
The Daemo Crowdsourcing Marketplace

Snehalkumar (Neil) S. Gaikwad, Mark E. Whiting, Dilrukshi Gamage, Catherine A. Mullings, Dinesh Majeti, Shirish Goyal, Aaron Gilbee, Nalin Chhibber, Adam Ginzberg, Angela Richmond-Fuller, Sekandar Matin, Vibhor Sehgal, Tejas Seshadri Sarma, Ahmed Nasser, Aipta Ballav, Jeff Regino, Sharon Zhou, Kamila Mananova, Preethi Srinivas, Karolina Ziulkoski, Dinesh Dhakal, Alexander Stolzoff, Senadhipathige S. Niranga, Mohamed Salih, Akshansh Sinha, Rajan Vaish, Michael S. Bernstein

Stanford Crowd Research Collective
Stanford University
daemo@cs.stanford.edu

Permission to make digital or hard copies of part or all 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 third-party components of this work must be honored. For all other uses, contact the Owner/Author. Copyright is held by the owner/author(s). CSCW '17 Companion, February 25 - March 01, 2017, Portland, OR, USA
ACM 978-1-4503-4688-7/17/02.
<http://dx.doi.org/10.1145/3022198.3023270>

Abstract

The success of crowdsourcing markets is dependent on a strong foundation of trust between workers and requesters. In current marketplaces, workers and requesters are often unable to trust each other's quality, and their mental models of tasks are misaligned due to ambiguous instructions or confusing edge cases. This breakdown of trust typically arises from (1) flawed reputation systems which do not accurately reflect worker and requester quality, and from (2) poorly designed tasks. In this demo, we present how Boomerang and Prototype Tasks, the fundamental building blocks of the Daemo crowdsourcing marketplace, help restore trust between workers and requesters. Daemo's Boomerang reputation system incentivizes alignment between opinion and ratings by determining the likelihood that workers and requesters will work together in the future based on how they rate each other. Daemo's Prototype tasks require that new tasks go through a feedback iteration phase with a small number of workers so that requesters can revise their instructions and task designs before launch.

Author Keywords

Crowdsourcing Markets, Trust

ACM Classification Keywords

H.5.3. [Group and Organization Interfaces]

Introduction

Crowdsourcing markets such as Amazon Mechanical Turk and Upwork function optimally when trust exists between workers and requesters. Requesters come expecting that workers will quickly produce high-quality results, and workers expect that their hard work will be rewarded fairly [6]. However, requesters are often surprised by low-quality results [5]—as a result, they rerun tasks, discard collected data, and deploy complex worker filters. Conversely, workers are surprised to receive negative feedback or no pay for their hard work. This produces a downward spiral whereby workers feel justified in producing lower quality work and requesters drive down wages, producing a “market for lemons” [4]. Unfortunately, without the ability to change the mechanics of the platform, most research today focuses on *downstream* solutions that minimize the damage done.

We will demonstrate the *Daemo* crowdsourcing marketplace, and in particular how Boomerang [2] and Prototype tasks [1], two of its fundamental building blocks, help create feedback loops to address trust breakdowns early and *upstream*, before they cascade and amplify in effect. *Daemo* focuses on two instances in the task lifecycle when trust is especially vulnerable.

The first instance is when a task is on the marketplace, waiting for workers to accept it. In this state, both workers and requesters must rely on *reputation scores* such as acceptance rates or star ratings to make a decision whether to accept the work. However, these reputation scores are often significantly inflated due to social pressure [3]. The result is that a worker or requester may agree to work with a highly-rated partner, only to find out that the partner is, in fact, fairly mediocre.

The second vulnerable moment is when workers execute the task. In this state, workers must build a mental model of

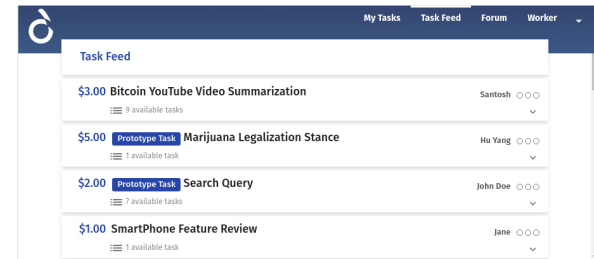


Figure 1: Daemo’s task feed lists all available tasks, their prices, and reputation of the requester.

what the requester wants. However, requesters are domain experts and not designers, leading them to believe that their *task authorship* is high-quality when it may in fact be quite the opposite, assuming tacit knowledge and ignoring edge cases [6]. When this occurs, even hard work may result in rejections because workers and requesters had no chance to reach a shared mental model.

First, addressing the breakdown in reputation, *Daemo* leverages *Boomerang*, a reputation system that differs from traditional rate-and-leave systems by “boomerang-ing” the accuracy of the rating decision back to directly impact the user [2]. This feedback loop means that giving someone a high rating increases the likelihood of working with that individual again, while giving a low rating reduces that likelihood. First, requester ratings determine early access to tasks. If a requester rates a worker highly, for example, that worker gets early access to the requester’s tasks. Rating a mediocre worker highly means that worker is likely to return and perform (mediocre) work for that requester again. Second, worker ratings determine the ranking of the list of tasks available to them (the task feed). If a worker rates a requester highly, that requester’s tasks will be ranked at the

Current State of Daemo

We are building up demand to make Daemo a vibrant platform.

Requester Side: On the requester side, Daemo has been used to build Stanford Question Answering Dataset (SQuAD), reading comprehension dataset consisting of 100,000+ questions posed by crowdworkers on a set of Wikipedia articles.

Worker Side: On the worker side, we are optimistic: many workers from Mechanical Turk, Upwork, etc., are part of our collective and act as a seed crowd of workers for Daemo.

top of their task feed. Rating a mediocre requester highly will add (mediocre) tasks to the top of the feed and make it difficult to find work. This approach aims to align reputation more with true opinion, reducing the chances that a user is surprised to find a high-rated worker or requester who is of poor quality [2].

Second, addressing the breakdown in task authorship, Daemo leverages *prototype tasks*, requiring that tasks go through iteration driven by feedback from workers before launching to the marketplace [1]. Inspired by best practice in the user-centered design process [7], prototype tasks pay 3–5 workers to complete a small percentage of the overall work and provide feedback on how to improve the task interface or clarify it. Requesters use the feedback, and any differences between their expected results and the prototype task results, to iterate. They identify unanticipated edge cases, clarify instructions and iterate before launching to the general platform. In doing so, requesters increase the probability that their tasks deliver the desired mental model and thus produce work that matches the requester’s expectations.

Years ago, the Web was a morass of information, with mediocre and high-quality sources mixed in every search query result. The main advance of search engines’ authority and centrality metrics was to increase users’ trust in the results they saw, to the point that we now trust search engines to make the world’s information readily accessible. Daemo is built by a worldwide group of workers, requesters, and researchers [1]. It represents an attempt to make collective intelligence as readily accessible as search engines made our information.

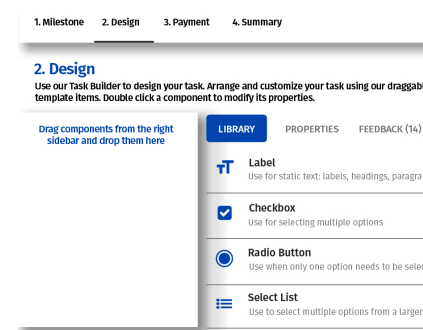


Figure 2: Daemo’s drag-and-drop task authoring interface includes a library of interactive controls that allows requesters to design prototype tasks.

Daemo

Daemo’s main page for workers is a *task feed* that lists all available tasks, their prices, and reputation of the requester. The feed lists all tasks that the worker has a sufficiently high reputation score to complete. Workers browse this feed to pick tasks and claim them (Figure 1). They complete the tasks, optionally rate the requester on a \checkmark^- , \checkmark , \checkmark^+ scale, and submit the work to the requester. When the work is accepted, they receive payment.

Requesters post prototype tasks using the Daemo interface. They upload a comma-separated value (CSV) file with any inputs that the task requires, and use a task builder to create the task interface (Figure 2). Requesters estimate the length of time that the task takes, which helps price the task. When ready, requesters post the task and monitor progress. Once work is submitted, requesters can choose to Accept each task, Return the task for revision, or (in exceptional cases, and after revision has failed) Reject the work. They can likewise rate workers on a \checkmark^- , \checkmark , \checkmark^+ scale.

Evaluation and Market Governance

To gather evidence for whether our technical approaches (Boomerang and Prototype tasks) help establish trust, we have run field experiments with workers and requesters. So far, evidence suggests that Boomerang successfully deflates reputation ratings and produces scores that are consistent with workers' and requesters' private opinions [2]. Likewise, requesters reported that task results are significantly better with the use of prototype tasks.

Our approach to all these challenges—technical, research, and policy—is highly iterative through collaboration with the crowdsourcing community. We are currently collectively designing an open governance structure and constitution for Daemo that brings all the constituent parties—worker, requester and platform—into the administration, thereby facilitating a greater level of representation in the system. As a first step toward building a self-governing community, we have drawn inspiration from historical worker guilds to design and implement *crowd guilds*: centralized groups of crowd workers who collectively certify each other's quality through double-blind peer assessment and participate in an integrated forum [8].

Conclusion

Daemo is a crowdsourcing marketplace that attempts to improve trust in crowdsourcing through upstream fixes. Our goal is to develop a strong community that balances power amongst workers, requesters, and researchers, and to demonstrate the opportunity to grow a crowdsourcing platform that does not succumb to the market for lemons.

References

- [1] Gaikwad, S., et al. Daemo: A self-governed crowdsourcing marketplace. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pp. 101–102. ACM, 2015.
- [2] Gaikwad, S., et al. Boomerang: Rebounding the consequences of reputation feedback on crowdsourcing platforms. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pp. 625–637. ACM, 2016.
- [3] Horton, J.J. and Golden, J.M. Reputation inflation: Evidence from an online labor market. 2015.
- [4] Ipeirotis, P.G., Provost, F., and Wang, J. Quality management on amazon mechanical turk. In *Proceedings of the ACM SIGKDD workshop on human computation*, pp. 64–67. ACM, 2010.
- [5] Kittur, A., Chi, E.H., and Suh, B. Crowdsourcing user studies with mechanical turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 453–456. ACM, 2008.
- [6] Martin, D., Hanrahan, B.V., O'Neill, J., and Gupta, N. Being a turker. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pp. 224–235. ACM, 2014.
- [7] Nielsen, J. Why you only need to test with 5 users, 2000.
- [8] Whiting, M.E., et al. Crowd guilds: Worker-led reputation and feedback on crowdsourcing platforms. In *Proceedings of the 2016 ACM conference on Computer supported cooperative work*. ACM, 2016.