

### Identifying the Complexity Limits of Crowd Work

A key question to the future of crowd work is *what* precisely will become part of this economy. Paid crowdsourcing began with simple microtasks on platforms such as Amazon Mechanical Turk, but microtasks are only helpful if they build up to a larger whole. So, our first question: how complex can the work outcomes from crowd work be?

#### *Crowd work's perspective*

**Crowdsourcing research has spent the better part of a decade proving the viability of crowdsourcing in complex work.** Unless crowdsourcing can demonstrate viability for meaningfully complex tasks, the argument runs, it will be incapable of ensuring a pro-social outcome for work and workers [25]. Kittur et al. first opened the question of whether crowdsourcing could be used for goals that are not simple parallel tasks [24]. Their work demonstrated proof-of-concept crowdsourcing of a simple encyclopedia article and news summary — tasks which could be verified or repeated with reasonable expectations of similar outcomes. Seeking to raise the complexity ceiling [38], researchers have since created additional proof-of-concept applications and techniques, including conversational assistants [28], medical data interpreters [28], and idea generation workflows [56, 54, 55], to name a few examples.

**To achieve complex work, this body of research has often applied ideas from Computer Science to design new crowdsourcing workflows.** Beginning with a goal that has presented significant challenges for computers, the researcher leverages an insight from Computer Science (for example, MapReduce [24] or sequence alignment algorithms [27]) and arranges humans as computational black boxes within those approaches. This approach has proven a compelling one because it leverages the in-built advantages of scale, automation, and programmability that software affords.

It is now clear that this computational workflow approach works with focused complex tasks, but the broader wicked problems largely remain unsolved [42]. As a first example, idea generation shows promise [56, 54, 55], but there is as yet no general crowdsourced solution for the broader goal of invention and innovation [14]. Second, focused writing tasks are now feasible [23, 5, 39, 48, 1], but there is no general solution to create a cross-domain, high-quality crowd-powered author. Third, data analysis tasks such as clustering [11], categorization [4], and outlining [33] are possible, but there is no general solution for sensemaking. It is not yet clear what insights would be required to enable crowdsourced solutions for these broader wicked problems.

Restricting attention to non-expert, microtask workers proved limiting. So, Retelny et al. introduced the idea of crowdsourcing with online paid *experts* from platforms such as Upwork. Expert crowdsourcing enables access to a much broader set of workers, for example designers and programmers. The same ideas can then be applied to expert “macro-tasks” [10, 18], enabling the crowdsourcing of goals such as user-centered design [40], programming [29, 13, 9], and mentorship [47]. However, there remains the open question of how complex the work outcomes from expert crowds can be.

#### *Piecework's perspective*

[a]2: I'm not sure that this really stays close to the three case studies that emerged as major threads (now in the piecework lit review). Should I refactor/rewrite this to make it more relevant? Measurement also limited the complexity of piecework: only tasks that could be measured and priced could be completed via piecework. When Brown investigated what limited the adoption of piecework in industries that otherwise gravitated toward it (e.g., railway engineers), the homogeneity of tasks arose as a major contributing factor [7]. Graves concurs via a case study of the Santa Fe Railway, which used “efficiency experts” to develop a “standard time” to determine pay for each task at the company informed by “thousands of individual operations” [16]. One might conclude from Graves's observations that complex, creative work — which is inherently heterogeneous and difficult to routinize — would be unsuitable for piecework.

**Piecework was limited to tasks that could be clearly evaluated.** For example, the roles required to facilitate piecework in the early 20th century included “piecework clerks, inspectors, and ‘experts’” [16]. Hart argue that evaluation is the ultimate complexity limit: at some point, evaluating multidimensional work for quality (rather than for quantity) becomes infeasible. In their words, “if the quality of the output is more difficult to measure than the quantity [...] then a piecework system is likely to encourage an over-emphasis on quantity produced and an under-emphasis on quality” [20]. Complex work, which is often subjective to evaluate, falls victim to this criteria.

**This focus on measurement and tracking had consequences.** Graves suggests that the first sparks of scientific management could be found in piecework: the approach of paying workers for each piece of output necessitated the rigorous tracking, measurement, and training of workers for which scientific management became famous [16]. If true, the concurrent upswing of scientific management and Fordism through the first two-thirds of the 20th century alongside piecework was not only understandable, but predictable [19].

**Piecework researchers also argue that, in addition to constraints on the kind of work that's amenable to piecework, only certain kinds of organizations were amenable to piecework.** Researchers detail three organizational criteria. First, Brown argues that piecework “is less likely in jobs with a variety of duties than in jobs with a narrow set of routinized duties” [7]. Agell points out the phenomenon here as a market effect: “in an environment with multi-tasking, pay schemes based on tightly specified performance may induce workers to neglect tasks that are less easy to measure” [2]. Second, complexity was limited by access to capital to create the necessary infrastructure. As Graves reports, only the largest and most wealthy railroads had the resources necessary [16]. Third, organizations required capable managers in charge of the pieceworkers. The West Virginia mines, for example, hired foremen to be the intermediary between upper management and the workers [6]. These foremen were responsible for allocating resources and understanding when and how to modify work as necessary [53]. So, in sum, organizations

historically could only take advantage of piecework if they had homogeneous work to be done, access to capital to purchase the necessary equipment, and the ability to hire people who could serve as intermediaries between pieceworkers and management.

**The research seems to suggest that it was difficult to apply piecework to more skilled work, particularly because maximizing the advantages of piecework seemed to reward smaller, more constrained, more narrowly-trained tasks, and only in organizations that could pay for the equipment and people to enable it.** For most of the 19th century, piecework was applied almost exclusively to farm and textile work. Work was simple and widely understood — farm workers didn't need to be trained on how to plow fields, or birth foals; seamstresses knew how to sew together denim [8, 41].

#### *Comparing the phenomena*

The research on piecework tells us that we should expect piecework to thrive in industries where the nature of the work is limited in complexity [7]. Given the flourishing of on-demand labor platforms such as Uber, AMT, and others, we ask ourselves what — if anything — has changed. We argue that the internet has trivialized the costs and challenges of the earlier limiting factors because technology makes it easier 1) for workers to do complex work without training, 2) to manage workers in doing complex work, and 3) to create the infrastructure necessary to manage the workers.

**Technology increases non-experts' levels of expertise by giving access to information that would otherwise be unavailable.** For example, taxi drivers in London endure rigorous training to pass a test known as “The Knowledge” — a demonstration of the driver's comprehensive familiarity with the city's roads. This test is so challenging that veteran drivers develop significantly larger the regions of the brain associated with spatial functions such as navigation [35, 34, 45, 46, 52, 51]. In contrast, with on-demand platforms such as Uber, services such as Google Maps & Waze make it possible for people entirely unfamiliar with a city to operate professionally [44, 21]. Other examples include search engines enabling information retrieval, and word processors enabling spelling and grammar checking. By augmenting the human intellect [12], computing has shifted the complexity of work that is possible without training.

Algorithms have automated some tasks that previously fell to management. Computational systems hire workers [32, 49], as well as direct their activities [30], and act as “piecework clerks” [16] to inspect, modify and combine work [22, 37]. In many cases, the intermediary function has been removed as well, leading workers to need to directly email requesters for clarification and feedback [36]. These algorithms, however, are less able than human managers to manage contingencies that were not programmed into them.

Finally, the organizational limit on infrastructure creation is somewhat lessened. Writing web scripts takes fewer people and fewer hours than creating physical equipment for piecework. Little et al.'s vision was that any user with basic pro-

gramming skills could tap into on-demand human intelligence. As better toolkits lower this threshold [38] and computational thinking diffuses, a broader population will be able to use crowd work.

#### *Implications for crowd work*

Technology's ability to support human cognition will enable stronger assumptions about workers' abilities, increasing the complexity of crowd work outcomes. Just as the shift to expert crowdsourcing increased complexity, so too will workers with better tools increase the set of tasks possible. Beyond this, further improvements would most likely come from replicating the success of narrowly-slicing education for expert work as Hart and Roberts and Grier described in their piecework examples of human computation [17] and drastically reformulating macro-tasks given the constraints of piecework [19]. To some extent, an argument can be made that MOOCs and other online education resources provide crowd workers with the resources that they need, but it remains to be seen whether that work will be appropriately valued, let alone properly interpreted by task solicitors [3]. If we can overcome this obstacle, we might be able to empower more crowd workers to do complex work such as engineering and metalworking, rather than doom them to “uneducated” match girl reputations [43]. However, many such experts are already available on platforms such as Upwork, so training may not directly increase the complexity accessible to crowd work unless it makes common expertise more broadly available.

Will the shift from human managers to Turing-complete algorithms raise the complexity ceiling? By the Turing test, the algorithms would be at best indistinguishable from human piecework clerks and foremen. So in terms of enabling coordination, algorithmic management is unlikely to directly raise the ceiling beyond what piecework could achieve. However, as a resource constraint, algorithms are a fixed cost and not a per-person cost like human managers. So in terms of accessibility, algorithms will allow a broader class of organizations and individuals to afford crowd work. This shift may enable complex goals that were not cost-effective before to become feasible. However, because algorithms remain far from replicating all of the foremen's responsibilities, most likely is a middle ground in which crowd work re-introduces the human element to management in a more targeted way (e.g., [18, 26, 50]). This move will require resolving the tension between workers and perilously antagonistic managers, as Boal and Pencavel suggest, to break a toxic cycle of mistrustful requesters [15].

Finally, the cost of creating piecework infrastructure has dropped. Expensive manufacturing equipment has been largely replaced by computer code [31]. As with lowered costs of management, lowered infrastructure costs will make crowd work accessible to a broader set of people and organizations. This in and of itself does not raise the complexity ceiling, but by broadening the potential market for crowd work, it may enable a new set of goals and needs take part.