- Abstract ✓
- Acknowledgements ✓
- Introduction ✓
- Background ✓
- Literature Review ✓
- Jakarta
  - Survey Description ✓
  - Descriptives ✓
  - Earnings difference ✓
  - $\rightarrow$ NLP analysis
- Chengdu
  - Description of Data
    * DiDi Dataset ✓
    * Repositioning paper ✓
    * Earnings table ✓
  - Estimation ✓
  - Discussion ✓
- Information Design
  1. Model Information Design ✓
  2. $\rightarrow$ Model Cheap talk
  3. Public Messages ✓
  4. $\rightarrow$ Main Theorem
  5. Remedies ✓
- Policy ✓

# Information Design for Platform Drivers

by

Andreas Haupt

B.Sc., University of Bonn (2014)
M.Sc., University of Bonn (2017, 2018)
B.Sc., University of Frankfurt (2019)

Submitted to the Institute for Data, Systems, and Society
in partial fulfillment of the requirements for the degree of

Master of Science in Social and Engineering Systems

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2021



Author . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Institute for Data, Systems, and Society
August 7, 2021

Certified by . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Jinhua Zhao
Associate Professor
Thesis Supervisor

Accepted by . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Fotini Christia
Program Head, Social and Engineering Systems

**Information Design for Platform Drivers**

by

Andreas Haupt

Submitted to the Institute for Data, Systems, and Society
on August 7, 2021, in partial fulfillment of the
requirements for the degree of
Master of Science in Social and Engineering Systems


## Abstract

With the rise of app-based matching platforms, gig workers become important suppliers of labor to transportation and food delivery. Where and when these workers—platform drivers—work depends on their expected earnings, and hence, in turn, on their information on demand. This thesis gathers evidence why information on demand is important for driver labor supply and in which ways it should be allocated. We make four contributions.

First, using a large-scale survey of platform drivers in Jakarta, Indonesia, we document significant deviations from earnings maximization and extract, using Natural Language Processing techniques, driver's reasons for their labor supply choices. By adopting a fine-tuned transformer model, we are able to analyze Indonesian free-text answers, which might be applicable in other scenarios as well.

Second, we estimate the potential earnings effect of optimal information on demand in Chengdu, China. We use a repositioning challenge that asked for algorithms to reposition drivers with the goal to maximize their average earnings rate. The dataset given gives the earnings rate in artificial units, not in terms of RMB. We recover, using public information on driver earnings structure and regression analysis, the earnings rate equivalents in terms of RMB and, as an application, recover the inter-temporal distribution of surge prices.

Third, we introduce a theoretical model to highlight qualitative features of information design to platforms taking into account the roles of commitment of the platform to particular information policies and asymmetric information, i.e. information that only some agents receive. We show that both platform and driver payoffs are the highest under asymmetric information that the platform can commit to, that both are decreased under no commitment and might, if driving to a high-demand location is costly enough result in no information being transmitted. The inferior result is if the platform is restricted to only public information, in which case commitment is irrelevant. We show that in cases where only a subset of drivers is targeted by public information, efficiency can (approximately) be restored. We comment on two challenges, a fairness and a commitment challenge, arising for the platform.

We close this thesis by assembling challenges and opportunities for platforms,

regulators and transportation engineers.

Thesis Supervisor: Jinhua Zhao
Title: Associate Professor

# Acknowledgments

I thank Helga Haupt (AES), Katrin Haupt (University of Leipzig), Sophia Bussey, Isabell Goldaracena and Juan-Pablo Rodriguez (155P) for support when sitting writing on this thesis. I thank Stephen Zoepf (Lacuna.ai, Stanford University), Ryan Green (Gridwise), Dan Calacci (MIT Human Dynamics), Xiaotong Guo (MIT JTL), James Parrott (New School), and Hai Wang (Singapore Management University) for helpful conversations. I thank my reader John Horton (MIT Sloan) for his open feedback. Finally, I thank my supervisor Jinhua Zhao (MIT JTL) for his visions, and for being such a kind lab leader.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

> [Gig economy,] the collection of markets that match providers to consumers on a [job] basis in support of on-demand commerce. [. . . ] Prospective clients request services through an Internet-based technological platform or smartphone application that allows them to search for providers or to specify jobs. Providers (gig workers) engaged by the on-demand company provide the requested service and are compensated for the jobs. – Congressional Information Service, Donovan, Bradley, and Shimabukuro 2016

## 1.1 Information Design for Platform Drivers

This thesis studies the relevance of providing demand information to platform drivers and how it should be allocated.

Platform drivers, as subset of all gig workers, include both grocery delivery drivers and ride-sourcing drivers. While food delivery, grocery delivery and ride-sourcing are distinct services, the groups share many similarities, and are sufficiently distinct from other sectors of the gig economy: Given demand information, both groups take on very short gigs that depend dynamically on demand by riders or food delivery customers. This dynamic nature makes information of different relevance to them compared to other parts of the gig economy such as accomodation (AirBnB, Vrbo) and freelancing (Amazon Mechanical Turk, Airtasker) which do not have to make such dynamic decisions because of the decreased time-sensitivity of their work. Also, many platforms operating in ridesourcing, so-called Transportation Network Companies (TNCs) such as Grab, Gojek, DiDi or Uber, offer also food and grocery delivery services.

The Congressional Information Service's definition of gig work given above highlights the opportunity of search for consumers: "Prospective clients request services through an Internet-based technological platform or smartphone application that allows them to *search* for providers or to specify jobs" (emphasis added by the author). The definition does not include search by gig workers as a characteristic. As we argue for platform drivers, dynamically reacting, or *searching and finding* demand is affecting driver earnings, and an important platform design parameter.

### 1.1.1 Three Levels of Platform Design

Information Design can be seen as a third level in the design of platforms.

#### Matching

Matching is the assignment of drivers to orders, i.e. rides, shopping trips, or deliveries. In the grander scheme of mechanism design, matching can be seen as the allocative dimension of ridehailing systems, and it answers the question of which driver is going to get work at all.

#### Pricing

Pricing refers to fares for customers and driver earnings for trips. It is the transfer domain in mechanism design.

#### Information Design

This thesis studies Information provision. While matching and pricing are real outcomes, behavior is shaped by forward-looking behavior and *expectations*. Expectations are based on information. Informing drivers about when and where to work can affect their repositioning, their working hours, and location of their log-in, and hence help the platform.

Because of strategic responses by drivers, however, more information is not always better, neither for platforms nor drivers,. For example, if too many drivers get the information of a high expected demand in some area, this area might become congested, drivers might not get orders as frequently or might, because there is no supply-demand mismatch, not benefit from pricing strategies by platforms such as *surge pricing*, compare chapter 2. This expectation leads drivers to less frequently visit high-demand areas.

The sophistication by drivers needed for such reasoning about demand has been demonstrated in other parts of the gig economy. For example, knowledge of demand and expectations have been demonstrated in the market for freelance software development Horton and Tambe 2019, when, with the announcement of the abandonment of Flash by iOS, freelance software developers in this language migrated to other freelance work so that wages for Flash developers remained unchanged.

After positioning platform drivers in the spheres of urban transportation and the gig economy in the rest of this chapter, and reviewing important dimensions of platform design which interact with information design and information provision through third parties in chapter 2, we make four contributions.

The first part of our study, chapter 3 investigates which other factors besides expected earnings affect driver labor supply decisions. In a large survey of platform drivers in Jakarta, Indonesia, we find significant inconsistencies with earnings maximization of drivers in terms of reported earnings, and find cultural, as well as informational aspects in drivers' labor supply decisions. We use Natural Language Processing techniques to analyze free-text answers in a foreign language without translating answers pre-analysis, which might be of independent interest.

We supplement our observations on the relevance of information design questions in chapter 4 by comparing the earnings effect of different driver repositioning strategies. We provide an approximate value of demand knowledge using earnings differences under optimal *repositioning*, i.e. movement of idle drivers to other areas in the city. We make this estimate by inferring real earnings for platform driver trips undertaken for TNC DiDi in November 2016 in Chengdu, China and comparing the earnings differences to average earnings of delivery drivers on the platform. We find that drivers with very good demand information can earn significantly more than other drivers, as long as only few drivers have such information.

We relax our assumption that only a small group is repositioned in our theoretical analysis in **??**. In a stylized information design model we show that only a few drivers receiving demand information is a robust property of optimal platform information design. We show that platforms that cannot commit to information policies to drivers have incentives to give demand information to many drivers, who, in turn act less on this information. We study solutions to this commitment problem via loyalty ranks, third-party aggregators, and audits.

The final chapter 6 collects policy recommendations both for platforms such as Grab, GoJek, Didi, Uber or Lyft and regulators such as the New York City Taxi and Limousine Commission. We first argue that information design is of different significance for platforms and regulators. While the former will need to rely on third-party aggregators to commit to disclosing information in a way that drivers will not ignore, the latter should think more carefully about equity in the proposed models—Information needs to be provided asymmetrically on a per-trip basis, but symmetrically on an aggregate level. We conclude with engineering challenges arising from our analysis.

While we provide several facets of information design for platform drivers, we largely abstract both matching and pricing platform design. Real-world platforms solve joint optimization problems of pricing and information, and, as a further complication, often do so in competition with other platforms. While this means that our focus on information design underestimates the ability to design a market for platforms, we highlight that information design should be an additional pillar besides matching and pricing platform design.

Our analysis is also not restricted to human drivers. Depending on Autonomous

Vehicle governance structure, conflicts of interest and congestion might have even more pronounced effect in a future of automated mobility. Our investigations, in particular in chapter 4 and chapter 5 directly feed also into this design question: Information Design has potentially large effects, and needs to be allocated thoughtfully to avoid congestion.

## 1.2 Background

Platform drivers take an increasingly important role in urban mobility (subsection 1.2.1), and have special characteristics compared to other forms of gig work (subsection 1.2.2). Some of these characteristics have led to more prominent regulatory interventions into platform driving (subsection 1.2.3) compared to other parts of the gig economy. Regulators have for the biggest part not intervened into information provision to drivers, and drivers find themselves in an informational environment between their platforms' apps and third parties aggregating information (subsection 1.2.4).

### 1.2.1 Platform Drivers and Urban Mobility

As of 2021, services provided by platform drivers now make up 1/3 of the global taxi market Bryan and Gans 2019. With the rise of the first TNCs about 15 years ago, the rapid changes leaves many of the design dimensions of this market open.

An important difference to other travel modes such as public transit is the need for dynamic rebalancing, i.e. ensuring a distribution of drivers such that in each part of the urban area expected ride demand can be met at each point in time, and rider waiting times are approximately constant throughout an urban area and time. While dynamic rebalancing is an important question in public transit operations, schedules can be designed with much less uncertainty as compared to the dynamic optimization required in TNC operations.

Part of the uncertainty making dynamic rebalancing for platforms hard comes from the need to design incentives that ensure participation from drivers. Even with accurate demand models for riders, platform drivers usually have the opportunity to log off their platform or reject (some of) the rides requested by a customer. If they are not incentivized to continue working on the platform, they might not.

### 1.2.2 Platform Drivers as Gig Workers

Platform drivers are also gig workers *qua* the definition introducing this chapter. Customers search on the platform and are matched to a driver for their ride or delivery. Nevertheless, platform drivers are extreme in some of their characteristics compared to other gig workers.

First, platform driving is, compared to other parts of the gig economy, very time-sensitive. This characteristic comes from the nature of deliveries or rides being short gigs. Compared to other parts of the gig economy, e.g. accomodation (e.g., on AirBnB), and services for care (e.g., care.com), technology (e.g., Andela), design (e.g., 99designs) and home services (e.g., Porch), whose gigs typically require scheduling or completion within at least a day, platform drivers have short gigs which are arranged within seconds. The shortness of gigs and scheduling makes dynamic driver incentive management particularly crucial for platform drivers.

Second, the terms of the contract are, more than in other parts of the gig economy, more strongly determined by the platform. While drivers have flexibility on when and where to work, which orders they will get, and also the pay they will receive, is not determined by them. This is in contrast to other platforms such as in accommodation, technology and design, where gig workers can set their own prices.

Third, reaching out to customers outside of the platform is much more challenging for platform drivers than for other gig workers. As argued above, compared to other gig work, platform driver gigs are short, and as both the complete contract and all payments are processed by the platform, there are few opportunities for drivers to get connection with riders that might lead to follow-up work. The absence of work options outside of platforms gives the platform additional bargaining power in negotiations with drivers.

Finally, many of the drivers are full-time and work many hours a day. This both raises concerns over work safety, but also over wage bounds, some of which have been introduced as regulations.

### 1.2.3 Regulation of Platform Driver Work

The combination of time-sensitive gigs, pay determination by platforms, missing outside options for drivers, and full-time platform work led to policy interventions regarding minimum wages and the employment status of platform drivers.

The New York City Transport and Limousine Commission introduced an earnings standard which guaranteed a *proxy* minimum wage for ridesourcing drivers in New York City Parrott and Reich 2018. Drivers earn for rides an additional amount to increase the average as calibrated against historical demand data to the minimum wage level.

In 2020, California accepted via a referendum a special provision for ride-sourcing drivers to be excluded from legal sufficient conditions for an employment relationship Padilla 2020.

In contrast, the UK Supreme Court decided that ride-sourcing drivers *are* employees of TNCs, with labor implications for drivers, with effects for payment and insurance of drivers Arden et al. 2021.

These regulations do not include provisions on the information that drivers get from TNCs. Drivers' information environment is given by different players, and interacts with drivers' payment structure.

### 1.2.4 Platform Drivers' Payment and Information Environment

Most platform drivers interact with the platform through a smartphone app, which offers them potential gigs, which they can accept or reject. At each point in time, drivers can select to log off the system. Among U.S. driver, roughly half multi-home (Valderrama 2020, p.48). In our survey conducted in Jakarta (chapter 3), we find that a smaller number of drivers multi-home in this market, and that even fewer have several apps open at the same time. Hence, many drivers rely in getting demand

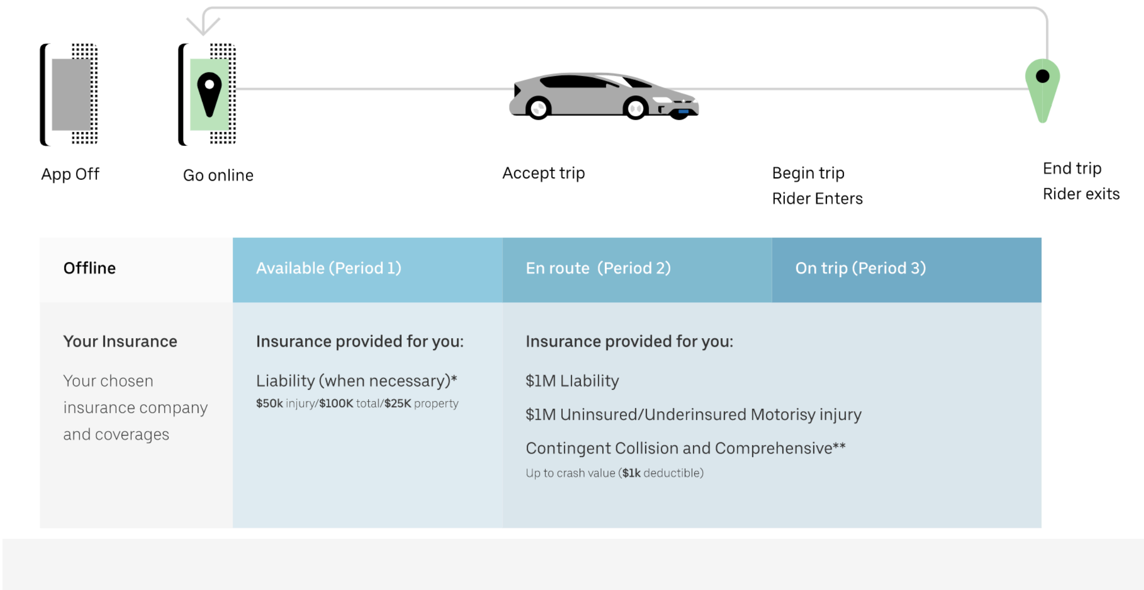



Figure 1-1: Nomenclature for different periods in coverage for Uber as presented in Inc. 2021.

information on one platform's app, as well as driver assistants.

Driver assistants such as the Surge app or Gridwise give information on demand across platforms and allow to track earnings information. These are outside of the contractual information, and purely provide drivers with information.

The informational environment becomes relevant for drivers through the payment structure of gig drivers. An important abstraction will be three time periods, which we call, in accordance with nomenclature introduced by Uber *phases 1, 2, and 3*, compare Figure 1-1.

Period 0 corresponds to times off the platform. Period 1 corresponds to times where the driver is online but idle. In these times, drivers are not paid.

Phase 2 is after the acceptance of an order, and *en route* to fulfilling the order. Except for flat payments for cancellations, drivers are not paid for this part of their trips as well.

Period 3 refers to the time between the acceptance of a ride request and the pickup of the passenger, and the ride, respectively. In these phases, riders are paid according to a combination of base prices, minutes and distance driven, as well as surge multipliers.

The goal of this thesis is to establish the relevance of providing drivers with demand information, and will only significantly affect driver behavior in phase 1. Drivers that are picking up or driving a rider have a clear task, and even a contract for their work. The main question we ask in our theoretical section will be under which the platform can shape behavior outside times where a contract between the platform driver and the customer is in place.

# Chapter 2

# Related Work

> [Traffic participants] make route choices based on their private beliefs about the state and other populations' signals. The question then arises, "How does the presence of asymmetric and incomplete information affect the [participants]' equilibrium route choices and costs?"—Wu, Amin, and Ozdaglar 2021

We contribute to several different literatures regarding platform design and to methodologies in Natural Language Processing, work with platform-produced data, and mechanism design. We start with the three pillars of platform design—matching, pricing and information provision—in section 2.1. In **??** we review two more topics that our contributions relate to: asymmetric treatment in platform design and platform competition. We situate our methodology in existing work in **??**.

## 2.1 Platform Design

We introduced the three pillars of platform design in chapter 1. Here, we name recent contributions, and relate them to information design for platform drivers.

### 2.1.1 Matching

Dynamic matching is a topic in the operations research. Recent contributions to this literature are Aouad and Saritaç 2020 and Vazifeh et al. 2018.

A main goal in matching is *dynamic rebalancing*. Dynamic rebalancing refers to moving vehicles to other parts of an urban area when the (expected) demand in this area is higher. Order dispatch algorithms with foresight optimize for this objective Zhe Xu et al. 2018, and other transportation modes have similar challenges, e.g., for bikesharing, Barabonkov et al. 2020.

Participation decisions by drivers determine the group of drivers to be matched. While drivers and platform have a shared objective to have many drivers in high-demand areas, they diverge on how many: While platforms do not have incentives to avoid congestion, as we show in chapter 5 drivers do, as oversupply in areas leads to reduced earnings. Hence, taking into account information provision to drivers has the potential to improve matching.

### 2.1.2 Pricing

A second important dimension of platform design is pricing.

The move towards non-uniform prices across time is associated to the *Wild Goose Chase* Castillo, Knoepfle, and Weyl 2017. A wild goose chase in ride-sourcing—the same phenomenon is, however, possible for other platform drivers as well—refers to the reduction of supply when supply is low and prices are constant through a vicious cycle of expectations: Given that supply is very low, drivers will quickly be thinly spread throughout an urban area. This means that the average pickup time for drivers is increased and driver earnings decreased. This leads to an expectation of drivers to earn less, and hence to even lower supply. Zhengtian Xu, Yin, and Ye 2020 shows that for many matching technologies and demand structures, Wild Goose Chases occur.

As an intervention into the market that reduces the appearance of the wild goose chase, *surge prices*, i.e. higher prices when demand is high are proposed Castillo, Knoepfle, and Weyl 2017. Further research has considered the effects of surge prices in a spatial setting Lee et al. 2019.

In case of the Wild Goose Chase, other interventions than price interventions are possible: The reduced demand comes from the fact that drivers cannot coordinate on when to be on the street. Quota-based interventions (allowing drivers only to log in during certain times) or informational interventions (targeting some drivers during some time with information) might help mitigate the Wild Goose Chase. We consider aspects of an information-based remedy in chapter 5.

Other models such as a recent contribution Zhou et al. 2020 solve the optimal dynamic matching problem with prices. Zhou et al. 2020 proposes a queuing model to optimize platforms' profit while considering price-sensitive customers and earning-sensitive drivers. Each driver has a reservation earning rate based on his outside option and the driver will provide services only if earning rate exceeds the reserved value.

### 2.1.3 Information

Information on demand or recommendations on where to reposition to maximize earnings were shown in a focus group study with platform drivers Ashkrof et al. 2020 as important points of improvement for TNCs.

However, except for in routing games (Wu, Amin, and Ozdaglar 2021; Systems 2021; Wu 2017), information design for transportation systems has, to the best of our knowledge, not been thoroughly studied. Wu, Amin, and Ozdaglar 2021 studies a model of competing information providers, which are restricted to public information from one of several information sources. As they, we find inefficiencies arising from only providing public information to drivers. While their model is on congestion levels on roads, we consider congestion in areas for drivers. Added to their study, we highlight the incentive mismatch between platform and drivers, which shape many of the incentives.

Shapiro 2018 shows using data from New York City that the highest welfare gains from ride-sourcing arise in less dense areas. In particular, these are areas where the informational component of platform design is particularly important, as drivers need to take into account idle repositioning time when accepting an order. We complement this point in chapter 5 by showing that, besides platform and rider gains, also driver earnings are positively affected by information such as the one provided by platform apps.

Jullien and Pavan 2019 study a two-sided market in which groups are uncertain about the joining decision of the other side of the market and how allocating information helps efficiency. We do not model demand for platform drivers as strategic, and only consider platform drivers. In this sense, our model studied in chapter 5 can be seen as a special case of the model in Jullien 2011.

Ong, Freund, and Crapis 2021 combines information design with pricing by incentivizing drivers to move to a new area with a monetary incentive. Closest to our contribution, **Chaudhari_Byers_Terzi_2018** show that a carefully designed repositioning strategy can change earnings by a factor of 2 chapter 4. Our results are much more modest and show improvements of approximately 20% for drivers.

## 2.2 Related Topics in Platform Design

In our contributions we also relate to other topics of platform competition. One is platform competition *for the market*, another is asymmetric treatment of participants *in the market*.

### 2.2.1 Competition

Our first contribution relates to a literature on platform competition, in particular whether one side of the market multi-homes, i.e. participates in more than one platform at a time.

Multi-homing has attracted much attention in ridesharing Liu, Loginova, and X. H. Wang 2018, regarding the effect on questions of pricing when both sides multi-home Bakos and Halaburda 2018, the welfare effects of multi-homing Belleflamme and Peitz 2019, how multi-homing can happen dynamically Biglaiser 2019 for platform migration, and questions on participants that cannot multi-home Jeitschko and

Tremblay 2014.

An important assumption in these models is that drivers have expectations on expected demand in all areas. Our analysis about information design plays into this environment by investigating how different information structures in the market affect market participants Liu Tat-How Teh Julian Wright Junjie Zhou 2019.

### 2.2.2 Asymmetric Treatment

Asymmetry to leverage network externalities are well-known in the network externality literature. Jullien 2011 studies a model of a one-sided platform with network externalities, and devises a pricing strategy that, at limited cost, incentivizes agents to stay on the platform. Fainmesser and Galeotti 2016; Fainmesser and Galeotti 2020 study this with players that have some heterogeneity.

More generally, asymmetry of information being welfare optimal in a class of games called *submodular games*, also called *games of strategic substitutes*, is well known Bergemann and S. E. Morris n.d. Our results in chapter 5 cannot be directly deduced from these insights, as we model incentive mismatch between platform and drivers.

## 2.3 Methodology

### 2.3.1 Labor Supply of Drivers

Our estimate of implicit costs of multi-homing in chapter 3 estimates preferences relating to driver labor supply. A classical study on taxi driver labor supply in New York City Camerer et al. 1997 estimated negative labor supply elasticities of cab drivers, i.e. drivers working less hours on days with higher earnings. They concluded that drivers work until they reach fixed earnings, and not for a particular number of hours. This behavior is in contrast with pure earnings maximization, as drivers that have high demand and can (to some extent) flexibly work, would work more on days with high demand and less on days with low demand. Other studies contest this claim Farber 2015; Fehr and Goette 2007; Farber 2005.

Our analysis, as Camerer et al. 1997 shows driver behavior that is incompatible with earnings maximization.

### 2.3.2 Gig Worker Data

Studies on platform behavior need data on the contracts that the platform intermediates. Getting such information has been challenging due to the complex pricing, matching and information environment, and that riders are not informed on payments to drivers and *vice versa*.

Information on contracts comes from primarily three sources, two of which have been exploited for research purposes: cooperations with platforms, automatic extraction of data through APIs or app simulators, and data collection from drivers.

Several studies in collaboration with TNCs have found large-scale behaviors. Among others, Sun, H. Wang, and Wan 2019 on the labor supply of platform drivers, Cook et al. 2020 for the gender pay gap, and Athey, Castillo, and Chandar 2019 for a comparison in driving quality of TNCs and Taxis. The literature on market structure or regulation-relevant topics is limited with the use of TNC datasets given contractual restrictions on platform data use by researchers.

Other studies use APIs to query for prices, such as Rosaia 2020 which evaluates counterfactual regulations in an urban mobility market in New York City and Shapiro 2018 which estimates the consumer surplus difference from TNCs compared to Lyft in the same market. Chen, Mislove, and Wilson 2015 uses large-scale simulated apps to elicit prices and estimate pricing fairness questions.

A third approach is given by eliciting information from drivers. Worker Information Exchange Worker Info Exchange 2021, an initiative in the UK, gathers driver information via Data Subject Access requests under a data protection regulation. Similarly, MIT Media Lab initiative gigbox Calacci 2021 collects information from drivers. The companies Gridwise and Surge gather driver data and give drivers information on pay-related information.

Our study falls between the first and the second category. We use openly (for researchers) available platform data, which we combine with other publicly available data to answer a question relating to driver earnings.

### 2.3.3 Mechanism and Information Design

Mechanism design is an field that studies the creation of strategic environments to reach social goals. Pioneered by Leonid Hurwicz and students of Kenneth Arrow, Eric S. Masking and Roger B. Myerson, mechanism designs where used for assigning students to schools, donors to receivers of organs, matching medical residents to hospitals, and internet ad auctions. For an introduction to the topic, see, for example, Eric Maskin's Nobel Memorial Prize lecture Maskin 2007.

A main goal of mechanism design is the design of allocation rules of scarce resources by a central entity when information is dispersed among different strategic actors. In the context of platform drivers, for example, the exact preferences of drivers for when and where to work are not known to the platform.

Mechanism design is the design of rules. A main assumption is that the designer, in our case the platform, can *commit* to a particular action it will take given information given a state of the world such as demand on a platform.

The commitment assumption is often violated, also, as we argue, in context of platform drivers. We compare the model to a setting where the platform cannot commit to some course of action in a model of *cheap talk*, which was introduced in B. Y. V. P. Crawford and Sobeli 2016. Our exact setup is influenced by the multi-receiver version Goltsman and Pavlov 2011.

While mechanism design pertains to the allocation of scarce resources, another allocation problem arises when information shapes behavior and a platform can control information flows.

Information is not scarce—it is freely replicable—but can affect the allocation of scarce resources: If every driver in an urban space is informed that a particular part of the city experiences high demand, congestion, too many drivers in this area, might be the result of providing drivers with this information. In particular in situations where congestion might be a result of information, information needs to be carefully designed.

Information design is mechanism design where the good being allocated is information. To make precise what information means, the timeline of interaction is important. *Before* some state of the world—such as the distribution of demand for drivers during rush hour—is realized, the mechanism designer, the platform in the platform driver case, commits which driver to reveal which part of this state. This commitment is important for the outcome, as drivers can make their decisions knowing, for example, which information other driver do *not* have.

Information design started with single agent environments without strategic interactions Kamenica 2019, building on notions of communication revelation principle Myerson 1983. In later works, the full information design with strategic interactions was developed, Bergemann and S. E. Morris n.d.; Bergemann, Brooks, and S. E. Morris 2020. We will use for a relevant part of our analysis a model in Bergemann and S. E. Morris n.d.

With strategic interactions, V. Crawford et al. 2013 study a strategic environments with several players and derive as a robust insight that in games where, if more players decide to participate, this increases the incentive for others to participate (so-called *games of strategic substitutes* or *submodular games*), asymmetric information improves welfare compared to public information.

The real platforms' problems are more complicated than merely providing information. The platform faces the challenge to co-design matching, pricing and information provision. The literature on mechanism-information co-design is limited, to simple downstream games Dworczak 2020, and an integration of matching, pricing and information is an exciting area for future work.

# Chapter 3

# Platform Driver Labor Supply Beyond Earnings Maximization

> Since Uber' s introduction in 2009, ridesharing platforms like Uber, Didi, Grab, and Lyft have radically transformed the taxi and limo industry. These services, which allow consumers to order a car to their location via a smartphone application, now control roughly 1/3 of the international taxi market. In other words, a ridesharing firm acts as a platform matching drivers to riders and setting the pricing terms between them. —Bryan and Gans 2019

This chapter analyzes data from a survey in Jakarta and shows that drivers do not purely maximize their earnings in their labor supply choices. In an analysis of free-text answers, we highlight understanding of the app and information as important determinants of driver behavior.

## 3.1 Survey

Our data comes from a survey conducted in spring 2021 in Jakarta, Indonesia.

### 3.1.1 Sample

Participants were asked to complete an online survey hosted on Qualtrics. 846 complete survey responses were submitted. We excluded from the analysis drivers that stated more than 16 hours of work per day and stated earnings smaller than 10,000 IRB or larger than 2,000,000 IRB. The survey participants were recruited through social media (platform driver WhatsApp groups).

Figure 3-2a shows the age distribution of participants, where around 50% of drivers are within the age group of 30 to 39. Figure 3-2b reveals that over 97% of drivers who participated in the survey are male and around 97% are motorbike (delivery) drivers.

Three major ride-hailing companies operate in Jakarta: Grab, GoJek and Shopee. While Grab and GoJek are incumbents in this market, Shoppee entered the market only in 2021. Figure 3-1 shows the percentage of drivers working for each of the platforms, double-counting multi-homers. Around 94% of drivers work for the incumbents Grab or GoJek. In our sample of drivers, only 12% of drivers questioned in the survey are multi-homers.

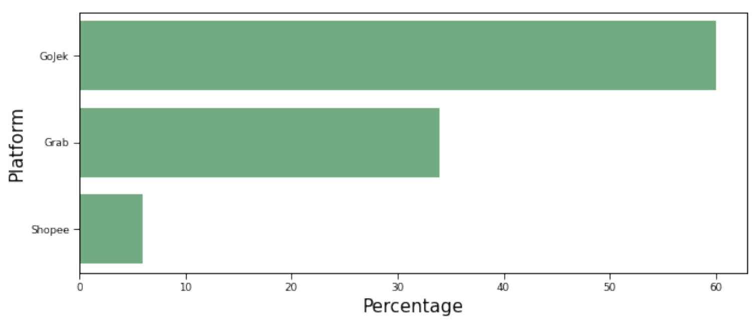
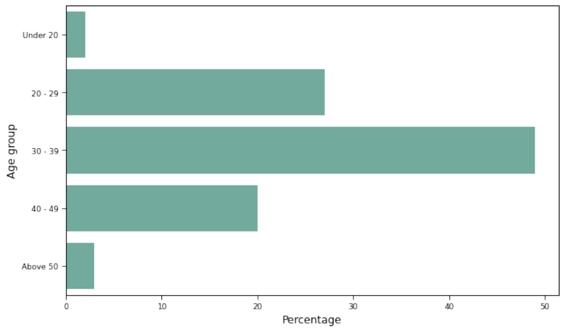



Figure 3-1: Platform distribution of drivers. (GoJek: 60%, Grab: 34%, Shopee: 6%)

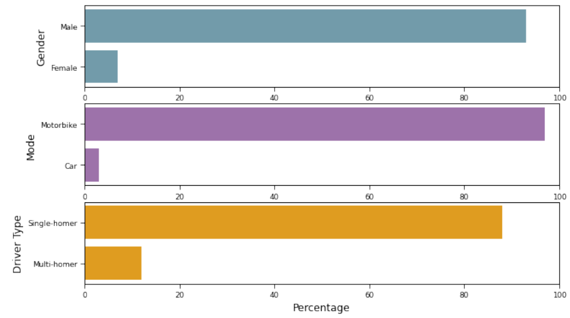
### 3.1.2 Questions

The survey consists of multiple questions blocks and is reproduced in its entirety in Appendix A.1.

In a first block, participants were asked to check all TNCs they have worked for in the past 30 days. A survey respondent is defined as a *multi-homer* if they have




(a) Age distribution of drivers. (Below 20: 2%, 20 to 29: 27%, 30 to 39: 49%, 40 to 49: 20%, above 50: 3%)




(b) Gender, mode and driver type distributions of drivers. (Male: 97%, female: 3%; motorbike: 97%, car: 3%; single-homer: 88%, multi-homer: 12%)

Figure 3-2: Sociodemographic information and driving behaviors of survey participants.

been working for at least two platforms in the past 30 days. In this first block, basic information and driving behavior are asked as well, e.g., vehicle type used, previous occupation, behaviors while waiting for the next order.

The second block's questions differ for multi-homers and non-multi-homers. Multi-homers are asked how they switch between different platforms, allocations of their working hours and reasons for multi-homing. For non-multi-homers, reasons for non-multi-homing and factors that would make them multi-home are elicited.

In a third block, for each ride-hailing platform survey participants have worked for in the last 30 days, respondents are asked questions about their labor supply choices. Type of work (delivery, ridesourcing), part of the city of highest activity, and daily working hours, distance, and earnings.

A final block elicited socio-demographic information, including age, gender, educational status, income level, living areas, and weekly expenses.

## 3.2 Descriptive Evidence

A first observation we make is that there is significant correlation between experience working as a taxi driver before and multi-homing behaviors. The OLS regression model shown in Table 3.1 suggests that drivers who has been motorcycle taxi drivers

Table 3.1: Results of regression models predicting multi-homing behaviors of ride-hailing drivers in Jakarta based on previous occupation

| Variable | coef | std err | t | P> \|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **Intercept** | 0.1276*** | 0.012 | 10.735 | 0.000 | 0.104 | 0.151 |
| **Former taxi driver** | -0.0796** | 0.031 | -2.574 | 0.010 | -0.140 | -0.019 |
| **R-squared:** | 0.008 | **Adjusted R-squared:** | | | 0.007 | |

Table 3.2: Results of regression models predicting multi-homing behaviors of ride-hailing drivers in Jakarta based on whether being part-time ride-hailing drivers

| Variable | coef | std err | t | P> \|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **Intercept** | 0.1097*** | 0.012 | 9.258 | 0.000 | 0.086 | 0.133 |
| **Part-Time** | 0.0441 | 0.032 | 1.384 | 0.167 | -0.018 | 0.107 |
| **R-squared:** | 0.002 | **Adjusted R-squared:** | | | 0.001 | |

before are less likely to multi-home in Jakarta market.

Finally, we turn our attention to the results of our regression models exploring the impact of being a part-time ride-hailing driver on drivers' multihoming behaviors.

From the model results shown in Table 3.2, we find that part-time ride-hailing drivers are more likely to multi-home in our sample. However, it is a less significant predictor of multihoming behaviors compared to other variables above.

### 3.2.1 Earnings for Different Platforms

Our main finding concerns the earnings for different platforms. Figure 3-3 displays the hourly salary distribution of all survey respondents in Jakarta. The median respondent makes less than 20,000 IRB (approximately 1.4 USD) per hour, which is significantly less than the average hourly salary 120,175 IRB (approximately 8.4 USD) in Jakarta **Jakarta_hourly_salary**. This stands even when controlling for sociodemographics.

We estimate, among all single-homers, the correlation between working for GoJek and hourly salary in Table 3.4. We exclude the 3% of non-motorcycle drivers due to their significantly different earnings. The model results suggest that GoJek drivers

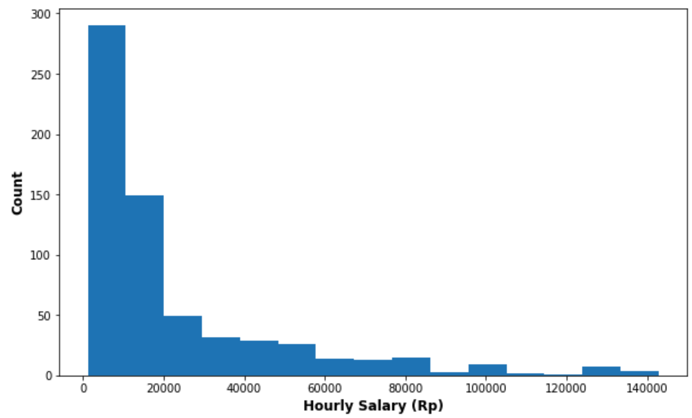



Figure 3-3: Hourly salary distribution of survey respondents (ride-hailing drivers in Jakarta)

earn significantly less than drivers from other platforms, 11,000 IRB (approximately 0.77 USD) less per hour. While a weak effect ($R^2 = 0.002$), it is significant.

For the same group of drivers, we consider the difference of total working hours between ride-hailing platforms. The model results in Table 3.3 show that GoJek drivers work significantly longer than drivers from other platforms, 1.0335 hours more per day. Both regression results imply that there existing a large switching cost between GoJek and other platforms (Grab and Shopee) in Jakarta ride-hailing market.

GoJek motorcycle drivers work longer hours but earn less compared to other drivers. We use Natural Language Processing tools to gain additional insights on what, besides earnings maximization, might drive the choices to work for Grab as compared to GoJek.

## 3.3 Free-Text

In this section, we use a topic model (Latent Dirichlet Analysis) to discover potential non-monetary reasons for the choice of TNC that drivers work for. All of the analysis is use answers to question 16 (compare Appendix A.1) “Anything else we should know about why you decided not to use multiple platforms?”.

We fine-tuned a natural language model BERT Devlin et al. 2019 trained on the Indonesian Wikipedia to predict the choice of platform worked for based on the

| | *Dependent variable:* | |
|---|---|---|
| | (1) | (2) |
| Intercept | 30713.817*** | 32141.158*** |
| | (1927.295) | (10919.276) |
| Q54[T.30 - 39] | | 599.344 |
| | | (9829.024) |
| Q54[T.30 - 39]:Q58[T.Yes] | | -1459.150 |
| | | (10215.505) |
| Q54[T.40 - 49] | | -5300.705 |
| | | (13705.572) |
| Q54[T.40 - 49]:Q58[T.Yes] | | 2649.267 |
| | | (14135.068) |
| Q54[T.Above 50] | | -655.081 |
| | | (17545.153) |
| Q54[T.Above 50]:Q58[T.Yes] | | 3628.347 |
| | | (19003.614) |
| Q54[T.Under 20] | | -689.077 |
| | | (4819.356) |
| Q54[T.Under 20]:Q58[T.Yes] | | -689.077 |
| | | (4819.356) |
| Q55[T.Male] | | -5813.134 |
| | | (4617.561) |
| Q55[T.Prefer not to answer] | | -17966.204 |
| | | (19349.857) |
| Q57[T.None of the above] | | -0.000 |
| | | (0.000) |
| Q57[T.S1 or higher] | | -5956.143 |
| | | (7460.034) |
| Q57[T.SD] | | 6745.220 |
| | | (8008.309) |
| Q57[T.SMA] | | -338.772 |
| | | (5669.506) |
| Q57[T.SMP] | | 1058.908 |
| | | (6598.707) |
| Q58[T.Yes] | | 5170.436 |
| | | (8691.783) |
| gojek[T.True] | -11001.520*** | -10506.619*** |
| | (2366.886) | (2426.682) |
| Observations | 552 | 552 |
| $R^2$ | 0.038 | 0.052 |
| Adjusted $R^2$ | 0.036 | 0.025 |
| Residual Std. Error | 26284.803(df = 550) | 26435.066(df = 536) |
| F Statistic | 21.605*** (df = 1.0; 550.0) | 1.942** (df = 15.0; 536.0) |
| *Note:* | | *p<0.1; **p<0.05; ***p<0.01 |



Table 3.3: Results of regression models predicting hourly earnings of platform drivers drivers in Jakarta based on whether working for GoJek

| | *Dependent variable:* | |
|---|---|---|
| | (1) | (2) |
| Intercept | 10.664*** | 8.328*** |
| | (0.162) | (0.874) |
| Q54[T.30 - 39] | | -0.761 |
| | | (0.791) |
| Q54[T.30 - 39]:Q58[T.Yes] | | 0.970 |
| | | (0.829) |
| Q54[T.40 - 49] | | -1.880* |
| | | (1.072) |
| Q54[T.40 - 49]:Q58[T.Yes] | | 1.916* |
| | | (1.116) |
| Q54[T.Above 50] | | -1.056 |
| | | (1.440) |
| Q54[T.Above 50]:Q58[T.Yes] | | 1.595 |
| | | (1.572) |
| Q54[T.Under 20] | | 1.204 |
| | | (2.642) |
| Q54[T.Under 20]:Q58[T.Yes] | | -2.903 |
| | | (2.785) |
| Q55[T.Male] | | 0.928** |
| | | (0.395) |
| Q55[T.Prefer not to answer] | | 1.378 |
| | | (1.523) |
| Q57[T.None of the above] | | -0.000 |
| | | (0.000) |
| Q57[T.S1 or higher] | | -0.198 |
| | | (0.603) |
| Q57[T.SD] | | 0.073 |
| | | (0.698) |
| Q57[T.SMA] | | -0.032 |
| | | (0.480) |
| Q57[T.SMP] | | 0.063 |
| | | (0.568) |
| Q58[T.Yes] | | 1.749** |
| | | (0.689) |
| gojek[T.True] | 1.033*** | 0.866*** |
| | (0.210) | (0.206) |
| Observations | 655 | 655 |
| $R^2$ | 0.036 | 0.136 |
| Adjusted $R^2$ | 0.034 | 0.114 |
| Residual Std. Error | 2.648(df = 653) | 2.536(df = 638) |
| F Statistic | 24.120*** (df = 1.0; 653.0) | 6.273*** (df = 16.0; 638.0) |
| *Note:* | | *p<0.1; **p<0.05; ***p<0.01 |



Table 3.4: Results of regression models predicting daily working hours of ride-hailing drivers in Jakarta based on whether working for GoJek

Table 3.5: Sum of Feature Importance in SHAP scores

answer to question 16. We list global feature importance in our training data set for the prediction of the platform in **??**.

The words most indicative of working for GoJek, the platform with lower reported earnigns, are

## 3.4 Discussion

We find significant earnings and working hours differences between single-homing drivers. The majority of respondents works for a platform where they report longer working hours and less hourly wages.

In our analysis of free-text answers, we find that drivers working for a lower-paying platform cite more frequently reasons for

This gives introductory evidence for features beyond matching and pricing, relating to the informational environment of drivers. The value of demand information, whose design we are going to study in chapter 5 is part of our analysis the next chapter.

## 3.5 Conclusion

# Chapter 4

# The Value of Information

While the last chapter investigated stated aspects of relevance for labor supply, this section shows the *potential* effects of platform design beyond matching and pricing in the case of information.

We present evidence that demand information can significantly increase earnings for platform drivers. The earnings gains are aligned with claims by companies that offer information services to drivers such as Gridwise or Surge that additional information can increase earnings, compare Weed 2019. We first describe the data used in our estimation in section 4.1. We then estimate driver earnings in optimized repositioning in section 4.2 and close with comparing this with driver income under non-optimized behavior and discuss our results in section 4.3.

## 4.1 Description of Data

The DiDi KDD Challenge 2020, sponsored by TNC DiDi, consisted of two challenges. An order dispatch challenge, that asked to match orders to drivers with the goal of maximizing the value of all orders completed, and a driver repositioning challenge that asked to reposition a small number of drivers to maximize the earnings rate of this group of drivers. We will work with data and submissions of the latter challenge.

### 4.1.1 Annotated Order Data

We use a dataset of 14,131,874 orders dispatched in Chengdu, China, in November 2016. An order consists of an origin-destination pair, a timestamp for beginning and end time, an identifier for the driver that took the order, and a number of *reward units*. Reward units are a proxy for driver income provided by DiDi. As we argue statistically below, reward units are linear transformations of real currency. We recover an approximate translation of reward units into RMB, the local currency. As of the submission time of this thesis, the data is not available anymore, and the terms do not allow hosting of the data. Hence, we cannot provide replication for this part of our analysis.

### 4.1.2 Evaluation in Paper

The KDD 2020 reinforcement learning challenge solves an optimal repositioning problem together with an optimal order dispatch problem. We make use of the repositioning scores in the paper Tang et al. 2021, which solves two dynamic planning problems

For the order dispatch algorithm, the system's observable state consists of a list of orders to be dispatched at each point in time. An order is an origin-destination pair, and for each driver timestamps of *estimated* order pickup and arrival times, as well as the distance for the driver. The available actions are matchings of drivers to orders, which are one-to-one, as the challenge does not consider carpooling. Orders need to be dispatched within two seconds, otherwise they are lost. (We specify the complete state in in section A.2. As part of the state transitions, there is estimated data for cancellation probabilities when sending a driver to pick up orders from different locations.

The objective in the order dispatch challenge is to maximize average driver income.

The system's observable state for the driver repositioning challenge, which we are using in this chapter, consists of a set of drivers to be repositioned. A driver becomes eligible for repositioning after five minutes of being idle if they are in a small group of *treated drivers*. In addition to a timestamp, the only information given on drivers is a coarse position on a hexagonal grid of the urban area of Chengdu. The actions are matchings of drivers to destination locations. Drivers move to their destination at 3m/s in the spherical/great arc distance. Idle drivers move according to historical transition probabilities between regions before being idle for 5 minutes or non-targeted.

The objective in the driver repositioning problem is to maximize mean driver income rate for drivers. Denote $L_k^n(\pi)$ the online time for driver $k$ at day $n$ in hours.[1] Denote driver $k$'d income on day $n$ under policy $\pi$ by $J_k^n(\pi)$. For the set of all targeted drivers $K$ and days $N$, the goal of the challenge is to maximize the earnings rate of repositioned drivers,

$$\frac{1}{|K|}\sum_{k\in K}\frac{\sum_{n\in N} J_k^n(\pi)}{\sum_{n\in N} L_k^n(\pi)}. \tag{4.1}$$

The optimal repositioning score for drivers is informative for the value of demand information for drivers. If drivers have sufficient information about the demand throughout the city, they can reposition themselves to another location in expectation of getting higher earnings. In contrast to the platform's incentive of maximizing the volume of orders completed, the repositioning challenges maximizes driver earnings taking into account periods 1, 2, and 3, and hence reflects the earnings rate that a perfectly informed driver would receive.

In this argument, the smallness number of repositioned drivers is important. A challenge for drivers, but not for the platform, is that a knowledge on high demand in some area might lead to congestive effects—too many drivers enter high-demand areas. We will discuss this point more thoroughly in chapter 5.

Tang et al. 2021 finds, compare their Fig. 3(a), that a policy that samples transition probabilities for a time of day from historical data reaches a repositioning score of about 8.5, whereas the earnings from their best tested repositioning policy gave a score of 9.

[1] While the existing documentation do not specify the unit of this measure, our calculations below show that assuming online time is measured in hours leads to correct results.

| | Base Price | Price/km | Price/min |
|---|---|---|---|
| $22:00-07:00$ | 12.40 | 2.70 | 0.48 |
| $07:00-10:00$ | 12.40 | 2.55 | 0.48 |
| $10:00-16:00$ | 11.40 | 1.95 | 0.42 |
| $16:00-19:00$ | 12.40 | 2.33 | 0.48 |
| $19:00-22:00$ | 11.40 | 1.95 | 0.42 |

Table 4.1: Earnings for DiDi Economic Express drivers.

### 4.1.3 Earnings Table

We use a publicly available table for earnings from 2021 Icauto.com 2021. In 2017, DiDi introduced two tiers of service for their drivers, Economic Express and Economic Premium. After personal conversations with experts in the market, we will use the earnings for the Economic Express tier as proxy for November 2016 earnings. We reproduce the table in **??**.

The earnings table contains, dependent of the time of the day, a base price $b_t$, earnings per kilometer $l_t$ from the third cilometer and minute driven $d_t$ from the fifth minute. At the time of the data, November 2016, additionally, a surge multiplier $s_t$ which is not part of the table. The total earnings $e_t$ for a ride of $L$ kilometers and duration $D$ are given by

$$e_t = s_t(b_t + l_t L + d_t D). \tag{4.2}$$

## 4.2 Estimation

We will estimate a linear relationship of earnings and reward units, and test the linearity of the relationship.

### 4.2.1 Estimation

We estimate the equation (4.2),

$$e_t = b_t + l_t L + d_t D.$$

for $t$ dependent on the different time periods in Table 4.1, $t \in \{22{:}00\text{-}07{:}00, 07{:}00\text{-}10{:}00, 10{:}00\text{-}16{:}00, 16{:}0$
as well as a simpler model

$$e = b + lL + dD$$

which disregards these periods. We use the Google Maps API to estimate driving distance associated with trips and the DiDi provided duration as duration. The estimation results are presented in Table 4.2. We find highly significant and consistent estimates on the coefficients for distance and duration and a very strong effect ($R^2 = 0.887$). The Harvey-Collier test for linearity of a model Harvey and Collier 1977, we are unable to reject the Null Hypothesis of linearity ($p = 0.56$). This means, that our assumption of a linear relationship between reward units and earnings cannot be rejected in the data. The $F$-statistic of the smaller is much larger than the one of the larger model (2911.7 vs. 1074.8), which means that all differences in hour-dependent pay are non-significant. As we do not have access to data from Changdu, we cannot determine whether this is a result of surge prices cancelling earnings variations or caused by the absence of price variations between hours. We will proceed by using the smaller model.

This allows us to estimate driver earnings for orders. By comparing the base price, per-kilometer and per-minute prices with our earnings, we get three estimates of RMB/reward unit: $11.4/0.751 = 15.2$, $1.95/0.29 = 6.7$, and $0.42/0.0065 = 6.5$. We will, in our estimations, use the average of the latter two, 6.6, assuming that the base price for drivers in Chengdu has decreased since 2017. Note that this is a very conservative estimate given that the rate would be much higher when calibrating against the base rate.

We assume that the earnings rate as given in (4.1) are estimated per hour. Given

| | Dependent variable: Reward Units | |
|---|---|---|
| | (1) | (2) |
| Intercept | 0.751*** | 0.703*** |
| | (0.042) | (0.041) |
| distance | 0.290*** | 0.317*** |
| | (0.014) | (0.076) |
| duration | 0.065*** | 0.116*** |
| | (0.005) | (0.041) |
| hour[T.22-7]:distance | | -0.043 |
| | | (0.077) |
| hour[T.22-7]:duration | | -0.050 |
| | | (0.041) |
| hour[T.7-10]:distance | | -0.067 |
| | | (0.089) |
| hour[T.7-10]:duration | | -0.037 |
| | | (0.043) |
| hour[T.offhour]:distance | | 0.003 |
| | | (0.078) |
| hour[T.offhour]:duration | | -0.057 |
| | | (0.041) |
| Observations | 3,997 | 3,997 |
| $R^2$ | 0.883 | 0.894 |
| Adjusted $R^2$ | 0.883 | 0.894 |
| *Note:* | | $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$ |

Table 4.2: Reward unit regression

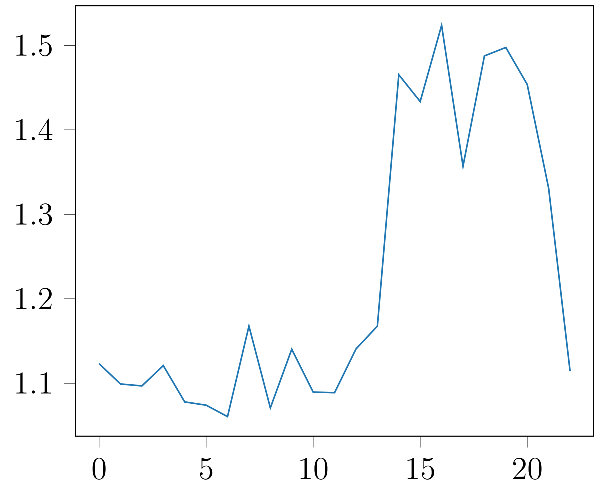



Figure 4-1: Estimate average surge multipliers throughout the day.

the data available, we are not able to calibrate whether this is correct.

Using our estimation, we can estimate the performance of the human expert policy as 8.5reward units/hour $\cdot$ $6.6\frac{\text{RMB}}{\text{reward units}} = 56.1\text{RMB}$. This lies in the band of first and second tier Chinese cities, for which average hourly income of 40-60 RMB are reported DayDayNews 2019. The optimal repositioning gives earnings of 9reward units/hour $\cdot$ $6.6\frac{\text{RMB}}{\text{reward units}} = 59.4\text{RMB}$ which is on the upper end of the interval, which means a significant earnings increase of 3.3 RMB per hour.

As another check of our model fit and immediate use of our model, we infer the distribution of surge prices throughout the day. We use the estimated values Table 4.2 to calculate a quantity of raw earnings $r_t = \bar{l}0.290L + \bar{d}0.065D$. We then regress earnings

$$e_t = \beta_t r_t$$

and find the surge multipliers given in Figure 4-1, which emphasize the evening rush hour, but do not show as pronounced of a morning rush hour.

## 4.3 Discussion

We find that optimal driver repositioning has a significant, but not too large effect on driver earnings if only a few drivers are treated to this additional information.

In a model with more drivers being treated, informed drivers might *crowd out* each other, and earnings differential might be smaller.

This does not mean that information design is not possible to improve platforms, but that, for each trip, only a subset of drivers may be informed of it. This robust property is one of our findings in the next chapter.

# Chapter 5

# Information Design

> The information design problem has a literal interpretation: there really is an information designer (or mediator, or sender) who can commit to provide extra information to players to serve her own interests. While the commitment assumption may be problematic in many settings, it provides a useful benchmark.
>
> —Bergemann and S. E. Morris n.d.

In this section, we study *pure* information design for platform drivers. We refer to it as *pure* as we study information provision while keeping the driver payoffs fixed. We show several limitations for the platform: Drivers, in equilibrium, ignore information if it cannot commit to not giving a particular driver some part of the information, it gives

## 5.1 Model

We present here, in a self-contained manner, a model of information design with an *omniscient* designer. Omniscience refers to the TNC, the information designer, having strictly more information than all drivers, the agents in our game. This assumption is not without loss, as existing challenges to improve Operations Research models, e.g., in last-mile logistics, compare Amazon Inc and Technology 2021, show. Our results, qualitatively do not differ, as we show in appendix **??**.

We consider a finite set of agents $i = 1, 2, \ldots, n$. There are states of the world $\Theta$. We consider a basic game given by, for each player $i$ an action set $A_i$ and a utility function

$$u_i \colon A \times \Theta \to \mathbb{R}$$

where $A = A_1 \times \cdots \times A_n$ and a prior distribution $F \in \Delta(\Theta)$ which we assume has full support. We assume that this prior is shared by all players and the designer. $((A_i, u_i)_{i=1,2,\ldots,n}, F)$ hence specifies a standard game.

Our main model will consider $n = 2$ drivers that simultaneously decide whether to drive to a part of a city ("go") or to follow other business on or off the platform ("not go"), i.e. $A_1 = A_2 = \{\text{go}, \text{not go}\}$. The utility functions are given by the tables in Figure 5-1. We assume that the prior is that with probability $l \in [0, 1]$ demand is realized. This corresponds to an outside option for the drivers of $\sigma$ and a cost to go to the high-demand area of $\varepsilon$.

We call this game the *basic game*.

The information designer has a utility function

$$v \colon A \times \Theta \to \mathbb{R}.$$

In our example, the information designer is a TNC. It will be an information designer, and will get utility 1 if at least one driver is able to satisfy demand, otherwise a utility of 0.

The platform can designs an *information structure* for each driver. This can be an arbitrary piece of information that drivers take into account. Under the assumption that the platform is omniscient and can commit to a rule, a *revelation*-type argument can be used to simplify the design space: For any messages that the designer sends to agents, given that the platform knows the best response of the agents to this information, the platform can recommend the agents their best responses. Hence,

| Driver 1 \ Driver 2 | go | not go |
| --- | --- | --- |
| go | $\frac{1}{2}-\varepsilon$, $\frac{1}{2}-\varepsilon$ | $1-\varepsilon$, $\sigma$ |
| not go | $\sigma$, $1-\varepsilon$ | $\sigma$, $\sigma$ |

(a) Demand, $\theta = 1$

| Driver 1 \ Driver 2 | go | not go |
| --- | --- | --- |
| go | $-\varepsilon$, $-\varepsilon$ | $-\varepsilon$, $\sigma$ |
| not go | $\sigma$, $-\varepsilon$ | $\sigma$, $\sigma$ |

(b) No, $\theta = 0$

Figure 5-1: Payoff Matrix for the two-driver game.

any information structure has a particularly simple structure as *recommendations*

$$\sigma \colon \Theta \mapsto A.$$

For a formal statement and a proof of the revelation function, see Bergemann and S. Morris 2019.

The timeline of the model is:

1. The information designer commits to an information structure $S = \sigma$.

2. The state of the world is realized.

3. The players each receive their action recommendations $\sigma_i(\theta)$.

4. The agents select their action $a_i$.

5. The payoffs are realized.

Mathematically, the remaining constraint of *obedience* can be written as

$$\sum_{a_i,\theta} u_i((a_i, a_{-i}); \theta)\sigma((a_i, a_{-i})|\theta)F(\theta) \geq \sum_{a_i,\theta} u_i((a_i', a_{-i}); \theta)\sigma((a_i, a_{-i})|\theta)F(\theta)$$

for any $a_i' \in A_i$. A profile of decision rules $\sigma$ that is obedient is called a Bayes Correlated Equilibrium (BCE). $\sigma$ is a coarse correlated equilibrium if an agent that

| $\theta = 0$ | go | not go |
|---|---|---|
| go | $\sigma$ | x |
| not go | x | 0 |

| $\theta = 0$ | go | not go |
|---|---|---|
| go | $\sigma$ | x |
| not go | x | 0 |

Table 5.1: Optimal Information Design

| $\theta = 0$ | go | not go |
|---|---|---|
| go | $r_0$ | $p_0 - r_0$ |
| not go | $p_0 - r_0$ | $1 + r_0 - 2p_0$ |

| $\theta = 0$ | go | not go |
|---|---|---|
| go | $r_1$ | $p_1 - r_1$ |
| not go | $p_1 - r_1$ | $1 + r_1 - 2p_1$ |

Table 5.2: Parameterized Information Design

updates their information on the state of the world given their recommendation does not want to deviate from said recommendation.

The platform's maximizes a utility function

$$V(\sigma) = \sum_{a,t,\theta} v(a, \theta)\sigma(a|\theta)F(\theta). \tag{5.1}$$

The platform's optimal information design is maximizing (5.1) among all BCEs $\sigma$.

We make the assumption that getting a ride, even after driving to a high-demand area is better than the outside option, but the outside option is preferrable to to getting a ride with $\frac{1}{2}$ probability, $1 - \varepsilon \geq \sigma \geq \frac{1}{2} - \varepsilon$.

**Proposition 5.1.** *An optimal information design for the platform is given by*

Conditional on the good state, this minimizes the correlation of

*Proof.* Observe that every recommendation $\sigma$ that is not symmetric, i.e. $\sigma_1(\theta) \overset{d}{\neq} \sigma_2(\theta)$ for some $\theta$, because the game is symmetric, also $\sigma'$, $\sigma_2' = \sigma_1$, $\sigma_1' = \sigma_2$, is a Bayes-correlated equilibrium, and so is ${}^{(\sigma+\sigma')}/{}_2$. Therefore, an optimal information design for the platform can be chosen to be symmetric.

Hence, the optimal information structure is characterized by the probability the probability $r_\theta$ that both drivers drive to the area and by $p_\theta$, the probability that each of the drivers goes. This gives rise to the information structure in Table 5.3. For all entried in Table 5.3 to be probabilities, it needs to hold that

$$\max\{0, 2p_\theta - 1\} \leq r_\theta \leq p_\theta$$

For $a = \text{go}$, the obedience constraint is

$$lp_0\sigma$$

The platform would like to minimize the probability that the demand cannot be met, i.e. minimize $l(1-p_B)^2 + (1-l)(1-p_B)^2$. □

### 5.1.1 Cheap Talk

We compare this to a model in which the platform cannot commit ot In cheap talk, the platform discloses some information to a driver, which, given the update, updates their behavior. More concretely, there is an abstract set of *messages* $m \in \mathcal{M}$ that the platform can send. The players receive the message and play optimally. In particular, the timeline of the game is:

1. The demand $\theta$ at the area is realized.

2. The platform decides to send messages $(m_1, m_2)$ to the drivers.

3. The drivers decide to drive or not drive to the area.

We solve for perfect Bayesian (signalling) equilibria of this game.

### 5.1.2 Public and Private

We call a recommendation $\sigma$ public if $\sigma_1(\theta) \overset{\text{a.s.}}{=} \sigma_2(\theta)$ for $\theta = 0, 1$,i.e. if both agents receive the same information. In this case, this leads to the following result:

**Proposition 5.2.** *The optimal information design when the platform is restricted to public messages is given in* **??**

| $\theta = 0$ | go | not go |
|---|---|---|
| go | $r_0$ | $p_0 - r_0$ |
| not go | $p_0 - r_0$ | $1 + r_0 - 2p_0$ |

| $\theta = 0$ | go | not go |
|---|---|---|
| go | $r_1$ | $p_1 - r_1$ |
| not go | $p_1 - r_1$ | $1 + r_1 - 2p_1$ |

Table 5.3: Parameterized Information Design

### 5.1.3 Results

**Theorem 5.3.** *In the cheap talk model, Perfect Bayesian equilibria have the following form:*

1. *If $\sigma < \frac{1}{2} - \varepsilon$ or $\sigma > 1 - \varepsilon$, then the platform is indifferent between any message, the drivers will visit their dominant choice.*

2. *If $\frac{1}{2} - \varepsilon \leq \sigma \leq 1 - \varepsilon$, the platform is indifferent between any message, and the drivers ignore the information.*

We show a proof to this statement in **??** If the general outside option is very low ($\sigma < \frac{1}{2} - \varepsilon$) or very high ($\sigma > 1 - \varepsilon$), drivers go to the area in hope to get a ride or stay away from it, respectively, even without demand information. The platform is indifferent between any demand information given that it won't influence the driver's decisions. This is an example where agents know that there is a high

In the other cases, the platform cannot transmit any information. The

In this environment, hence, if the not internalized cost from

## 5.2 Inefficiencies

After analysing the predictions, we identify two sources of inefficiencies: a potential lack of commitment (a comparison between our cheap talk and the commitment solution) and a potential lack of being able to provide asymmetric information.

### 5.2.1 Inefficiencies through lack of commitment

**Proposition 5.4.** *The cheap talk version of the game has lower welfare. This statement is uniform: Both platform and driver surplus are lower in the cheap talk game.*

The statement shows that neither side benefits from limited commitment of the platform. The drivers have to rely on less information, and cannot trust the platform. The equilibrium is "babbling".

### 5.2.2 Inefficiencies through public information

Comparing our second two theorems, we find the following loss from public information.

**Theorem 5.5.** *In the commitment regime, only allowing public information reduces platform and driver surplus. In the cheap talk regime, this restriction increases welfare*

This result might look somewhat surprising.

## 5.3 Discussion and Remedies

Given the inefficiencies observed in the last section, some remedies for the market should be considered. We introduce two potential solutions to the platform's commitment problem, and one solution to the allocation problem.

### 5.3.1 Commitment via third parties

A main driver of the inefficiencies in our model was that there is a conflict of incentives between the platform and the drivers. In this environment, if drivers "pay" the platform, there is not much.

### 5.3.2 Commitment via reputation

A second opportunity for the platform to commit is via reputation. This assumes that the game of information provision studied in this section is repeated sufficiently often.

**Theorem 5.6** (Friedman 1973)**.** *In an infinitely repeated game with sufficiently patient agents, all individually rational payoffs can be achieved as a result of equilibrium play.*

This statement is allowed by agents playing a "punishment" equilibrium from the platform. It is unlikely in the setting of platform drivers that they can individually punish the platform, which, in return, does not incentivize the platform to give recommendations according to a commitment problem.

### 5.3.3 Approximate Efficiency of Public Information to Few Drivers

For our last statement, we will need a more general model. There are $n$ drivers, which learn a state $\theta \in \{0, 1, 2, \ldots, n\}$. Drivers get a utility $1-\varepsilon$ if matched to a ride and they get a utility of $-\varepsilon$ if not matched. We assume that $n$ is large, and a small fraction of drivers can get a public message.

**Proposition 5.7.** *Assume that drivers are*

# Chapter 6

# Conclusion and Policy Recommendations

> There' s fragmentation and a lack of transparency of information on what' s happening in real time and what' s going to happen. We' re connecting bread crumbs across data areas. —Ryan Green, quoted in Weed 2019

This thesis studied the relevance and optimal design of information provided to platform drivers. Our argument proceeded in three steps: chapter 3 showed that drivers deviate from earnings maximization in their labor supply decisions. chapter 4 showed that there are high potential earnings gains from information on demand. We found in chapter 5 that a lack of commitment and public information decrease platform and driver earning when drivers are strategic. In this final chapter, we consider three stakeholder groups, platforms, regulators and transportation engineers and list open problems and recommendations for each of these groups.

## 6.1 The Commitment Challenge for Platforms

We observed in chapter 5 that the limited commitment of the platform can lead to, in equilibrium, limits to how much information the platform can transmit such that drivers will act on the information. More specifically, the platform's commitment problem arises when it has high demand and, in an optimal information design, would show the information of high demand only to a subset of drivers. Without commitment, the platform has incentives to reveal this information to more drivers, which would maximize the likelihood that a rider is picked up, but, in turn, makes it less attractive for drivers to follow this recommendation.

As the software underlying a TNC does mostly not change, the platform has an opportunity to mechanistically commit to a particular information structure. To reach this goal, however, it is crucial that drivers *understand* this commitment. There are several ways to give such asymmetry, which we outline next, in order of decreasing applicability inside of the platform.

### 6.1.1 Designating some drivers as informationally advantaged

If some drivers are openly designated as having access to additional information compared to other drivers, commitment to only giving some drivers information is transparent and easily communicable with drivers. Whether linked to performance on the platform or not, the main problem of a lack of commitment as identified in chapter 5 disappear if the drivers being recommended to go to a high-demand area do not foresee the congestion in this place.

### 6.1.2 Audits

A second opportunity for platforms is to achieve their commitment is via audits and reporting from trustworthy sources. This could come, for example, via a publication of the numbers of drivers informed of high demand in an area.

### 6.1.3 External Aggregators

A third approach is what we observed in the marketplace already in chapter 1. Third parties such as Gridwise or Surge aggregate and provide information. In contrast to a TNC, these companies do, depending on their business model, not have incentives that are misaligned with platform drivers and can hence provide optimal information: The question becomes one of optimal cooperative, and not strategic information provision.

We highlight that in our analysis, we disregard the pricing aspect of the platform design problem. Pricing in forms of costly signaling, such as paying small amounts to drivers to go to a particular area, could alleviate commitment challenges as well, but are outside of the scope of this thesis.

## 6.2 Challenges for Transportation Regulators

As a prerequisite for understanding the regulation of information provision, understanding the informational needs of a regulator is crucial.

The access to information from TNCs is regulated under different legislation. New York, as an example, gives their regulator, the Taxi and Limousine Commission (TLC) particularly strong powers: Section 2302 of the Charter of New York City empowers them to information access to origin-destination pairs and fares. In its executive practice, the TLC publishes data on fares set by transportation platforms and has access to disaggregated data. Importantly in the case of New York City, the power of the regulator comes from times pre-dating TNCs, when the regulator purely regulated taxis and limousines.

Other cities have different regulations, and might face platform emigration when trying to get more data access. The case of Austin, that faced emigration by Uber and Lyft when trying to introduce driver background checks, Texas showed that potential regulation might lead to platforms abandoning some areas, compare Zeitlin 2019.

The informational demands on regulating information to drivers are much more stringent than regulation on earnings or matching: In addition to trips and fares, regulators need to get access to which information is shown to drivers, and, hence, insights into the platform.

We propose, in three steps, a path forward that would allow regulation of information provided to drivers.

### 6.2.1 Information Provision as a Preliminary Agreement

A first important step is that, without changing the payment structure of TNCs (no payment in phase 2), to re-classify information on demand and offers to drivers as parts of a negotiation between the platform and riders. This can have legal implications by making statements of the platform subject to pre-contractual liability, compare Myerson 1983. In particular, it would allow the driver to contest the veracity of demand information.

### 6.2.2 Using Existing Information Access to Track Information Provision to Drivers

The classification of information provision to drivers as pre-contractual negotiation would also allow existing regulators to require more information by platforms. It could then be that the list of drivers receiving information, the *targeted* drivers, are shown along with the information.

### 6.2.3 Trip Level vs. Aggregate Level Fairness

Having access to data on information provision to platform drivers would allow to test one of the main challenges outlined in our theoretical analysis in chapter 5: Asymmetry. Regulators would be able to regulate on which drivers get which demand information at which time.

## 6.3 Challenges for Transportation Engineers

The challenges for TNCs and regulators come with challenges for transportation engineers, with which we conclude this thesis.

### 6.3.1 Dynamic Rebalancing with Information Constraints

Taking into account the effect of information on the reaction of drivers gives rise to challenges for operations researchers. As we saw in chapter 5, information provision to drivers needs to balance the expected congestion caused by information with the likelihood of satisfying a demand for rides. Varying between revenue maximizing and welfare maximizing platforms in information provision might give rise to interesting parameterized algorithmic challenges.

### 6.3.2 Information Design as Market Design

A second problem for engineers is the integration of information into pricing. Effectively targeting information to drivers with payments incentivizing them might be more efficient than information design alone. The use of combined pricing-information designs might help the functioning of urban mobility systems.

### 6.3.3 Data Specifications

To allow for effective regulation, the development of exchange formats for information provision is a relevant challenge. As origin-destination pairs, fares and time stamps are standard data structures for matching and pricing problems, a data format for recommendations to drivers, might allow for the inclusion of such information into high-performance algorithms.

Depending on the development and ownership structure of Autonomous Vehicles (AVs), such specifications and information designs might even be relevant in a future without (human) drivers. Coordinating the demand of AVs through proper information design requires effective data specifications to track them.

### 6.3.4 Operationalizing Fairness in Information Provision

A last challenge is the operationalization of fairness in information provision and its interlinkage with pricing and matching. Describing the expected earnings for different groups based on their differential treatment by a TNC algorithm, be it in terms of information, matching, or pricing can allow policymakers and the public to discuss equity issues and distributional consequences of platforms, in addition to other (environmental and congestion, compare Diao, Kong, and Zhao 2021) externalities of TNCs.

We leave these challenges for the future, in hope of informed and efficient urban mobility environments.

# Appendix A

# Supplemental Material

## A.1 Driver Survey

### A.1.1 Basic Driving Information and Behavior

1. Which ride hailing platform companies did you work for in the last 30 days? (Select all that apply) (a) Grab (b) GoJek (c) Shopee

2. What type of vehicle do you use the most for work? (a) Car (b) Motorbike

3. How did you acquire this vehicle?

   (a) I or a family member purchased it with full payment

   (b) I or a family member purchased it on credit

   (c) I leased it through the ride-hailing platform

   (d) I leased it myself through a third party

   (e) Others (please specify)

4. Did you drive for a taxi company in the last 30 days, e.g. for bluebird? (a) Yes (b) No

5. Did you work in any other jobs in addition to either ride-hailing and/or taxi driving in the last 30 days? (a) Yes (please specify job type) (b) No

6. Did this job give you more money per hour than money for an hour you are on bid? (a) Yes (b) No
7. Do you like this job more or less than working for Grab/GoJek/Shopee? (a) Yes (b) No (c) Indifferent
8. What job did you have before starting as an ojol? (a) Opang (b) Other (c) Ojol is my first job
9. Did this job give you more money per hour than money for an hour you are on bid? (a) Yes (b) No
10. Why did you change jobs? (Select all that apply)
    (a) Better pay
    (b) Lost my previous job
    (c) More flexible hours
    (d) My friends were joining ojol
    (e) I like this job better
    (f) More respected/higher status as an ojol
    (g) Other (Please specify)
11. Where do you prefer to wait when not on an order (most frequently)?
    (a) Wait in the area where last ride/order ended
    (b) Drive to nearest hangout/resting spots with other drivers
    (c) Driver to high demand areas even if I have to drive there
    (d) Others (please specify)
12. When do you usually stop working (most frequently)?
    (a) When I reach a certain income
    (b) When I get enough points/tupo

(c) At a set time I set for myself

(d) When I get tired

(e) Others (please specify)

### A.1.2 Multihoming Behavior

**For Non-Multi-Homers**

13. Why do you not use multiple platforms and only use one platform? Check whether you agree with the following statements.

    (a) I have enough work on one so do not need to work for both (Disagree/Agree)

    (b) I fear retaliation by the platform (Disagree/Agree)

    (c) I understand one system much better than the other (Disagree/Agree)

    (d) I feel loyal to my platform (Disagree/Agree)

    (e) I want to maintain my ranking on one of the platform (Disagree/Agree)

    (f) I don’t like the other platform (Disagree/Agree)

    (g) My friends/community are on this platform (Disagree/Agree)

    (h) I want to concentrate/focus on one platform for better performance (Disagree/Agree)

14. Would the following factors be important in your decision to start working for multiple platforms at the same time?

    (a) Higher bonus on the other platform (Yes/No)

    (b) Higher income on the other platform (Yes/No)

    (c) No penalty for multihoming (Yes/No)

    (d) My friends/community began to multihome (Yes/No)

    (e) More orders/demand on the other platform (Yes/No)

    (f) Understanding the other system better (Yes/No)

15. If you were to leave your current platform and ONLY WORK for the other platform, would the following factors be important in the decision to switch?
    (a) Higher bonus on the other platform (Yes/No)
    (b) Higher Income on the other platform (Yes/No)
    (c) More orders/demand on the other platform (Yes/No)
    (d) All your friends/community shifting to the other platform (Yes/No)
    (e) Understanding the other system better (Yes/No)
    (f) None of these will make me change to the other platform and leave my current application (Yes/No)
16. Anything else we should know about why you decided to not use multiple platforms?

**For Multi-Homers**

17. How do you usually switch between multiple companies?
    (a) Have multiple phones open at the same time
    (b) I block times, i.e. have one company open on my phone for each time period
    (c) Have multiple apps running on the same phone throughout the day
    (d) Only check the other company when I don't get orders for some time
    (e) Others (please specify)
18. What percentage of your working day do you have multiple phones open at the same time? (0 - 100)
19. Were the following factors important in your decision to start working for multiple platforms?
    (a) Getting a higher bonus on one of the platforms (Yes/No)

(b) Making more income on one of the platform (Yes/No)

(c) Knowing there is no penalty for multihoming (Yes/No)

(d) More demand on one of the platforms (Yes/No)

20. Any other reasons you started working for multiple platforms?

21. When working for multiple platforms, are you worried about the following?

    (a) I fear being penalized by platforms for multihoming (Yes/No)

    (b) I understand one system much better than the other (Yes/No)

    (c) I feel disloyal to the company working for both (Yes/No)

    (d) I feel disloyal to the my friends/community working for both (Yes/No)

    (e) Working for both platforms is distracting/I can't focus on one (Yes/No)

    (f) It is difficult to maintain ranking on one platform (Yes/No)

22. What would make you work EXCLUSIVELY for a platform and not multihome? (Select all that apply)

    (a) Higher bonus on this platform

    (b) Higher Income on this platform

    (c) More orders/demand on this platform

    (d) No penalty to multihoming

    (e) All my friends/community shifting to this platform

    (f) Others (please specify)

    (g) None of these will make me change to working for only one platform

23. Anything else we should know about your decision to use multiple platforms?

### A.1.3 Driving Activities

**Note.** This set of questions is also asked for Grab and GoJek if respondents select Grab or GoJek in Question 1. We only show questions for Shopee in this section, which has one additional Shopee-specific question compared to Grab and GoJek.

**For the next set of questions think about your activity only for Shopee for the last 30 days**

24. Which Shopee service did you take most orders from? (a) Food (b) Other (c) No specialization

25. (Shopee-specific question) What has changed when Shopee appears?

    (a) I drive less for GoJek

    (b) I drive less for Grab

    (c) I drive more for GoJek

    (d) I drive more for Grab

    (e) No changes

26. What type of area do you spend most of your online search time for Shopee? (a) Central areas (b) Outskirts

27. Out of the following, Which area do you spend most of your online/search time for Shopee in? (a) North Jakarta (b) South Jakarta (c) East Jakarta (d) West Jakarta (e) Central Jakarta (f) Other area (Bodetabek)

28. What days a week do you usually drive for Shopee at least one trip? (Select all that apply) (a) Monday (b) Tuesday (c) Wednesday (d) Thursday (e) Friday (f) Saturday (g) Sunday

29. How many hours do you typically work for Shopee on these days?

    (a) Weekdays (0 - 24)

(b) Weekends (0 - 24)

30. For the last two weeks: What has been your average daily salary from Shopee on weekdays? (in Rp)

31. For the last two weeks: What has been your average daily salary from Shopee on Saturdays and Sundays? (in Rp)

32. How many kilometers have you driven on average daily for Shopee when on bid [If more than 300, move slide to maximum value]? (0 - 300 KM)

33. How satisfied are you with Shopee's:

    (a) Bonus Scheme (1 - 5)

    (b) Daily Income you make on the platform (1 - 5)

    (c) Matching system (1 - 5)

    (d) Responsiveness to driver complaints/problems (1 - 5)

34. When working for Shopee, approximately how many days do you hit your bonus? (a) All the time (b) Around half the time (c) Less than half the time

### A.1.4 Sociodemographic Information

35. What is your age? (a) Under 20 (b) 20 - 29 (c) 30 - 39 (d) 40 - 49 (e) Above 50

36. What is your gender? (a) Male (b) Female (c) Non-binary / third gender (d) Prefer not to answer

37. What is the highest degree you obtained? (a) SD (b) SMP (c) SMA (d) D3 (e) S1 or higher (f) None of the above

38. Is Grab/GoJek your main source of income? (a) Yes (b) No

39. How many people does your income support (not including yourself)? (0 - 10)

40. Approximately how much do you spend on food each week? (in Rp)

41. How much do you spend on rent each week? (in Rp)

42. What is your kecamatan?

43. What is your kelurahan?

## A.2 Decision Environment for Reinforcement Learning Agents

Here, we describe the observed state and action for the Reinforcement Learning challenge, compare Xie et al. 2020.

### A.2.1 Dispatch Challenge

The order dispatch, a matching challenge, observes as an observable state for each incoming order and each idle driver in the platform during each time step

- The estimated distance from driver to order
- The origin and destination of the order
- The estimated order start and completion time assuming the order was matched to the driver
- The reward units as an earnings proxy for the driver.

The state of the system contains all driver's locations at each time.

### A.2.2 Repositioning Challenge

The repositioning challenge observes as an observable state for each driver in a *targeted* group of drivers that is idle for at least 5 minutes and each time step at which repositioning is possible a coarse location of the driver on a coarse grid of Chengdu.

# Bibliography


[ACC19] Susan Athey, Juan Camilo Castillo, and Bharat Chandar. "Service Quality in the Gig Economy: Empirical Evidence about Driving Quality at Uber". In: *SSRN Electron. J.* (2019). ISSN: 1556-5068. DOI: 10.2139/ssrn.3499781.

[Ard+21] Lady Arden et al. "Uber BV and others ( Appellants ) v Aslam and others ( Respondents ) [ 2021 ] UKSC 5 On appeal from : [ 2018 ] EWCA Civ". In: (2021).

[AS20] Ali Aouad and Ömer Saritaç. "Dynamic Stochastic Matching under Limited Time". In: *EC 2020 - Proc. 21st ACM Conf. Econ. Comput.* (2020), pp. 789–790. DOI: 10.1145/3391403.3399524.

[Ash+20] Peyman Ashkrof et al. "Research in Transportation Business & Management Understanding ride-sourcing drivers ' behaviour and preferences : Insights from focus groups analysis". In: *Res. Transp. Bus. Manag.* 37.July (2020), p. 100516. ISSN: 2210-5395. DOI: 10.1016/j.rtbm.2020.100516. URL: https://doi.org/10.1016/j.rtbm.2020.100516.

[AT21] Amazon Inc and Massachusetts Institute of Technology. *Last-Mile Routing Challenge*. 2021. URL: https://routingchallenge.mit.edu/.

[Bar+20] Damian Barabonkov et al. "Simulating and Evaluating Rebalancing Strategies for Dockless Bike-Sharing Systems". In: (2020), pp. 1–12. arXiv: 2004.11565. URL: http://arxiv.org/abs/2004.11565.

[BBM20] Dirk Bergemann, Benjamin Brooks, and Stephen Edward Morris. "Competition and Public Information: A Note". In: *SSRN Electron. J.* 2234 (2020). ISSN: 1556-5068. DOI: 10.2139/ssrn.3599648.

[BG19] Kevin A. Bryan and Joshua S. Gans. "A theory of multihoming in rideshare competition". In: *J. Econ. Manag. Strateg.* 28.1 (2019), pp. 89–96. ISSN: 15309134. DOI: 10.1111/jems.12306.

[BH18] Yannis Bakos and Hanna Halaburda. "Platform Competition with Multihoming on Both Sides: Subsidize or Not?" In: (2018). URL: https://cdn.questromworld.bu.edu/platformstrategy/files/2018/06/bakos@stern.nyu%7B%5C_%7D.edu%7B%5C_%7D.pdf.

[Big19] Gary Biglaiser. "Migration between platforms". 2019.

[BM] Dirk Bergemann and Stephen Edward Morris. *Information Design: A Unified Approach.*

[BM19] Dirk Bergemann and Stephen Morris. "Information design: A Unified Perspective". In: *J. Econ. Lit.* 57.1 (2019), pp. 44–95. ISSN: 14769344. DOI: 10.1057/ejis.2011.22.

[BP19] Paul Belleflamme and Martin Peitz. "Platform competition: Who benefits from multihoming?" In: *Int. J. Ind. Organ.* 64 (2019), pp. 1–26. ISSN: 01677187. DOI: 10.1016/j.ijindorg.2018.03.014.

[Cal21] Dan Calacci. *Gigbox*. 2021. URL: https://github.com/dcalacci/gigbox.

[Cam+97] Colin Camerer et al. "Labor Supply of New York City Cabdrivers: One Day at a Time". In: *Q. J. Econ.* 112.2 (1997), pp. 406–441. ISSN: 00335533. DOI: 10.1162/003355397555244.

[CKW17] Juan Camilo Castillo, Dan Knoepfle, and Glen Weyl. "Surge pricing solves the wild goose chase". In: *EC 2017 - Proc. 2017 ACM Conf. Econ. Comput.* December (2017), pp. 241–242. DOI: 10.1145/3033274.3085098.

[CMW15] Le Chen, Alan Mislove, and Christo Wilson. "Peeking beneath the hood of uber". In: *Proc. ACM SIGCOMM Internet Meas. Conf. IMC* 2015-Octob (2015), pp. 495–508. DOI: 10.1145/2815675.2815681.

[Coo+20] Cody Cook et al. "The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers". In: *Rev. Econ. Stud.* (2020). ISSN: 0034-6527. DOI: 10.1093/restud/rdaa081.

[Cra+13] Vincent Crawford et al. "Robust Predictions in Games With Incomplete Information". In: *Econometrica* 81.4 (2013), pp. 1251–1308. ISSN: 0012-9682. DOI: 10.3982/ecta11105.

[CS16] B Y Vincent P Crawford and Joel Sobeli. "Strategic Information Transmission Author ( s ): Vincent P . Crawford and Joel Sobel Published by : The Econometric Society Stable URL : http://www.jstor.org/stable/1913390 Accessed : 14-04-2016 18 : 08 UTC Your use of the JSTOR archive indicates your acc". In: 50.6 (2016), pp. 1431–1451.

[Day19] DayDayNews. *The full version of Didi ordering skills , see how Didi drivers earn more than 10 , 000 yuan a month?* 2019. URL: https://daydaynews.cc/en/technology/211658.html.

[DBS16] Sarah A Donovan, David H Bradley, and Jon O Shimabukuro. "What Does the Gig Economy Mean for Workers?" In: *Congr. Res. Serv. Rep.* (2016), pp. 1–16. URL: http://search.ebscohost.com/login.aspx?direct=true%7B%5C&%7Ddb=tsh%7B%5C&%7DAN=112855546%7B%5C&%7Dsite=ehost-live.

[Dev+19] Jacob Devlin et al. "BERT: Pre-training of deep bidirectional transformers for language understanding". In: *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.* 1.Mlm (2019), pp. 4171–4186. arXiv: 1810.04805.

[DKZ21] Mi Diao, Hui Kong, and Jinhua Zhao. "Impacts of transportation network companies on urban mobility". In: *Nat. Sustain.* (2021). ISSN: 23989629. DOI: 10.1038/s41893-020-00678-z. URL: http://dx.doi.org/10.1038/s41893-020-00678-z.

[Dwo20] Piotr Dworczak. "Mechanism Design With Aftermarkets: Cutoff Mechanisms". In: *Econometrica* 88.6 (2020), pp. 2629–2661. ISSN: 0012-9682. DOI: 10.3982/ecta15768.

[Far05] Henry S. Farber. "Is tomorrow another day? The labor supply of New York City cabdrivers". In: *J. Polit. Econ.* 113.1 (2005), pp. 46–82. ISSN: 00223808. DOI: 10.1086/426040.

[Far15] Henry S. Farber. "Why you can't find a taxi in the rain and other labor supply lessons from cab drivers". In: *Q. J. Econ.* 130.4 (2015), pp. 1975–2026. ISSN: 15314650. DOI: 10.1093/qje/qjv026.

[FG07] Ernst Fehr and Lorenz Goette. "Do workers work more if wages are high? Evidence from a randomized field experiment". In: *Am. Econ. Rev.* 97.1 (2007), pp. 298–317. ISSN: 00028282. DOI: 10.1257/aer.97.1.298.

[FG16] Itay P. Fainmesser and Andrea Galeotti. "Pricing network effects". In: *Rev. Econ. Stud.* 83.1 (2016), pp. 165–198. ISSN: 1467937X. DOI: 10.1093/restud/rdv032.

[FG20] Itay P. Fainmesser and Andrea Galeotti. "Pricing network effects: Competition". In: *Am. Econ. J. Microeconomics* 12.3 (2020), pp. 1–32. ISSN: 19457685. DOI: 10.1257/MIC.20170226.

[Fri73] James W. Friedman. "A non-cooperative equilibrium for supergames: A correction". In: *Rev. Econ. Stud.* 40.3 (1973), p. 435. ISSN: 1467937X. DOI: 10.2307/2296463.

[GP11] Maria Goltsman and Gregory Pavlov. "How to talk to multiple audiences". In: *Games Econ. Behav.* 72.1 (2011), pp. 100–122. ISSN: 08998256. DOI: 10.1016/j.geb.2010.08.007. URL: http://dx.doi.org/10.1016/j.geb.2010.08.007.

[HC77] Andrew C. Harvey and Patrick Collier. "Testing for functional misspecification in regression analysis". In: *J. Econom.* 6.1 (1977), pp. 103–119. ISSN: 03044076. DOI: 10.1016/0304-4076(77)90057-4.

[HT19] John J Horton and Prasanna Tambe. "The Death of a Technical Skill". In: *Work. Pap.* (2019), pp. 1–61.

[Ica21] Icauto.com. *2021 滴滴价格收费标准，滴滴垫付最高垫付多少*. 2021. URL: https://www.icauto.com.cn/baike/67/671106.html (visited on ).

[Inc21] Uber Inc. *Uber Period Overview*. 2021. URL: https://www.uber.com/us/en/u/steven%7B%5C_%7Dbc2%7B%5C_%7Dtest/ (visited on 07/28/2021).

[JP19] Bruno Jullien and Alessandro Pavan. "Information management and pricing in platform markets". In: *Rev. Econ. Stud.* 86.4 (2019), pp. 1666–1703. ISSN: 1467937X. DOI: 10.1093/restud/rdy040.

[JT14] Thomas D. Jeitschko and Mark Tremblay. "Platform Competition with Endogenous Homing". 2014.

[Jul11] Bruno Jullien. "Competition in multi-sided markets: Divide and conquer". In: *Am. Econ. J. Microeconomics* (2011). ISSN: 19457669. DOI: 10.1257/mic.3.4.186.

[Kam19] Emir Kamenica. "Bayesian Persuasion and Information Design". In: *Annu. Rev. Econom.* 11 (2019), pp. 249–272. ISSN: 19411391. DOI: 10.1146/annurev-economics-080218-025739.

[Lee+19] Kyungmin (Brad) Lee et al. "Surge Pricing on A Service Platform under Spatial Spillovers: Evidence from Uber". In: *Acad. Manag. Proc.* 2019.1 (2019), p. 16279. ISSN: 0065-0668. DOI: 10.5465/ambpp.2019.16279abstract.

[Liu19] Chunchun Liu Tat-How Teh Julian Wright Junjie Zhou. *Multihoming and oligopolistic platform competition*. Tech. rep. 2019.

[LLW18] Qihong Liu, Oksana Loginova, and X. Henry Wang. "The Impact of Multi-Homing in a Ride-Hailing Market". In: *SSRN Electron. J.* (2018), pp. 1–15. ISSN: 1556-5068. DOI: 10.2139/ssrn.2968504.

[Mas07] Eric S. Maskin. *Nobel Prize Lecture: Mechanism Design: How to Implement Social Goals*. 2007. DOI: 10.1057/9780230280847_22.

[Mye83] Roger B. Myerson. "Mechanism Design by an Informed Principle". In: *Econometrica* 51.6 (1983), pp. 1767–1797.

[OFC21] Hao Yi Ong, Daniel Freund, and Davide Crapis. "Driver Positioning and Incentive Budgeting with an Escrow Mechanism for Ridesharing Platforms". In: (2021), pp. 1–35. arXiv: 2104.14740. URL: http://arxiv.org/abs/2104.14740.

[Pad20] Alex Padilla. *General Election November 3, 2020*. Tech. rep. 2020. URL: https://elections.cdn.sos.ca.gov/sov/2020-general/sov/complete-sov.pdf.

[PR18] James A Parrott and Michael Reich. *An Earnings Standard for New York City's App-based Drivers*. Tech. rep. July. 2018.

[Ros20] Nicola Rosaia. "Competing Platforms and Transport Equilibrium: Evidence from New York City". 2020.

[Sha18] Matthew H Shapiro. "Density of Demand and the Benefit of Uber". In: *Work. Pap.* May (2018). URL: http://www.shapiromh.com/uploads/8/6/4/0/8640674/mshapiro%7B%5C_%7Djmp.pdf%7B%5C%%7D0Ahttp://www.shapiromh.com..

[SWW19] Hao Sun, Hai Wang, and Zhixi Wan. "Model and analysis of labor supply for ride-sharing platforms in the presence of sample self-selection and endogeneity". In: *Transp. Res. Part B Methodol.* 125 (2019), pp. 76–93. ISSN: 01912615. DOI: 10.1016/j.trb.2019.04.004.

[Sys21] Engineering Systems. "Information , Learning and Incentive Design for Urban Transportation Networks Manxi Wu". In: (2021).

[Tan+21] Xiaocheng Tang et al. "Value Function is All You Need: A Unified Learning Framework for Ride Hailing Platforms". In: (2021). arXiv: 2105.08791. URL: http://arxiv.org/abs/2105.08791.

[Val20] Daniel X Valderrama. "Rider Multihoming in the United States Rideshare Market". PhD thesis. 2020.

[Vaz+18] M. M. Vazifeh et al. "Addressing the minimum fleet problem in on-demand urban mobility". In: *Nature* 557.7706 (2018), pp. 534–538. ISSN: 14764687. DOI: 10.1038/s41586-018-0095-1.

[WAO21] Manxi Wu, Saurabh Amin, and Asuman E. Ozdaglar. "Value of information in bayesian routing games". In: *Oper. Res.* 69.1 (2021), pp. 148–163. ISSN: 15265463. DOI: 10.1287/OPRE.2020.1999. arXiv: 1808.10590.

[Wee19] Julie Weed. *These Apps Are an Uber Driver's Co-Pilot*. 2019.

[Wor21] Worker Info Exchange. *Your Data. Your Power*. 2021. URL: https://www.workerinfoexchange.org/.

[Wu17] Manxi Wu. "Effect of Information in Bayesian Congestion Games Signature redacted". In: (2017). URL: http://iibraries.mit.edu/ask.

[Xie+20] Qiaomin Xie et al. "Learning Zero-Sum Simultaneous-Move Markov Games Using Function Approximation and Correlated Equilibrium". In: (2020). arXiv: 2002.07066. URL: http://arxiv.org/abs/2002.07066.

[Xu+18] Zhe Xu et al. "Large-scale order dispatch in on-demand ride-hailing platforms: A learning and planning approach". In: *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* (2018), pp. 905–913. DOI: 10.1145/3219819.3219824.

[XYY20] Zhengtian Xu, Yafeng Yin, and Jieping Ye. "On the supply curve of ride-hailing systems". In: *Transp. Res. Part B Methodol.* 132.xxxx (2020), pp. 29–43. ISSN: 01912615. DOI: 10.1016/j.trb.2019.02.011. URL: https://doi.org/10.1016/j.trb.2019.02.011.

[Zei19] Matthew Zeitlin. *How Austin's failed attempt to regulate Uber and Lyft foreshadowed today's ride-hailing controversies*. 2019. URL: https://www.vox.com/the-highlight/2019/9/6/20851575/uber-lyft-drivers-austin-regulation-rideshare.

[Zho+20] Yaqian Zhou et al. "Competitive ride-sourcing market with a third-party integrator". In: *arXiv* (2020), pp. 1–42. ISSN: 23318422.