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. We demonstrate the relevance and study the design of providing demand information to gig workers. We find in a large-scale survey in Jakarta tha This thesis makes a case for the importance of another dimension of platform design for platform drivers: Demand information given to drivers. Using a large-scale survey and platform-provided data, we give evidence for the relevance of demand information for drivers' labor supply decisions. We then start our investigation into how information should be designed, and derive policy recommendations for platforms and their regulators.

Thesis Supervisor: Jinhua Zhao
Title: Associate Professor

Acknowledgments

I thank my family, Helga and Katrin Haupt, my roommates, Isabell, Juan-Pablo, Sophia, Salome, Bernardo, and friends and mentors, Christina, Marieke and Roberta, for the emotional support I needed in this time. I thank Stephen Zoepf (Lacuna.ai, Stanford University), Ryan Green (Gridwise), Dan Calacci (MIT Human Dynamics), Xiaotong Guo (MIT JTL), 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, [19]

Uber highlights flexibility of drivers [47].

[18] study the effect of ridesharing on urban mobility and find that ridesharing intensifies urban challenges.

Preference for flexibility is documented in several studies, also empirical of nature. [1] uses a choice experiment with virtual license plates for gig workers. Gig workers give up significant payments for working with . See also [35].

[**Jiang_Kong_Zhang_2021**] show that regret aversion and ignorance of suggestion are the two major behavioral factors which influence drivers' re-positioning decisions.

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.

In our definition, we 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: Given demand information, both groups position themselves in an urban space waiting for orders. Also, several Transportation Network Companies (TNCs) specializing in ride-sourcing, 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). In contrast, search by gig workers, or an equivalent of waiting for orders for platform drivers, is not present in this decision.

We argue that information provision to platform drivers, and to gig economy workers more broadly, is important for competition of workers on platforms, as well as for regulation. Gig workers make labor supply decisions which shape the performance of transportation platforms: When to work, where to work, and for how long to work. Not only do these decisions have an impact on driver earning, but also on TNC operations.

1.1.1 Three Levels of Demand Information

We introduce the allocation of information as a third level of incentive for platform drivers, beyond matching and pricing.

Matching is the assignment of drivers to orders, i.e. rides, shopping trips, or deliveries. Matching can be seen as the allocative dimension of platform design for drivers. Allocative efficiency often relates to the

Pricing is the determination of prices for orders.

This thesis studies Information provision.

that could affect decisions of gig workers to supply or not supply their labor often takes the form of demand information. For example, information on an area or a time period of high demand for rides can affect drivers' earnings.

This information becomes particularly crucial as platform drivers usually do not set prices for their services.

While information provision to platform driver influences decision-making of drivers, information provision becomes also important for other parts of the gig economy. For example, knowledge of demand and expectations have been demonstrated in the market for freelance software development [2] when the market environment changed.

This thesis focuses on the design of information to drivers. It highlights that merely considering *pricing*, i.e. the setting of prices to consumers and gig workers, might be beneficially complemented by considering *information provision*, both for TNC operations, a market for third-party information providers and regulation. 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 three 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 also find that drivers often do not reposition themselves, but follow the algorithm whenever possible.

We supplement our observations on the relevance of non-pricing TNC design questions in chapter 4 using earnings differences in platform-provided data. 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 Yuching in November 2016 in Chengdu, China and comparing the earnings differences to average earnings of delivery drivers on the platform. We find that optimal repositioning, which can be seen as a very strong form of knowledge of demand, can lead to significant (at least two-fold) increase in earnings if a small group of drivers is perfectly informed about demand.

Third, we take first steps into the design of information for platform drivers in **??**. In a benchmark model we formalize properties of the informational environment. We first show that not providing the same demand information to all drivers is welfare-maximizing. We show that in our model, an informational Braess' paradox arises, which is distinct from existing notions of an informational Braess' paradox.[1] We expose that platform drivers might ignore information presented to them by the platform as a result of limited commitment of a TNC which does not allow them to credibly commit to not maximize their revenue with a choice of information. We show that third-party information aggregators can remedy this inefficiency, as can a change in the payment structure of gig drivers, and that aggregators can do so even with information provided to all subscribed drivers, as long as these are a small proportion of all drivers.

The final chapter 6 collects policy recommendations both for TNC operators and regulators. We first show that information design is of different significance for TNC operators 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, and are in this endeavour aligned with drivers, the latter should think more carefully about equity in the proposed models: Information needs to be provided asymmetrically. While asymmetrically, it still needs to require that drivers are equally treated. We conclude with engineering challenges arising from our observations.

While we provide several facets of information design for platform drivers, we largely abstract both pricing (who pays what) and matching (which driver gets to serve which rider) decisions and competition concerns away. Real-world platforms solve joint optimization problems of pricing and information in competition with

[1] [51] studies this in road networks; Manish Raghavan studies it for several classification technologies [36].

other platforms. While this means that all of our analysis needs to be seen as having a narrow view of the design possibilities of platforms, we highlight, on the contrary, that viewing platform design purely as a *pricing* question might fall short of optimal mobility platform design.

Our analysis can also only be seen as mid-term, as information design to Autonomous Vehicles will depend on whether all vehicles are centrally controlled and cooperative or not. Throughout this thesis, we maintain that platform drivers are human.

1.2 Background

- Drivers and Urban Mobility; 1/3 important - Three important points: Matching (Dynamic Rebalancing), pricing (Wild Goose Chase), and information.

- Drivers and Gig work - What are other examples - Driving is characterized by particularly strong informational advantage - Other points: Full time, flexibility

- Therefore interventions: - Earnings - Prop. 22 - UK Supreme Court - No interventions into informational environment

- Informational and Payment environment for drivers - Apps - Periods - Multi-Homing - Demand information in app - Third parties

- Impact of informational structure for regulation.

Platform drivers take an increasing 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). Regulators have for the biggest part not intervened into information provision to drivers, which puts drivers into a status quo of an information environment given by their TNC 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 [11]. With the rise of the first transportation network companies 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. While dynamic rebalancing is also an important question in public transit operations, schedules can be designed with much less uncertainty.

Part of the uncertainty making dynamic rebalancing hard comes from the need to design incentives that ensure participation. 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 will not.

1.2.2 Platform Drivers as Gig Workers

Platform drivers are also gig workers *qua* the definition introducing this chapter. Riders 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.

First, platform driving is, compared to other parts of the gig economy very flexible. This characteristic comes from the nature of gigs being short. Compared to other parts of the gig economy, e.g. lodging (e.g., on AirBnB, gigs usually at least one day), and services for care (e.g., care.com), technology (e.g., Andela), design (e.g., 99designs) and home services (e.g., Porch), typically require scheduling or completion within at least a day. This makes management of incentives compared to drivers more crucial.

Second, the terms of the contract are fully determined by the platform. This is in contrast to other platforms such as in lodging, 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. Compared to other gig work, giving a ride on the platform is short, and as both the complete contract and all payments are processed by the platform, it is hard for drivers to build gigs independently of the platform. This gives the platform additional bargaining power in negotiations with the driver.

Fourth, 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

Flexibility, limitations on contracting and communication, and the ubiquity non-wage controlled, full-time 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 [41]. 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 ridesourcing drivers to be excluded from legal sufficient conditions for an employment relationship.

In contrast, the UK Supreme Court decided that ridesourcing drivers *are* employees of TNCs, with labor implications for drivers, with effects for payment and insurance of dirvers.

These regulations do not include provisions on the information that drivers get from TNCs, and other actors work in this sphere.

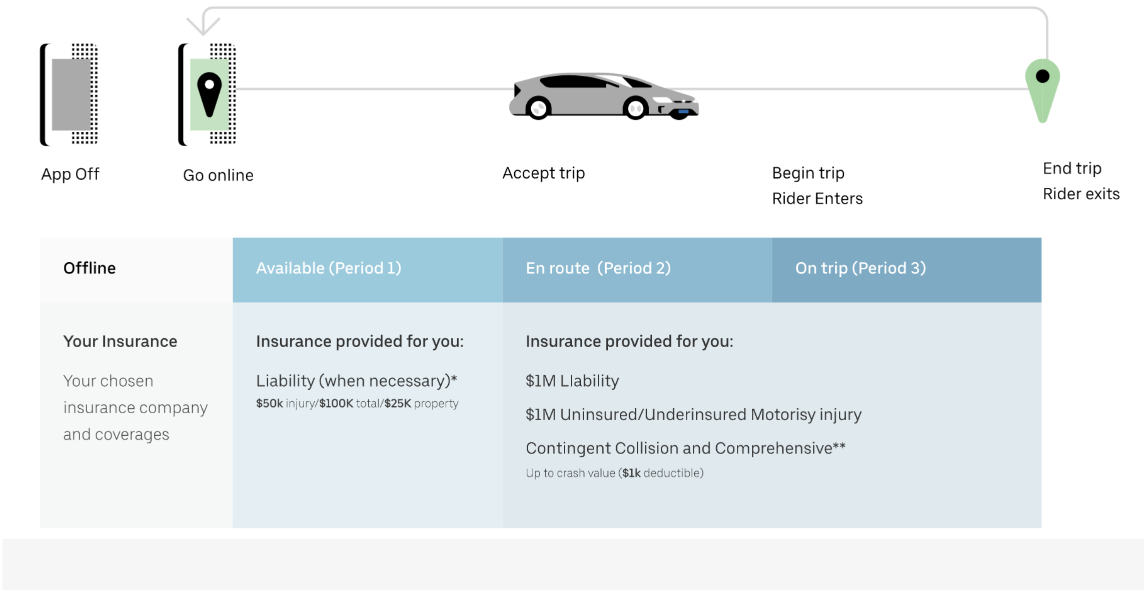


Figure 1-1: Nomenclature for different periods in coverage for Uber as presented in [29] .

1.2.4 Platform Drivers' Payment and Information Environment

Most platforms drivers face an app on a smartphone, 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 ([48, p.48]). In our survey conducted in Jakarta (chapter 3), we find that a minority of drivers has several apps open at the same time. Hence, drivers are limited in their ability to observe demand on the other platforms.

This interacts also with payment design of platforms. Many platforms associate bonuses with the acceptance of rides, or punishments up to being blocked on the platform for not accepting ride requests. This makes being online in different platforms complicated for many drivers.

An important distinction will be three time periods, which we call, in accordance with nomenclature introduced by Uber *phases 1, 2, and 3*, compare Figure 1-1.

Phase 1 corresponds to times where the driver is online but idle. In these times, the driver is not sent ride requests and does (except for bonus payments for repositioning) not earn money.

Phases 2 and 3 refer 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.

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 when the platform can transmit any information that will not be ignored.

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 travelers' equilibrium route choices and costs?"—[51]

Information can add another dimension to driver behavior. 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

Besides pricing and matching, we highlight a third component of platform design: information provision. As an example of why information provision might not be appreciated in platform design in mobility, questions about information design cannot be expressed in unified models of ridesourcing [28]. We give references for the three tiers of matching, and relate them to information design questions.

2.1.1 Matching

Dynamic matching is solved in the operations literature. Recent contributions to this literature are [2] and [49].

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 [53], and other transportation modes have similar challenges, compare for bikesharing [6].

From a perspective of information design for platforms, drivers and the platform have a shared objective to have many drivers in high-demand areas, and dynamic rebalancing goals could be reached via information provision. We explore design considerations in chapter 5.

2.1.2 Pricing

A second literature is on pricing.

An important phenomenon related to pricing is the *Wild Goose Chase* [13]. A wild goose chase refers to the reduction of supply when supply is low and prices are constant. Drivers go on the platform, and are matched to a free rider which might be very far away. The expectation of a long pickup distance reduces earnings expectations and further reduces driver labor supply. [52] shows in high generality that this insight holds.

As an intervention into the market that reduces the appearance of the wild goose chase, *surge prices*, i.e. higher prices when the market is high are proposed [13]. Further research has considered the effects of surge prices in a spatial setting [37].

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. Matching-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.

2.1.3 Information

Information on demand or, equivalently, recommendations on where to move, were shown in a focus group study [3] as important points of improvement for TNCs from the side of ride-hailing drivers.

However, except for in routing games ([51, 46, 50]), information design for transportation systems has not taken an important role. [51] 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 demand for platform drivers.

[44] shows using data from New York City that the highest welfare gains arise in less dense areas. In particular, these are areas where the informational component of platform design is particularly important. We complement this point in chapter 5 by showing that also driver's surplus can change with sufficient allocated information.

[33] study a two-sided market in which groups are uncertain about the joining decision of the other side of the market and how this helps efficiency. We do not model demand for platform drivers as strategic, and only consider platform drivers, but do so in a setting with a spatial structure.

[40] 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, which we also find in our analysis.

2.2 Related Topics in Platform Design

In our contributions we also relate to other topics of platform competition which we do not further target. 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 [38], regarding the effect on questions of pricing when both sides multi-home [5], the welfare effects of multi-homing [7], how multi-homing can happen dynamically [10] for platform migration, and questions on participants that cannot multi-home [30].

assume exogenous models of [**castillo2017surge**] or do not model multi-homing decisions [**liu2017impact**].

We view information design mostly for a monopolistic setting and view it for future work. Views about third-party integration are orthogonal to our question [54].

Other papers assume that in moments where agents join a platform, they have exact knowledge of the surplus that is gained and the prices that both sides receive [**liu2019multihoming**].

[**Wang_Yang_2019**] offers a comprehensive review on the ride-hailing system including demand, supply and platform perspectives. Driver behavior is modeled differently in the ride-hailing literature. Regarding operations and optimizations for ride-hailing platforms, e.g., driver-customer matching, idle vehicle rebalancing and vehicle routing, drivers are assumed to follow the instructions of platforms ([**Bertsimas_Jaillet_Martin_20** **Alonso-Mora_2017**, **Braverman_Dai_Liu_Ying_2019**, **Wen_Zhao_Jaillet_2017**]).

Simple driver behavior models are proposed in the pricing literature of the ride-hailing system. [**Bai_So_Tang_Chen_Wang_2019**] propose 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. [**Taylor_2018**] evaluate the impact of uncertainty in passenger delay sensitivity and driver independence on platforms' optimal per-service price and wage. Each driver has an opportunity cost and the driver will participate in the platform when having non-negative expected utility.

2.2.2 Asymmetry in Platform Design

Asymmetry to leverage network externalities are well-known in the network externality literature. [32] 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. [21, 22] study this with players that have some heterogeneity.

2.3 Methodology

2.3.1 Estimation of Implicit Costs of Certain Behaviors

Our estimate of implicit costs of multi-homing in chapter 3 estimates a preference for a particular platform. This relates to a literature on such information preferences in the case of flexibility.

We also relate in our quantitative evidence to cab driver labor supply in New York City [12], which reported negative labor supply elasticities of cab drivers, i.e. drivers working less hours on days with higher earnings. This study led to an introduction of reference-dependence into labor supply, and was followed up by studies replicating the claim [24, 25, 23].

Our analysis merely points towards other factors besides earnings maximization and does not yield conclusive evidence on what these factors are.

2.3.2 Gig Worker Data

Studies on algorithmic behavior need data on the contracts being intermediated. In the case of platforms, this 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: cooperation with TNCs, automatic extraction of data through APIs, and gathering data from drivers.

Several studies in collaboration with TNCs have found large-scale behaviors. Among others, [45] on the labor supply of platform drivers, [15] for the gender pay gap, and [4] 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.

Other studies use APIs to query for prices, such as [43] who evaluates counterfactual regulations in an urban mobility market in New York City and [44] who estimates the consumer surplus difference from TNCs compared to Lyft in the same market. [14] uses large-scale simulated apps to elicit prices and estimate pricing fairness questions.

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

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

2.3.3 Mechanism and Information Design

Mechanism design is an old 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, the study led to designs of economic environments from school assignment to transplantation matching, internet ad auctions. For an introduction to the topic, see, for example, Eric Maskin's Nobel lecture [39].

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 a ridesharing context, for example, the exact preferences of drivers for when and where to work are not known to a 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 actions by strategic actors, in our application drivers.

The commitment assumption is often violated, as we argue in particular in the context of platform drivers. We therefore 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 [16]. Our exact setup is influenced by the multi-receiver version [27].

While mechanism design pertains to the allocation of scarce resources, another allocation problem arises when a platform has particularly strong informational effects and 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 spot has many orders, congestion or the absence of drivers, depending on luck and the reaction of the riders to the information. Information needs to be carefully designed, in particular when strategic interaction between different agents are present.

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 [34]. In later works, the full information design with strategic interactions was developed, [9, 8]. We will use for a relevant part of our analysis a model in [9].

With strategic interactions, [17] 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 complementarity*), asymmetric information can be optimal.

The real platform's problem is 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 [20], and an integration of matching, pricing and information is an exciting area for future research.

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. —[11]

This chapter analyzes results 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

We analyze data 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 and received 846 complete survey responses after being distributed and circulated to drivers through known social media (WhatsApp groups) used by drivers.

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 drivers.

There exist three major ride-hailing companies 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.

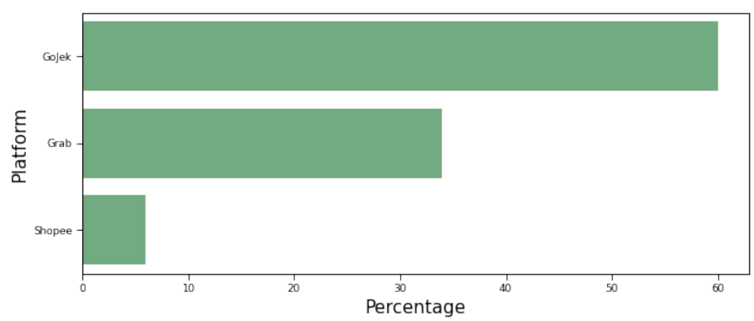
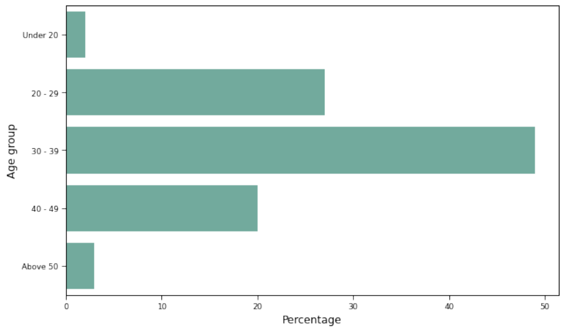


Figure 3-1: Platform distribution of drivers. (GoJek: 60%, Grab: 34%, Shopee: 6%)

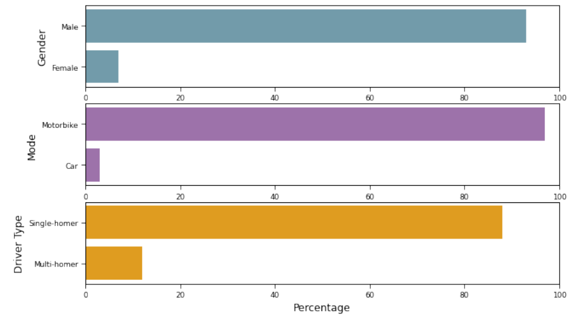
3.1.2 Questions

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

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



(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.

The second block depends on whether respondents were identified in the first block as multi-homers or not. 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 asked.

In a third block, for each ride-hailing platform survey participants have worked for in the last 30 days, we asked a number of questions about driving activities during the last two weeks. These questions included whether drivers mostly worked on delivering food or transporting passengers, their most frequent driving area, days and hours to work, total driving distance, and daily salaries.

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

3.2 Findings

To understand the multi-homing behavior and who works for multiple ride-hailing platforms, we first constructed a Lasso regression model with the regularization coefficient 0.001 based on drivers' socio-demographic information, which were shown in Table 3.1.

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

Variable	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	0.3168***	0.052	6.040	0.000	0.214	0.420
Age (30 - 39)	-0.0455*	0.025	-1.786	0.075	-0.095	0.005
Age (above 50)	-0.1189*	0.069	-1.721	0.086	-0.255	0.017
Male	-0.1213***	0.042	-2.911	0.004	-0.203	-0.040
Degree (bachelor's or higher)	0.1258***	0.044	2.840	0.005	0.039	0.213
Degree (junior high school)	0.0327	0.035	0.939	0.348	-0.036	0.101
Ride-hailing as main source of income	-0.0733*	0.038	-1.925	0.055	-0.148	0.001
Age (40 - 49) and ride-hailing as main source of income	-0.0480	0.033	-1.474	0.141	-0.112	0.016
R-squared:	0.033	**Adjusted R-squared:**			0.025	

Table 3.2: **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
Motorcycle taxi driver before	-0.0796**	0.031	-2.574	0.010	-0.140	-0.019
R-squared:	0.008	**Adjusted R-squared:**			0.007	

We find that male drivers whose age is between 30 to 39 and above 50 are less likely to multi-home in Jakarta market. Also, drivers are less likely to multi-home if working for ride-hailing platforms is the main source of their income. On the other hand, we find that drivers with higher degrees are significantly more likely to multi-home. Next, we observe the correlation between previous work as a motorcycle taxi driver before on multi-homing behaviors. The OLS regression model shown in Table 3.2 suggests that drivers who has been motorcycle taxi drivers 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 3.3, we find that part-time ride-hailing drivers are more likely to multihome in our sample. However, it is a less significant

Table 3.3: **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
Being part-time ride-hailing drivers	0.0441	0.032	1.384	0.167	-0.018	0.107
R-squared:	0.002	**Adjusted R-squared:**		0.001		

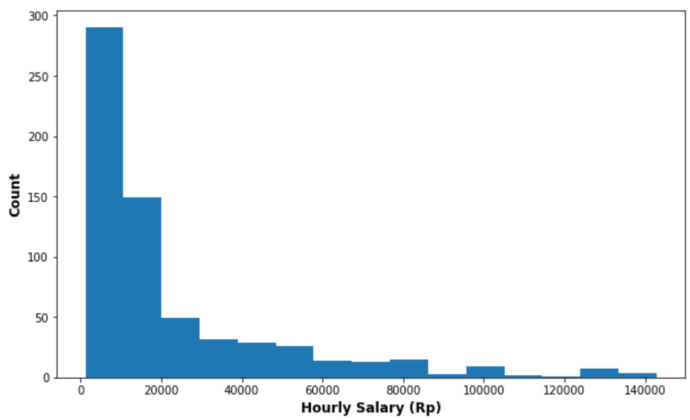


Figure 3-3: Hourly salary distribution of survey respondents (ride-hailing drivers in Jakarta)

predictor of multihoming behaviors compared to other variables above.

3.2.1 Earnings for Different Platforms

Then, we consider the Jakarta ride-hailing market and drivers' hourly salary and total working hours. Figure 3-3 displays the hourly salary distribution of all survey respondents in Jakarta, which fits into a negative binomial distribution. Most ride-hailing drivers earn less than 20,000 Rp (approximately 1.4 USD) per hour, which is significantly less than the average hourly salary 120,175 Rp (approximately 8.4 USD) in Jakarta [**Jakarta_hourly_salary**]. This stands even when controlling for sociodemographics.

To understand whether drivers working for different ride-hailing platforms are payed differently, we constructed a regression model on the impact of working for GoJek on hourly salary, shown in Table 3.4. In this model, we only consider motor-

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

Variable	**coef**	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
Intercept	30710	2215.544	13.863	0.000	26400	35100
Only GoJek	-11000	2547.137	-4.319	0.000	-16000	-6009.224
R-squared:	0.038	**Adjusted R-squared:**			0.036	

Table 3.5: **Results of regression models predicting daily working hours of ride-hailing drivers in Jakarta based on whether working for GoJek**

Variable	**coef**	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	10.6642***	0.162	65.928	0.000	10.347	10.982
Only GoJek	1.0335***	0.210	4.911	0.000	0.620	1.447
R-squared:	0.036	**Adjusted R-squared:**			0.034	

cycle drivers who work exclusively for one platform. The model results suggest that GoJek drivers earn significantly less than drivers from other platforms, 11,000 Rp (approximately 0.77 USD) less per hour.

Furthermore, we built another regression model to investigate the difference of total working hours between ride-hailing platforms. The model results in Table 3.5 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.

3.2.2 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?".

In a topic analysis, the word "focus" is highly descriptive of one of the clusters.

Respondents indicate that “focus was needed for best results” and that they would not like to “confuse the apps”.

Others refer to the character of one of the platforms as being an Indonesian incumbent Gojek that competes with an (in our sample) better-paying incumbent from Singapore. Respondents make reference to “children of the nation” to underline their choices.

We use survey evidence showing that in an a significant share of non-multi homing drivers in an oligopolistic market for ridehailing in Jakarta. Our findings hint at substantial frictions to multi-homing, non-monetary incentives in the decision what TNC drivers choose to drive for, and when to drive.

3.2.3 Discussion

We found in our analysis that

3.3 Conclusion

Chapter 4

The Value of Information

After evidence from a survey, we move on with evidence derived from platform data.

we give evidence here on the value of information to drivers. This is aligned with claims by companies that offer information services to drivers such as Gridwise or the Mileage tracker.

4.1 Description of Data

The DiDi KDD Challenge 2020, sponsored by DiDi Yuching, consisted of two challenge. An order dispatch challenge, and a driver repositioning challenge. We will work with data and submissions of the latter.

4.1.1 Annotated Order Data

We are given for November 2016 about 10 Million orders. An order consists of an origin-destination pair, a timestamp for beginning and end, an identifier for the driver that took the order, and a number of *reward units*. As the data is not publicly available anymore, we are unable to provide replication data for the following analysis.

A main challenge in our analysis is to recover the number of reward units corresponding to a RMB. Given our estimation below, we can transfer repositioning scores to monetary values, which gives us an estimate of

4.1.2 Reinforcement Learning Challenge Leaderboard

The KDD 2020 reinforcement learning challenge solves an optimal repositioning problem together with an optimal order dispatch problem. The problem is a partial observed semi-Markov Decision Process. In Reinforcement Learning language, which we review in **??**.

The system's observable state for the order dispatch algorithm consists of a list of orders to be dispatched at each point in time. This means a list of origin-destination pairs for orders, and for each driver timestamps of *estimated* order pickup and arrival times, as well as the distance for the driver. The reinforcement learning algorithm's actions are matchings of drivers to orders, the challenge does not consider carpooling. Orders need to be dispatched within two seconds, otherwise they are lost. (We specify the complete environment in **??**. 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 dipatch algorithm is to maximize average driver income, which, as the system does not model drivers logging off strategically, is proportional to the total driver revenue in terms of orders completed.

The system's observable state for the driver repositioning challenge, which we are most interested in, consists of a set of drivers to be repositioned. There is a targeted group of drivers, which, after their first five minutes of being idle move according to historical transition probabilities between regions, can be repositioned. 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, for each driver to be repositioned, a destination location. Agents are then repositioned at 3m/s in the spherical/great arc distance.

The objective in the driver repositioning problem is to maximize mean driver income rate for drivers. Denote $J_k^n(\pi)$ the online time for driver k at day n in hours.[1] Denote driver k'd income on day n under policy π. For the set of all targeted drivers

[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.

K and days N, the goal of the challenge was to maximize

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

We claim that the optimal repositioning score for drivers can give insight into the value of information for these drivers, as long as there are not too many of them.

If drivers have sufficient information about the demand throughout the city, they can reposition themselves to another location in expectation of getting higher earnings. As the maximization takes into account all online time, in particular the time moving to another area of the city, the earnings increase from a performant repositioning algorithm in this challenge can also be seen as a proxy for earnings of informed drivers.

In this argument, the small 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.

But even with more drivers, private messages, which we discuss in chapter 5 help. Private messages is demand information (or, as we show, equivalently recommendations of to which area to move) only given to particular drivers, solving their

4.1.3 Earnings Table

We are using a publicly available table for earnings from 2021. In 2017, DiDi introduced two tiers of service for their drivers, Express and Premium. After personal conversations with experts, we use the table for express, reproduced in **??**.

The earnings table contains, dependent of the time of the day, a base price b_t, earnings per kilometer l_t and minute driven d_t. At the time of the data, November 2016, additionally, a surge multiplier s_t. The total earnings e_t for a ride of L kilometers and duration D are given by

$$e_t = s_t(b_t + l_tL + d_tD). \tag{4.1}$$

Table 4.1: Earnings for express drivers

4.2 Estimation

We assume in the following that reward units are a constant multiple of RMB. The underlying assumption is the following proposition.

Proposition 4.1. *Assume that reward units r are a function f of the earnings from an order in RMB. Assume that any policy generated for reward units is also optimal for policies in RMB. Then, the function f is a constant multiplication, $f(x) = cx$ for some $c > 0$.*

4.2.1 Regression

We assume that a negligible fraction of rides uses a surge price and therefore estimate an equation simplifying (4.1),

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

Our estimate of time is the times given in the order data. We use the Google Maps API to estimate distance in driving. The estimation results are presented in Table 4.2. We find highly significant and consistent estimates on the coefficients for distance and duration.

4.2.2 Checks

4.2.3 Model Fit

We observe that much of the variance can be explained by our model ($R^2 = 0.887$)

Dep. Variable:	reward_units	**R-squared:**	0.887
Model:	OLS	**Adj. R-squared:**	0.887
Method:	Least Squares	**F-statistic:**	1380.
Date:	Sun, 18 Jul 2021	**Prob (F-statistic):**	0.00
Time:	14:45:52	**Log-Likelihood:**	-2998.9
No. Observations:	2000	**AIC:**	6004.
Df Residuals:	1997	**BIC:**	6021.
Df Model:	2		

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
Intercept	-0.1664	0.077	-2.155	0.031	-0.318	-0.015
distance	0.3191	0.021	15.449	0.000	0.279	0.360
duration	0.0635	0.007	9.645	0.000	0.051	0.076

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
Intercept	-0.2569	0.262	-0.981	0.326	-0.770	0.256
rush_hour[T.True]	0.0095	0.320	0.030	0.976	-0.618	0.637
distance	0.3318	0.047	7.113	0.000	0.240	0.423
rush_hour[T.True]:distance	-0.0292	0.048	-0.602	0.547	-0.124	0.066
duration	0.0653	0.011	6.170	0.000	0.045	0.086
rush_hour[T.True]:duration	0.0028	0.014	0.200	0.841	-0.024	0.030

Table 4.2: OLS Regression Results

4.3 Discussion

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.–[9]

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

In our stylized model, two agents simultaneously decide whether to drive to a part of a city ("go") or to follow other business on or off the platform ("not go"). The platform observes a demand state $\theta \in \{0, 1\}$. We can view this as the outcome of an accurate demand prediction model. Before each of the drivers makes her decision on whether to drive, the system can send a message $m_1(\theta)$ to driver 1 and $m_2(\theta)$ to driver 2 on whether there is demand on the platform. Drivers get utility 1 if they get an order, Driver 2

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

(a) Demand, $\theta = 1$

Driver 2

Driver 1	go	not go
go	$-\varepsilon$, $-\varepsilon$	$-\varepsilon$, σ
not go	σ, $-\varepsilon$	σ, σ

(b) No, $\theta = 0$

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

and incur a cost of driving to the location of ε. This gives, depending on the demand state, the game tables in Figure 5-1

The platform gets utility of 1 if at least one of the drivers gets to the desired location, and 0 utility otherwise. The divergence of platform and driver utility stems from the payment structure for drivers, as we described in the introduction: Drivers only get paid for phase 2, when they are matched to a ride.

The platform can send messages to drivers. The platform can either send a message to drivers without committing to any particular form of information revelation (subsection 5.1.1) such as “I recommend you to go to this place, and I will not recommend this to others”, or does not have the power to credibly commit to this (subsection 5.1.3).

We assume that the drivers best respond to the messages sent by the platform. In our first model, we assume that the platform can commit to a particular mapping from demand to messages, in the latter they cannot.

5.1.1 Information Design

We first study a setting where the platform can commit to an arbitrary mapping of demand θ to tuple of messages $(m_1(\theta), m_2(\theta))$. Before stating our theorem, we go into public or private information.

5.1.2 Public and Private Messages

Some information is shown to all drivers, while other is shown only to a few drivers. We call messages *public* if $m_1(\theta) = m_2(\theta)$ for $\theta = 0, 1$ and all other messages *private*. As is well-known in the information design literature [9], the game that agents play is one of *strategic substitutes*, andp private information will increase the total utility for all players. In our case, this means, drivers. Our theorem adds that this is also the optimal choice of messages for the platform.

Theorem 5.1. *All maximizers of platform revenue are $m_1(\theta) = \theta$,*

5.1.3 Cheap Talk

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 θ 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.

Theorem 5.2. *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 5.3. *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 5.4. *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 5.5 ([26])**.** *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 5.6. *Assume that drivers are*

Chapter 6

Conclusion and Policy Implications

This thesis studied the relevance and optimal design of information provided to ridesourcing drivers. chapter 3 showed that besides earnings maximization, drivers also take into account cultural and infromational considerations in their labor supply decisions. chapter 4 showed that there are high potential earnings gains from additional information. We then went on to study commitment and privacy of messages as two challenges. In this last chapter, we consider three stakeholder groups, platforms, regulators and transportation engineers with policy recommendations and additional complications arising out of this work.

6.1 The Commitment Challenge for Platforms

We observed in chapter 5 that the limited commitment of the platform to providing different platform can lead to not information being transmitted and we proposed reputation

6.2 Challenges for Transportation Regulators

Before thinking about regulation of drivers

6.3 Challenges for Transportation Engineers

The above two challenges give interesting challenges for transportation engineers.

6.3.1 Managing Fairness in Information Provision

Appendix A

Datasets and Surveys

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 Complete Earnings Information for DiDi Platform Drivers

In this brief appendix, we reproduce payment information for platform drivers and describes more concretely the decision environment for the Reinforcement Learning agent in the KDD 2020 Callenge.

A.2.1 Payment Information for Drivers

A.2.2 Decision Environment for Reinforcement Learning Agents

Here, we describe the exact state and action spaces for Reinforcement Learning agents in the KDD 2020 challenge.

Appendix B

Omitted Proofs

Appendix C

Mathematical Background

C.1 On Markov Decision Processes

Reinforcement learning is used in ridesharing [42].

[31] use the KDD data to estimate their value.

Here we introduce a notation on Markov Decision Processes

C.2 On Statistical Experiments and Bayes-Correlated Equilibria

Bibliography

[1] J Angrist, S Caldwell, and J Hall. "Uber versus taxi: A driver's eye view". In: 13.1729822 (2017), pp. 272–308. URL: https://voxeu.org/article/uber-versus-taxi-driver-s-eye-view.

[2] 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.

[3] 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.

[4] 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.

[5] Yannis Bakos and Hanna Halaburda. "Platform competition with multihoming on both sides: Subsidize or not?" In: *Manage. Sci.* 66.12 (2020), pp. 5599–5607. ISSN: 15265501. DOI: 10.1287/mnsc.2020.3636. URL: https://cdn.questromworld.bu.edu/platformstrategy/files/2018/06/bakos@stern.nyu%7B%5C_%7D.edu%7B%5C_%7D.pdf.

[6] 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.

[7] Paul Belleflamme and Martin Peitz. "Platform competition: Who benefits from multihoming?" In: *Int. J. Ind. Organ.* 64.November 2017 (2019), pp. 1–26. ISSN: 01677187. DOI: 10.1016/j.ijindorg.2018.03.014.

[8] 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.

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

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

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

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

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] Xiaotong Guo et al. "Dissolving the Segmentation of Shared Mobility Markets: A Unified Theoretical Framework and Four Examples". 2020.

[29] 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).

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

[31] Yan Jiao et al. "Real-world Ride-hailing Vehicle Repositioning using Deep Reinforcement Learning". In: (2021). arXiv: 2103.04555. URL: http://arxiv.org/abs/2103.04555.

[32] 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.

[33] 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.

[34] 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.

[35] M. Keith Chen et al. "The value of flexible work: Evidence from uber drivers". In: *J. Polit. Econ.* 127.6 (2019), pp. 2735–2794. ISSN: 1537534X. DOI: 10.1086/702171.

[36] Jon Kleinberg. "Algorithmic Monoculture and Social Welfare". In: (), pp. 1–35. arXiv: arXiv:2101.05853v1.

[37] 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.

[38] Qihong Liu et al. "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.

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

[40] 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.

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

[42] Zhiwei Qin, Hongtu Zhu, and Jieping Ye. "Reinforcement Learning for Ridesharing: A Survey". In: (2021). arXiv: 2105.01099. URL: http://arxiv.org/abs/2105.01099.

[43] Nicola Rosaia. "Competing Platforms and Transport Equilibrium: Evidence from New York City". 2020. URL: https://scholar.harvard.edu/files/rosaia/files/draft.pdf.

[44] 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..

[45] 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. URL: https://doi.org/10.1016/j.trb.2019.04.004.

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

[47] Uber. *Uber Rider Survey Results*. 2020.

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

[49] 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.

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

[51] 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.

[52] 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.

[53] 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.

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