Identifying the Limits of Crowd Work

Crowd work's perspective. **Crowdsourcing research has spent the better part of a decade attempting to prove the viability of crowdsourcing in increasingly complex work.** Kittur et al. map the discussion toward this goal in their work on crowdsourcing complex work [13]. The broader body of work has varied significantly in type — providing conversational assistants, interpreting medical data, and telling coherent and compelling stories, to name a few examples [15, 20, 12].

This body of research has involved similar approaches to problems, often involving insights made in Computer Science and applied to human work–flows. The crowd work literature typically identifies target milestones in computer science that have presented significant challenges for researchers, leverages some of the approaches and insights that Computer Science researchers have already made (for example, MapReduce in the case of Kittur et al.'s *CrowdForge*), and arranges humans as computational black boxes within those approaches and processes [13, 22, and others]. This approach has proven a compelling one because it leverages the in–built advantages that technology and digital media afford. For example, *Foundry*'s tools for managing and arranging expert groups into a cohort allow researchers to convincingly argue that expert teams can be rapidly formed, just like non–expert teams [22].

Piecework's perspective. **Piecework researchers have found themselves trying to understand what characteristics limit piecework, or more precisely what has prevented piecework from becoming more prevalent.** The research into piecework makes the case that piecework has been limited principally by the challenges of human management and oversight. Graves describes a case study in Santa Fe Railway, which deployed scientific management and a piecework regime in an attempt to stymie rising repair costs [7]. Graves reports on the hiring of Harrington Emerson, an "efficiency expert", who went on to develop a "standard time" for each task at the company informed by "thousands of individual operations at the Topeka shops". The cost of measuring workers in such excruciating detail at the turn of the 20th century was undoubtedly immense, but this "standard time" value, which determined the pay that workers would earn for each task they do, was the only viable approach at the time to determine appropriate pay given the task [7]. But the repeated measurement of workers' time to complete tasks had shortcomings; for one thing, pay rates for rarer tasks were necessarily less certain than for more common tasks, which had the simple benefit of a larger sample size. [al2: Do I need "for another thing…"? This paragraph is kind of huge already. Not sure how to break it up/down as it is.] One might conclude from Graves's observations that complex, creative work — which is inherently heterogeneous and difficult to routinize — would be unsuitable for piecework.

Determining appropriate pay rates, informed by the careful measurement of workers, isn't the only major challenge piecework faced; evaluation proved a limiting factor as well. Graves enumerates some of the roles required to facilitate piecework in the early 20th century — among them "… piecework clerks, inspectors, and 'experts'…" [7]. Graves further recognizes that it's necessary for a successful piecework shop to employ clerks, inspectors, and other experts to properly design and evaluate complex work. Hart later makes a more concrete observation of this hurdle, as he argues an ultimate limit to how far this can go; at some point, evaluating multidimensional work output 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, perhaps because of 'difficult–to–observe' production techniques, then a piecework system is likely to encourage an over–emphasis on quantity produced and an under-emphasis on quality" [10]. This, Hart argues, may have fundamentally hamstrung piecework, and ultimately precipitated its downfall, especially with the increasing complexity of manufacturing work over the course of the 20th century. [al2: just to be clear, this whole point is queuing up so that later I can be like "Oh hey computers are terrific at evaluating some stuff really quickly. And researchers in crowd work have done some work on workers evaluating other workers (e.g. Find–Fix–Verify, arguably *PeerStudio* does something along this line) [2, 14]"]

The research seems to suggest that it was difficult to apply piecework to more skilled work, particularly because maximizing on the advantages of piecework seemed to reward smaller, more constrained, more narrowly trained tasks. 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 [5, 23].

This isn't to say that complex work is outside of the realm of piecework; indeed, we've discussed complex applications of crowd work already. While Hart and Roberts described a flourishing of ingenious piecework design, much of it arose out of necessity — it was infeasible to provide new workers with the comprehensive education that was familiar to men through apprenticeships [9]. While this constraint led to much more tightly scoped work who now had to specialize in extremely narrowly defined roles. The same could be said of Airy and his *computers* — young boys whose preparations consisted principally of a relatively specific mathematics curriculum [8]. Instead, we argue that the literature suggests that piecework is tightly limited only when the application of piecework follows a direct, perhaps even unimaginative, mapping from an time–based regime to an output–based one. When the work is redesigned from the ground up — as we see with mathematicians in the 19th century and with the metalworking industry during the Second World War — it seems that we don't yet know the limits of complexity with regard to piecework. [al2: This is *almost* making a prediction, but I wanted to take something punchy away from the stuff I cite here.]

Piecework researchers also argue that, in addition to constraints on the kind of *work* that's amenable to piecework, only certain kinds of *organizations* are amenable to piecework. Brown discusses the organizational factors necessary for piecework to thrive, arguing that piecework "… is less likely in jobs with a variety of duties than in jobs with a narrow

set of routinized duties" [4] Graves adds further, that successful cases of piecework owed themselves in part to the fact that "... only [the largest and most wealthy railroads] had the resources to ... pay the overhead involved in installing work reorganization" [7]. Together, Graves and Brown make a persuasive argument that piecework is limited in complexity by the capacity to endure managerial overhead while transitioning to a new system. [MSB: wait, wasn't that in a previous paragraph? that should be joined with the managerial overhead text above] [al2: I'm trying to make this argument that [7] is making two claims; the first is that there are limits on doing hard, "difficult–to–observe" work, and the second is that the cost propagates up to management, putting limits on who can run piecework systems. So...

- ☐ Rewrite
- ☐ Refactor
- ☐ Drop

Maybe I should just merge it with the next paragraph?]

There are other characteristics to effective complex piecework institutions, such as appropriately designed management practices. Boal and Pencavel describe the role of the foreman in West Virginia coal mines under the piecework model: "The foreman had the power to hire and fire workers and allocate workplaces, but then left the face–worker largely free to his own efforts so that often he went all day without seeing the foreman" [3]. The general approach adopted by these West Virginia mines was, as in other factories with active foremen, to let the foreman be the intermediary between management and the worker. Specifically, foremen were responsible for allocating resources and understanding when and how to modify work as necessary [30]. The management of pieceworkers demanded people in positions akin to foremen — intermediate managers people who were 1) familiar with and even sympathetic to the needs of workers, 2) empowered by higher level management to make decisions, and 3) relaxed enough in day–to–day work to allow workers to go about their work [30, 3].

What's different about crowd work. [MSB: Before you get into this, summarize what I am supposed to have learned from the prior section on piecework. You're about to draw on those points to make your argument, so they need to be at the top of my mind here.] **Piecework makes a number of observations leading to the conclusion that piecework's complexity is fundamentally bounded by several limitations, chief among them the costs of managerial overhead and the transition thereto.** Brown and Graves's claims that organizations can't adopt piecework unless they're sufficiently large to absorb the cost of transitioning to a piecework system; Boal and Pencavel and Wray's observations for the importance of competent, effective managerial oversight — a human resource, which made the scaling cost prohibitively expensive for many [3, 30, 7, 4]. [al2: Something like this? Should I dig deeper? I can line these points up in a way to make the next paragraph sort of obvious or inevitably, depending on how actively engaged someone is while reading this...]

Digital media have expanded the scope of viable piecework by pushing drastically on the limits cited by piecework researchers. 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 [4]. 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 make it easier 1) to do complex work aided by computers and 2) to evaluate and manage workers as they do increasingly complex work, even observing their work to an otherwise unprecedented granularity.

Technology has made it possible for non–experts to do work that was once considered within the domain of experts. [MSB: I don't yet buy the following argument. If the point is that technology makes us more expert, I disagree that the CrowdCrit/Voyant systems are using technology to do this. They are building the smarts into their OWN workflows, rather than giving workers EXTERNAL tools that make them smarter. Giving workers a calculator is an external tool; the mathematical tables project already demonstrated that you can build smarts into the workflow if you don't have one.] Yuan et al. builds on the work of others (*Voyant* and, more relevantly, *CrowdCrit*) to design workflows that yield "expert–level feedback" [32, 31, 17]. This body of work identifies ways to transform a variety of duties comprising complex tasks and distills them into "a narrow set of routinized duties", informed in part by researchers — acting as inspectors — and experts [quotations from 7]. Where Graves would call additionally for the identification of crowdsourcing's version of "piecework clerks", we point out that today algorithms manage workers as pieceworkers once did [16, 7].

Furthermore, technology more directly facilitates the subversion of expertise requirements by giving non–experts access to information that would otherwise be unavailable. Taxi drivers in London endure rigorous training to pass a test known as "The Knowledge" — a demonstration of the driver's comprehensive familiarity. Researchers have identified significant growth of the hippocampal regions of the brains in veteran drivers, generally understood to be responsible for spatial functions such as navigation [19, 18, 26, 27, 29, 28]. Services such as Google Maps & Waze make it possible for people entirely unfamiliar with a city to know more about a city even than experts through the collective data generated by other users ranging topics such as police activity, congestion, construction, etc... [25, 11]. [MSB: what's the insight I should take away from this paragraph? what does this say about crowd work?]

Implications for crowd work research. [MSB: This paragraph can be expanded to make a more concrete argument. What will be possible? What won't?] **The piecework literature gives us a template for pushing the boundaries of complexity in piecework, but it also signals some of the ultimate limitations of crowd work and piecework in general.** While the threshold preventing task requesters from utilizing piecework has dropped thanks to affordances of the Internet, the ceiling on task complexity hasn't moved significantly. [MSB: is that your prediction? I would argue against the fact that it hasn't

moved significantly, crowdforge did far more complex work, as did flash teams and flash orgs] If we're to make use of Brown's prescriptions, we would benefit from finding ways to decompose varied tasks into homogeneous microtasks. [MSB: isn't that what we've been doing all along?]

[MSB: this doesn't seem like a concrete prediction. what would piecework say will happen if we didn't resolve the tension?] We should also consider exploring the limitations that algorithmic management bring along more carefully. While research has touched on this subject, we've yet to make out the bigger picture of this theme [16]. If we can resolve the tension between workers and perilously antagonistic managers, as Boal and Pencavel suggest, then we may be able to break a toxic cycle of mistrustful requesters [for example 6] and develop more considerate platforms as McInnis et al. advocate [21].

Finally, and perhaps most importantly, we need to replicate the success of narrowly slicing education and training for expert work as Hart and Roberts and Grier described in their piecework examples [9, 8] [MSB: remind us of what those were]. That is, we need to identify new ways to train crowdworkers for uniquely narrowly defined work. [MSB: I don't understand: why?] 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 [1]. If we can overcome this obstacle, we might be able to empower crowd workers to do complex work such as engineering and metalworking, rather than doom them to match girl reputations: "brash, irregular, immoral, and uneducated" [24]. [MSB: how will that reduce any of the problems except the last one?]

References

[1] J Ignacio Aguaded-Gómez. "The MOOC Revolution: A new form of education from the technological paradigm". In: *Comunicar* 41.21 (2013), pp. 7–8.

[2] Michael S. Bernstein et al. "Soylent: A Word Processor with a Crowd Inside". In: UIST '10 (2010), pp. 313–322. DOI: 10.1145/1866029.1866078. URL: http://doi.acm.org/10.1145/1866029.1866078.

[3] William M. Boal and John Pencavel. "The Effects of Labor Unions on Employment, Wages, and Days of Operation: Coal Mining in West Virginia". In: *The Quarterly Journal of Economics* 109.1 (1994), pp. 267–298. ISSN: 00335533, 15314650. URL: http://www.jstor.org/stable/2118435.

[4] Charles Brown. "Firms' Choice of Method of Pay". In: *Industrial & Labor Relations Review* 43.3 (1990), 165S–182S. DOI: 10.1177/001979399004300311. eprint: http://ilr.sagepub.com/content/43/3/165S.full.pdf+html. URL: http://ilr.sagepub.com/content/43/3/165S.abstract.

[5] Edwin Chadwick. "Opening Address of the President of the Department of Economy and Trade, at the Meeting of the National Association for the Promotion of Social Science, held at York, in September, 1864". In: *Journal of the Statistical Society of London* 28.1 (1865), pp. 1–33. ISSN: 09595341. URL: http://www.jstor.org/stable/2338394.

[6] Ujwal Gadiraju et al. "Understanding Malicious Behavior in Crowdsourcing Platforms: The Case of Online Surveys". In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI '15. New York, NY, USA: ACM, 2015, pp. 1631–1640. ISBN: 978–1-4503–3145–6. DOI: 10.1145/2702123.2702443. URL: http://doi.acm.org/10.1145/2702123.2702443.

[7] Carl Graves. "Applying Scientific Management Principles to Railroad Repair Shops — the Santa Fe Experience, 1904-18". In: *Business and Economic History* 10 (1981), pp. 124–136. ISSN: 08946825. URL: http://www.jstor.org/stable/23702539.

[8] David Alan Grier. *When computers were human*. Princeton University Press, 2013.

[9] Robert A Hart and J Elizabeth Roberts. "The rise and fall of piecework–timework wage differentials: market volatility, labor heterogeneity, and output pricing". In: (2013).

[10] Robert A Hart et al. "the rise and fall of piecework". In: *IZA World of Labor* (2016).

[11] Sam Hind and Alex Gekker. "'Outsmarting Traffic, Together': Driving as Social Navigation". In: *Exchanges: the Warwick Research Journal* 1.2 (2014), pp. 165–180.

[12] Joy Kim and Andrés Monroy-Hernández. "Storia: Summarizing Social Media Content Based on Narrative Theory Using Crowdsourcing". In: *Proceedings of the 19th ACM Conference on Computer–Supported Cooperative Work & Social Computing*. CSCW '16. New York, NY, USA: ACM, 2016, pp. 1018–1027. ISBN: 978–1-4503–3592–8. DOI: 10.1145/2818048.2820072. URL: http://doi.acm.org/10.1145/2818048.2820072.

[13] Aniket Kittur et al. "CrowdForge: Crowdsourcing Complex Work". In: *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*. UIST '11. New York, NY, USA: ACM, 2011, pp. 43–52. ISBN: 978–1-4503–0716–1. DOI: 10.1145/2047196.2047202. URL: http://doi.acm.org/10.1145/2047196.2047202.

[14] Chinmay E. Kulkarni, Michael S. Bernstein, and Scott R. Klemmer. "PeerStudio: Rapid Peer Feedback Emphasizes Revision and Improves Performance". In: *Proceedings of the Second (2015) ACM Conference on Learning @ Scale*. L@S '15. Vancouver, BC, Canada: ACM, 2015, pp. 75–84. ISBN: 978-1-4503-3411-2. DOI: 10.1145/2724660.2724670. URL: http://doi.acm.org/10.1145/2724660.2724670.

[15] Walter S. Lasecki et al. "Chorus: A Crowd-powered Conversational Assistant". In: *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*. UIST '13. St. Andrews, Scotland, United Kingdom: ACM, 2013, pp. 151–162. ISBN: 978-1-4503-2268-3. DOI: 10.1145/2501988.2502057. URL: http://doi.acm.org/10.1145/2501988.2502057.

[16] Min Kyung Lee et al. “Working with Machines: The Impact of Algorithmic and Data–Driven Management on Human Workers”. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI ’15. New York, NY, USA: ACM, 2015, pp. 1603–1612. ISBN: 978–1-4503–3145–6. DOI: 10.1145/2702123.2702548. URL: http://doi.acm.org/10.1145/2702123.2702548.

[17] Kurt Luther et al. “CrowdCrit: Crowdsourcing and Aggregating Visual Design Critique”. In: *Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. CSCW Companion ’14. Baltimore, Maryland, USA: ACM, 2014, pp. 21–24. ISBN: 978-1-4503-2541-7. DOI: 10.1145/2556420.2556788. URL: http://doi.acm.org/10.1145/2556420.2556788.

[18] Eleanor A. Maguire, Rory Nannery, and Hugo J. Spiers. “Navigation around London by a taxi driver with bilateral hippocampal lesions”. In: *Brain* 129.11 (2006), pp. 2894–2907. ISSN: 0006-8950. DOI: 10.1093/brain/awl286. eprint: http://brain.oxfordjournals.org/content/129/11/2894.pdf. URL: http://brain.oxfordjournals.org/content/129/11/2894.

[19] Eleanor A. Maguire et al. “Navigation-related structural change in the hippocampi of taxi drivers”. In: *Proceedings of the National Academy of Sciences* 97.8 (2000), pp. 4398–4403. DOI: 10.1073/pnas.070039597. eprint: http://www.pnas.org/content/97/8/4398.full.pdf. URL: http://www.pnas.org/content/97/8/4398.abstract.

[20] Sam Mavandadi et al. “Distributed medical image analysis and diagnosis through crowd-sourced games: a malaria case study”. In: *PloS one* 7.5 (2012), e37245.

[21] Brian McInnis et al. “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. New York, NY, USA: ACM, 2016, pp. 2271–2282. ISBN: 978–1-4503–3362–7. DOI: 10.1145/2858036.2858539. URL: http://doi.acm.org/10.1145/2858036.2858539.

[22] Daniela Retelny et al. “Expert Crowdsourcing with Flash Teams”. In: *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*. UIST ’14. New York, NY, USA: ACM, 2014, pp. 75–85. ISBN: 978–1-4503–3069–5. DOI: 10.1145/2642918.2647409. URL: http://doi.acm.org/10.1145/2642918.2647409.

[23] Jacob August Riis. *How the other half lives: Studies among the tenements of New York*. Penguin, 1901.

[24] Lowell J. Satre. “After the Match Girls’ Strike: Bryant and May in the 1890s”. In: *Victorian Studies* 26.1 (1982), pp. 7–31. ISSN: 00425222, 15272052. URL: http://www.jstor.org/stable/3827491.

[25] Thiago H Silva et al. “Traffic condition is more than colored lines on a map: characterization of waze alerts”. In: *International Conference on Social Informatics*. Springer. 2013, pp. 309–318.

[26] Walter Skok. “Knowledge Management: London Taxi Cabs Case Study”. In: *Proceedings of the 1999 ACM SIGCPR Conference on Computer Personnel Research*. SIGCPR ’99. New Orleans, Louisiana, USA: ACM, 1999, pp. 94–101. ISBN: 1-58113-063-5. DOI: 10.1145/299513.299625. URL: http://doi.acm.org/10.1145/299513.299625.

[27] Walter Skok. “Managing knowledge within the London taxi cab service”. In: *Knowledge and Process Management* 7.4 (2000), p. 224.

[28] Katherine Woollett and Eleanor A Maguire. “Acquiring “the Knowledge” of London’s layout drives structural brain changes”. In: *Current biology* 21.24 (2011), pp. 2109–2114.

[29] Katherine Woollett, Hugo J. Spiers, and Eleanor A. Maguire. “Talent in the taxi: a model system for exploring expertise”. In: *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 364.1522 (2009), pp. 1407–1416. ISSN: 0962-8436. DOI: 10.1098/rstb.2008.0288. eprint: http://rstb.royalsocietypublishing.org/content/364/1522/1407.full.pdf. URL: http://rstb.royalsocietypublishing.org/content/364/1522/1407.

[30] Donald E Wray. “Marginal men of industry: The foremen”. In: *American Journal of Sociology* (1949), pp. 298–301.

[31] Anbang Xu, Shih-Wen Huang, and Brian Bailey. “Voyant: Generating Structured Feedback on Visual Designs Using a Crowd of Non-experts”. In: *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. CSCW ’14. Baltimore, Maryland, USA: ACM, 2014, pp. 1433–1444. ISBN: 978-1-4503-2540-0. DOI: 10.1145/2531602.2531604. URL: http://doi.acm.org/10.1145/2531602.2531604.

[32] Alvin Yuan et al. “Almost an Expert: The Effects of Rubrics and Expertise on Perceived Value of Crowdsourced Design Critiques”. In: *Proceedings of the 19th ACM Conference on Computer–Supported Cooperative Work & Social Computing*. CSCW ’16. New York, NY, USA: ACM, 2016, pp. 1005–1017. ISBN: 978–1-4503–3592–8. DOI: 10.1145/2818048.2819953. URL: http://doi.acm.org/10.1145/2818048.2819953.