

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 [11]. 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 [12, 18, 10].

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 [11, 21, 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 [21].

Piecework's perspective. [MSB: needs a transition to set the tone here. e.g., The piecework literature also examined the complexity limits possible with piecework.] **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 [5]. Returning to Hart's reflections on piecework's limitations, we recall the multidimensional problem — tasks comprising of numerous, sometimes conflicting, goals [8]. It would be reasonable, then, to infer that work like this — reasonably highly skilled work where quality is difficult to assess — would be unsuitable for piecework. [MSB: that conclusion came way too fast. give a little detail on the railway. how did they come to that conclusion?]

[MSB: this paragraph needs a topic sentence. it's a wandering paragraph currently. is the claim here that evaluation is the limiting factor?] **Hart and Graves, without acknowledging one another, seemingly corroborate one another's conclusions at different levels of observation.** Graves enumerates some of the roles required to facilitate piecework in the early 20th century: "... piecework clerks, inspectors, and 'experts'..." [8, 5]. Graves and Hart may seem to be making differing claims about the limitations of piecework, but we argue that Graves is simply making a more concrete observation illustrating the insight that Hart later makes. [MSB: what insight? I've lost the thread here.] Graves 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 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.

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. As Hart and Roberts described, the 1930s saw a flourishing of clever piecework job design out of necessity due to the fact that it was infeasible to provide new workers with the comprehensive education that was familiar to men [7]. This constraint led to much more tightly scoped work, and (perhaps surprisingly at the time) more efficient allocations of workers, who could now 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 [6]. [MSB: so what are we to take out of this? that there is no limit? just our creativity?]

Piecework researchers also make claims regarding the organizations that benefit from piecework in the first place. [MSB: I think a reader wouldn't understand why that claim is relevant. Let's invert it, something like how the researchers point out that only certain kinds of organizations can effectively make use of piecework.] Brown discusses the factors necessary for piecework to thrive: "... incentive pay is less likely in jobs with a variety of duties than in jobs with a narrow set of routinized duties" [3] [MSB: that's confusing, is incentive pay the same as piecework? can you cut the first part of the quote and give us the noun instead of forcing us to figure it out?]. 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" [5]. Together, Graves and Brown make a persuasive argument that piecework is limited in complexity by managerial overhead [MSB: wait, wasn't that in a previous paragraph? that should be joined with the managerial overhead text above] and the fixed cost of adopting a piecework payment regime [MSB: that seems like it's the new point here].

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" [2]. 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 [28]. [MSB: Currently unclear what I'm supposed to take out of this. Summarize for me. What does this say about the complexity limits of piecework?]

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.] **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 [3] [MSB: I don't recall this point in the prior section: I remember management overhead and fixed costs of materials, I don't remember a paragraph about complexity. If we want to draw on that point, make it in the earlier section]. 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 for two reasons: 1) Technology make it much easier to do complex work aided by computers; and 2) The Internet allows us to leverage the benefits of “economies of scale” at very little cost to the system-designer [14, 20] [MSB: I don't understand that argument. What is scaling here? The people? The tools for complex work?].

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” [30, 29, 15]. 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 5]. 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 [13, 5].

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 [17, 16, 24, 25, 27, 26]. 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. . . [23, 9]. [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 [13]. 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 4] and develop more considerate platforms as McInnis et al. advocate [19].

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 [7, 6] [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” [22]. [MSB: how will that reduce any of the problems except the last one?]