


OPEN

Understanding Inequalities in Ride-Hailing Services Through Simulations

Eszter Bokányi^{1,2}  & Anikó Hannák^{3,4,5*}

Despite the potential of ride-hailing services to democratize the labor market, they are often accused of fostering unfair working conditions and low wages. This paper investigates the effect of algorithm design decisions on wage inequality in ride-hailing platforms. We create a simplified city environment where taxis serve passengers to emulate a working week in a worker's life. Our simulation approach overcomes the difficulties stemming from both the complexity of transportation systems and the lack of data and algorithmic transparency. We calibrate the model based on empirical data, including conditions about locations of drivers and passengers, traffic, the layout of the city, and the algorithm that matches requests with drivers. Our results show that small changes in the system parameters can cause large deviations in the income distributions of drivers, leading to an unpredictable system that often distributes vastly different incomes to identically performing drivers. As suggested by recent studies about feedback loops in algorithmic systems, these short-term income differences may result in enforced and long-term wage gaps.

As they grow in popularity, ride-hailing and food-delivery services such as Uber, Lyft, Ola or Foodora are quickly transforming urban transportation ecosystems^{1,2}. These services have revolutionized most aspects of the transportation market. By managing the rides through a mobile application, they lower the entry barriers to the service for both users or passengers and drivers. The rating system facilitates trust between drivers and users, and the flexible working hours make ride-hailing services a popular choice for people starting a new career or a side-job.

A key feature of these services is that an algorithm replaces human dispatchers in the task of matching available drivers to the incoming requests. Companies are now able to optimize the matching with unprecedented precision using data they possess on cars, drivers, and traffic conditions³, resulting in better service availability, shorter waiting times, and ultimately a boost in efficiency or company profits⁴ (optimization refers to maximizing efficiency in a given system, with given supply and demand parameters). On the other hand, in the process of maximizing effectiveness or minimizing waiting times, drivers' interests get sidelined, also, undesirable social outcomes might emerge^{5–8}.

Recent studies and media articles raise concerns about the risks threatening workers' well-being, including racial bias, worker safety, fairness to workers, and asymmetries of information and power. As documented in case studies^{9–13}, workers are struggling to obtain remedies through official channels^{12,14,15}, and strikes have become common in the past years (see Chapter 2 in¹⁵) with drivers of Uber, Lyft, Ola, Foodora demanding higher fares, job security, and livable incomes all over the world¹⁶.

The dispatcher systems of traditional taxi services allowed drivers to hear the same information and receive updates about traffic conditions¹³, moreover, drivers were in direct contact with their employers. However, as described by Rosenblat and Stark in their article "Case Study of Uber Drivers", Uber-like modern systems subject drivers to algorithmic management methods coupled with an almost compulsory blind acceptance of destinations. This results in a hierarchical information flow, in which the company decides the content and the means of disclosing information to the drivers¹². The app-based communication, helpdesk-like assistance, and the automated algorithmic features act as barriers between employees and management, annihilating this relationship¹⁵. Additionally, given the proprietary nature of these systems, there is limited access to data or the rules of the system^{5,17,18}. As a result, drivers don't have a detailed insight into the underlying processes and consequences of

¹Eötvös Loránd University, Budapest, Hungary. ²Agglomeration and Social Networks Lendület Research Group, Centre for Economic and Regional Studies of the Hungarian Academy of Sciences, Budapest, Hungary. ³University of Zürich, Zürich, Switzerland. ⁴Vienna University of Economics and Business, Wien, Austria. ⁵Centre for Economic and Regional Studies of the Hungarian Academy of Sciences, Budapest, Hungary. *email: hannak@ifi.uzh.ch

frequent condition changes^{12,13,19–21}, seek communication channels outside of the companies' systems^{14,22}, and regulatory oversight is hindered^{23,24}.

Most existing literature in the area of taxi matching algorithms is concerned with optimizing aggregate outcomes for the whole system^{25–31}. Such approaches aim to maximize the benefits for the company or to minimize the adverse effects such as CO₂ emissions, overall distances driven, or the passenger waiting times. Following the line of fairness measurement literature^{32–35}, we instead focus on the fair distribution of income from the drivers perspective, because current systems do not guarantee the same income for the same amount of work, neither across workers nor over time^{9,10,19,21}.

We use an agent-based simulation to systematically study the mechanisms in ride-hailing and delivery systems from the perspective of the drivers for a fixed timeframe of one week. Given the social context of the problem, the timeframe of interest is defined to be relatively short: workers of ride-hailing companies have to focus on daily and weekly income targets because of regular fees and payments^{15,36,37}. Moreover, initial wage differences might be amplified through different feedback loop mechanisms^{11,38}, that hinder the balancing effect which evens out inequalities in the long run. We first quantify the income inequality level of systems that use a company-level efficiency maximizing approach, and explore how the system-level behavior changes as a function of input parameters like the number and distribution of taxis and passengers, city layout, and driving strategies. Next, we investigate the trade-off between fairness of driver incomes and loss in overall revenue through an algorithm, designed to integrate the fairness perspective into the matching of drivers with requests.

Income Inequality in Ride Hailing

We create a simulated city environment with drivers, passengers and various parameters which we will introduce throughout the section. Our goal is to simulate a diverse range of cities and real-world traffic conditions, thus we derive our model parameters from real-world data sets.

Taxis in our city drive along a grid, moving one block at each time step, with constant speed. In our basic setting, both request and drop-off locations are most likely to be in the city center, in line with previous studies on real-world data^{39–42}. The matching algorithm between drivers and requests is similar to the algorithm that Uber and most taxi companies use: passengers are matched with the closest available car. Additional parameters capture various real-world scenarios such as city layout, changes in supply and demand, driver strategies, and different settings of the matching algorithm. The pricing scheme is similar to that of UberX in Boston⁴³, and fuel costs are accounted for⁴⁴. We run all simulations for what equals a 40-hour workweek (see Section 3 for details), and calculate income as the earning over the time a driver spends online. Initially, the closest available driver is assigned to the longest waiting passenger (see details of the *nearest* algorithm in Section 3), and every driver remains at the drop-off location of their latest passenger.

According to interviews with Uber drivers¹⁵, the main concerns of drivers are their hourly wage, the overall rates with which they operate, and the number of hours they need to work until they reach a given daily/weekly income target. In our simulation, inequality will manifest as a varying hourly wage, as drivers spend the same amount of time online. In real life, the same often manifests as a highly varying number of hours needed to achieve a given amount of income. Since Uber drivers have limited possibilities to communicate with each other (in contrast to traditional taxi companies, see p. 94¹⁵, it prohibits them from comparing their wages and strategies. Therefore, while drivers might not recognize it, their lack of information and the information asymmetry towards to company are the greatest problems in such a system. Thus, our goal is to create a system of full information, helping us investigate the distribution of incomes across workers for various settings.

Changes in Supply and Demand. We use two variables to calibrate the supply and the demand in our system: the supply d (density) is defined as the number of taxis per square kilometer, while R captures demand-to-supply ratio, which is the fraction of the demanded travel distance over the supplied travel distance (see Section 3 for details). To pick parameter spaces for d and R that cover actual real-world scenarios, we calculate the number of cars and passengers using empirical data from the NYC Taxi and Limousine Commission⁴⁵.

The dots in Fig. 1A depict each hour of the year 2013 in NYC in the $d - R$ parameter space. On top, in white, we place an example day, 15 January 2013, with its hourly parameter values. We see that even within one day and one city, both parameters strongly vary. While such detailed data is not available from other cities, we analyzed aggregate statistics as well as extreme events such as public transportation failures, strikes, or bad weather conditions (see Section 3). We conclude that the taxi density range $d \in [3, 30]$ [1/km²] and the demand-to-supply ratio range $R \in [0.1, 1]$ will capture a wide range of city sizes, seasonal changes, varying traffic at different times of the day, and even extraordinary events that cause sudden changes in the demand or supply.

While Uber uses increased fares for high demand locations or periods called *surge pricing*, we chose not to implement this feature in our system. The timescale of surge pricing is very small compared to one shift or to the timescale of the whole simulation⁴⁶, and we omit rapidly changing temporal pricing parameters. According to the literature and Uber's own research, higher surge multipliers result in fewer requests, thus, surge pricing decreases R ^{47–49}. On the drivers' side, the surge has the long-term effect of drivers learning the most locking times and areas, as such, surge pricing increases supply in the long run, which decreases R and increases d ⁴⁷. Our parameter ranges of R and d are wide enough to cover these cases of decreased demand and increased supply that could be the result of an ongoing surge.

Figure 1B shows how the average income changes with growing number of requests, but constant taxi numbers, that is, a constant density d . Each marker denotes the averaged result of 10 simulations, ran with the associated parameters. As expected, the income is directly proportional to the demand, regardless of d . For example, by doubling the demand-to-supply ratio R from 0.3 to 0.6, the income roughly increases from \$4700 to \$9400 for all three d values. Thus, as long as the system is able to serve all the passengers, demand determines the total income.

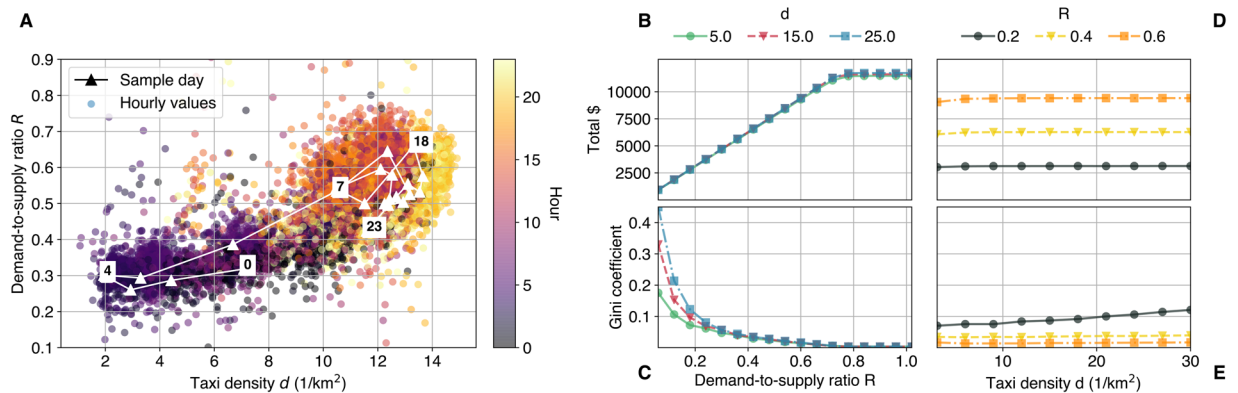


Figure 1. (A) Hourly values of the taxi density d and the demand-to-supply ratio R throughout the year 2013 in NYC. Parameter changes within one selected day, 15 January are plotted as white connected triangles. The total income (B) and Gini coefficient (C) for varying demand-to-supply ratio, at three fixed density values. We see that income grows linearly as the demand-to-supply ratio grows. Higher demand leads to a decrease in inequality, for all density levels. The total income (D) and Gini coefficient (E) for varying density values, at three fixed demand-to-supply ratio values. Higher demand generates higher incomes. Inequality grows with more cars in a city.

After a certain point, incomes saturate as taxis are not able to serve all requests, and the system reaches its maximum capacity around $R = 0.78$. (We caution the reader that fares highly vary across cities, service providers, and over time even within the same company. While we present the incomes in \$, the values are approximations of the fares and the emphasis is on the relative value and the shape of the distributions).

Figure 1C measures the inequality for the same parameters using the Gini coefficient of the incomes at the end of the simulation. Gini coefficient is an inequality measure that captures the deviation of the Lorenz curve of the income distribution from that of an ideal one, where a given cumulative percentile of the population holds the same percentile amount from the incomes (see Section 3). For low demand, that is, low R , the Gini coefficient starts at high values, with 0.18 for $d = 5$ [1/km²], 0.32 for $d = 15$ [1/km²] and 0.45 for $d = 25$ [1/km²]. As the demand increases, the Gini coefficient decreases and converges for different taxi densities, with the Gini at $R = 0.6$ equalling to as little as 0.01 for all three measured d values.

Figure 1D–E depicts the effect of increasing traffic (more demand and more supply), that is, constant R , but increasing taxi density d . Again, higher demand generates higher average incomes, that is roughly \$3100 for $R = 0.2$, \$6200 for $R = 0.4$, and \$9400 for $R = 0.6$. The income is only very slightly affected by the taxi density in a city, as it increases by only 3% for a tenfold increase of $d = 3$ [1/km²] to $d = 30$ [1/km²]. However, higher density leads to a more unequal distribution of those same average incomes, as seen from Fig. 1E. For $R = 0.2$, the Gini coefficient almost doubles with the same a tenfold increase of taxi density $d = 3$ [1/km²] to $d = 30$ [1/km²] that caused the 3% increase in the incomes. For higher demand, this inequality shift is smaller: while there is a 14% increase with changing $d = 3$ [1/km²] to $d = 30$ [1/km²] for $R = 0.4$, the Gini is almost constant for $R = 0.6$.

Since the phenomenon that inequality increases with increasing supply is not intuitive, we expand on the explanation a little: let us consider the probability of a driver at a certain position being assigned to the next ride. Using the nearest algorithm, our chosen driver gets the job if no other drivers are closer to the request origin. In practice, this means that with a higher density d , taxis on the outskirts of the city are even less probable to be assigned to a ride because other available drivers might “stand in their way”. On the other hand, a lower taxi density means that the potential reach of drivers who are further away from the center increases, which equalizes the ride assignment probabilities and lowers the Gini coefficients.

Finally, we note that in certain parameter ranges of our system, we find inequalities that are as high as for the entire income distribution of a whole country⁵⁰. As opposed to our simulation, people in a country work in different positions, different sectors, and might have very different skills. Therefore, it is remarkable that in an artificially equalized situation (drivers with equal working hours and equal skills), the Gini index of incomes is of the same magnitude.

Varying Spatial Activity. Next, we consider variations in the city layout and the traffic flow patterns. Apart from the simplest city center scenario, we run simulations for i.) a city with a larger center, ii.) a city with 2 centers, and iii.) cases where the pickup and destination locations do not overlap (see the detailed description of these layouts in Section 3, and the distributions illustrated in Fig. 1 of the SI). Our motivation is explained by Fig. 2A which shows the temporal and spatial changes in the difference between trip origin and destination location distributions in NYC. We can clearly see the average flow of passengers in the morning towards the center for work (Fig. 2B) and away from the center in the evening (Fig. 2C).

Before we turn our attention to income inequality in the above layouts, it is worth investigating whether and how they affect the overall income, since they all share the same R given identical traffic, i.e., the same demand-to-supply ratio. Figure 2D shows that the average income strongly depends on the distribution of spatial

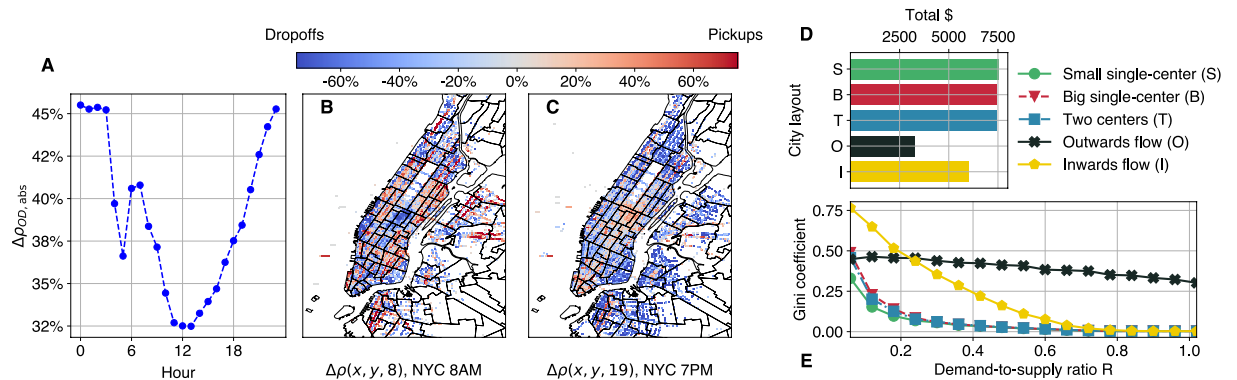


Figure 2. (A–C) Temporal and spatial changes in the difference between trip origin and destination locations in NYC. (A) Hourly average of the unsigned symmetric difference $\Delta\rho_{OD,abs}$ in NYC pixel areas. The figure indicates that origin and destination patterns vary throughout one day. (B and C) Spatial distribution of ρ_{OD} at 8 AM and 7 PM in NYC. Red values indicate more pickups, blue values more drop-offs in a pixel area. At 8 AM, there is an influx into the center of Manhattan and other business and work areas. At 7 PM, the direction of the taxi flow is reversed, as people travel towards the outskirts from the center. See Section 3 for details on $\Delta\rho_{OD,abs}$ and ρ_{OD} . (D) Average income of drivers depending on city layout. Generally, the overlapping cases generate more income, and movement towards the city center as well. (E) Gini coefficient of the seven layouts with varying demand-to-supply ratios. As demand grows, inequality decreases for every layout. Non-overlapping layouts (flows) in the city generate higher inequality.

activity, when examining a setting with fixed $d = 15$ [$1/\text{km}^2$] and $R = 0.48$. If the distribution of origins and destinations overlap completely, more income is generated regardless of the details, e.g. the spreading of the city center or having multiple centers.

On the other hand, in the cases where there is a dominating flow of vehicles due to asymmetries in the origin and destination demands, the income is lowered to roughly 50% (Outwards flow) and 20% (Inwards flow). Moreover, inequality is significantly higher over the whole demand-to-supply R range (Fig. 2E). In the Inwards flow case, the Gini coefficient starts from as high as 0.76, but decreases fast thereafter, and reaches 0.07 for $R = 0.6$. When the traffic flows mostly outwards from the city center, the initial Gini of 0.45 decreases very slowly, though, and stays as high as 0.38 for $R = 0.6$. The above results indicate that the distribution of spatial activity strongly influences both the average income and income inequality. This might be mitigated by modifying the idling strategy of drivers, as we will describe in the next subsection. However, drivers have no real-time information about the spatial distribution of requests, thus, they can only rely on their intuition or past experiences when choosing their idling strategy. This creates a situation where only platform control operations⁵¹, or shared knowledge about driver and passenger positions could decrease the effect of the spatial imbalances.

Waiting vs Cruising. An important decision every driver faces many times throughout the day is what to do in the time period until the next request comes in, which might amount to a considerable share from the overall time spent online^{3,24}. So far in our simulations, the cars were waiting in place after dropping off their passengers. Another obvious idling strategy is to cruise back towards the center to meet more demand. While in a real system, drivers presumably use a mix of these two strategies, here, we investigate the outcome of the two extreme cases, namely when all taxi drivers either wait or head towards the city center.

Intuitively, we would expect that the cruising strategy leads to similar throughput in the cases of overlapping pick-up and drop-off distributions, smaller overall throughput in the cases where traffic flows towards the center, and higher overall throughput when population flows towards the outskirts. But does higher throughput also lead to lower Gini coefficients?

Figure 3A–E shows the waiting/cruising scenarios side by side for $d = 15$ [$1/\text{km}^2$] and $R = 0.4$. We observe significant differences in fairness between the waiting and cruising strategies in all city layouts. In the case of overlapping pick-up/drop-off location distributions, the waiting strategy is fairer, as illustrated by the narrower distributions on Fig. 3A–C corresponding to lower Gini coefficients while the average income is untouched. In the case of the asymmetrical layouts, the strategy of cruising back to the center increases fairness, and in the Outwards flow layout, it even raises the average income (decrease of the Gini from 0.22 to 0.07 and average income increases by almost 200%), see Fig. 3D.

These results underline the importance of transparency and the direct effect of information asymmetry on drivers, who in the current setup of ride-hailing systems can not make informed decisions about their strategies. Moreover, it shows that a seemingly small change in the system settings can lead to large differences in the fairness guarantees of the overall system.

Fixing the algorithm. Lastly, we examine whether we can incorporate the fairness perspective into our system and achieve more equal incomes while conserving the overall revenue, similarly to the perspective of³³. Our goal is to keep track of drivers' income throughout the day and take it into account when assigning rides.

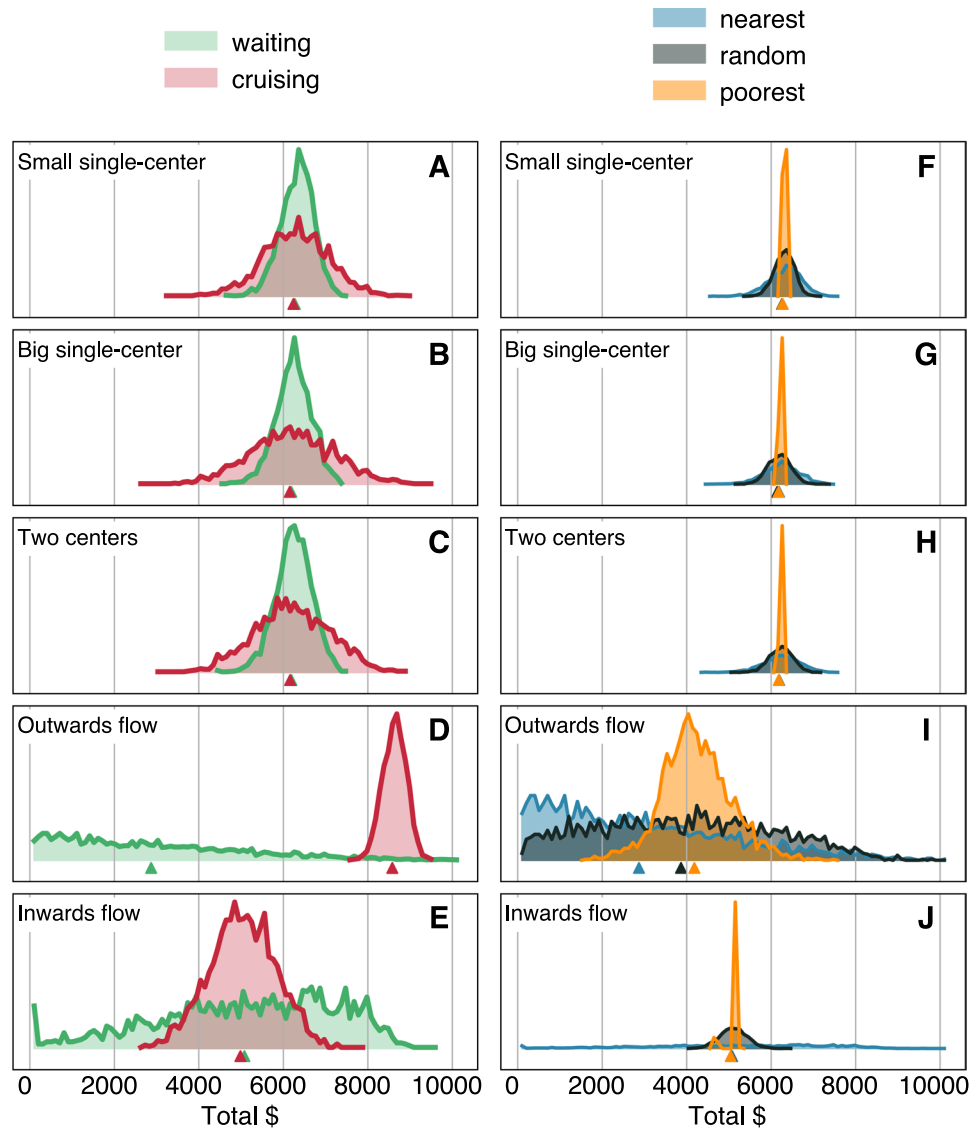


Figure 3. (A–E) Distribution of incomes of different cruising strategies for different city layouts at $R = 0.4$ and $d = 15$ [1/km²]. In symmetrical layouts, the cruising strategy leads to more unequal distribution of incomes with similar means. Among the asymmetrical layouts, in the case of outward flows, cruising towards the center leads to higher incomes. (F–J) Distribution of incomes of different matching algorithms in different city layouts and with the waiting strategy, at fixed $d = 15$ [1/km²] and $R = 0.4$. The poorest matching strategy creates a more equal distribution for all of the investigated geometries. (A–J) The vertical scale of the distributions is omitted for better readability, but all distributions are normalized such that the areas under the curve are equal to 1. Triangles mark the means below the distribution curves.

With this idea in mind, we create the *poorest* algorithm: a modification of the current matching algorithm which keeps track of drivers' income at each point in time and assigns taxis based on the money they made so far. In this *poorest* case, the pool of drivers available for a certain passenger is still limited to be within a certain distance to avoid unreasonably far matches (see details in Section 3). To meaningfully compare the algorithms, we include a baseline algorithm that assigns drivers to passengers randomly within a given radius. This *random* setup should create higher fairness but lower total income than picking the nearest available vehicle.

Figure 3F–J shows the income distribution for the three algorithms for all city layouts. The narrow distributions of the *poorest* algorithm in the symmetric cases show that this poorest correction effectively mitigates the adverse effects of the *nearest* algorithm. In these cases, *poorest* performs even better than the *random* assignment that we consider as a baseline for fair conditions. Moreover, the strategy also helps mitigating inequalities on the Inwards flow layout. While not as strongly as with the other layouts, *poorest* significantly increases fairness with the Outwards flow layout, and even increases the mean income. Since the nearest algorithm mostly assigns drivers from the center, more and more drivers end up in the outskirts without close-by rides. As the poorest algorithm is more likely to pick a driver stranded on the outskirts, it compensates for this undesired process and ultimately leads to higher mean income.

Concluding Discussion

We presented results of an agent-based automated taxi system simulation from a so-far unexplored angle, namely, the fairness of incomes among drivers. Our simulation environment allows us to cover a wide spectrum of city layouts and traffic conditions, and to test arbitrary matching algorithms. We find that a low demand-to-supply ratio (more taxis/fewer passenger requests) leads to greater inequality and that the inequality is largely dependent on the spatial distribution of request origins and destinations, on the chosen idling strategy of the drivers, and the matching algorithm.

We also proposed a new matching algorithm that attempts to equalize incomes by promoting drivers on the lower end of the current income distribution. This method significantly boosts fairness in driver incomes in all investigated scenarios, while not impairing, and sometimes even increasing average incomes through stimulating optimal spatial redistribution of the taxis. This proposed approach is just a showcase to demonstrate how a small modification in the matching algorithm can lead to significantly fairer income distribution. It also shows the validity of a simulation model for testing interface or policy changes. While regulators are increasingly aware of the problems such platforms create, they still lack the tools for monitoring the large-scale effects. As research in algorithmic fairness suggests, policies could include certain target metrics that such stochastic, complex, algorithmically-driven systems should achieve when aiming for better social outcomes.

Notably, our most surprising result is that drivers with the same qualifications and working hours can end up with vastly different incomes by chance after simulating one week. One might argue that these income differences, observed over a short time period, disappear in the long run. This would indeed be the case under our model if used for long-term predictions (even if the regression to the mean could take a considerable amount of time, see Fig. 6 in the SI). However, literature suggests that workers have to focus on daily and weekly income targets because of regular fees and payments^{15,36,37}, and also, their income is mostly reported weekly by Uber¹⁵. Thus, short-term income variance creates unpredictable circumstances and may cause drivers to take risks such as unsafe amounts of overtime¹², or quitting the job^{2,15,37,52}.

Moreover, a large body of sociological and economic literature has shown that emerging inequalities are amplified through feedback loops and other processes that are not directly captured by the simulation in this paper^{11,38}. These initial income differences combined with emerging feedback loops and varying skills, working hours, and ratings can result in increased and continuing wage gaps. For example, introducing driver skills into the simulation would most likely have created a cumulative disadvantage between drivers that persists over time.

While hourly wage would align with our fairness criteria and would solve most of the precarity issues of the drivers, it would also eliminate the gamified and algorithmized incentives with which ride-hailing companies try to enhance the performance and efficiency of their systems. Related work suggests that the current setup of ride-hailing companies includes multiple mechanisms that help the companies keep the wages down, which makes it unlikely that they would switch to an hourly payment system. In the current setup, least earning drivers who drop out of the system can be easily replaced by newcomers since the barriers to entry the work are low^{13,15}. Moreover, since drivers bear all the liabilities and responsibilities related to their work, the company is not forced to compensate them fairly^{15,53}. Unionizing is also out of the question, given the independent contractor status^{52–54}.

Due to the sheer size and complexity of the problem, our study is limited in different ways. First, our city model simplifies complex road maps to a square grid, and the traffic model averages over highly dynamic conditions such as the true taxi velocity, which might depend on the time of day, road type, and physical obstacles. For example, Uber's surge pricing, which aims to steer supply and demand, is an important factor in driver earnings and another source of unpredictability. In contrast, our algorithms are only elementary dispatching systems that are likely not identical to the proprietary solutions. Finally, we neglect any possible effect of the difference in worker skills, such as varying driving skills or differences in app usage. Incorporating such measures into the system design would make it very hard to disentangle the income differences coming from geometry, supply, and demand or the initial skill distribution. On the other hand, if implemented, they would extend the simulation timescale to confirm the existence of long-term inequalities.

Impact. Our results underline the urgency to establish the necessary means for drivers to make informed decisions while working. In our view, giving more control to workers is only helpful when coupled with insights that help them make informed decisions. Currently, drivers' basic decisions, such as when and where to work, are based on limited and local information. Our systematic analysis aims to fill this gap by connecting local knowledge with global outcomes while focusing on drivers' interests. We derive results that directly help drivers align short-term strategies with their long-term goals based on a few key system indicators such as demand, supply or the spatial distribution of requests.

Our work will be among the firsts to take fairness metrics into account in the evaluation of system performance. Currently, companies focus on maximizing their own efficiency⁵, which does not align with worker welfare and fairness. Technical research has followed the former demand, considering only optimization problems of system-level efficiency. By sharing our simulation system on GitHub (<https://github.com/bokae/taxi>), we hope to encourage the research community to work towards a thorough understanding of the effects of algorithm choice on social outcomes, and that eventually service designers will be provided with specific technical recommendations on how to implement a possible fairness perspective. Ultimately, fairness guarantees, either provided by the platforms themselves or via external monitoring of policy agencies, are necessary to increase the trust between ride-hailing companies and their workers.

Lastly, we want to emphasize the advantages of complex systems methods in algorithmic fairness research. Currently, most algorithmic auditing research projects focus on identifying issues with individual platforms at particular points in time^{55,56}. We presented an enhanced approach to algorithmic auditing that takes into account system dynamics, complex phenomena, and downstream consequences of algorithms. This allows the generalization of insights across platforms, pointing towards the fundamental properties of socio-technical systems.

	Origin			Destination		
	μ	σ	f	μ	σ	f
Small city center	[20, 20]	10	1.0	–		
Large city center	[20, 20]	20	1.0	–		
Multi-centered city	[12, 12]	8	0.5	–		
	[28, 28]	8	0.5			
Outwards flow	[20, 20]	6	1.0	[20, 20]	12	1.0
Inwards flow	[20, 20]	12	1.0	[20, 20]	6	1.0

Table 1. List of different origin and destination distributions in the simulations. Each three-column row represents a 2D Gaussian distribution with a mean at μ in simulation coordinates, a standard deviation σ , and a weight f . Distributions having more than one row are a mixture of the listed Gaussians with a weight f . The – symbol denotes that the destination distribution is the identical to the origin distribution.

Complex algorithmic auditing has the potential to be adapted to research harmful phenomena beyond this project, such as the filter bubble effect, discrimination, or polarization in algorithmically aided systems^{57–59}.

Materials and Methods

Simulated city. We investigate generalized and simplified ride-hailing systems with an agent-based simulation. In this simulation, the city is represented as a square grid similarly to⁶⁰, consisting of n times ($n = 20$) m pixels ($m = 20$). To associate simulation units to real-world dimensions, the side length of the pixels is set to $du = 100$ m in SI distance units.

Time, movement, and distances. In each simulation time step, taxis move exactly one distance unit on the grid, and they keep this constant average velocity v throughout the whole simulation. We set v to be 10 m/s, a hypothetical average daytime speed as in²⁵, which implicitly defines simulation time unit: $tu = a/v = 10$ s. All presented results throughout the paper correspond to simulations ran for what equals a 40-hour workweek, that is, until $5 \cdot 8 \cdot 3600/tu = 14400$ simulation time units. Literature suggests one workweek to be a reasonable time frame for which workers might expect predictable and stable incomes because of weekly fixed costs³⁶.

Supply and demand. There are T taxis in the system. We control the number of taxis through a density parameter d , which is the number of taxis per total area of the system. The dimension of d is [$1/\text{km}^2$].

$$d = \frac{T}{A} = \frac{T}{n \cdot m \cdot du^2}. \quad (1)$$

The demand-to-supply ratio is characterized by a dimensionless measure R , which is the fraction of the total demanded travel distance over the distance provided by the taxis during the same time window:

$$R = \frac{\lambda \cdot l \cdot t}{T \cdot v \cdot t} = \frac{\lambda \cdot l}{T \cdot v}, \quad (2)$$

where l is the average trip length requested by passengers in the system, t is the elapsed time, and λ is the average number of requests per time unit.

Traffic flow and city layout. Since the spatial distribution of the demand is typically not uniform in a city^{39–41}, we introduce different probability distributions that characterize the request origins and destinations. In the simplest case, the origin and destination distributions are identical 2D Gaussians, mimicking scenarios where demand is higher in the city center, and gradually decreases further from the core^{61,62}. The scenarios where origin and destination distributions are not overlapping, such as in the case of morning rush hours or late-night rides to residential areas, are simulated by defining separate Gaussian distributions for request origin and destination, respectively, with different standard deviations or centers. Table 1 provides the different distributions in our simulations, with f being the weight of Gaussian mixtures, μ their centers, and σ their standard deviations. We compared the distribution of trip lengths generated from these Gaussian distributions to that of a real-world dataset⁶³. We concluded that the distribution from a real day is the composition of many different system states and spatial activities, therefore, we do not require a strict match between the statistics of the two. We also generated layouts with different cutoffs (exponential, sigmoid) in the decrease of the density of the pick-ups and drop-offs towards the outskirts, but the results did not differ qualitatively from those obtained with Gaussian distributions.

Implementation. Once we have the taxis and incoming requests in the system, we have to define how these elements behave while running the simulation. If a passenger makes a request, the origin and destination location is submitted to the waiting queue, and the request is first marked as pending. At each time step, we try to assign an available taxi to all requests in the queue, from oldest to newest. If the assignment is successful, the request is marked as confirmed, and the assigned taxi is marked as booked, and it ceases to be available. When the taxi reaches the request origin, it picks up the passenger, and the taxi and the request both enter into serving state. A path is created for the taxi, along which it moves towards the request destination as the time is running. Once the destination is reached, the taxi drops off the passenger, the request is marked as completed, and the taxi becomes

available again. Requests that pend too long, that is, for more than 30 simulation units (corresponding to 5 min in real time units) in the pending queue are canceled, as passengers tend to not waiting very long.

Income. Because the pricing scheme might differ between different service providers, cities, or it might change rapidly even within the same system (similar to surge pricing in Uber⁴⁶), we only introduce illustrative USD values based on real UberX fares from an example US city, Boston⁴³. Thus, the income has a fixed \$2/trip and a distance-based \$1/km part, that already includes the mile-based and minute-based fees calculated using an average velocity of 10 m/s. Note that the final income values only serve as indicators of general income trends in mean incomes and inequalities, and they shall not be interpreted as realistic incomes. Fixed costs are assumed to be identical for each driver, thus they are omitted from the calculations. Fuel consumption is proportional to the total distance covered (going for and serving a request, cruising while empty). With an average price of fuel of \$3 per gallon⁴⁴, an estimated 10 l per 100 km³⁵ consumption leads to $\approx \$0.008/du$ costs.

Matching algorithm. The essence of the assignments between taxis and requests is the matching algorithm. All presented algorithms first consider a pool of available drivers in a 1 km range around the request origin. We note here that the 1 km pool limit is arbitrary, and the total income loss that the algorithm causes is directly related to this distance we allow here. First, we implement the *nearest* algorithm, that looks for the nearest available taxi from the pool. This algorithm is the basis for many existing solutions^{29,31,64,65}. We also propose a mechanism that takes the drivers' previous incomes into account and chooses the driver with the least income so far (*poorest*). As a baseline, we also include a method with uniform random assignment from the pool (*random*).

We can also set different initial conditions and different idle behavior for the taxi drivers. They either start the day from a "taxi base" in the center location or from a random location on the grid that mimics their "home" location. After a completed ride, they can either be *waiting* for their next ride in place, or they can start *cruising* back towards the city center. There is also a possibility for simulating a realistic 8-hour shift, where after the end of the shift, every taxi completes its last assignment (if it has one) and then starts the new shift again from their "home/taxi base".

The code for the simulation is available at <https://github.com/bokae/taxi>.

Supply and demand parameter ranges. To match our simulation to real-world parameter values, we obtained the only comprehensive ride-hailing vehicle dataset that is available for the vehicles operating in New York City from the year 2013⁶³. We calculate average trip length l , average velocity v , number of distinct taxis operating T , and number of served requests per second λ for each hour of the year from the dataset. Using Eqs. (1) and (2), we calculate the two main system parameters d and R for these hours using $A = 783 \text{ km}^2$ as the area for NYC. Figure 1A shows the calculated values, and the changes in d and R within one sample day on 15 January 2013. According to the figure, d varies between 2 and 15 [$1/\text{km}^2$], and R between 0.2 and 0.75. However, we do not have data on the number of canceled requests. Therefore, it might happen even within this dataset, that the real value of R is higher than 0.75 during peak hours. We also calculate approximations of d and R based on aggregated metrics for other towns. For example, we take a typical example for an oversupplied taxi system, the traditional taxi service in Berlin, with $d \approx 1 - 5$ [$1/\text{km}^2$] throughout one day, and a typical $R = 0.5$ in peak hours⁶⁶. An opposite scenario of undersupply is present in the city of Barcelona, where d varies between 15 [$1/\text{km}^2$] and 35 [$1/\text{km}^2$], and R between 0.15 and 0.5⁶⁷. Though representing very different real-life situations, in terms of overall numbers, both cities fall into our simulated parameter range. Extreme events such as public transportation failures or bad weather conditions might cause a sudden increase in demand, sometimes as high as doubling R ⁶⁷. Therefore, we conduct our simulations for up to $R = 1$.

City layouts. We also investigate whether an expected shift in the direction of the trips happens during a day. First, we divide NYC into rectangular pixels of 0.01° longitude and 0.01° latitude, and we calculate the normalized distribution of pick-ups $\rho_O(x, y, h)$ and drop-offs $\rho_D(x, y, h)$ for each hour $h = 0, \dots, 23$ averaged over one year. We filter those x, y longitude-latitude pixels in the dataset, where the overall number of both pick-ups and drop-offs is below 5, and where $x \notin [-74.05, -73.9]$, and $y \notin [40.67, 40.85]$.

Normalization means that for each h ,

$$\sum_{x,y} \rho_O(x, y, h) = \sum_{x,y} \rho_D(x, y, h) = 1. \quad (3)$$

We then define the symmetric difference between these two distributions:

$$\Delta \rho_{OD}(x, y, h) = \frac{\rho_O(x, y, h) - \rho_D(x, y, h)}{\rho_O(x, y, h) + \rho_D(x, y, h)} \quad (4)$$

Figure 2A shows the daily mean of the absolute values of the symmetric differences in the pixels

$$\Delta \rho_{OD,abs}(h) = \frac{1}{N} \sum_{x,y} |\Delta \rho_{OD}(x, y, h)|, \quad (5)$$

where N is the number of x, y pairs in the analysis. Figure 2B,C show $\Delta\rho_{OD}(x, y, 8)$ and $\Delta\rho_{OD}(x, y, 19)$, respectively, on the taxi zone map of NYC. We can observe the inflow of people into the city center in the morning, and the outflow of the traffic around 7 PM.

Fairness metric and its interpretation. To assess the inequality in the income distributions of the drivers after 1 week's worth of work, we calculate the Gini coefficient⁶⁸. The Gini coefficient is an important tool in economics for analyzing income distributions over countries or regions. Its typical value for countries (being between 0.2–0.4) is summarized e.g. in OECD reports⁵⁰. The value of the Gini coefficient is independent of the mean of the distribution, therefore, differences in the inequalities can be compared for scenarios generating different mean incomes in the simulation.

Received: 11 September 2019; Accepted: 9 January 2020;

Published online: 16 April 2020

References

1. Stone, B. Uber is winning over Americans' expense accounts, <https://www.bloomberg.com/news/articles/2015-04-07/uber-is-winning-over-americans-expense-accounts> (2015).
2. Hall, J. V. & Krueger, A. B. An analysis of the labor market for Uber's driver-partners in the United States. *ILR Review* **71**, 705–732, <https://doi.org/10.1177/0019793917717222> (2018).
3. Cheng, S.-F. & Nguyen, T. D. TaxiSim: A multiagent simulation platform for evaluating taxi fleet operations. In *2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, 14–21, <https://doi.org/10.1109/WI-IAT.2011.138> (IEEE, 2011).
4. Kedmy, D. Uber's surge pricing: A quick and dirty guide, <http://time.com/3633469/uber-surge-pricing> (2014).
5. Popper, B. Uber kept new drivers off the road to encourage surge pricing and increase fares, <https://www.theverge.com/2014/2/26/5445210/in-san-diego-uber-kept-drivers-off-the-road-to-encourage-surge> (2014).
6. Shepardson, D. & Somerville, H. Uber not criminally liable in fatal 2018 Arizona self-driving crash: prosecutors, <https://www.reuters.com/article/us-uber-crash-autonomous/uber-not-criminally-liable-in-fatal-2018-arizona-self-driving-crash-prosecutors-idUSKCN1QM2O8> (2019).
7. Oh, I. Uber will stop charging ridiculous prices during emergencies, https://www.huffpost.com/entry/uber-surge-pricing_n_5568087 (2014).
8. Griswold, A. Uber price surging on Halloween: Does it take unfair advantage of drunk people?, <https://slate.com/business/2014/11/uber-price-surfing-on-halloween-does-it-take-unfair-advantage-of-drunk-people.html> (2014).
9. Graham, M. *et al.* The Fairwork Foundation: Strategies for improving platform work in a global context. *Geoforum* **29**, 242–253, <https://linkinghub.elsevier.com/retrieve/pii/S0016718520300257> (2020).
10. Lee, M. K., Kusbit, D., Metsky, E. & Dabbish, L. 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*, 1603–1612, <https://doi.org/10.1145/2702123.2702548> (ACM Press, New York, New York, USA, 2015).
11. Cook, C., Diamond, R., Hall, J., List, J. & Oyer, P. The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. Tech. Rep., National Bureau of Economic Research, Cambridge, MA, USA, <http://www.nber.org/papers/w24732.pdf> (2018).
12. Rosenblat, A. & Stark, L. Algorithmic labor and information asymmetries: A case study of Uber's drivers. *International Journal of Communication* **10**, <https://doi.org/10.2139/ssrn.2686227> (2016).
13. Calo, R. & Rosenblat, A. The taking economy: Uber, information, and power. *Columbia Law Review* **117**, 1623–1690, <https://doi.org/10.2139/ssrn.2929643> (2017).
14. Confessions of an Uber driver – empty promises, <http://uberconfession.tumblr.com/post/108296995865/empty-promises> (2015).
15. Robinson, H. C. *Making a digital working class - Uber drivers in Boston*. Ph.D. thesis, MIT (2017).
16. Dickey, M. R. Uber and Lyft drivers are striking ahead of Uber's IPO, <https://techcrunch.com/2019/05/06/uber-and-lyft-drivers-are-striking-ahead-of-ubers-ipo> (2019).
17. Clark, L. Uber denies researchers' 'phantom cars' map claim, <http://www.wired.co.uk/article/uber-cars-always-in-real-time> (2015).
18. Wilson, M. Uber fudges the position of local drivers, but they've got a pretty good reason why, <https://www.fastcodesign.com/3049169/uber-fudges-the-position-of-local-drivers-but-theyve-got-a-pretty-good-reason-why> (2015).
19. Fieseler, C., Bucher, E. & Hoffmann, C. P. Unfairness by design? The perceived fairness of digital labor on crowdworking platforms. *Journal of Business Ethics* **156**, 987–1005, <https://doi.org/10.1007/s10551-017-3607-2> (2019).
20. Ivanova, M., Bronowicka, J., Kocher, E. & Degner, A. The app as the boss? Control and autonomy in application-based management. *Arbeit – Grenze – Fluss* 1–28 (2018).
21. Graham, M. & Shaw, J. *Towards a Fairer Gig Economy* (Meatspace Press, 2017).
22. The Uber Driver's Subreddit. <https://www.reddit.com/r/uberdrivers/>.
23. Chen, L., Mislove, A. & Wilson, C. Peeking beneath the hood of Uber. In *Proceedings of the 2015 ACM Conference on Internet Measurement Conference - IMC '15*, 495–508, <https://doi.org/10.1145/2815675.2815681> (ACM Press, New York, New York, USA, 2015).
24. Jiang, S., Chen, L., Mislove, A. & Wilson, C. On ridesharing competition and accessibility: Evidence from Uber, Lyft, and Taxi. In *Proceedings of the 2018 World Wide Web Conference, WWW '18*, 863–872, <https://doi.org/10.1145/3178876.3186134> (International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 2018).
25. Kümmel, M., Busch, F. & Wang, D. Z. Taxi dispatching and stable marriage. In *Procedia Computer Science*, vol. 83, 163–170, <https://doi.org/10.1016/j.procs.2016.04.112> (Elsevier, 2016).
26. Jorgensen, R. M., Larsen, J. & Bergvinsdottir, K. B. Solving the dial-a-ride problem using genetic algorithms. *Journal of the Operational Research Society* **58**, 1321–1331, <https://doi.org/10.1057/palgrave.jors.2602287> (2007).
27. Alshamsi, A., Abdallah, S. & Rahwan, I. Multiagent self-organization for a taxi dispatch system. In *8th international conference on autonomous agents and multiagent systems*, 21–28 (2009).
28. Bailey, W. A. & Clark, T. D. A simulation analysis of demand and fleet size effects on taxicab service rates. In *Proceedings of the 19th Conference on Winter Simulation*, 838–844 (1987).
29. Maciejewski, M. & Nagel, K. Simulation and dynamic optimization of taxi services in MATSim. *Investment Management and Financial Innovations* **5**, 7–24 (2014).
30. Maciejewski, M. & Bischoff, J. Large-scale microscopic simulation of taxi services. In *Procedia Computer Science*, vol. 52, 358–364, <https://doi.org/10.1016/j.procs.2015.05.107> (2015).
31. Maciejewski, M., Bischoff, J. & Nagel, K. An assignment-based approach to efficient real-time city-scale taxi dispatching. *IEEE Intelligent Systems* **31**, 68–77, <https://doi.org/10.1109/MIS.2016.2> (2016).
32. Hutchinson, B. & Mitchell, M. 50 years of test (un)fairness. In *Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19*, 49–58, <https://doi.org/10.1145/3287560.3287600> (ACM Press, New York, New York, USA, 2019).

33. Chakraborty, A., Biega, A. J., Hannak, A. & Gummadi, K. P. Fair sharing for sharing economy platforms. *Fair Sharing for Sharing Economy Platforms Abhijnan* 2–5, <https://doi.org/10.18122/B2BX2S> (2017).
34. Zehlike, M. et al. FA*IR. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management - CIKM '17*, 1569–1578, <https://doi.org/10.1145/3132847.3132938> (ACM Press, New York, New York, USA, 2017).
35. Weng, J., Liang, Q., Qiao, G., Chen, Z. & Rong, J. Taxi fuel consumption and emissions estimation model based on the reconstruction of driving trajectory. *Advances in Mechanical Engineering* 9, 168781401770870, <https://doi.org/10.1177/1687814017708708> (2017).
36. Farber, H. S. Is tomorrow another day? The labor supply of New York City cabdrivers. *Journal of Political Economy* 113, 46–82, <https://doi.org/10.1086/426040> (2005).
37. Hua, J. & Ray, K. Beyond the precariat: race, gender, and labor in the taxi and Uber economy. *Social Identities* 24, 271–289, <https://doi.org/10.1080/13504630.2017.1321721> (2018).
38. Van de Rijt, A., Kang, S. M., Restivo, M. & Patil, A. Field experiments of success-breeds-success dynamics. *Proceedings of the National Academy of Sciences* 111, 6934–6939, <https://doi.org/10.1073/pnas.1316836111> (2014).
39. Qian, X. & Ukkusuri, S. V. Spatial variation of the urban taxi ridership using GPS data. *Applied Geography* 59, 31–42, <https://doi.org/10.1016/j.apgeog.2015.02.011> (2015).
40. Peng, C., Jin, X., Wong, K.-C., Shi, M. & Liò, P. Collective human mobility pattern from taxi trips in urban area. *PLoS One* 7, e34487, <https://doi.org/10.1371/journal.pone.0034487> (2012).
41. Liu, Y., Wang, F., Xiao, Y. & Gao, S. Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai. *Landscape and Urban Planning* 106, 73–87, <https://doi.org/10.1016/j.landurbplan.2012.02.012> (2012).
42. Santi, P. et al. Quantifying the benefits of vehicle pooling with shareability networks. *Proceedings of the National Academy of Sciences of the United States of America* 111, 13290–13294, <https://doi.org/10.1073/pnas.1403657111> (2014).
43. Uber Boston — Uber prices & services, <http://www.alvia.com/uber-city/uber-boston-2> (2019).
44. U.S. average gas prices by year 1990–2018 — Statista, <https://www.statista.com/statistics/204740/retail-price-of-gasoline-in-the-united-states-since-1990/> (2019).
45. NYC taxi trips by andresmh, <http://www.andresmh.com/nyctaxitrips> (2017).
46. Diakopoulos, N. How Uber surge pricing really works, <https://www.washingtonpost.com/news/wonk/wp/2015/04/17/how-uber-surge-pricing-really-works> (2015).
47. Chen, M. K. Dynamic pricing in a labor market. In *Proceedings of the 2016 ACM Conference on Economics and Computation - EC '16*, 455–455, <https://doi.org/10.1145/2940716.2940798> (ACM Press, New York, New York, USA, 2016).
48. Castillo, J. C., Knoepfle, D. & Weyl, G. Surge pricing solves the wild goose chase. *EC 2017 - Proceedings of the 2017 ACM Conference on Economics and Computation* 241–242, <https://doi.org/10.1145/3033274.3085098> (2017).
49. Hall, J. V., Horton, J. J. & Knoepfle, D. T. Labor market equilibration: Evidence from Uber, http://john-joseph-horton.com/papers/uber_price.pdf (2017).
50. Gasparini, L. & Tornarolli, L. A review of the OECD income distribution database. *The Journal of Economic Inequality* 13, 579–602, <https://doi.org/10.1007/s10888-015-9299-x> (2015).
51. Afeche, P., Liu, Z. & Maglaras, C. Ride-hailing networks with strategic drivers: The impact of platform control capabilities on performance. *SSRN Electronic Journal* 1–56, <https://doi.org/10.2139/ssrn.3120544> (2018).
52. Dubal, V. B. Wage slave or entrepreneur? Contesting the dualism of legal worker identities. *California Law Review* 105, 65–123, <https://doi.org/10.15779/Z38M84X> (2017).
53. Kaltner, J. Employment status of Uber and Lyft drivers: Unsettlingly settled. *Hastings Women's Law Journal* 29, 29 (2018).
54. Dubal, V. The drive to precarity: A political history of work, regulation, & labor advocacy in San Francisco's taxi & Uber economies. *Berkeley Journal of Employment and Labor Law* 38, 73, <https://doi.org/10.15779/Z387P8TD1D> (2017).
55. Sandvig, C., Hamilton, K., Karahalios, K. & Langbort, C. Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and discrimination: converting critical concerns into productive inquiry* 22 (2014).
56. Hannak, A. et al. Measuring personalization of web search. In *Proceedings of the 22nd international conference on World Wide Web*, 527–538, <https://doi.org/10.1145/2488388.2488435> (ACM, 2013).
57. Flaxman, S., Goel, S. & Rao, J. M. Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly* 80, 298–320, <https://doi.org/10.1093/poq/nfw006> (2016).
58. Hannak, A. et al. Bias in online freelance marketplaces: Evidence from TaskRabbit and Fiverr. In *20th ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW 2017)*, <https://doi.org/10.1145/2998181.2998327> (Portland, OR, USA, 2017).
59. Conover, M. D. et al. Political polarization on twitter. In *Fifth International AAAI Conference on Weblogs and Social Media* (2011).
60. Ranjit, S., Witayangkurn, A., Nagai, M. & Shibasaki, R. Agent-based modeling of taxi behavior simulation with probe vehicle data. *ISPRS International Journal of Geo-Information* 7, 177, <https://doi.org/10.3390/ijgi7050177> (2018).
61. Yuan, J., Zheng, Y., Xie, X. & Sun, G. Driving with knowledge from the physical world. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '11*, 316, <https://doi.org/10.1145/2020408.2020462> (ACM Press, New York, New York, USA, 2011).
62. Liu, Y., Kang, C., Gao, S., Xiao, Y. & Tian, Y. Understanding intra-urban trip patterns from taxi trajectory data. *Journal of Geographical Systems* 14, 463–483, <https://doi.org/10.1007/s10109-012-0166-z> (2012).
63. Wong, C. FOILing NYC's taxi trip data, https://chriswhong.com/open-data/foil_nyc_taxi (2014).
64. Novak, K. M. & Kalanick, T. C. System and method for dynamically adjusting prices for services, <https://patentimages.storage.googleapis.com/3a/a7/19/82903f8e9761cc/US20130246207A1.pdf> (2013).
65. Rocca, C. R. L. & Cordeau, J.-F. Heuristics for electric taxi fleet management at Teo Taxi. *INFOR: Information Systems and Operational Research* 0, 1–25, <https://doi.org/10.1080/03155986.2019.1607808> (2019).
66. Bischoff, J., Maciejewski, M. & Sohr, A. Analysis of Berlin's taxi services by exploring GPS traces. In *2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 209–215, <https://doi.org/10.1109/MTITS.2015.7223258> (IEEE, 2015).
67. Maciejewski, M., Salanova, J. M., Bischoff, J. & Estrada, M. Large-scale microscopic simulation of taxi services. berlin and barcelona case studies. *Journal of Ambient Intelligence and Humanized Computing* 7, 385–393, <https://doi.org/10.1007/s12652-016-0366-3> (2016).
68. Champnowne, D. G. A comparison of measures of inequality of income distribution. *The Economic Journal* 84, 787, <https://doi.org/10.2307/2230566> (1974).

Acknowledgements

We would like to thank the helpful comments and insights of Johannes Wachs, Christoph Stadtfeld, Sándor Juhász, Gábor Hannák, Kenneth Joseph, Piotr Sapiezynski and David Garcia. During this research, Eszter Bokányi was supported by the ÚNKP-18-3 New National Excellence Program of the Hungarian Ministry of Human Capacities.

Author contributions

A.H. conceived the study, E.B. wrote the simulation, A.H. and E.B. analyzed the data and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-63171-9>.

Correspondence and requests for materials should be addressed to A.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020