Complexity Limits of On–Demand Work

A key question to the future of on–demand 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 on–demand work be?

The perspective of on–demand work

Kittur et al. first opened the question of whether crowdsourcing could be used for goals that are not simple parallel tasks [23]. Their work demonstrated proof–of–concept crowdsourcing of an encyclopedia article and news summary — tasks which could be verified or repeated with reasonable expectations of similar outcomes. Seeking to raise the complexity ceiling, researchers have since created additional proof–of–concept applications and techniques, including conversational assistants [25], medical data interpreters [25], and idea generation work–flows [46, 44, 45].

To achieve complex work, this body of research has often applied ideas from Computer Science to design new crowdsourcing work–flows. The system designer leverages techniques such as MapReduce [23] and sequence alignment algorithms [24], arranging humans as computational black boxes within. 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 some classes of complex tasks, but the broader wicked problems largely remain unsolved [35]. As a first example, idea generation shows promise [46, 44, 45], but there is as yet no general crowdsourced solution for the broader goal of invention and innovation [13]. Second, focused writing tasks are now feasible [22, 4, 31, 41, 1], but there is no general solution to create a cross–domain, high–quality crowd–powered author. Third, data analysis tasks such as clustering [8], categorization [3], and outlining [27] are possible, but there is no general solution for sense–making. 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" [7, 16], enabling the crowdsourcing of goals such as user–centered design [33], programming [26, 12, 6], and mentorship [40]. However, there remains the open question of how complex the work outcomes from expert crowds can be.

The perspective of piecework

Piecework's body of research most squarely addresses complexity in two of the cases we looked at earlier: Airy's human computers and among industrial workers.

Airy's work on astronomical charts opened the door to greater task complexity by encoding the intelligence into the process rather than the people. Airy's computers had relatively limited education in mathematics, but by combining individual simple mathematical operations, Airy was able to create a complex composite outcome [15]. Likewise, in Ford's factories, no individual could build the entire car, but the process could emergently produce one.

However, when piecework intially entered the American economy, it was not used for complex work. Without having designed complex work processes, piecework managers were restricted to available workers' skills such as sewing: it was infeasible to provide new pieceworkers with the comprehensive education that apprenticeships imparted [17]. So, initially piecework arose for farm work, and as Raynbird and others discuss, the practice remained relatively obscure until it blossomed in the textile industry [32]. Complexity levels remained low at the turn of the 20th century as piecework saturated New York City [34].

Measurement also limited the complexity of piecework: only tasks that could be measured and priced could be completed via piecework. Earlier we discussed Graves's and later Brown's analysis of railway workers. They identified task homogeneity and measurement as key requirements for piecework. However, complex, creative work — which is inherently heterogeneous and difficult to routinize — was unsuitable [14]. Brown's description of "efficiency experts" would corroborate this; efficiency experts can effectively gauge how long known tasks should take, but would find themselves overwhelmed if they attempted to assess creative tasks like scientific research, which can take an arbitrary number of iterations before proceeding to a subsequent step.

Moreover, piecework was limited to tasks that could be quickly and accurately evaluated. Hart argue that evaluation limited piecework's complexity: at some point, evaluating multidimensional work for quality (rather than for quantity) becomes infeasible. In his 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 . . . and an under–emphasis on quality" [18]. Complex work, which is often subjective to evaluate, falls victim to this criteria.

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 [5], and become less common as work becomes more complex. However, has computation shifted piecework's previous limits of expertise, measurement, and evaluation?

In some ways, yes: 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 [29, 28, 38, 39, 43, 42]. In contrast, with on–demand platforms such as Uber, services such as Google Maps and Waze make it possible for people entirely unfamiliar with a city to oper-

ate professionally [37, 19]. Other examples include search engines enabling information retrieval, and word processors enabling spelling and grammar checking. By augmenting the human intellect [10], computing has shifted the complexity of work that is possible without training.

Algorithms have automated some tasks that previously fell to management. Computational systems now act as "piecework clerks" [14] to inspect, modify and combine work [21, 30]. However, these algorithms are less able than human managers to evaluate subjective work.

Implications for on–demand work
Algorithms are undoubtably capable of shepherding more complex work than the linear processes available to Airy and Ford. However, as work becomes more complex, it becomes increasingly difficult to codify a process to achieve it [11, 9]. So, while algorithms will increase the complexity ceiling beyond what was possible previously with piecework, there is a fundamental limit to how complex such work can become.

Technology's ability to support human cognition will enable stronger assumptions about workers' abilities, increasing the complexity of on–demand 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 [15] and drastically reformulating macro–tasks given the constraints of piecework [17]. An argument might 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 [2]. If we can overcome this obstacle, we might be able to empower more of these workers to do complex work such as engineering, rather than doom them to "uneducated" match–girl reputations [36]. However, many such experts are already available on platforms such as Upwork, so training may not directly increase the complexity accessible to on–demand work unless it makes common expertise more broadly available.

Evaluation remains as difficult for crowd work as it did for the efficiency experts. Reputation systems for crowdsourcing platforms remain notoriously inflated [20]. Ultimately, many aspects of assessment remain subjective: whether a logo made for a client is fantastic or terrible may depend on taste.

So, in the case of complexity, the history of piecework does not yet offer compelling evidence that on–demand work will achieve far more complex outcomes than piecework did. Improvements in workflows, measurement and evaluation have already been made, and it is not clear that the remaining challenges are readily solveable. However, on–demand work will be far more broadly distributed than piecework historically was—reaching many more tasks and areas of expertise by virtue of the internet.