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. were among the first to ask whether crowdsourcing could be used for more than parallelizing tasks [23]. Their work showed that it could, with proof–of–concept crowdsourcing of encyclopedia articles and news summaries — tasks which could be verified or repeated with reasonable expectations of similar results. Seeking to raise the complexity ceiling, researchers have since created yet more applications and techniques, including conversational assistants [25], medical data interpreters [25], and idea generation [43, 44].

To achieve complex work, this body of research has often applied ideas from Computer Science to design new workflows. System designers leverage techniques such as MapReduce [23] and sequence alignment algorithms [24], arranging humans as computational black boxes. This approach has proven a compelling one because it leverages the inherent 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. As a first example, idea generation shows promise [43, 44], 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, 40, 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 [39]. 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 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.

But 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 [18]. 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 densely populated urban areas such as London and 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 to be successful. 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 argues 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" [17]. Complex work, which is often subjective to evaluate, falls victim to this pitfall.

Comparing the phenomena

The research on piecework tells us that we should expect it to thrive in industries where the nature of the work is limited in complexity [5], and become less common as work becomes more complex. 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 [28, 29, 37, 38, 42, 41]. 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 operate professionally [36, 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 with minimal or no training.

Algorithms have automated some tasks that previously fell to management. Computational systems now act as "piecework clerks" [14] to inspect and modify work [21, 30]. However, these algorithms are less competent than humans at evaluating subjective work, as well as in their ability to exercise discretion, causing new problems for workers and managers.

Implications for on–demand work

Algorithms are undoubtedly 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 later Grier described in their piecework examples of human computation [15] and drastically reformulating macro–tasks given the constraints of piecework [18]. 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 [35]. 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's not immediately clear that the remaining challenges are readily solvable. 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.

REFERENCES

1. Elena Agapie, Jaime Teevan, and Andrés Monroy-Hernández. 2015. Crowdsourcing in the field: A case study using local crowds for event reporting. In *Third AAAI Conference on Human Computation and Crowdsourcing*.
2. J Ignacio Aguaded-Gómez. 2013. The MOOC Revolution: A new form of education from the technological paradigm. *Comunicar* 41, 21 (2013), 7–8.
3. Paul André, Aniket Kittur, and Steven P. Dow. 2014. Crowd Synthesis: Extracting Categories and Clusters from Complex Data. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, 989–998. DOI: http://dx.doi.org/10.1145/2531602.2531653
4. Michael S. Bernstein, Greg Little, Robert C. Miller, Björn Hartmann, Mark S. Ackerman, David R. Karger, David Crowell, and Katrina Panovich. 2010. Soylent: A Word Processor with a Crowd Inside. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology (UIST '10)*. ACM, 313–322. DOI:http://dx.doi.org/10.1145/1866029.1866078
5. Charles Brown. 1990. Firms' Choice of Method of Pay. *Industrial & Labor Relations Review* 43, 3 (1990), 165S–182S. DOI: http://dx.doi.org/10.1177/001979399004300311
6. Yan Chen, Steve Oney, and Walter S. Lasecki. 2016. Towards Providing On-Demand Expert Support for Software Developers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, 3192–3203. DOI: http://dx.doi.org/10.1145/2858036.2858512
7. Justin Cheng, Jaime Teevan, Shamsi T. Iqbal, and Michael S. Bernstein. 2015. Break It Down: A Comparison of Macro- and Microtasks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, 4061–4064. DOI: http://dx.doi.org/10.1145/2702123.2702146
8. Lydia B. Chilton, Greg Little, Darren Edge, Daniel S. Weld, and James A. Landay. 2013. Cascade: Crowdsourcing Taxonomy Creation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, 1999–2008. DOI: http://dx.doi.org/10.1145/2470654.2466265
9. Amy C. Edmondson. 2012. *Teaming: How organizations learn, innovate, and compete in the knowledge economy*. John Wiley & Sons, San Francisco, California.
10. Douglas C Engelbart. 2001. Augmenting human intellect: a conceptual framework (1962). *PACKER, Randall and JORDAN, Ken. Multimedia. From Wagner to Virtual Reality. New York: WW Norton & Company* (2001), 64–90.
11. Samer Faraj and Yan Xiao. 2006. Coordination in Fast-Response Organizations. *Management Science* 52, 8 (2006), 1155–1169. http://mansci.journal.informs.org/content/52/8/1155.short

12. Ethan Fast and Michael S. Bernstein. 2016. Meta: Enabling Programming Languages to Learn from the Crowd. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, 259–270. DOI: http://dx.doi.org/10.1145/2984511.2984532
13. Mark Fuge, Kevin Tee, Alice Agogino, and Nathan Maton. 2014. Analysis of collaborative design networks: A case study of openideo. *Journal of Computing and Information Science in Engineering* 14, 2 (2014), 021009.
14. Carl Graves. 1981. Applying Scientific Management Principles to Railroad Repair Shops — the Santa Fe Experience, 1904-18. *Business and Economic History* 10 (1981), 124–136. http://www.jstor.org/stable/23702539
15. David Alan Grier. 2013. *When computers were human*. Princeton University Press.
16. Daniel Haas, Jason Ansel, Lydia Gu, and Adam Marcus. 2015. Argonaut: macrotask crowdsourcing for complex data processing. *Proceedings of the VLDB Endowment* 8, 12 (2015), 1642–1653.
17. Robert A Hart and others. 2016. the rise and fall of piecework. *IZA World of Labor* (2016).
18. Robert A Hart and J Elizabeth Roberts. 2013. The rise and fall of piecework–timework wage differentials: market volatility, labor heterogeneity, and output pricing. (2013).
19. Sam Hind and Alex Gekker. 2014. 'Outsmarting Traffic, Together': Driving as Social Navigation. *Exchanges: the Warwick Research Journal* 1, 2 (2014), 165–180.
20. John J. Horton, Leonard N. Stern, and Joseph M. Golden. 2015. Reputation Inflation: Evidence from an Online Labor Market. (2015).
21. 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 (CHI '13)*. ACM, 611–620. DOI:http://dx.doi.org/10.1145/2470654.2470742
22. Joy Kim, Sarah Sterman, Allegra Argent Beal Cohen, and Michael S Bernstein. 2017. Mechanical Novel: Crowdsourcing Complex Work through Revision. In *Proceedings of the 20th ACM Conference on Computer Supported Cooperative Work \& Social Computing*.
23. Aniket Kittur, Boris Smus, Susheel Khamkar, and Robert E. Kraut. 2011. CrowdForge: Crowdsourcing Complex Work. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST '11)*. ACM, 43–52. DOI: http://dx.doi.org/10.1145/2047196.2047202
24. Walter Lasecki, Christopher Miller, Adam Sadilek, Andrew Abumoussa, Donato Borrello, Raja Kushalnagar, and Jeffrey Bigham. 2012. Real-time Captioning by Groups of Non-experts. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology (UIST '12)*. ACM, 23–34. DOI: http://dx.doi.org/10.1145/2380116.2380122
25. Walter S. Lasecki, Rachel Wesley, Jeffrey Nichols, Anand Kulkarni, James F. Allen, and Jeffrey P. Bigham. 2013. Chorus: A Crowd-powered Conversational Assistant. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. ACM, 151–162. DOI:http://dx.doi.org/10.1145/2501988.2502057
26. Thomas D. LaToza, W. Ben Towne, Christian M. Adriano, and André van der Hoek. 2014. Microtask Programming: Building Software with a Crowd. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)*. ACM, 43–54. DOI:http://dx.doi.org/10.1145/2642918.2647349
27. Kurt Luther, Nathan Hahn, Steven P Dow, and Aniket Kittur. 2015. Crowdlines: Supporting Synthesis of Diverse Information Sources through Crowdsourced Outlines. In *Third AAAI Conference on Human Computation and Crowdsourcing*.
28. Eleanor A. Maguire, David G. Gadian, Ingrid S. Johnsrude, Catriona D. Good, John Ashburner, Richard S. J. Frackowiak, and Christopher D. Frith. 2000. Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences* 97, 8 (2000), 4398–4403. DOI: http://dx.doi.org/10.1073/pnas.070039597
29. Eleanor A. Maguire, Rory Nannery, and Hugo J. Spiers. 2006. Navigation around London by a taxi driver with bilateral hippocampal lesions. *Brain* 129, 11 (2006), 2894–2907. DOI:http://dx.doi.org/10.1093/brain/awl286
30. Brian McInnis, Dan Cosley, Chaebong Nam, and Gilly Leshed. 2016. Taking a HIT: Designing Around Rejection, Mistrust, Risk, and Workers' Experiences in Amazon Mechanical Turk. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, 2271–2282. DOI: http://dx.doi.org/10.1145/2858036.2858539
31. Michael Nebeling, Alexandra To, Anhong Guo, Adrian A. de Freitas, Jaime Teevan, Steven P. Dow, and Jeffrey P. Bigham. 2016. WearWrite: Crowd–Assisted Writing from Smartwatches. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, 3834–3846. DOI: http://dx.doi.org/10.1145/2858036.2858169
32. Hugh Raynbird. 1847. *Essay on Measure Work, locally known as task, piece, job, or grate work (in its application to agricultural labour)*.
33. Daniela Retelny, Sébastien Robaszkiewicz, Alexandra To, Walter S. Lasecki, Jay Patel, Negar Rahmati, Tulsee Doshi, Melissa Valentine, and Michael S. Bernstein. 2014. Expert Crowdsourcing with Flash Teams. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)*. ACM, 75–85. DOI:http://dx.doi.org/10.1145/2642918.2647409
34. Jacob August Riis. 1901. *How the other half lives: Studies among the tenements of New York*. Penguin.

35. Lowell J. Satre. 1982. After the Match Girls' Strike: Bryant and May in the 1890s. *Victorian Studies* 26, 1 (1982), 7–31. http://www.jstor.org/stable/3827491

36. Thiago H Silva, Pedro OS Vaz de Melo, Aline Carneiro Viana, Jussara M Almeida, Juliana Salles, and Antonio AF Loureiro. 2013. Traffic condition is more than colored lines on a map: characterization of waze alerts. In *International Conference on Social Informatics*. Springer, 309–318.

37. Walter Skok. 1999. Knowledge Management: London Taxi Cabs Case Study. In *Proceedings of the 1999 ACM SIGCPR Conference on Computer Personnel Research (SIGCPR '99)*. ACM, 94–101. DOI: http://dx.doi.org/10.1145/299513.299625

38. Walter Skok. 2000. Managing knowledge within the London taxi cab service. *Knowledge and Process Management* 7, 4 (2000), 224.

39. 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 (CHI '16)*. ACM, 2645–2656. DOI: http://dx.doi.org/10.1145/2858036.2858121

40. Jaime Teevan, Shamsi T. Iqbal, and Curtis von Veh. 2016. Supporting Collaborative Writing with Microtasks. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, 2657–2668. DOI: http://dx.doi.org/10.1145/2858036.2858108

41. Katherine Woollett and Eleanor A Maguire. 2011. Acquiring "the Knowledge" of London's layout drives structural brain changes. *Current biology* 21, 24 (2011), 2109–2114.

42. Katherine Woollett, Hugo J. Spiers, and Eleanor A. Maguire. 2009. Talent in the taxi: a model system for exploring expertise. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 364, 1522 (2009), 1407–1416. DOI: http://dx.doi.org/10.1098/rstb.2008.0288

43. Lixiu Yu, Aniket Kittur, and Robert E. Kraut. 2016a. Distributed Analogical Idea Generation with Multiple Constraints. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, 1236–1245. DOI: http://dx.doi.org/10.1145/2556288.2557371

44. Lixiu Yu, Aniket Kittur, and Robert E. Kraut. 2016b. Encouraging "Outside- The- Box" Thinking in Crowd Innovation Through Identifying Domains of Expertise. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, 1214–1222. DOI: http://dx.doi.org/10.1145/2818048.2820025