONSIDERATIONS

Subcontracting of microwork fundamentally alters many of the assumptions currently underlying crowd work platforms, such as economic incentive models and the efficacy of some prevailing workflows. However, subcontracting also legitimizes and codifies some existing informal practices that currently take place off-platform [15]. While addressing the challenges of introducing formal subcontracting support may be complex, the benefits offered in terms of moving some unofficial activity into a legitimate and structured realm, as well as the potential benefits to workers, requesters, and platform operators mentioned in the Introduction, makes grappling with these challenges a worthwhile endeavor. Here, we identify five key issues crucial to creating a successful subcontracting structure, and reflect on design alternatives for each: *incentive models*, *reputation models*, *transparency*, *quality control*, and *ethical considerations*.

Incentive Models

Developing a fair and understandable compensation model for primary and secondary workers is integral to the success of any subcontracting scheme for crowdwork. Three possible design alternatives are job-based, flat-rate and altruistic.

Job-based compensation is the most flexible model, in which the economic arrangement for how the initial wage offered per HIT will be flexibly divided amongst primary and secondary workers and is based on the properties of the work. One possibility is that the primary worker will have the ability to negotiate the HIT wage with the requester, which may be relevant in cases such as task improvement subcontracting, wherein significant work beyond what the requester initially envisioned is required to improve the quality of a task. Given a set wage for the original HIT, another possibility is that platforms provide infrastructure for secondary workers to negotiate subcontracting wages with primary workers.

Flat-rate compensation avoids negotiation; instead, either the platform or the requester can specify fee splits for different types of subcontracting arrangements (e.g., real-time assistance = 50% of the initial HIT wage multiplied by the percent of the HIT duration for which the assistance lasts).

Altruistic compensation schemes may apply to particular subcontracting scenarios, in which workers may volunteer their services in a subcontracting capacity through intrinsic motivation [29]. Workers may complete microtasks for causes they care about (e.g., a worker interested in disability rights may voluntarily fix HITs to make them more accessible to screen reader users). Some secondary workers may want to enhance the experience of fellow workers whom they care for (exchanging social capital rather than cash). Research on unpaid volunteers contributing to open source projects show how people volunteered based on the concept of *generalized exchange* in which they helped other people because others helped them before, and they would want

someone to do the same if they encountered a problem in the future [23]. Or, some secondary workers may be motivated to complete subcontracting tasks for non-monetary rewards, such as to build reputation within a system or learn new skills.

Reputation Models

Reputation models are an important component of existing crowd platforms. The introduction of formal subcontracting systems may necessitate changing current reputation models, as current models typically award reputation points to a worker based on the outcome of a given HIT; however, if multiple workers are able to contribute to a single HIT via subcontracting, the allocation of reputation becomes less straightforward.

One possibility is that primary workers may take on the reputation risk of work done by secondary workers to whom they subcontract work, which may help create strong incentives for careful choice of when and with whom to engage in subcontracting relationships and incentivize quality-checking of subcontracted work outcomes by primary workers.

If secondary workers directly earn reputation for their work, one consideration is whether they should be rated by the primary worker or by the original task requester (or by a third party, such as peer-comprised Crowd Guilds [38]). The transparency of the subcontracting relationship (see next section) may influence the choice in this regard.

Alternately, there could be new and different categories of reputation within crowd work platforms, such as separate reputation point systems for different task types, including meta-tasks associated with new subcontracting forms such as assisting other workers, metadata creation, task restructuring, etc.

The worker community strongly advocates for formal reputation models for requesters, and as many platforms do not yet support requester rating, this is often done unofficially on third-party sites like Turker Nation and Turkopticon [18]. In platforms supporting requester rating, workers taking on the primary worker role may need to earn reputation both as requesters (as rated by the secondary workers they recruit) as well as having worker reputation (as rated by the original HIT requester).

Transparency

Transparency concerns the extent to which parties need to be aware of subcontracting practices. Two key design choices are the degree of transparency of subcontracting to requesters and the degree of transparency to workers. To some extent, transparency choices are interrelated with choices regarding reputation and incentive models.

If work quality is unchanged from non-subcontracting systems (or improved, as we postulate it may be in many cases), then requesters may not need or desire transparency

regarding any subcontracting arrangements. However, in some cases it may be important for requesters to know whether subcontracting has occurred, or even to be able to prevent subcontracting on certain tasks (perhaps by setting a special flag within the system). For example, some academic research HITs, such as psychology studies, may have the fundamental assumption that a single worker is completing a questionnaire (although even without formal subcontracting infrastructure in place, such assumptions may be invalid due to informal, offline worker practices of collaboration and information sharing [15]).

It may also be beneficial in some cases for requesters to understand if and why subcontracting has occurred, such as task improvement work – this may help the requester to create better tasks themselves in the future, or to be alerted to any unintended changes in the nature of the HIT that may result from well-intentioned task improvement subcontracting. In addition to passing back the final HIT result to the original requester, it may also be valuable to allow primary workers to pass back additional meta-information, including information about why subcontracting may have been needed, what improvements (if any) were made to the task design, perhaps even re-usable templates for replicating that design in the future. Requesters may also wish to give bonus payments to primary or secondary workers who enhance the outcomes of their tasks through high-quality assistance, task management, or task improvement labor. On the other hand, if task improvement work actually results in low quality secondary HITs – as has been observed in some task decompositions performed by workers [22] – transparency may protect requesters from negative reviews that could damage their reputation and affect their ability to recruit future workers.

It may also be desirable to include a system flag that specifically lets requesters indicate a wish to have certain types of subcontracting work (such as task improvement) performed for their HIT, such as in instances where requesters are piloting a new task or are unsure of how to best design or break down work or of how to advertise to the right worker set.

Determining whether secondary workers need to know whether they are taking on a HIT from an original requester versus a subcontracted HIT from a primary worker is another key design decision. In general, we recommend transparency to secondary workers, to help avoid worker exploitation. It may also be important to task outcome/quality for secondary workers to understand their role within a larger task structure.

Quality Control

Requesters submit work for crowd workers to complete because they assume they will receive quality responses in return. The quality of these responses is already a concern when task workflows are more direct and one person is working on one task [20]; paradigms such as replication +

majority vote or embedding of gold-standard tasks are fairly standard for quality control in status quo microtasking environments [8, 24]. However, subcontracting can lead to multiple people collaborating on the same task, adding more uncertainty to the system and thereby making quality control become a more pressing, or at least more complex, issue. Prior work has examined this issue by comparing different types of quality control techniques [20] and how workers give feedback on task submissions [12].

Correctness is only one part of quality control. Requester feedback has also been shown to improve result quality [12, 26]. Further, prior work shows that when comparing self- and expert-generated feedback on submitted tasks, there was no difference in worker performance. However, workers who received expert feedback did revise their work more [12] and thought the feedback improved their work [26]. Regardless of the type of feedback and correctness checks, any system using a subcontracting model would need some form of quality control because multiple workers contributing to the same task may have different perceptions of completion quality.

Ethical Considerations

While subcontracting has the possibility of making crowd work more interesting and challenging for primary workers, it may carry the risk, depending upon implementation, of creating even more menial, micro-microtasks for secondary workers. While such tasks may be desired by some workers, extreme levels of task decomposition may dehumanize and deskill work to inappropriate levels. Our proposed models of subcontracting (real-time assistance, task management, and task improvement) aim to support improved workflows overall rather than promote extreme task decomposition for decomposition's sake; however, the possibility for subcontracting to move crowdwork further along the piecework continuum is an ethical consideration to be aware of, and to attempt to prevent through careful platform design.

The economic model of subcontracting also warrants ethical consideration; fair pay continues to be a problem for many crowd workers who struggle to earn an hourly wage that meets the legal standards set in the U.S. for traditional forms of labor [3, 13]. It may be important for platforms to include checks against efforts to game the system, such as a primary worker merely reposting a task, unaltered, to secondary workers at a very low wage, thereby taking a large cut of pay without having contributed value.

There is the risk that secondary workers could be an even further de-valued type of laborer than today's crowd workers, particularly if secondary work is not well-designed (e.g., extreme deskilling) or is not fairly paid. To avoid such ethical pitfalls, it may be important that platform designers implement features to safeguard against such possibilities rather leaving the mechanics of subcontracting completely in the hands of task requesters or primary workers, whose incentive models may not produce behavior that aligns with

a platform operator’s ethical standards (better legal protections for crowd workers would, of course, also be of benefit). For instance, platform features to enable reporting and adjudication of unfair pay practices may help address some of these concerns, as would legal rules to establish and enforce fair pay guidelines for microwork. Of course, striking an appropriate balance between the autonomy afforded requesters and workers versus the controls put in place by the platform may itself present an additional ethical issue.

Zyskowski et al [39] suggested that real-time assistance subcontracting may be vital to supporting the full participation of people with disabilities in crowd work. The United Nations’ Convention on the Rights of Persons with Disabilities [36, Article 27] affirms that people with disabilities have the right to work and that labor markets and work environments should be “open, inclusive, and accessible to persons with disabilities.” The potential for subcontracting workflows to expand participation in this emerging form of digital labor to diverse, under-represented, and under-served populations is also an important ethical consideration.

SUBCONTRACTING IN PRACTICE

To explore the idea of subcontracting, we released two tasks on Amazon Mechanical Turk. The first was designed to explore the feasibility of one type of task improvement subcontracting (HIT metadata creation), and the second aimed to provide insight into task management by exploring worker motivations and decision-making when provided the chance to outsource components of a task to other workers.

Task Improvement HIT

We designed and released a HIT on Mechanical Turk to explore one type of task improvement subcontracting, labeling other HITs with metadata about task properties. Our HIT took the form of a survey asking workers to accept a *different* HIT (not ours) to work on simultaneously in another browser tab, and then to answer several questions about that HIT in our survey. These questions included providing descriptive tags about the topic of the task, indicating what abilities or technical skills the task required, and identifying any accessibility issues that might make it difficult for people with disabilities to complete the task (a category inspired by Zyskowski et al. [39] and Brewer et al. [6]). Note that this HIT is designed to illustrate (within the constraints of a status quo platform) the potential benefit of task improvement workflows and the feasibility of having workers label tasks for this purpose (i.e., can workers identify inaccessible tasks); however, this HIT does not necessarily represent an ideal task improvement workflow, which might include additional subcontracted tasks to identify relevant metadata categories, evaluate the efficacy of the added metadata, etc. Figure 1a gives an overview of the HIT’s workflow.

We recruited 100 workers to complete our HIT, which took less than 5 minutes on average paid \$0.95 USD (about \$11/hour). In practice, the time and cost could likely be optimized by asking fewer, more targeted questions; optimizing templates and workflows for task improvement work is an important area for further study.

All workers were able to finish our HIT and provide feedback about another microtask. Worker labels indicated that a substantial portion of tasks may have accessibility concerns for certain user groups, a finding consistent with prior, qualitative work [6, 39]. For instance, 76% of tasks labeled by workers during our HIT were reported as containing visual material that would not be accessible to a blind worker, and 13% necessitated listening to audio materials that would not be accessible to a worker who was deaf or hard of hearing.

If a platform were designed to support collecting and using task labels such as those gathered in our example HIT, this could offer several benefits; for instance, workers could use these labels to more easily search and filter HITS to find work that is a good match for their skills, interests, and abilities, and task requesters could use these labels to gain insight into improvement areas for their tasks (in our example, the labels identify accessibility issues that requesters should be interested in rectifying in order to attract a larger worker base, improve task outcome quality, and/or comply with potential future legal regulations regarding the accessibility of digital labor).

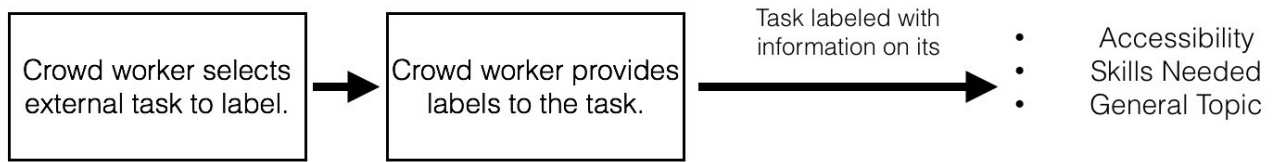
Of course, platform support for task improvement microtasking could also allow better, more integrated interfaces for task labeling than the method employed for our exploratory task. Note also that we did not provide any training for workers in our labeling task (nor did we verify the accuracy of their labels); for straightforward labeling tasks such as HIT topic, such training is likely unnecessary, but more nuanced labels such as our example of identifying accessibility bugs might necessitate selecting workers with existing expertise (or willing to take training), and/or implementing other quality control measures.

Task Management HIT

Our second HIT was designed to give workers on Mechanical Turk a chance to engage in task management, since due to the platform’s capabilities and the existing norms around task types and workflows on mTurk, most existing HITs are so decomposed as to not be suitable for task management subcontracting.

In our HIT, we presented a task with three components: (1) watching a one-minute video (excerpted from a TED talk) and captioning it by transcribing its audio track, (2) writing a 100-word response to the video, and (3) reading the 100-word response aloud and uploading that recording. We felt this task was especially well-suited for exploring subcontracting because each subtask would require several

(a) Task Improvement HIT Workflow:



(b) Task Management HIT Workflow:

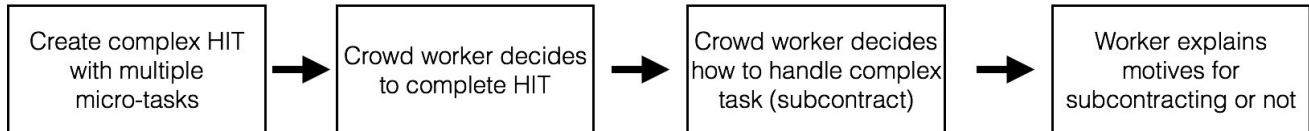


Figure 1. (a) Overview of our Task Improvement HIT workflow, aimed at assessing the feasibility of workers adding valuable content to pre-existing HITs. (b) Overview of our Task Management HIT workflow, aimed at exploring whether workers would choose to engage in subcontracting decomposed tasks, and assessing their motivation behind this choice. These sample HITs are meant to give initial insights into the feasibility and appeal of subcontracting workflows rather than to exemplify ideal subcontracting HIT design.

minutes to complete. Furthermore, each subtask required substantially different skills, which may appeal to different workers, and also may require special equipment (speakers to play back audio, a microphone to record it, etc.).

Workers were offered \$0.50 USD for the base task and were randomly assigned to one of three bonus conditions (\$0.50, \$1.00, or \$2.00 USD) for each of the three components they chose to complete; primary workers could therefore earn up to a total of \$2.00, \$3.50, or \$6.50 depending on their pay condition assignment. Because the Mechanical Turk platform does not offer subcontracting support, workers could not actually assign any of the three task components to a secondary worker; instead, we simulated subcontracting capabilities by telling workers that they could either earn money by doing each task themselves (the bonuses), or indicate to us which (if any) of the sub-tasks they'd prefer we reassign to other workers. Workers were required to complete at least one of the three sub-tasks themselves. We also asked workers to fill in a brief free-response box indicating why they did or did not choose to carry out each subtask. 200 workers completed this HIT. Figure 1b gives an overview of this task's workflow.

We hypothesized that offering different monetary incentives would affect workers' decisions about whether or not to subcontract. However, a chi-square test of independence did not find a statistically significant effect on bonus price for subcontracting any of the three subtasks at the 0.05 significance level ($\chi^2(2, N = 200) = 1.77, p = .41$) for the captioning subtask; $\chi^2(2, N = 200) = 1.19, p = .55$) for the writing subtask; and $\chi^2(2, N = 200) = 0.15, p = .93$) for the audio recording subtask). Thus, the remainder of our analysis considers the aggregate data from all three bonus-price

conditions. The following qualitative analysis suggests that other factors dominated the variation in bonus amount that we offered.

72% of workers chose that they would want to subcontract at least one of the three sub-tasks. 11% of workers marked the captioning task for subcontracting, 23% wanted to subcontract the written response, and 65% indicated they would subcontract creating the audio recording. Below, we explore workers' motivations for these choices, by using Latent Dirichlet Allocation (LDA) topic models [4] to extract keywords from the open-ended responses and cluster them into themes, then manually reviewing and interpreting these clusters; through this analysis, we identified three primary themes (skills, money, and interests) that influenced workers' choices. The relative prevalence of these motivating factors is summarized in Figure 2.

Choosing to Subcontract

Skills: 22% of the choices to subcontract sub-tasks were because workers did not have the necessary skills or technology to execute that portion of the task. For instance, some workers did not have a microphone to record audio; others did, but were unfamiliar with how to use it ("I don't know how to record and upload voice files"). One expressed that a speech disability made him choose not to record audio, noting, "I do not like my voice, and I have a slight speech impediment. I thought the recording would not turn out well, and it would ruin the impact of my written critique..." Another worker expressed not having the right skill-set to effectively complete the critique-writing portion of the task: "I didn't think I could come up with a 100-word critique of the video. She had presented scientific evidence [in the video], and there wasn't much I could say to either back it

up or refute it, as I have limited knowledge of brain science...”

Money: When choosing to subcontract part of a task, 38% of the time workers mentioned that they felt the time a subtask would take was not worth the money they would earn for that portion of the task. For example, one worker noted that a task he subcontracted “*would have taken more time than I thought it was worth.*” Another worker noted that he had a microphone, but that the time to configure it for the audio recording task would have changed the time/earnings tradeoff for him.

Interests: For 40% of subcontracting choices, workers indicated that personal interests and preferences impacted their choice to subcontract portions of our HIT. For example, one noted, “*I didn’t feel like thinking and writing a critique*”; another said, “*I don’t like writing non-fiction.*”

Choosing Not to Subcontract

Skills: In 9% of cases where workers chose to complete a task themselves, they referred to their skills as being a reason not to subcontract out a particular task, noting that they felt they were particularly strong at certain types of tasks and therefore it made sense to perform them themselves. For example, “*[I decided not to subcontract the recording part because] I have done HITs with my microphone before and already had the software installed in order to do it.*”

Money: In 64% of cases where workers did not choose to outsource a subtask, they noted that money was the main influence in deciding not to subcontract; in these cases, the workers chose to do the work themselves so that they would not have to share earnings with others. “*I wanted \$2 [bonus],*” one noted. Another said, “*I’m currently trying to earn enough money to buy an expensive DSLR camera that I want really badly... and this task paid extra money [that I need].*”

Interests: Finally, in 27% of cases of not subcontracting, workers indicated that personal interests influenced their choice of completing sub-tasks on their own, specifically that they enjoyed doing certain types of work. For instance, one worker said, “*I like recording stuff,*” and another said, “*I love English, so I wanted to express [myself].*” Some workers in this category specifically noted how doing certain task types would benefit them by helping them acquire needed skills; for instance, one chose to do the video captioning himself because “*I love doing transcription tasks. It helps me in learning the art to type or write faster while listening.*” Another felt that watching and responding to the video would benefit them educationally because “*I also generally enjoy TED talks presentations... I believed I would further enjoy performing the task while I, hopefully, learned something new.*”

Reflection

These findings indicate that money was not the only factor in whether primary workers chose to complete subtasks



Figure 2. Overview of the motives crowd workers reported for their decisions to subcontract or not subcontract portions of the larger HIT. Sacrificing income was one of the main motives for not subcontracting.

versus mark them as being for subcontracting; rather, our findings show that workers chose to assign tasks to others based on their *interests* or disinterest in different components of a task (and desire to grow their interests and knowledge) and on their *skills* or lack thereof (in this case, having or lacking audio equipment, being strong in English or science, or even having disabilities). This supports the idea that subcontracting (if implemented in a manner mindful of the considerations we identified relating to incentives, reputation, transparency, quality, and ethics) can be an important route to realizing Kittur et al.’s [20] vision of a more engaging, meaningful, and challenging future of crowd work, as well as supporting broader participation in crowd work by people with varying technical, physical, or cognitive capabilities [6, 39].

DISCUSSION

In this paper, we synthesized a set of work styles and user needs articulated in prior research in order to formalize the concept of subcontracting as it applies to microwork, identifying three distinct classes of subcontracting (real-time assistance, task improvement, and task management), and reflecting on the potential advantages and challenges associated with this style of work. We also presented two sample HITs that offered some initial insights into the feasibility and motivation issues likely to impact subcontracting’s success. In the “Implementation Considerations” section, we discussed several topics (incentive models, reputation models, transparency, quality control, and ethical considerations) that must be addressed in the instantiation of subcontracting. Implementing and exploring the parameter space of these concepts is a key area for future research. Building platforms that support subcontracting workflows in an intentional manner will enable the crowdwork research community to evaluate the efficacy of these choices and further refine this concept. We particularly stress the importance of the ethical considerations component, as our intent in introducing and formalizing concepts related to subcontracting microwork is to facilitate more inclusive, satisfying, efficient, and high-

quality work, rather than to facilitate extreme task decomposition strategies that may result in deskilling or untenable wages.

Our two example tasks illustrated that some limited styles of subcontracting may be feasible to retrofit onto status quo platforms (though we advocate for intentional feature support within platforms as a preferred implementation), and that many workers have interest in subcontracting work. However, these example HITs are merely first steps in evaluating this style of microwork. Larger, longer-term, and more varied evaluations are necessary to understand the nuance and efficacy of each of the three types of subcontracting. Measuring the impact of subcontracting workflows on variables such as worker satisfaction, requester satisfaction, work efficiency and quality, impact on worker reputation and pay, factors influencing worker participation in various subcontracting roles, etc. are important topics for future investigation. Employing multiple methods (e.g., surveys and interviews to gain insight on stakeholder perspectives; design, building, and testing of new platform features; and/or execution of HITs exemplifying different subcontracting algorithms) will be key to understanding this emerging topic.

It is also important to note that our proposed model of subcontracting may be incomplete. Indeed, the process of building and deploying platforms that explicitly support or even encourage subcontracting will likely shed light on unconsidered nuances of this work style that we were unable to anticipate based on our experiences with status quo platforms and tasks. For example, what should be done if a primary worker drops out of a task before it is complete – should they be replaced with an entirely new worker, or perhaps should one of the secondary workers be elevated to fill their place? What should the cascading impact be on any agreements (i.e., regarding pay allocation) that the now-absent primary made with any secondary workers? The relative pros and cons of such choices remain hypothetical in the absence of concrete implementations and scenarios.

CONCLUSION

In this paper, we proposed that *subcontracting*, i.e. outsourcing one or more aspects of microtasks from a primary worker to one or more secondary workers, is a work model that has potential to benefit crowd workers, task requesters, and platform operators. We defined three subcontracting styles: real-time assistance, task improvement, and task management, and drew on crowd work literature to identify sample scenarios suited to each approach. However, current microtasks and microwork platforms are not necessarily designed in ways that facilitate subcontracting; we identified important considerations, such as incentive models, reputation models, transparency, quality control, and ethics that should be taken into account when implementing subcontracting support. We also presented the results of two microtasks deployed on Mechanical Turk that

explored aspects of subcontracting. These task outcomes lend support to our arguments regarding the feasibility and desirability of subcontracting microwork. This work contributes a practical foundation on which to design new platforms and workflows that can facilitate new forms of engagement for all members of the crowdwork ecosystem. The ideas and findings in this work also set the stage for a rich body of future research on subcontracting microwork, such as building subcontracting support into platforms, testing the outcomes of different incentive structures, reputation models, transparency policies, or quality control workflows, and measuring the impact of subcontracting-related practices on metrics such as output quality, platform efficiency, and worker and requester satisfaction.

REFERENCES

1. Bederson, B.B., Quinn A.J. Web workers unite! Addressing challenges of online laborers. *Proceedings of CHI 2011*.
2. Belkin, N.J. Helping People Find What They Don't Know. *Communications of the ACM*, 43(8), August 2000, p. 58 – 61.
3. Berg, J. Income Security in the On-Demand Economy: Findings and Policy Lessons from a Survey of Crowdworkers. *Comparative Labor Law & Policy Journal*, Vol. 37, No. 3, March 2016.
4. Blei, D., Ng, A., and Jordan, M. 2003. Latent Dirichlet Allocation. *The Journal of Machine Learning Research*, 3, 993-1022.
5. Brady, E., Morris, M.R., and Bigham, J.P. Gauging Receptiveness to Social Microvolunteering. *Proceedings of CHI 2015*.
6. Brewer, R., Morris, M.R., and Piper, A.M. "Why would anybody do this?": Older Adults' Understanding of and Experiences with Crowdwork. *Proceedings of CHI 2016*.
7. Callison-Burch, C. Crowd-Workers: Aggregating Information Across Turkers To Help Them Find Higher Paying Work. *HCOMP 2014 Extended Abstracts*.
8. Callison-Burch, C. Fast, cheap, and creative: evaluating translation quality using Amazon's Mechanical Turk. *Proceedings of EMNLP 2009*.
9. Chilana, P.K., Ko, A.J., and Wobbrock, J.O. LemonAid: Selection-Based Crowdsourced Contextual Help for Web Applications. *Proceedings of CHI 2012*.
10. Clery, D. Galaxy Zoo Volunteers Share Pain and Glory of Research. *Science*, July 2011.
11. Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., Leaver-Fay, A., Baker, D., Popovic, Z., and Foldit Players. Predicting Structures with a Multiplayer Online Game. *Nature* (2010).

12. Dow, S.P., Kulkarni, A., Klemmer, S.R., and Hartmann, B. Shepherding the Crowd Yields Better Work. *Proceedings of CSCW 2012*.
13. Dynamo Wiki. "Fair Payment". Retrieved December 12, 2016 at http://wiki.wearedynamo.org/index.php?title=Fair_payment
14. Flanagan, J. The critical incident technique. *Psychological Bulletin*, 51:327-358, 1954.
15. Gray, M.L., Suri, S., Ali, S.S., Kulkarni, D. The Crowd is a Collaborative Network. *Proceedings of CSCW 2016*.
16. Hanrahan, B.V., Willamowski, J.K., Swaminathan, S., and Martin, D.B. TurkBench: Rendering the Market for Turkers. *Proceedings of CHI 2015*.
17. Ipeirotis, P.G. Analyzing the Amazon Mechanical Turk Marketplace. *ACM Crossroads*, 17(2), Winter 2010, p. 16 – 21.
18. Irani, L. and Silberman, M.S. Turkothon: Interrupting Worker Invisibility in Amazon Mechanical Turk. *Proceedings of CHI 2013*.
19. Kittur, A., Khamkar, S., André, P., and Kraut, R. CrowdWeaver: Visually Managing Complex Crowd Work. *Proceedings of CSCW 2012*.
20. Kittur, A., Nickerson, J.V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., Lease, M., and Horton, J. The Future of Crowd Work. *Proceedings of CSCW 2013*.
21. Kobayashi, M., Ishihara, T., and Itoko, T. Age-Based Task Specialization for Crowdsourced Proofreading. *Proceedings of HCII 2013*.
22. Kulkarni, A., Can, M., and Hartmann, B. Collaboratively Crowdsourcing Workflows with Turkomatic. *Proceedings of CSCW 2012*.
23. Lakhani, K.R. and von Hippel, E. How Open Source Software Works: "Free" User-to-User Assistance. *Research Policy*, 32, 2003, p. 923-943.
24. Lasecki, W.S., Murray, K.I., White, S., Miller, R.C., and Bigham, J.P. Real-Time Crowd Control of Existing Interfaces. *Proceedings of UIST 2011*.
25. Luther, K., Fiesler, C., and Bruckman, A. Redistributing Leadership in Online Creative Collaboration. *Proceedings of CSCW 2013*.
26. Luther, K., Tolentino, J., Wu, W., Pavel, A., Agrawala, M., Bailey, B., Hartmann, B., and Dow, S. Structuring, Aggregating, and Evaluating Crowdsourced Design Critique. *Proceedings of CSCW 2015*.
27. Martin, D., Hanrahan, B.V., O'Neill, J., and Gupta, N. Being a Turker. *Proceedings of CSCW 2014*.
28. McInnis, B., Cosley, D., Nam, C., and Leshed, G. Taking a HIT: Designing around Rejection, Mistrust, Risk, and Workers' Experiences in Amazon Mechanical Turk. *Proceedings of CHI 2016*.
29. Morris, M.R., Teevan, J., and Panovich, K. What Do People Ask Their Social Networks, and Why? A Survey Study of Status Message Q&A Behavior. *Proceedings of CHI 2010*.
30. Papoutsaki, A., Guo, H., Metaxa-Kakavouli, D., Gramazio, C., Rasley, J., Xie, W., Wang, G., and Huang, J. Crowdsourcing from Scratch: A Pragmatic Experiment in Data Collection by Novice Requesters. *Proceedings of HCOMP 2015*.
31. Plante, C. Doritos and the Decade-Long Scam for Free Super Bowl Commercials. *The Verge*, Feb 3., 2016.
32. Rzeszutarski, J. and Kittur, A. CrowdScope: Interactively Visualizing User Behavior and Output. *Proceedings of UIST 2012*.
33. Singh, V., Twidale, M. B., Rathi D. Open Source Technical Support: A Look at Peer Help-Giving. *Proceedings of HICSS 2006*.
34. Stanford Crowd Research Collective. Daemo: A Self-Governed Crowdsourcing Marketplace. *UIST 2015 Extended Abstracts*.
35. Suzuki, R., Salehi, N., Lam, M.S., Marroquin, J.C., and Bernstein, M.S. Atelier: Repurposing Expert Crowdsourcing Tasks as Micro-internships. *Proceedings of CHI 2016*.
36. United Nations. Convention on the Rights of Persons with Disabilities. December 2006.
37. Vakharia, D. and Lease, M. Beyond Mechanical Turk: An Analysis of Paid Crowd Work Platforms. *Proceedings of iConference 2015*.
38. Whiting, M. et al. Crowd Guilds: Worker-led Reputation and Feedback on Crowdsourcing Platforms. *Proceedings of CSCW 2017*.
39. Zyskowski, K., Morris, M.R., Bigham, J.P., Gray, M.L., and Kane, S. Accessible Crowdwork? Understanding the Value in and Challenge of Microtask Employment for People with Disabilities. *Proceedings of CSCW 2015*.