MAJOR RESEARCH QUESTIONS

[al2: 4 paragraphs on what happened in the above contexts; 2 paragraphs on best/worst outcomes & why; ??? paragraphs on predictions for crowdsourcing]

What are the limits of crowdsourcing?

For more than a decade research in crowd work has explored ways of expanding the work that crowdsourcing can do by identifying challenges to this form of labor, overcoming them through novel designs of work–flows and processes, and repeating the process [4, etc. . .]. The question that has emerged among these researchers and through the work that they have produced then has been whether there are limits to crowdsourcing (and, if so, what bounds them).

In the field of piecework, the research covering this topic has both explored a breadth of tasks that might be rendered doable by piecemeal workers *as well as* longitudinally documented the success of these approaches.

Points to make:

- Blossoming of piecework
    - high point: consultants?
    - most cases: auto workers/etc. . .
    - worst case: sweatshops (especially in developing nations)
- what led to these outcomes?
    - "consultant" work came out well because the work was complex; this made it difficult to turn into a mass market commodity. We see consultants ranging skill levels like oDesk (implementing modules) to Accenture (on–demand teams of consultants).
    - auto workers, working in settings where capital couldn't be moved as easily found themselves in the same workspace as a direct — if multi–stepped — result to the benefits of putting people in the same place to consolidate resources. Moreover, workers had leverage over factory owners as a result of that consolidated capital; operating equipment required training.
    - sweatshop workers fared the worst, for reasons that may seem obvious with hindsight. Source materials for textiles are easier to ship than mechanical components such as engines, making it easier for factories to relocate to developing nations (where cost–of–living, and consequently wages, would be lower). As wages, Co-LAs, and QoL rise, workers begin asking for (and later demanding) higher wages and better conditions. But where the auto workers have leverage, textile workers find only a precarious economic balance now tipped by their collective action, spurring manufacturers to move to a new locale

What forms of work design and worker management are viable?

- researchers have looked at how to increase worker productivity (e.g. finding the maximal speed at which gig workers can be expected to work before making errors) [8].
- we've also seen people "embrace error" [30].

- but scholarship looking into the design and management of work and workers isn't new; lots of research into getting pieceworkers to do work more quickly [48].
- Researchers have even asked the age old question of *what motivates* pieceworkers (echoing similar research on Wikipedia and Mechanical Turk) [45, 38, 27]

What will work and the place of work look like for workers?

The metaphorical mechanics of these dynamics are still at play; workers and managers continue to interact in adversarial manners, despite substantive work into aligning the motivations of workers and requesters

The existing body of research has shed light on on–demand labor from various perspectives, and revealed a number of topics that, through our framing, are clearly situated together. Those topics are, at a high level, as follows:

1. the **processes** involved in making work into tasks, or discretization;
2. the outcomes (and indeed the **fallout**) of that discretization, both on the work itself as well as the workers; and finally
3. the **relationships** between workers and requesters of the work — both *cooperative* and *adversarial* cases.

The Processes of Making Gig(s) Work

The HCI community is perhaps most familiar with examples of task–based work such as 99designs, Amazon Mechanical Turk (AMT), and increasingly Uber & Lyft, which all allow requesters in various forms to tap into resources such as cars, computers, and above all "cognitive surplus" with relative ease [21, 13, 25, 50]. This insight, that workers can be geographically distributed and tasks decomposed, has proven remarkably compelling and an effective fulcrum for leveraging the Internet for highly scalable work [7, 30].

This section will largely discuss the processes at work that make distributed, digitally managed work both possible and indeed preferable for "requesters" (in other words, the employers who solicit workers). This body of research spans a broad field within the CSCW and broader HCI community. In this context, we'll look at the body of research through the lens of highlighting contributions which expand what we (as "requesters" of work) can do by managing workers in novel ways. This work broadly consists of three areas: 1) Decomposition, 2) Work Abstraction, and 3) Flexibility. As we explore this work, we'll attempt to relate the advances in the design of crowd work to the research contributions made in the research of assembly line manufacturing, Taylorism, and scientific management during the 20th century [22].

Decomposition

Piecework may be thought of as vertically slicing work such that each person is responsible for the whole task — making a whole garment, in this case. Broken down in this way, work could grow to unprecedented scales, but the quality of the work would remain relatively variable [36]. Textile work being a salient example, it took time for workers to acquire sufficient skill to do every aspect of the work so that the garment would be accepted by the company soliciting that work [61].

A compelling solution emerged in the early 20th century to break tasks down into discrete, manageable routines that could be taught relatively easily, and whose work output could be evaluated in abstraction from the rest of the work [3]. In Ford's assembly line, this meant that workers were not responsible for building a whole car, but a single very narrowly defined action that needed to be done on every car [34]. By the mid–20th century, Schoenberger writes, "... the intensification of the labor process is argued to have hit mental, physical, and social limits." [47].

This approach, "Fordism" (and its better–known contemporary "Taylorism" of similar ethos), can be seen today in crowd work and on–demand labor through the application of micro–tasks.

Teevan, Iqbal, and Veh highlight some of the advantages of breaking work into pieces, facilitating evaluation and parallelization [54]. By decomposing and recomposing tasks, and in particular by assigning similarly natured work to the same workers, workers could become "experts" in a small aspect of the work that they did, speeding their work dramatically [31]. Perhaps more important, however, was that the breaking down of work into tasks has made it more practical to evaluate work at each stage [43].

Scholarship describing and exploring the decomposition of tasks is perhaps the most established of the above areas among HCI researchers; Kittur et al. specifically drive at this goal by addressing the possibility of "crowdsourcing complex work" [28]. Cheng et al. found that microtasks — though not necessarily *faster* than "macrotasks" — yield higher quality work, particularly when that work might be readily interrupted [9]. Teevan, Liebling, and Lasecki further push the boundaries of decomposed work by exploring "selfsourcing", and further this work with Teevan et al. [55, 56]. While this work doesn't strictly fall under "crowdsourcing", the major contributions here seem uncontroversially to be inspired by the design of crowd work.

Much of the research in the space of designing crowd work has sought to illustrate the potential to take highly creative or skilled work and generate high–quality results. Perhaps the most notable case study here can be found in Retelny et al.'s *Foundry*, which employed "flash teams" to achieve expert–level outcomes via thoughtful decomposition of work as "modular tasks" [41].

Work decomposition, then, is far from new; "decomposition" generally illustrates the same concepts of work that "Taylorism" and scientific management sought to embody — Silberman, Irani, and Ross in particular foresaw this danger and warned of it in 2010 [29, 51, 37]. In both the historical and contemporary cases of decomposed work, work was, at least initially, distributed in the form of tasks to the homes of workers; Riis captured this in his documentary work *How the other half lives: Studies among the tenements of New York* in 1901 [42].

Work Abstraction

Decomposition allows requesters to assign tasks without concern for the broader context. While we'll discuss this aspect of crowd work more critically later, it's worth pointing out that discrete blocks of work containing all the relevant context for a worker allows workers to engage with virtually any component of work without worrying that their lack of higher–level awareness of the goals of the requester might negatively affect their work.

Chilton et al. perhaps best illustrated this with *Cascade* by demonstrating that it's possible to break certain classes of tasks apart in such a way that they yield taxonomies of various subjects, a task previously thought to be safely within the domain of expert workers with top–down awareness of the context of the work as a whole [10]. Verroios and Bernstein further illustrate this potential by forming a task one might consider highly contextually dependent — summarizing the

contents of a movie — in such a way that crowd workers could contribute small pieces of work without needing to know the content of the rest of the project [60].

Here, Hu's work, saying of assembly line work that "it is assumed that men are of equal ability and every man can do any of the *n* jobs", parallels the approach that dominated early research into crowd work — namely, using non–expert crowds for complex work [22]. This mindset in Hu's analysis, and indeed the study of factory and mass manufacturing labor through the 20th century, substantively owes its existence to scientific management and the rigorous decomposition of work into tasks, discussed earlier, and persists to this day as it colors researchers' goals and objectives in the study and design of crowd work.

Piecework's influence on the abstraction of work into tasks, described above, is more than just caused by the decomposition of work; work abstraction itself makes it possible for workers to come and go flexibly, prompting work requesters to consider ways to design these now discrete tasks in ways that maximize flexibility, both by allowing (and even anticipating) some inconsistency in worker availability *and* allowing and anticipating some inconsistency in the quality of the work output itself. It's to this area that we now turn our attention.

Flexibility
Earlier we discussed Cheng et al.'s work measuring the impact that interruption has on worker performance. This work both points to and embodies a broader sentiment in both the study and practice of crowd work that microtasks should be designed resiliently against the variability of workers, fully exploiting the abstracted nature of each piece of work [23, 31, 59]. That is to say, micro–tasks should be designed such that a single worker's poor performance, or a good worker's sudden departure, would not significantly impact the agenda of the work as a whole. While Cheng et al. identified costs with breaking tasks into smaller components in the form of higher cumulative time to complete (albeit much shorter real time to complete, owing to parallelization), Lasecki et al. found that at least *some* performance can be recouped by stringing similar tasks together.

Given the importance of consistent work results, one might intuit that requesters would prefer high–quality workers who can be relied upon to be available (even for contextually independent tasks), which would appear to contradict the benefits of flexibility already discussed; requesters have thus made significant headway toward "embracing error" to allow requesters to maximize the benefits of a flexible, even transient, workforce.

Krishna et al. offer orders–of–magnitude improvements in various binary classification tasks on the principle that diverse workers complete these tasks in order to accurately inform the model on the variety of delays in response times. And rather than building tasks to *tolerate* worker drop–off and attrition, some researchers have designed work predicated on the expectation of this phenomenon: Celis et al. describe ways of assigning tasks in such a way that crowd workers would never be given enough information to piece together sensitive information about any single topic [7].

Flexibility has been explored through the lens of Fordism, perhaps best illustrated by Tolliday and Zeitlin's treatment describing turnover rates rising above 300% in the decade leading to the introduction of the assembly line in 1913. Specifically, the utilization of "... 'semi–special' machine tools which could be adapted [and] ... added flexibility through seasonal layoffs for production workers and the use of piece rates ... rather than a day wage system" [57].

The Fallout of Crowd Work
Irani and Silberman point out the disillusion that companies such as Amazon foster on platforms for work like AMT (see also Salehi et al.'s work continuing in the spirit of this observation to generate collective action to improve worker conditions) [26, 46]. Lee et al. find similarly that workers on gig work platforms are frustrated by the systems on which they work, to say little of the policies which these systems enforce [32].

We discussed the benefits of flexibility (both in the sense of having arbitrary workers perform tasks and in the sense that we can design tasks to be more resilient to poor work) in the previous section. It's from that point in the literature that we turn our attention to the perhaps unintended effects of crowd work and the affordances for transience that we build into this mode of work. We'll address two major areas of work under this subject: 1) Low Pay; and 2) Variable quality work.

Low Pay
Horton and Chilton identified problems with crowd work wages relatively early on, attempting to address this imbalance from a behavioral economic perspective — that is, identifying and presenting a model that describes a worker's "*reservation wage*" [20]. This work has largely informed much of the research into and practice of estimating crowd work compensation [49, 39].

But we turn to Irani and Silberman's discussion of "*Turkopticon*", a system they designed to interrogate worker invisibility and to promote better wages across several dimensions [26]. Of particular relevance here, Irani and Silberman call to attention that "Turkers" are ultimately vulnerable to wage theft and pay rates that translate to well under minimum wage. Returning to Horton and Chilton, we find that the median "reservation wage" in 2010 was $1.38, while the mean was $3.63 [20].

Understanding workers' motivations given these conditions has thus become a goal for some researchers [5]. Sun, Wang, and Peng conclude that "... solvers participate in online tasks not only for money but also for enjoyment or the sense of self–worth" [53]. This might have rung true in 2011, and certainly corroborates Ross et al.'s findings after investigating "who are the crowdworkers", but as Silberman points out "we [have since] learned that most tasks on AMT are done by a small group of professional Turkers..." [44, 52].

Now, Irani and Silberman and later Salehi et al. cite insufficient pay as a central point of frustration among workers, via Irani and Cushing's contributions in this space [46, 24, 12, 26].

On–demand workers were not the first to be exploited along the dimension of low pay rates. Frustration over low (and declining) pay was one of the chief grievances among then

nascent British labor unions in the early 20th century [58]. This, Ebbinghaus and Visser argued, fueled the rocketing union membership rates through the mid–20th century until 1980 (to which we'll return when we discuss Levi et al.'s reexamination of labor unions) [14, 33]. This realization has similarly fueled a body of research into the various incentive structures available to piecework employers [45].

The parallels between the complaints of low pay among crowd workers and other on–demand workers and the pieceworkers and later factory workers in the 20th century are inescapable. We argue further that the *causes* here — work decomposition, work abstraction, and flexibility — lead inexorably to low and declining pay for workers. Moreover, we point out that low pay leads to other negative outcomes both in on–demand work as well as in piecework and on assembly lines.

Variable quality work

Researchers have struggled with what we might generously call work of "variable quality" along two dimensions. The first, to use the characterization of one of these contributions, we can call "understanding malicious behavior" [16]. While some work has cast workers as "malicious" or at least adversarial parties, the evidence thus far suggests that workers behave in unexpected ways as they attempt to assert some control over their interaction with the system (a topic of discussion to which we'll return later) [32]. The second dimension of research in this space generally attempts to eke out the highest quality work possible from workers given the apparent difficulty in predicting work outcomes [30].

The effect low wages have had on piece work and factory workers is well–known; Gantt discuss this exact mechanism in his book on *Work, wages, and profits*, pointing out that "... where there is no union, the class wage is practically gauged by the wages the poor workman will accept, and the good workman soon becomes discouraged and *sets his pace by that of his less efficient neighbor*, with the result that the general tone of the shop is lowered" (emphasis added) [17].

This research is similar to, but subtly different from, the notion of the "market for 'lemons'" which Fort, Adda, and Cohen discuss; specifically, Akerlof's writing of a "market for 'lemons'" describes a marketplace where the quality of the product or service is unknown to the buyer [15, 1]. The effect of this *perceived* uncertainty is that the *actual* trustworthiness drops precipitously as all of the consistent, reliable, high–quality workers capable of leaving these markets do so, leaving only the ones who cannot or will not establish their trustworthiness.

Relationships Between Workers and Managers

Suffice it to say that poor pay and poor work are linked, and that we should not be surprised to find this relationship play out online as strongly as it does offline. But the poor treatment of workers by managers — both human and algorithmic — do more than affect the economic relationships between workers and employers. Here, then, we turn to examine this facet of on–demand work and how these dynamics strikingly replicate the relationships researchers in labor advocacy encountered in the study of piecework and factory work.

This topic can be condensed into two major areas: 1) external (scientific) management, and the evaluation of workers as functional modules; and 2) the consequential resistance workers express due to their perceived alienation and distance from managing forces.

External Management

We discussed Fordism and Taylorism earlier in our discussions of Decomposition and Flexibility, but here the core of these paradigmatic views — the scientific management of work — becomes relevant. We use "external" here instead of "scientific", however, to more broadly capture the disconnect between managers and workers. By describing it as thus, we can touch on the relationship that workers have with *researchers*, as well, even though that work is not strictly — or just not exclusively — of the same nature as the management and experience as when interacting with requesters.

First, intuitively, the variable–quality work we discussed previously has led to a large and growing body of research attempting to evaluate workers' performance and error rates across numerous dimensions; for example, Cheng, Teevan, and Bernstein explore the error rates of workers by operating on a sliding scale giving workers varying amounts of time to accomplish micro–tasks [8]. Irani and Silberman describe the treatment of workers as sorts of "human APIs" that can, importantly, be rigorously evaluated [25]. Gevins and Smith began to explore the neurophysiological effects of cognitively demanding tasks on workers, informing crowdsourcing research by suggesting the use of cognitive load assessments such as NASA Task Load Index surveys to evaluate workers pre and post–tasks [30, 8].

External management comes in other forms than scientific, as previously mentioned. Researchers in particular have noticed that their relationships with on–demand workers are, at the least, complex. Irani and Silberman point out that their relationships with Turkers are highly complex; specifically, their interactions with field sites in which they work as designers and mediators of change influence the relationships they have with Turkers [25].

The scientific management of pieceworkers has been well–studied under the umbrella of assembly line research, and even physiological study of pieceworkers closely resembles the research into cognitive loads and stress levels that we discussed among on–demand crowd workers [22, 6]. Even the complicated relationships between observers and workers themselves are not necessarily new; Riis's photodocumentary of pieceworkers has even been re–examined through an exercise asking crowd workers to photograph themselves for similar purposes as Riis's — to document and humanize an otherwise abstracted, invisible workforce [2, 26, 42].

Similarly, Pollard's words on the punishment factory workers faced — for example, that "unsatisfactory work was punished ... by fines or by dismissal" — seems especially relevant given the fears we now know to be ubiquitous on platforms such as AMT, Uber, and other on–demand markets [40, 32, 46, 26, 35].

Resistance

It shouldn't surprise us, then, that workers have resisted the management imposed on them both by other people and their systems, often without recourse or opportunity for feedback, let alone substantive input. Indeed, Lee et al. discover of Uber drivers that many toggle their availability to avoid being dispatched to more distant locations, resisting the intent of the designers of the systems and their "algorithmic and data–driven management" [32].

Resistance has sometimes been more coordinated, as well; we see this in Irani and Silberman's coverage on *Turkopticon* as workers collectively accumulated information about requesters, and in Salehi et al.'s work on *Dynamo*, which generated "Guidelines for Academic Requesters" written by crowd workers [26, 46].

Resistance against managers in piecework and factory labor settings are deeply well–explored, but perhaps the most relevant case study to draw on here is to be found in Waldinger et al.'s case study of "Justice for Janitors", where marginalized workers managed to raise awareness for their plight and secure support for badly needed reforms [62]. The achievements of labor advocacy groups such as labor unions as resistant, even adversarial organizations counter–balancing the management is somewhat well–understood [19, 11]. We argue that these threads of resistance against management in various forms are in fact one.