Working with Machines: The Impact of Algorithmic and Data-Driven Management on Human Workers

Min Kyung Lee[1], Daniel Kusbit[1], Evan Metsky[1], Laura Dabbish[1, 2]
[1]Human-Computer Interaction Institute, [2]Heinz College
Carnegie Mellon University
{mklee, dkusbit, emetsky, dabbish}@cmu.edu

ABSTRACT
Software algorithms are changing how people work in an ever-growing number of fields, managing distributed human workers at a large scale. In these work settings, human jobs are assigned, optimized, and evaluated through algorithms and tracked data. We explored the impact of this algorithmic, data-driven management on human workers and work practices in the context of Uber and Lyft, new ridesharing services. Our findings from a qualitative study describe how drivers responded when algorithms assigned work, provided informational support, and evaluated their performance, and how drivers used online forums to socially make sense of the algorithm features. Implications and future work are discussed.

Author Keywords
Algorithm; algorithmic management; human-centered algorithms; intelligent systems; CSCW; on-demand work; sharing economies; data-driven metrics; work assignment; performance evaluation; dynamic pricing; sensemaking.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
Increasingly, software algorithms allocate, optimize, and evaluate work of diverse populations ranging from traditional workers such as subway engineers [16], warehouse workers [28], Starbucks baristas [19], and UPS deliverymen [7] to new crowd-sourced workers in platforms like Uber, TaskRabbit, and Amazon mTurk [13]. How do human workers respond to these algorithms taking roles that human managers used to play?

We call software algorithms that assume managerial functions and surrounding institutional devices that support algorithms in practice *algorithmic management*. Algorithmic management allows companies to oversee myriads of workers in an optimized manner at a large scale, but its impact on human workers and work practices has been largely unexplored. In recent years, the press and many scholars have brought attention to the importance of studying the sociotechnical aspects of algorithms [2, 10, 37], yet to our knowledge, there has been little empirical work in this area.

We explored the impact of algorithmic management in the context of new ridesharing services Uber and Lyft. Algorithmic management is one of the core innovations that enables these services. Independent, distributed drivers with their own cars are algorithmically matched with passengers within seconds or minutes, and the fare dynamically changes based on where passenger demand surges, all through the app on their mobile phones. Drivers' performance is evaluated by passengers' rating of their service quality and drivers' level of cooperation with algorithmic assignment. Algorithmic management allows a few human managers in each city to oversee hundreds and thousands of drivers on a global scale. Drivers have little direct contact with company representatives, but can interact with each other through online forums to gain social knowledge of the rideshare systems. This setting allowed us to explore the practices that emerged when algorithms assigned work, optimized work behavior through information processing and evaluated job performance. Do human workers cooperate with algorithmically-assigned work? How much are people motivated or demotivated by algorithmic optimization? How effective is algorithmic, data-driven evaluation and how do workers feel about it?

As a first step toward answering these questions, we interviewed 21 drivers with Uber and Lyft and triangulated their experiences by interviewing 12 passengers and conducting archival analysis of online driver forums and official company communication materials. The findings highlight opportunities and challenges in designing human-centered algorithmic work assignment, information, and evaluation as well as the importance of supporting social sensemaking around algorithmic systems. We use the findings to discuss how algorithms and data-driven management should be designed to create a better workplace with intelligent machines, offering implications for future work.

Our study makes the following two contributions to human-computer interaction (HCI): 1) we describe the upside and

CHI 2015, April 18 - 23, 2015, Seoul, Republic of Korea

ACM 978-1-4503-3145-6/15/04...$15.00
http://dx.doi.org/10.1145/2702123.2702548

downside of algorithmic and data-driven management, its impact on human workers, and sensemaking strategies that workers developed; 2) our results highlight new areas of research for HCI, computer-supported cooperative work (CSCW) and intelligent systems.

IMPACT OF MACHINES ON WORK

Two threads of research are closely relevant to the topic of this paper: research on the impact of technology in the workplace and interaction with intelligent machines.

The impact of technology in the workplace

A long and rich stream of research in HCI and CSCW has investigated the impact of technology in the workplace: email [15], instant messaging [18], organizational information repositories [1], groupware [31], video conferencing [9], awareness technologies [12], desktop office software [8, 39], GPS for taxies [11], and robots in hospitals [30]. Collectively, this research shows how psychological, social, and organizational factors shape the adoption of new technologies and how new work practices and norms emerge in the workplace. To our knowledge, very little empirical research has investigated the impact of algorithmic management in the workplace, with the exception of emerging studies of new work-monitoring devices such as GPS for bus drivers [34].

Interaction with intelligent machines

Much research has investigated how people respond to intelligent machines such as automated manufacturing technologies [23, 36, 44], recommender systems [6], context-aware systems [3, 43], agents [35], and robots [30]. As intelligent machines are only recently being integrated into workplaces outside of factories, there is relatively little work that examines intelligent machines in work contexts with exceptions being systems in hospitals [30] and offices [22]. To our knowledge, our paper is one of the first studies that investigates how people respond when intelligent machines take on managerial roles in a workplace. Research on intelligent systems with different roles in other contexts such as home, entertainment, or education has identified important theoretical concepts and design principles for successful human interaction with automated and intelligent machines: establishing trust and cooperation [33], creating accurate mental models [21, 35], providing transparency and explanations [3, 6, 20, 24], and designing shared control between humans and intelligent machines [23, 36]. Our study significantly extends this work by probing the consequences of these design dimensions of intelligent systems in a new context, and by identifying new problems and implications arising from novel managerial roles of intelligent systems.

METHOD

We conducted a qualitative study on algorithmic management in Uber and Lyft. To understand drivers' experiences and perspectives, we interviewed 21 drivers and triangulated our findings by interviewing 12 passengers and analyzing 128 posts by drivers on online forums and 132 official blog posts and communication materials from both of the ridesharing companies.

Research context: Uber & Lyft ridesharing services

Uber and Lyft are currently the largest peer-to-peer ridesharing companies. Founded in 2009 and 2012 respectively, Uber and Lyft operate in over 100 cities in 37 countries. Lyft attempts to create a social culture among customers by encouraging passengers to sit in the front seat and greet the driver with a friendly "fistbump." Uber creates a more professional chauffeur environment where social experience with the driver is not stressed. Anyone over 21 years of age with a valid drivers license and a personal vehicle in good condition can apply to be a driver. Companies screen applicants with a background check, and new drivers go through brief video-based online orientations. Once accepted, new drivers become "independent contractors," not employees, and are in complete control over where, when, and how often to drive.

Algorithmic management in the ridesharing platforms

Three algorithmic features of Uber and Lyft – passenger-driver assignment, the dynamic display of surge-priced areas, and the data-driven evaluation that uses acceptance rates and ratings – respectively correspond to decisional, informational, and evaluation roles of human managers in organizations [29].

Work assignment: Driver-passenger assignment algorithms. Drivers need to turn on their ridesharing app to be able to receive and execute their work. According to Uber, after a rider requests a ride through the mobile application, "the closest driver to that rider automatically receives the trip request with a 15 second window to accept it [42]." (Uber and Lyft will not reveal details of proximity-based assignment algorithm, but in our study, we learned that other things could be factored into the algorithm.) The request includes information about the passenger's location, rating, picture, and name. If the driver accepts the request, the passenger is notified, and the driver drives to the passenger location to start the ride. With both Uber and Lyft, drivers cannot choose or set preferences for specific passengers or rides they wish to receive on their app. Lyft, however, does allow drivers to block assignments from passengers that do not pay the full suggested donation in areas where pricing is still donation based. Uber and Lyft allow only passive rejection of assigned passengers in that, if the driver does not wish to accept the request, they must wait out the allotted 15-second window. After this, the app goes into stand-by mode until there is a new request.

Informational support: Dynamic in-app display of surge-priced areas. Pricing is determined by a standard fare and fluctuates according to a dynamic pricing algorithm. The companies explain this feature to drivers and the public in broad terms. For example, "when demand outstrips supply, dynamic pricing algorithms increase prices to help the market reach equilibrium [41]." In this paper we will refer to this as "surge pricing," adopting the term that Uber uses

(Lyft uses the term "Prime Time"). Surge pricing can play a major role in shaping driver income, as their eighty percent commission remains constant through these periods of peak pricing. As of September 2014, both companies show surge-priced areas in-app with map areas shaded in different colors based on the price in real time. This is designed to motivate drivers to move to areas where demand (and price) is surging, in order to meet passenger demand and maximize the total number of transactions.

Performance evaluation: Rating systems and acceptance rates that track driver performance. After the ride both the passengers and drivers evaluate each other through a 5-star rating system. Lyft instructs that when rating a driver, passengers should "consider whether your driver was friendly, safe, a good navigator, and made you want to use Lyft again [25]." Drivers also have an acceptance rate that is calculated by the number of accepted rides divided by the total number of requests sent to the driver. Drivers are encouraged to keep a high ride acceptance rate through occasional promotions that offer a guaranteed hourly pay if the driver's acceptance rate is above a certain threshold.

Drivers with a low average passenger rating and acceptance rate may be subject to review or even immediate deactivation on the ridesharing platform. Likewise passengers who fall below a rating threshold risk rejection by drivers, as drivers have the ability to ignore incoming requests from passengers with ratings below their liking. Long-standing drivers who maintain a high passenger rating and acceptance rate are occasionally promoted to become mentors or recruiters. In addition to driving for the service, mentors and recruiters recruit new drivers and oversee the application process, while earning extra income for these activities.

Interviews

We conducted semi-structured interviews with drivers to understand their experience with algorithmic management. We also interviewed passengers to confirm drivers' opinions and perceptions about passenger behaviors expressed in their interviews.

Participant recruitment

Online postings and paper flyers were used to solicit current or former Uber or Lyft drivers and passengers. We posted ads on Facebook driver groups, volunteer sections of Craigslist, and relevant Reddit webpages. Paper flyers were posted in three major cities in the US. A $10 gift card was offered for participating in a 30-45 minute interview.

Driver interviews

We interviewed 21 drivers for ridesharing platforms who operated in 13 cities across the United States (15 Males; average age of 35 (SD=8.86)). Of the drivers interviewed, 5 drove for Uber only, 5 drove for Lyft only and 11 drove for both Uber and Lyft. The drivers worked an average of 23.5 hours (SD=21.4) per week and had a range of experience driving for a ridesharing platform (3 weeks to a year) with an average tenure of 149 days (SD=107). 19 of the drivers drove part time and 2 drove full time. Four of the 21 drivers were also working for, or had previous experience with similar driving services including taxis, trucks, and personal chauffeurs, and car services. We conducted additional 30-45 minute long interviews with these drivers to compare their experience with more traditional driving jobs to Uber and/or Lyft.

Interviews were done through video chat, phone, or in person, depending on the interviewee's location. The interviewer began with questions about the driver's last ride, best or worst assignment and ride experiences. Follow-up questions probed drivers' understanding of three algorithmic features and how their understanding influenced their work strategies. For the ridesharing drivers who also worked as taxi drivers, personal car service drivers, or chauffeurs, we asked them to compare work assignment and evaluation models in two driving jobs.

Passenger interviews

We interviewed 12 passengers who had used or currently use Uber and/or Lyft in 4 cities across US (8 Females, average age 24.2 years (SD=7.1)). On average, the passengers had used the services 2-3 times per month for about 5 months (SD=4.46). 8 of the passengers had used both Uber and Lyft, 3 only used Lyft, and 1 only used Uber. Interview questions focused on confirming or disconfirming drivers' perceptions of passengers' use of services, in particular, how they rate drivers and their attitudes toward and behaviors around surge pricing.

Archival data: online driver forums & company websites

Online driver forums. We analyzed postings on online driver forums as all drivers in our interviews mentioned that the forums were primary knowledge sources and places for socialization. We looked at two categories of online forums: groups unmoderated by Uber and Lyft, including various Facebook groups and Reddit pages, and official company Facebook "lounges." One author signed up to become a Lyft driver, and was given access to a "new driver" Facebook forum, hosted by Lyft, in which information was disseminated with direct relation to the company. We accessed other unmoderated private driver forums on Facebook by requesting to join as researchers, to avoid deceptively posing as drivers, and maintained an observation-only status. Following the approach used in [27], we sampled 128 online forum posts and comments mentioning the algorithmic features selected out of thousands made in a five-month period.

Company websites. We also looked at how companies officially educate and communicate with drivers in order to understand how much information they share about the underlying mechanisms of the platforms' algorithmic features. We analyzed information on the Uber and Lyft company websites and 132 official blog posts posted between 2012 and 2014.

Analysis

We triangulated our findings from interviews and archival data. We qualitatively analyzed [32, 38] interview transcripts, excerpts of online forum postings and comments, and company communication materials using Dedoose, a qualitative data coding software. We used three algorithmic features of the ridesharing services to divide the data set, and then open-coded the data about each feature at the sentence or paragraph level. We analyzed the rest of the data to identify important themes including social sensemaking and socialization. This resulted in a total of 372 concepts. We then categorized the concepts into 18 themes explaining emerging phenomena. In addition to the ones reported in the paper, themes such as employee (de)identification with company culture emerged but were excluded in further analysis. We focused on 8 categories relevant to our research questions around algorithmic management, and used modeling techniques and affinity diagrams in order to explain the relationship between the categories. The final coding scheme had good reliability across two coders when tested with 10% of the transcripts (Kappa=.71). Conflicts between coders were resolved through discussion.

FINDINGS

We describe how drivers responded when algorithms assigned work, provided informational support, and evaluated job performance, as well as how drivers used online forums to socially make sense of the algorithmic features of the system.

Background: driver motivation

According to drivers, one main advantage of working for a rideshare platform was the flexibility that the system affords in terms of where and when to work, and the low level of commitment that is required by signing up. Some individuals drove full-time, but many also drove for fun, out of curiosity, or on a part-time basis. Many drivers used the ridesharing app in collaboration with their own daily routine to earn extra income, turning the driver app on for the daily commute for example, or doing chores around the house while waiting for a ride request to come in. In addition to the added financial flexibility that rideshare work affords, many drivers we interviewed mentioned social motivations for rideshare driving. Several drivers, for example, weighed the fun of meeting and having conversations with new people and the desire to help out the community as greater than or equal to their motivation to earn extra income.

Algorithmic work assignment: proximity-based driver-passenger assignment

Our findings highlight how transparency of algorithmic assignment influences worker cooperation, work strategy and workaround creation, and describe the potential impacts of automating choices that workers used to have in similar work settings.

Accepting and cooperating with algorithmic assignments

Previous research suggests that people may cooperate less with work assignments made by machines rather than people [33]. In Uber and Lyft, the way that assignments were presented to drivers on their app and the regulation of acceptance rate cut-off influenced drivers to accept as many assignments made by the assignment algorithms as possible. *"I mean you can always decline to pick up a passenger if you can make that decision within 12 seconds. (Uber/Lyft) make it sort of difficult to say no for a couple of reasons. […] when they show the spot on the map where you're going to pick someone up its very zoomed in so if you're not immediately familiar with the area you probably wouldn't be able to discern within 12 seconds if its somewhere you want to go or not. They just tell you how far away it is in driving time (P4)."*

Interestingly, one factor that influenced driver cooperation was whether the assignment made sense to them. While the assignment was generally based on proximity, there were other factors that influenced assignment such as passenger-driver mutual rating and driver login time. This sometimes caused drivers to receive requests from distant passengers to which they were not the closest driver. When this happened, many drivers reported rejecting the unfavorable ride assignment given that they would have to drive a great distance (such as 15 minutes) to the pick-up location. For example, P23 stated: "*I'm one that keeps a close watch of where [other drivers] are when I'm not with a passenger. So if I'll see three [other drivers] over on the [area A] and I get a request from [area B] to the [area A] knowing that there should have been three [drivers] sitting right there ready to go, tells me one of two things happened. Either all three of them passed on the ride which is highly unlikely that they're sitting here and they pass on a ride that's right in front of them. Or the system didn't coordinate the GPS correctly and sent it to me over ten minutes away instead of somebody that was 30 seconds away."* In this quote, it is unclear whether the assignment was due to errors in the system, or for other legitimate reasons, because no explanation was given about how the assignment was decided on the drivers' app. P23 assumed that the assignment was made by mistake and rejected the assignment. On the other hand, even with distant and inconvenient requests, drivers accepted rides when the assignment made sense to them. For example, P13 stated: "*Distance wise, sometimes I've gone like 17 miles, but that's not really the [passenger's] fault; that's because there's just not that many drivers out right now and I just really am the closest."* This suggests that an explanation of why certain assignments were made might be an important, but currently missing feature.

Creating work strategies and workarounds for algorithmic assignments

Drivers used their understanding that assignments are based on proximity to create their work and workaround strategies that helped them maintain control that the automated

assignment did not support as part of the existing system functionality.

Drivers strategically controlled when and where to work and when to turn on the driver mode of the app to get the types of requests and clienteles that they preferred: limiting the area that they worked in by turning off the driver mode when returning from a long-ride, avoiding bad neighborhoods to avoid dangerous situations, going downtown for successive short-rides during the lunch time, not going to bar areas to avoid drunk passengers, and instead, staying in residential areas to drive people to bars. Drivers attracted repeat passengers by arranging rides via phone, asking passengers to request a ride once they were in the driver's car to get matched. Drivers used online forums to post about bad passengers so that other drivers could avoid them, similar to self-regulation strategies of mTurkers [17].

Drivers also distanced themselves from one another by checking other drivers' locations on the map so that they did not compete with each other for passenger requests. When drivers desired a break but did not want to turn off their driver applications to benefit from an hourly payment promotion, they parked in between the other ridesharing cars in order not to get any requests.

The companies communicated only general rules of assignment, e.g., "the closest drivers get assigned," and this general understanding helped drivers create their work strategies. The lack of details of the assignment algorithm, however, seemed to foster drivers' ambivalent, sometimes negative feelings toward the companies: *"Uber is very close lipped about what actually happens right I mean they say 'oh we route it to the closest driver' or whatever but who really knows what's going on behind the scene it's up to whoever engineers their iPhone app (P4).*"

More knowledge more advantage

Our findings suggest drivers benefited from deeper knowledge of the assignment algorithm. Drivers with more knowledge created workarounds to avoid undesirable assignments, whereas those with less knowledge rejected undesirable assignments, lowering their acceptance rating, or unwillingly fulfilled the uneconomical rides. For example, P2 had knowledge that Lyft's assignment algorithms take into consideration how long drivers have been online and that a driver's radius for pickups will increase as they wait for passenger assignments. He used his knowledge to periodically turn on and off his driver application while at traffic signals, so that he did not get distant requests. However, this information was not publicly made available to all drivers, and in our interviews, Lyft drivers who did not have this knowledge attributed the distant assignment to the error of the assignment system, or drivers with higher ratings getting priority. These drivers could not create workaround strategies to avoid distant requests.

Getting assigned versus choosing whom to pick up

Drivers were generally satisfied with their level of control over assignment algorithms, except for a few drivers who desired to have control over the radius that the assignment algorithm searches to assign the passenger. Interestingly, one driver P17 who was also a Yellow Cab driver preferred his Taxi dispatching system where he could see all the incoming requests and choose freely from among them. He explained that he could strategically choose the location of ride requests in the taxi system, and he had developed knowledge of how to best use them: *"At certain times of certain days, you know that they're usually a lot of really good trips happening in those areas. Like, Thursdays around four o'clock in, like, [area names] you know that there's gonna be a lot of airport trips, for instance. So you can focus on those. Another thing that's good about it is […] if it's […] a busy Friday night. You're just growing tired of mining a certain area. You can totally shift. And a good way to do that is to take something that's not in a close area that you think maybe going from far, but then coming back in. […] It gives us the option for a change of pace."* Uber and Lyft's automated assignment got rid of this fine level of control and predictability. He said that while he could try to be in those locations in Uber and Lyft's system, it did not guarantee that he would get requests in the area. Often, he would get requests outside of the area, and he did not want to drive to these areas just for a change of pace.

Algorithmic information support: dynamic in-app display of surge priced areas

Surge-pricing algorithms are used to optimize pricing in online, airline, and hotel markets, among others. Our findings show breakdowns when these algorithms are used to influence human behavior.

Algorithmic information not accommodating human abilities, emotion, and motivation

Some drivers in cities where surge pricing was applied citywide, instead of being neighborhood-based, reported that they would go out to drive when they received surge-pricing notifications. Other drivers reported that the times when they were available to drive were already in line with surge-priced timing.

More than half of interviewed drivers, however, were not influenced by surge pricing information as the supply-demand control algorithms failed to accommodate their abilities, emotion, and motivation. Surge pricing changed too rapidly and unexpectedly to utilize the information in a strategic way to boost their incomes. Surge areas were on and off, sometimes by the second, and being in the surge area did not guarantee requests from within the surge area. Drivers reported getting no requests or requests from outside of the surge area (which do not qualify for surge pricing), or the surge price disappearing when they arrived to the surge area.

The economic and rational assumptions of the supply-demand control algorithm did not always motivate drivers' behaviors, as it does not account for feelings of unfairness about dynamic surge pricing [14]. Most passengers reported that surge pricing was unfair, and they tried to avoid it if they could. Some drivers, in particular ones that used the ridesharing services as passengers, expressed that they thought that surge pricing was unfair and they did not try to chase surge-priced areas. The appeal of increased incomes in surge priced areas did not motivate some drivers who also drove primarily for social, rather than financial reasons.

Trusting their own knowledge more than algorithmic data

A couple of drivers had more trust in their own knowledge and experience driving in the city than in the unpredictable surge pricing algorithms. In part, the drivers did not have knowledge to evaluate how accurate surge-priced areas were. For example, P19 stated: *"They probably do have some kind of algorithm over people who open up the app to request the ride, and they might have noticed, but they don't tell us how those [surge-priced areas] work. I ignore them for the most part, because I'm from here. […] I've lived here 35 years."*

Algorithmic, data-driven evaluation: performance evaluation through acceptance rate and driver rating

The regulation of the acceptance rate threshold and the driver-passenger rating system offered many benefits to overall service functioning. However, these numeric systems that made drivers accountable for all interactions were sometimes seen as unfair and ineffective and created negative psychological feelings in drivers.

Unfairness in treating all assignment rejections equally

The regulation of the acceptance rate threshold encouraged drivers to accept most requests, enabling more passengers to get rides. Keeping the assignment acceptance rate high was important, placing pressure on drivers. For example, P13 stated in response to why he accepted a particular request: *"Because my acceptance rating has to be really high, and there's lots of pressure to do that. […] I had no reason not to accept it, so […] I did. Because if, you know, you miss those pings, it kind of really affects that rating and Lyft doesn't like that."*

Assignment algorithms penalized equally all drivers' rejections of passenger requests, which lowered the drivers' acceptance rates. Sometimes drivers, however, had legitimate reasons and circumstances for rejecting passengers. For example, female drivers did not accept male passengers without pictures at night because of safety concerns. Drivers often rejected passengers "blacklisted" for their misbehavior on online driver forums. Sometimes technical glitches in the app showed the request with only a few seconds left to accept. When they felt that they had legitimate reasons for rejecting the requests, drivers would sometimes send emails to company representatives, hoping that they would not get penalized for the legitimate assignment rejects, but often times they did not hear back from the companies.

Inaccuracy in using only numeric metrics of service quality

Our findings suggest the passenger-driver rating system established basic trust and service attitudes in the ridesharing systems, however they fell short when used for driver performance metrics.

Drivers used passenger ratings to decide whether to accept the request or not, trusting 5-star passengers and being cautious with 3-star passengers. While not paying equal attention to driver ratings, passengers reported that the existence of a rating system gave them a sense of security. The rating system also promoted a service mindset in all drivers. For example, P16 said: "*[…] I want to get all five's. So I try to be friendly and engaging with the passengers. And give them options when they get in, like, you know, 'Do you want A/C? Do you want the windows down? What kind of music do you want to listen to?' I even have a candy tray, gum, stuff like that."*

Drivers took their ratings seriously. High ratings such as 4.98 became a source of pride whereas a rating below 4.7 became a source of disappointment, frustration, and fear of losing their jobs. Being tracked, evaluated, and judged by each passenger seemed to have a negative psychological impact on drivers who did not have scores near 5. P11 said: "*[the rating system] makes you cautious that what you're doing is being judged and rated and if you're rated poorly enough over a period time then eventually the platform could ask you to stop driving for them."*

Many drivers felt that the average of the passengers' ratings of the drivers was not reflective of their driving performance and services, as P22 stated: *"It's like in baseball a stat line doesn't always tell the particulars of a player. A player could hit 35 homeruns and knock in 100 runs but if they're hitting .240 and strike out 150 times, that doesn't mean they were such a productive player."* Many reported that various physical and psychological states of passengers such as being in a hurry and late for a flight or being drunk, could influence them to give a lower rating after the ride. Additionally, drivers noticed that passengers misattributed system faults and negative experiences that drivers could not control to drivers themselves, which in turn resulted in lower ratings (e.g., surge pricing, traffic jams, GPS errors etc.). Drivers also often attributed their low rating to passengers reviewing them using inappropriate review rubrics. For example, drivers often perceived that passengers rated them as if it were a 5-star rating system for products, movies, or restaurants, where perfect ratings are rare. This led many drivers to conclude that passengers needed education for the rating system in ridesharing services. On the other hand, most passengers we interviewed reported that they are more lenient and positive when rating drivers, while they are more critical in their online reviews of other goods.

Because of their perception that there are many uncontrollable factors influencing driver ratings, drivers seemed to develop a detached, indifferent attitude once their scores were above a certain threshold of deactivation risk. The rating averages scores from one hundred or more rides. With this average, the impact of each ride is reduced. For example, P8 stated: "*Well I used to micromanage my rating so to speak. I used to sweat and be oh my gosh my rating is now going down - it's a 4.85 that kind of thing. Now I don't worry about it. I see there's a lot of error that can take place in the rating.*"

Online social sensemaking

As drivers worked independently in distributed locations, online driver forums became a primary avenue for the driver socialization and system sensemaking. Drivers discussed the workings of the ridesharing systems' algorithmic management. One of the successful online sensemaking examples was about improving and maintaining driver performance in ratings and acceptance rate. When novice drivers asked for tips to improve their ratings, other drivers shared strategies and lessons that they learned over time. For example, one driver posted questions about how to improve her rating (4.63) after giving 38 rides in three weeks. About 50 comments were made within two hours of her posting, empathizing with her feelings, disclosing common experiences of going through the initial hurdle, and sharing specific strategies that they developed – creating their own service information brochure for the backseat, going to downtown during the day for many short rides, etc. The experienced drivers also explained that the rating would stabilize over time, and advised that she should not stress about it too much. Often times, original question askers followed up on the posting, making comments that the strategies had worked.

On the other hand, sensemaking activities around assignment algorithms and surge pricing seemed less successful in terms of informational usefulness. Common posts were questions of how assignment algorithms and surge-pricing work, how to interpret dynamic visualization of surge pricing areas, and real-time questions of frustrating events—from getting no requests in surge priced areas or getting distant requests that required long driving.

When answers to drivers' questions went beyond the information that the company officially communicated, the social discussion on the online forums focused on providing social and emotional support, rather than informational support. For example, one driver posted frustration in real time that he just got an assignment from across the city from east to west even when he saw other drivers around the requesting passenger that were closer. Many comments were made to the posting that provided emotional support. For instance, another driver from the west chimed in to say "I should have logged in to save you (from driving from east to west)," and other drivers said "it sucks when it happens." But none of the comments provided an explanation as to why such an assignment was made. Postings that made jokes on the surge pricing areas that seemed wrong according to common sense (e.g. a surge area extending into a lake), or ones that humorously interpreted shapes of surge-priced areas as "Surgemon," also reflect an attempt to have control over the unknown through humor and emotional processes instead of rational, cognitive ones [40]. On rare occasions, company representatives came to the forums and answered drivers' questions, but their answers were usually washed away in the influx of other forum postings and comments.

DISCUSSION

We discuss how to use our findings to improve the design of algorithmic and data-driven management.

Designing algorithmic work assignment

Algorithmic passenger assignment in Uber and Lyft automatically distributes myriad ride requests to drivers in a matter of seconds. Drivers' quick and frequent acceptance of the assignments ensures the efficiency of the service, maximizing the number of passengers able to get a prompt ride. Our findings suggest that in algorithmic work assignment, not only the source of the assignment (i.e., human versus algorithm), but also how the assignment was presented and regulated, influences worker cooperation with the assignment. Choices of which information to present on screen, a short time limit to accept the ride, and the acceptance rate collectively reinforced drivers' cooperation with the assignments in our study.

Our findings also suggest that transparency of assignment process could elicit greater cooperation with assignments, especially undesirable ones. While the company explained that their assignments are based on proximity, there were various additional factors that the algorithm took into account in addition to passenger-driver distance. This sometimes resulted in assignments where passengers were not assigned to the closest driver. Providing explanations for [6], or allowing workers to ask questions about [20, 24] each assignment could reduce drivers rejection of distant assignments by reducing their attribution of such assignments to technical errors. Transparency may also improve some drivers' ambivalent feelings toward the companies. This finding is consistent with previous research on recommender systems where transparency improved people's trust and acceptance of recommendations [6]. Moreover, the study highlights new implications of transparency, which have received relatively little attention in previous research on intelligent systems: how transparency in algorithmic assignment helps people create better work strategies and workarounds. Drivers with more detailed knowledge about the assignment algorithm could create workarounds to avoid less economical rides whereas people only with a general understanding of proximity-based assignment could not.

The stakeholders involved with work platform apps (companies and workers) complicate providing

transparency. In previous work on transparency of intelligent systems, explaining a user model usually sufficed [3, 6, 20, 24]. Algorithmic work assignment offers new challenges in design transparency where fully disclosing the algorithm may not be a viable solution. Companies may be unwilling or unable to share the underlying mechanisms of their assignment algorithms, as they might be patented or proprietary assets. Companies may also desire a degree of user ignorance to prevent the system from being gamed.

We were surprised that ridesharing drivers desired little (or did not feel entitled to) direct control over the assignment algorithm, for example, specifying pick-up locations or being able to see and choose from all requests. We believe the organizational context of being independent contractors played an important role: the flexibility and choices that the ridesharing drivers have in work compensate for the lack of control in assignment algorithms. Another explanation for why drivers did not desire control could be the lack of experience with other systems. For example P17, who worked as both a taxi and rideshare driver, preferred the taxi assignment process where he could directly access and choose passenger requests. P17 did not like the ridesharing assignment systems because algorithms made decisions that he used to make himself, making him feel he lost agency regarding strategies that he had developed to maximize his income. This could be interpreted as resistance to change, but also raises open-ended, ethical questions about the trend in new technology that sacrifices individual control for the sake of overall system efficiency, and its implications for learning and development on the job.

Designing algorithmic information support

Supply-demand control algorithms were originally designed to solve mathematical optimization problems that involve non-human entities. In Uber and Lyft, however, they are used to motivate and control human behaviors. This causes problems, as the supply-demand control algorithm does not consider the pace at which drivers work. Consistent with previous research on a smart agent that tried to encourage sustainable behaviors [5], the algorithm failed to account for feelings of inequity people had toward surge pricing, and ignored the social and altruistic motivations of drivers. This highlights the importance of making algorithmic management accommodate: a) the speed and the way humans work, b) diverse types of motivations rather than only economic ones, and c) emotions that people feel about the decisions that algorithms make. In addition, some drivers did not trust the surge-priced areas as they trusted in their experiences more. Transparency in how the surge-priced area was computed in real-time could improve workers' trust toward the algorithmic information.

Designing algorithmic, data-driven performance evaluation

Using driver ratings and acceptance rate, companies are able to evaluate drivers at a large scale. Driver ratings in particular may seem to be a legitimate evaluation metric because customer satisfaction is an important metric of service success and human service provider quality. Using only the tracked performance data in evaluating workers, however, revealed many complications that can occur when one relies too heavily on quantified metrics without deeper consideration of their meanings and nuances. Consistent with previous research on letter-grading systems or numeric evaluation of teaching skills [4], many random factors that are out of drivers' control influence the way passengers rate drivers. The efficacy and accuracy of averaged collective evaluation, rather than an in-depth holistic evaluation done by a human manager or peer, is also in question. As P18 put it, "*you are at the mercy of random people, in [his other work], you are evaluated by people that you know.*" Our study also shows drawbacks of adopting the 5-star rating system shared with online products, content, or business reviews to review human workers. Drivers felt that passengers rated conservatively as they do in online reviews; yet interviews with passengers suggest that they are being more lenient and positive than drivers think. This misconception suggests that a 5-star rating metaphor and rubric may have brought up inappropriate associations. Finally, the long-term motivational effect of the rating is also in question. As the drivers' ratings were averaged over multiple rides, the impact of one positive or negative ride was minimized, and drivers in our study became less sensitive to the changes in their ratings once they were above a minimum threshold.

Successful management provides work protocols and allows improvisation in response to changes and exceptions [29]. On the other hand, assignment algorithms penalized equally all driver rejections of assignments even when certain drivers had legitimate reasons and circumstances for doing so. While we did not observe serious problems from this lack of flexibility in algorithmic assignment in our study, it brings up an open challenge in creating flexibility in algorithmic management.

In most online rating systems, review is optional, and many even skip the process. In the ridesharing services, all passengers were encouraged to rate their service encounter, and most of them did. Being held accountable for every interaction, drivers were very aware of the existence of this external evaluation. Trying to deliver good services for all service interactions could pose psychological stress to workers. Additionally, as extensive research on the impact of extrinsic rewards on intrinsic motivation suggests, the external device could weaken the intrinsic motivation that drivers might have and change the meaning that they attribute to their behaviors. From the passengers' point of view, the ambiguities in providers' motivation for friendliness and good services risks rendering provider-client interaction more superficial and perfunctory.

Supporting social sensemaking online

Our study showed that online forums became a main place where drivers socialized, asked questions of each other, and

exchanged knowledge and strategies. In most research on sensemaking and mental models of intelligent systems, the focus has been on individual sensemaking [20, 21, 24, 35, 43]. Our study shows that social sensemaking is another important activity that needs to be better understood and supported for intelligent systems to be successfully adopted. Social sensemaking activities on the driver forums followed the form of "fragmented social sensemaking [26]" where there were many active contributors but no central authority figure to synthesize different ideas and narratives into a coherent story. This type of sensemaking was effective for discussing rating improvement strategies where there were no right or wrong answers, and workers' experiences and learned and improvised strategies played critical roles. On the other hand, fragmented social sensemaking fell short on subject matters where only an authority figure had the right information. This highlights opportunities to design structured online social sensemaking of algorithmic features where individuals can build on each other's knowledge.

LIMITATIONS

Like any study, this work has some limitations. Our results are from interviews with a small sample of drivers, passengers and archival data analysis. We could not interview developers or official representatives of the companies as it was against company policy. Our findings should be complemented by future research that uses different research methods such as ethnography, surveys or experiments. Our study was done in the specific context of ridesharing (on-demand, independent contractor work) thus further work needs to be done in different organizational contexts such as with full-time or co-located employees.

IMPLICATIONS

HCI and CSCW have a long history of research on computational systems that support individual work and collaboration. Our work suggests that algorithmic management in the workplace is a new and fruitful ground for research, where the same effort can be made to establish theories and design principles for algorithmic management. The research presented in this paper also offers implications for research on intelligent systems. Much of the previous work on intelligent systems was done in the context of individual user and non-work settings. Our work suggests that important concepts and theories in the field of intelligent systems such as transparency and control of systems, user mental models, and sensemaking need to be updated to accommodate social and organizational contexts that involve multiple stakeholders and new roles of intelligent systems in workflows. Finally, this research raises the need for new methodological research in HCI and interaction design on designing human-centered algorithmic management. HCI and interaction design have established systematic processes and methods for designing human-centered interfaces and interactions. Compared to designing and building traditional user interfaces, designing algorithmic management will require different ways of specifying and evaluating requirements, states, and interactivity.

CONCLUSION

Increasingly, software algorithms allocate, optimize, and evaluate work. In this paper, we explored the impact of this algorithmic, data-driven management in the context of new ride sharing services, Uber and Lyft. Our findings from a qualitative study highlight opportunities and challenges in designing human-centered algorithmic work assignment, information, and evaluation and the importance of supporting social sensemaking around the algorithmic system. The implications for HCI, CSCW, and research on intelligent systems are discussed. We hope this research inspires future work, so that we support human workers to work with intelligent machines not only in an effective, but also a satisfying and meaningful way.

ACKNOWLEDGEMENT

This research was supported by NSF grants CNS-1205539 and ACI-1322278. We thank Su Baykal for helping us collect and analyze data, and the anonymous reviewers for their feedback that improved the paper.

REFERENCES

1. Ackerman, M. S. (1998). Augmenting organizational memory: a field study of answer garden. *TOIS, 16*(3), 203-224.
2. Barocas, S., Hood, S., & Ziewitz, M. (2013). Governing algorithms: A provocation piece. *SSRN.*
3. Bellotti, V., & Edwards, K. (2001). Intelligibility and accountability: human considerations in context-aware systems. *HCI*, *16*(2-4), 193-212.
4. Braga, M., Paccagnella, M., & Pellizzari, M. (2014). Evaluating students' evaluations of professors. *Economics of Education Review*, *41*, 71-88.
5. Costanza, E., Fischer, J. E., Colley, J. A., …& Jennings, N. R. (2014). Doing the laundry with agents: a field trial of a future smart energy system in the home. In *Proc. of CHI*, 813-822.
6. Cramer, H., Evers, V., Ramlal, S., ... & Wielinga, B. (2008). The effects of transparency on trust in and acceptance of a content-based art recommender. *UMUAI*, 18(5), 455-496.
7. Davidson, A. & Kestenbaum D. (2014). The Future of Work Looks Like a UPS Truck. *NPR.*
8. Dourish, P. (2003). The appropriation of interactive technologies: Some lessons from placeless documents. *CSCW* 12(4), 465-490.
9. Egido, C. (1988). Video conferencing as a technology to support group work: a review of its failures. In *Proc. of CSCW*, 13-24.
10. Gillespie, T. (2014). The Relevance of Algorithms. *Media Technologies: Essays on Communication, Materiality, and Society*, 167-193.

11. Girardin, F., & Blat, J. (2010). The co-evolution of taxi drivers and their in-car navigation systems. *PMC*, 6(4), 424-434.
12. Grudin, J. (1988). Why CSCW applications fail: problems in the design and evaluation of organizational interfaces. In *Proc. of CSCW,* 85-93.
13. Hassan, U., O'Riain, S., & Curry, E. (2013). Effects of expertise assessment on the quality of task routing in human computation. In *Proc. of Workshop on Social Media for Crowdsourcing and Human Computation.*
14. Haws, K. L., & Bearden, W. O. (2006). Dynamic pricing and consumer fairness perceptions. *Journal of Consumer Research*, *33*(3), 304-311.
15. Hinds, P., & Kiesler, S. (1995). Communication across boundaries: work, structure, and use of communication technologies in a large organization. *Org. Sci.*, 6(4), 373-393.
16. Hodson, H (2014). The AI Boss that Deploys Hong Kong's Subway Engineers. *New Scientist*
17. Irani, L., & Silberman, M. (2013). Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk. In *Proc. of CHI,* 611-620.
18. Isaacs, E., Walendowski, A., Whittaker, S., …& Kamm, C. (2002). The character, functions, and styles of instant messaging in the workplace. In *Proc. of CSCW*, 11-20.
19. Kantor, J. (2014). Working Anything But 9 to 5. *NYT*
20. Kay, J. & Kummerfeld, B. (2012). Creating personalized systems that people can scrutinize and control: Drivers, principles and experience. *TiiS*, 24-42.
21. Kulesza, T., Stumpf, S., Burnett, M., & Kwan, I. (2012). Tell me more: the effects of mental model soundness on personalizing an intelligent agent. In *Proc. of CHI*, 1-10.
22. Lee, M. K., Kiesler, S., Forlizzi, J., & Rybski, P. (2012). Ripple effects of an embedded social agent: a field study of a social robot in the workplace. In *Proc. of CHI*, 695-704.
23. Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics, 35*(10), 1243-1270.
24. Lim, B. Y., Dey, A. K., & Avrahami, D. (2009). Why and why not explanations improve the intelligibility of context-aware intelligent systems. In *Proc. of CHI,* 2119-2128.
25. Lyft (2014) Rating Your Driver. Retrieved from https://www.lyft.com/help/article/1453135
26. Maitlis, S. (2005). The social processes of organizational sensemaking. *Academy of Management Journal, 48*(1), 21-49.
27. Martin, D., Hanrahan, B. V., O'Neill, J., & Gupta, N. (2014). Being a turker. In *Proc. of CSCW*, 224-235.
28. McClelland, M. (2012). I Was a Warehouse Wage Slave. *Mothers Jones.*
29. Mintzberg, H. (1973). *The Nature of Managerial Work.* Harpercollins College Div.
30. Mutlu, B. & Forlizzi, J. (2008). Robots in organizations: the role of workflow, social, and environmental factors in human-robot interaction. In *Proc. of HRI,* 287-294.
31. Orlikowski, W. J. (1992). Learning from notes: Organizational issues in groupware implementation. In *Proc. of CSCW*, 362-369.
32. Patton, M. Q. (1990). *Qualitative evaluation and research methods.* SAGE Publications, inc.
33. Parise, S., Kiesler, S., Sproull, L., & Waters, K. (1999). Cooperating with life-like interface agents. *Computers in Human Behavior*, *15*(2), 123-142.
34. Pritchard, G., Vines, J., Briggs, P., … & Olivier, P. (2014). Digitally driven: how location based services impact the work practices of London bus drivers. In *Proc. of CHI*, 3617-3626.
35. Rodden, T. A., Fischer, J. E., Pantidi, N., … & Moran, S. (2013). At home with agents: exploring attitudes towards future smart energy infrastructures. In *Proc. of CHI*, 1173-1182.
36. Sheridan, T. B., & Parasuraman, R. (2005). Human-automation interaction. *Reviews of human factors and ergonomics*, *1*(1), 89-129.
37. Steiner, C., & Dixon, W. (2012). *Automate this: How algorithms came to rule our world.* New York.
38. Strauss, A., & Corbin, J. M. (1990). *Basics of qualitative research: grounded theory procedures and techniques.*
39. Suchman, L. A. (1983). Office procedure as practical action: models of work and system design. *TOIS*, 1(4), 320-328.
40. Tracy, S., Myers, K., Scott, C., (2007). Cracking jokes and crafting selves: Sensemaking and identity management among human service workers. *Communication Monographs,* 283-308.
41. Uber (2014). a Deeper Look at Uber's Dynamic Pricing Model. http://blog.uber.com/dynamicpricing
42. Uber (2014). How the Uber System Works [Video file] https://www.youtube.com/watch?v=makYbqd7mGA
43. Yang, R., & Newman, M. W. (2013). Learning from a learning thermostat: lessons for intelligent systems for the home. In *Proc. of PUC*, 93-102.
44. Walker, C. R. (1958). Life in the automatic factory. *Harvard Business Review,36*(1), 111-119.