Information and Incentives in Online Platforms

by

Emily Meigs

Submitted to the Sloan School of Management
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2022

Author .....
Sloan School of Management
May 17, 2022

Certified by .....
Asuman Ozdaglar
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by .....
Patrick Jaillet
Dugald C. Jackson Professor, Department of Electrical Engineering and Computer Science
Co-director, Operations Research Center

Information and Incentives in Online Platforms

by

Emily Meigs

Submitted to the Sloan School of Management
on May 17, 2022, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis studies the impact of information and design of services for online platforms in three settings: traffic routing, network games, and competition between streaming platforms. In the first part of this thesis, Chapters 2 and 3, we study game play in routing and network games, where it is reasonable to assume agents do not originally know their payoff functions. Specifically, in Chapter 2 we examine the outcome of the learning dynamics in traffic routing where the latency functions are unknown. We show that the combination of selfish routing and learning dynamics converges to the full-information Wardrop equilibrium, this supports the study of the Wardrop equilibrium even in settings where information must be learned over time. In Chapter 3 we use analogous learning dynamics in a different setting, network games where the agents' personal utility functions are not known. This may arise in games of local public goods provision or firm competition. We show that the combination of best response and learning dynamics converges to the Nash equilibrium.

In the second part of the thesis, Chapter 4, we study the problem of sharing information in traffic routing. We investigate whether a routing platform, for example Google Maps or Waze, should share full information, no information, or partial information. We characterize the optimal information strategy in a two-stage setting, where the platform is also learning of the road conditions from the users. We then extend the intuition to an infinite stage setting and find an information scheme that achieves a lower cost than full information.

In the final chapter of the thesis, we study bundling and pricing strategies in streaming platforms, for example Netflix or Hulu. We investigate why there are so many streaming platforms that are succeeding in the market. We first study the setting where a market leader creates a new product and has a monopoly on the market. We show in this case it is optimal in some cases for the platform to bundle their goods. Once another firm enters the market though we show that unbundling becomes the unique optimal strategy.

Thesis Supervisor: Asuman Ozdaglar
Title: Professor of Electrical Engineering and Computer Science

Acknowledgments

First and foremost, I must thank my advisor Asu Ozdaglar. Asu, your continual support and enthusiasm through the years has always inspired me. I appreciate all that you have taught me over the years. I would also like to extend my thanks to my committee members Patrick Jaillet and Saurabh Amin, who through their comments have helped make this thesis better. I also would like to thank Georgia Perakis who generously served on my generals committee during my second year.

I was extremely fortunate to work with some amazing collaborators during my time at MIT. Including Daron Acemoglu, Francesca Parise, Theo Diamandis, Ali Makhdoumi, and Azarakhsh Malekian. Each of you has been fundamental in helping me learn how to be a better researcher. I enjoyed each of our research conversations and am honored to get to work with such fantastic researchers.

I would like to thank the ORC and LIDS sta. In particular, Francisco Jaimes, Brian Jones, Laura Rose, and Andrew Carvalho for always helping me with administrative tasks. An extra special thanks to Roxana Hernandez for always helping me find time in Asu's busy schedule.

I was lucky to be a part of both the ORC and LIDS, where I have met many amazing students. The supportive nature of the ORC was invaluable to me. I thank all those that worked with me on psets, studied for quals with me, shared a happy hour conversation. I would also like to thank everyone in Asu's group that helped support me along the way and thank you for the many discussions, research and otherwise, that we shared in 32D-640.

I would like to thank my friends that have always believed in me. Specifically, I would like to thank Elisabeth and Fillipos for being great roommates, Aleyda for always making me laugh, Christopher and Jon for their great advice and good times, and Courtney for her unending support. Additionally, I must thank my friends and the community of Cambridge Running Club (CRC). They have continually inspired me, helped me become a better runner, and made the process extremely fun.

Although she cannot read, I would like to thank Lucy, my dog, for the constant

joy she brings to my life. She always reminds me that there is time for walks and play.

Last but certainly not least, I would like to my family for their continual love and support in everything I do. My parents Randyl and Zoe for giving me every opportunity possible and my sister Lisa for always making me laugh and planning our sister trips.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Every day people, users, all over the world interact with online platforms, e.g. Facebook, Google Maps, Netflix. Both the users and the platforms have separate aims of these interactions. For example, Google Maps may aim to manipulate traffic in order to minimize overall congestion, whereas individual drivers care only about their own travel times. In these settings users are interacting and learning both about each other's strategies and uncertain changing environments. Driven by applications in traffic routing, network aggregative games, and streaming services, this thesis investigates the effects of information and incentives on users of these platforms and the platforms themselves. In particular, the four main chapters are outlined as follows.

Learning Dynamics in Routing Games First, we consider a subclass of congestion games, traffic routing. Most of the works in the literature assume that the when agents make decisions they have perfect knowledge of their latency functions and of the other agents' strategies. This is however rarely the case and, in practice, agents need to learn such game primitives from (possibly noisy) observations that they gather over days of routing through the network. Thus, we study the problem of parametric learning, where agents know each latency function of the network up to some parameter and aim at learning these parameters via repeated iterations of routing.

We consider a variant of the repeated nonatomic network routing game where each agent controls a negligible amount of flow and routes herself selfishly. In our setting the network has affine stochastic edge latency functions whose slope is unknown

at the start. We consider a simple process of learning where agents share common observations of travel times, estimate the unknown edge slope parameters via ordinary least squares and, at every step, dispatch their flow on the network according to the Wardrop equilibrium computed with their most recent estimates. We prove that under these learning dynamics and under minimal assumptions on the users, the flow in the network converges almost surely to the full information Wardrop equilibrium. Moreover, the slope parameters of all the edges used in the full information Wardrop equilibrium are learned almost surely.

Learning Dynamics in Network Games Second, we consider another subclass of congestion games, network aggregative games. Again, most of the literature in this area assumes the players have perfect knowledge of their utility functions and of other agents' strategies. Since this is not necessarily the case, we consider a repeated setting, where agents must learn over repeated play of the game.

Specifically, we consider a repeated network aggregative game where agents are unsure about a parameter that weights their neighbors' actions in their utility function. We consider simple learning dynamics where agents iteratively play their best response, given previous information, and update their estimate of the network weight parameter according to ordinary least squares. We derive a su□cient condition dependent on the network and on the agents' utility function to guarantee that, under these dynamics, the agents' strategies converge almost surely to the full information Nash equilibrium. We illustrate our theoretical results on a local public good game where agents are uncertain about the level of substitutability of their goods.

Optimal Dynamic Information Provision in Traffic Routing In the previous chapters we consider when information is shared fully between agents. We now investigate what happens when a central planner can strategically share the information. For example, can Google Maps or Waze help mitigate tra□c through using information in the form of route recommendations.

We consider a setting where the central planner aims at minimizing tra□c congestion by providing individualized route recommendations to its users. Importantly, we assume that the planner itself is not informed about the state of the tra□c net-

work but needs to learn through users' experimentation, and users follow the received recommendation only if this is in their best interest given their past experiences. We focus on a two-road dynamic routing game where the state of one of the roads (the ïrisky road") is stochastic and may change over time.

Within this framework, we characterize the optimal incentive compatible recommendation system, first in a two-stage game and then in an infinite-horizon setting. Our analysis uncovers two main insights. First, under the optimal recommendation scheme the central planner does not provide full information about the state of the road to all agents. Thus, information design proves e□ective as a control tool for tra□c regulation. Second, since agents are strategic and long lived, in contrast to classic exploration-exploitation settings the central planner needs to limit the number of agents participating in the exploitation phase to generate enough incentives for experimentation. This aspect becomes particularly relevant in the infinite horizon setting where agents can learn not only from received recommendations but also from observations of previous tra□c flows.

(Sub)Optimality of Bundling in Streaming Platforms Finally, we turn to studying bundling in streaming platforms. In recent years, the competition among streaming platforms such as Netflix, HBO Max, and Hulu is on the rise as even more streaming platforms are being started. We aim in this chapter to study when platforms should bundle separate types of goods.

We model the competition among streaming platforms as a two-dimensional Hotelling model. Each axis is some type of content, potentially original content for each platform or comedies on each platform. Each user has a preference over the platforms and the two types of content. Each platform can choose whether to bundle these two types of content, by giving both for one fee, or provide them separately, each with a fee. We first study the monopoly case and show that in some settings it is more profitable for the monopoly platform to bundle the goods. We then study the duopoly setting, where another firm has entered the market, and show that now in all cases it is now an equilibrium for the platforms to unbundle the goods and sell them separately.

Chapter 2

Learning Dynamics in Routing Games

Tra□c congestion is a major problem for billions of commuters around the world. Research in this area has focused on developing tra□c equilibrium models in order to predict how flows get allocated in tra□c networks and hence analyze performance with di□erent tra□c management policies. While most of the literature assumes that, in making their routing decisions, users have all the relevant information (in particular travel times or delays), in practice a major problem of drivers is to estimate the delays they are going to incur on di□erent routes in the network. This problem of estimation is especially true for drivers that must continue to route themselves again and again in the same road network, which is usually the case for daily commuters.

In this chapter, we consider a model in which a large number of risk-neutral users are unsure about the edge latency (delay) functions and aim at learning them by using common observation of past travel experiences. Specifically, we assume that the users sequentially: i) send their tra□c demand over the network, ii) observe the corresponding (stochastic) travel time and iii) use these observations to update their estimates of the edge latency functions. Given the most recent estimates, we assume that at every step of the learning dynamics the users direct their tra□c demand according to a new slightly modified version of the well-known *Wardrop equilibrium* (see Wardrop [73]; Beckmann, McGuire, and Winsten [12]) based on mean estimates of the latency functions. The justification for this notion comes from two assumptions. The first is that, as mentioned above, users are risk-neutral and thus care only about

the mean of the stochastic travel times or delays they will be facing. The second is that the traffic adjustment takes place faster than information updates thus leading to a Wardrop adjustment at each step of the information update of the users [27]. Given the past observations, we model the estimation problem in point iii) without specifying tight priors about others knowledge or behavior and instead assume that users perform a simple least-squares type estimation.

We prove that this completely decentralized stochastic process with our notion of Wardrop equilibrium leads to convergence of the flow to the full information Wardrop equilibrium. Moreover, the estimates of the latency functions for all edges that are used in the full information Wardrop equilibrium converge to the true parameters. To this end, we build on previous results by Taylor [72] to relate the error terms of the parameter estimates to a martingale and we consequently prove almost sure convergence of the errors. We then use Wardrop equilibrium sensitivity results from Dafermos et al. [29] to prove that, because we have convergence of the errors, the flow under the estimated parameters converges to the full information Wardrop equilibrium.

Our work is related to the recent papers on *stochastic Wardrop equilibrium*, which assumes that the latency functions are stochastic (Chancelier et al. [21]; Cominetti [26]; Nikolova and Stier-Moses [64]). This literature focuses on developing a static description of traffic equilibrium in this stochastic environment with risk-averse users and provides characterization results. Our work is also related to the literature on learning dynamics in routing games. This literature studies a variety of dynamics for routing games including fictitious play (Monderer and Shapley [62]; Marden, Arslan, and Shamma [57]), adaptive sampling and replicator dynamics (Fischer [34] [33]), regret-minimizing or no-regret algorithms (Kalai and Vempala [46]; Blum et al. [16]; Krichene et al. [50]). These algorithms assume that users make adaptive routing decisions in order to minimize regret over time by observing latency values at various congestion levels, but do not assume any stochasticity related to the latency functions. Our work instead assumes that latency functions are stochastic and that there is an additive error term that accounts for the variability due to incidents or other exogenous factors (weather, time-of-day etc.). We claim this more realistically models

commuters day to day experience.

The rest of the chapter is organized as follows. In Section 2.1 we present the model and the learning dynamics we utilize. In Section 2.2 we discuss preliminary results that lead to our main result in Section 2.3 where we prove convergence to the full information Wardrop equilibrium in networks with a□ne stochastic latency functions through ordinary least squares updating of the parameters. In Section 2.4 we illustrate our theoretical results on a Wheatstone network and finally Section 2.5 concludes the chapter.

Notation: Given a set of indices $\mathcal{E}$, $[x_e]_{e\in\mathcal{E}}$ denotes the column vector with components x_e.

2.1 The model

2.1.1 The full information routing game

The network

We consider a directed graph (i.e. road network) with a finite vertex set $\mathcal{V}$ and a finite edge set $\mathcal{E}$, where each edge (i.e. road) $e \in \mathcal{E}$ connects two vertices (i.e. locations) $u, v \in \mathcal{V}$. We denote such a road network by $\mathcal{N}(\mathcal{V}, \mathcal{E})$ or $\mathcal{N}$ for brevity. In a network $\mathcal{N}$, a path r (i.e. route) is an alternating sequence of vertices and edges that begins and ends with vertices $(v_0, e_1, v_1, \ldots, e_n, v_n)$ such that all vertices are distinct and each edge in the sequence is such that $e_j = (v_{j-1}, v_j) \in \mathcal{E}$. We denote by $\mathcal{R}$ the set of all paths and we use the notation $e \in r$ to state that edge e belongs to the path r. We denote by $\mathcal{W}$ the set of origin-destination pairs considered in the road network and by $\mathcal{R}_w$ the set of paths connecting a certain origin-destination pair $w \in \mathcal{W}$, that is, the set of paths with v_0 equals to the origin and v_n equals to the destination in w.

Feasibility

We assume that for each origin-destination pair $w \in \mathcal{W}$ there is a tra□c demand $d_w \geq 0$ that needs to be allocated using the routes in $\mathcal{R}_w$. We use the symbol x_r to

denote the total flow assigned to route r. A travel assignment $\{x_r\}_{r\in\mathcal{R}}$ is feasible if $x_r \geq 0$ for all $r \in \mathcal{R}$ and $\sum_{r\in\mathcal{R}_w} x_r = d_w$ for all $w \in \mathcal{W}$, that is, if the tra□c demand is satisfied. We denote the associated total flow on edge e by $x_e := \sum_{r\in\mathcal{R}|e\in r} x_r$ and the total edge flow vector by $x := [x_e]_{e\in\mathcal{E}}$.

Travel time

We assume that the travel time on each edge $e \in \mathcal{E}$ depends on the congestion level x_e according to the following stochastic latency function.

Assumption 1 (Stochastic latency function). *For each edge $e \in \mathcal{E}$ and congestion level x_e the experienced travel time of the users is stochastic and is given by*

$$y_e(x_e) := l_e(x_e) + \epsilon_e$$

where $l_e(x_e)$ is a deterministic, continuous, non-decreasing positive function of x_e and ϵ_e is a zero mean random variable with variance σ_e^2.

Note that since ϵ_e is a zero mean random variable, the deterministic function $l_e(x_e)$ coincides with the expected travel time on edge e when the congestion level is x_e. For this reason we call $l_e(x_e)$ the *expected latency function*. The stochastic term ϵ_e, on the other hand, models an additive error term that accounts for the variability in travel time due to incidents or other exogenous factors (weather, time-of-day etc.).

Often, latency functions in tra□c networks are modeled through polynomial functions [28]. In this study we focus for simplicity on a□ne latency functions.

Assumption 2 (A□ne expected latency functions). *For each edge $e \in \mathcal{E}$ we consider a ne expected latency functions $l_e(x_e) := a_e x_e + b_e$ where $b_e \geq 0$ is the free flow travel time and the congestion term is modeled as $a_e x_e$ where x_e is the flow on edge e and $a_e > 0$ is the congestion coe cient.*

We denote the set of expected latency functions for all edges by $\mathcal{L} := \{l_e | e \in \mathcal{E}\}$ and we use $\mathcal{G}(\mathcal{N}, \mathcal{L})$ to denote a routing game with network $\mathcal{N}$ and expected

latency functions $\mathcal{L}$. Throughout the chapter we refer to the game $\mathcal{G}(\mathcal{N}, \mathcal{L})$ as the *full information* routing game.

Wardrop equilibrium

Given a routing game $\mathcal{G}(\mathcal{N}, \mathcal{L})$, we assume that the tra□c demand gets allocated across di□erent routes according to the following slightly modified definition of Wardrop equilibrium.

Definition 1 (Wardrop equilibrium). *A Wardrop equilibrium of a game $\mathcal{G}(\mathcal{N}, \mathcal{L})$ is a feasible flow allocation $\{x_r\}_{r \in \mathcal{R}}$ such that for any origin-destination pair $w \in \mathcal{W}$ and any route $\bar{r} \in \mathcal{R}_w$ with $x_{\bar{r}} > 0$ we have that for all $r \in \mathcal{R}_w$*

$$\mathbb{E}[y_{\bar{r}}(x)] \leq \mathbb{E}[y_r(x)]$$

where the latency of a route r under the total edge flow vector x is defined as the sum of the latencies of the edges on that route, that is,

$$y_r(x) := \sum_{e \in r} y_e(x_e).$$

In other words, the Wardrop equilibrium detailed in Definition 1 corresponds to a tra□c assignment where any route with positive flow has equal or lower *expected* travel time than any other possible route connecting the same origin-destination pair under the flow x. Definition 1 is a slightly modified version of the standard definition of Wardrop equilibrium which is formulated for routing games with *deterministic* latency functions. We note however that, since $\mathbb{E}[y_e(x_e)] = l_e(x_e)$, Definition 1 coincides with the standard definition of Wardrop equilibrium for a routing game with deterministic latency functions equal to the expected latency functions $l_e(x_e)$.[1] Because of this equivalence, it is immediate to show that, under Assumption 2, the Wardrop equilibrium in Definition 1 is unique.

[1] It is well known that this notion coincides with the Nash equilibrium of a routing game with infinitesimal (nonatomic) users (see [38]).

Lemma 1 (Uniqueness [60]). *Under Assumptions 1 and 2 the Wardrop equilibrium exists and is unique.*

2.1.2 Partial information, flow allocation and learning dynamics

Contrary to most of the literature on routing games, we assume that the users do not completely know the latency functions. Specifically, we assume that users know the free flow travel time b_e for each edge $e \in \mathcal{E}$, but do not know the congestion coefficient a_e. Our motivation is that while b_e depends on fixed parameters, (such as the road length, the speed limit, etc.) the coefficient a_e that models the effect of congestion is usually difficult to characterize a priori. In our model, the users estimate such parameters $\{a_e\}_{e\in\mathcal{E}}$ by using past observations of travel time.

Specifically, we assume that at the initial step $k = 1$ the flow gets allocated according to an initial feasible edge flow vector x^1, such that $x_e^1 > 0$ for all $e \in \mathcal{E}$. Based on the observation of the travel time in each edge, the users build a first estimate $\hat{a}_e^1$ of the congestion coefficient for each edge e. Note that since all users have the same observations they all produce the same estimates.

For each step $k > 1$ all users, based on the estimates $\hat{a}_e^{k-1}$ obtained with the previous $k-1$ observations,

1. allocate their flow according to a Wardrop equilibrium of the *partial information* game $\mathcal{G}(\mathcal{N}, \mathcal{L}^{k-1})$, where

$$\mathcal{L}^{k-1} := \{\hat{l}_e^{k-1}(x_e) := \hat{a}_e^{k-1}x_e + b_e \mid e \in \mathcal{E}\}$$

   is the set of estimated expected latency functions;

2. for each edge e with positive flow $x_e^k > 0$, observe a realization of the travel time for that edge at the current flow level, i.e. they observe

$$y_e^k := l_e(x_e^k) + \epsilon_e^k, \tag{2.1}$$

where ϵ_e^k is a realization of ϵ_e;

3. use the information on the experienced travel time at step k to update the congestion coefficient estimates to $\hat{a}_e^k$.

There are two assumptions that justify the allocation of the flow in point 1). First, users are risk neutral and therefore care about the expected travel time. Second, the traffic flow allocation takes place faster than information updates, leading to a Wardrop equilibrium at each step [27]. Regarding the observation of travel time in point 2) we make the following assumption of independence between different steps

Assumption 3. *For each edge $e \in \mathcal{E}$ and step $k \in \mathbb{N}$, ϵ_e^k are i.i.d. realizations of ϵ_e.*

In the next subsection we describe in more detail the update that each user performs in point 3). The resulting learning dynamics is then summarized in Table 2.1.

2.1.3 Parameter Estimation

We assume users perform ordinary least squares estimation to update the parameters given the information collected up to step k. This rule avoids the need to specify tight priors and compute posteriors on unknown parameters and lead to simple calculations for the users. In particular, given k observations of travel times

$$\{y_e^i = a_e x_e^i + b_e + \epsilon_e^i\}_{i=1}^k \tag{2.2}$$

where $\{x_e^i\}_{i=1}^k$, b_e are known and $\{\epsilon_e^i\}_{i=1}^k$ are i.i.d samples of the random variable ϵ_e, the least squares estimate of a_e at step k is given by

$$\begin{aligned}\hat{a}_e^{\mathsf{LS}}(\{x_e^i, y_e^i\}_{i=1}^k) &:= \arg\min_{a_e \in \mathbb{R}} \sum_{i=1}^k (y_e^i - a_e x_e^i - b_e)^2 \\ &= \arg\min_{a_e \in \mathbb{R}} \sum_{i=1}^k (a_e x_e^i)^2 - 2a_e x_e^i (y_e^i - b_e) \\ &= \arg\min_{a_e \in \mathbb{R}} a_e^2 \sum_{i=1}^k (x_e^i)^2 - 2a_e \sum_{i=1}^k x_e^i (y_e^i - b_e) \\ &= \frac{1}{\sum_{i=1}^k (x_e^i)^2} \sum_{i=1}^k x_e^i (y_e^i - b_e).\end{aligned} \tag{2.3}$$

Remark 1. *Ordinary least squares is used as an estimator because in many cases it is consistent, which means that the least squares estimate converges in probability to the actual parameter. One of the assumptions that guarantees this consistency property is that the samples x_e^i are drawn in an i.i.d. fashion. In our case x_e^k is dependent on the estimates vector $\hat{a}^{k-1}$ used to compute the Wardrop equilibrium at step k, which itself depends on (x^i, y^i) for all $i < k$. Thus, the samples $\{x_e^i\}_{i=1}^\infty$ are not i.i.d. and no immediate convergence result for the least squares estimator is available.*

Our goal in this chapter is to show that by using the least squares estimate in (2.3) the learning dynamics specified in Table 2.1 converges almost surely to the full information Wardrop equilibrium.

Table 2.1: The learning dynamics

Initialize: For each edge $e \in \mathcal{E}$ send the initial feasible flow $\hat{x}_e^1 \in \mathbb{R}_{>0}$, measure y_e^1 and set $\hat{a}_e^1 = \hat{a}_e^{\mathsf{LS}}(\{x_e^1, y_e^1\})$. Set $k = 2$.

Iterate until convergence:

1) Compute the Wardrop equilibrium

$$x^k = \text{ Wardrop equilibrium of the game } \mathcal{G}(\mathcal{N}, \mathcal{L}^{k-1}) \tag{2.4a}$$

2) Measurements

for each $e \in \mathcal{E}$
  if $x_e^k > 0$ measure
    $y_e^k = l_e(x_e^k) + \epsilon_e^k$
  else set
    $y_e^k = b_e$
end (2.4b)

3) Update the congestion coe cient estimates

for each $e \in \mathcal{E}$
  $\hat{a}_e^k = \hat{a}_e^{\mathsf{LS}}(\{x_e^i, y_e^i\}_{i=1}^k)$
end (2.4c)

$k \leftarrow k + 1$

Note that in step 2) if $x_e^k = 0$ the value of y_e^k is actually irrelevant and in any case leads to $\hat{a}_e^k = \hat{a}_e^{k-1}$ in step 3).

2.2 Preliminary results

In this section, we derive some preliminary results needed in Section 2.3 to prove convergence of the learning dynamics in Table 2.1. Note that the formula for the ordinary least squares estimator derived in (2.3) can be equivalently rewritten, by

plugging in the values for y_e^i, as

$$\begin{aligned}\hat{a}_e^k &= \frac{1}{\sum_{i=1}^k (x_e^i)^2} \sum_{i=1}^k x_e^i (y_e^i - b_e) \\ &= \frac{1}{\sum_{i=1}^k (x_e^i)^2} \sum_{i=1}^k x_e^i (a_e x_e^i + \epsilon_e^i) \\ &= \frac{1}{\sum_{i=1}^k (x_e^i)^2} \sum_{i=1}^k a_e (x_e^i)^2 + \epsilon_e^i x_e^i \\ &= a_e + \frac{\sum_{i=1}^k \epsilon_e^i x_e^i}{\sum_{i=1}^k (x_e^i)^2}.\end{aligned}$$

Hence the error in the estimate of the congestion coe□cient at step k is

$$\text{err}_e^k := \hat{a}_e^k - a_e = \frac{\sum_{i=1}^k \epsilon_e^i x_e^i}{\sum_{i=1}^k (x_e^i)^2}. \tag{2.5}$$

Our main preliminary result is to construct an auxiliary martingale s_e^k, related to the error term in (2.5), and show its almost sure convergence.

Definition 2 (Martingale pg. 474 in [69]). *A sequence of random variables s^k is a martingale if for all $k \geq 1$,*

1. $\mathbb{E}[s^k | s^{k-1}, ..., s^1] = s^{k-1}$ *and*
2. $\mathbb{E}[|s^k|] < \infty$.

Lemma 2. *For each edge $e \in \mathcal{E}$, let $\{x_e^i\}_{i=1}^\infty$ be as defined in Table 2.1. Then under Assumptions 1 and 3 for all $i \in \mathbb{N}$, ϵ_e^i is independent of $\{x_e^i, x_e^{i-1}, ..., x_e^1, \epsilon_e^{i-1}, ..., \epsilon_e^1\}$ and the stochastic process*

$$s_e^k = \sum_{i=1}^k \frac{x_e^i \epsilon_e^i}{\sum_{j=1}^i (x_e^j)^2} \tag{2.6}$$

is a martingale and converges almost surely to a finite value as $k \to \infty$.

Proof. We first show that ϵ_e^i is independent of $\{x_e^i, x_e^{i-1}, ..., x_e^1, \epsilon_e^{i-1}, ..., \epsilon_e^1\}$ for all i. By Assumption 3 ϵ_e^i is independent of $\epsilon_e^1, ..., \epsilon_e^{i-1}$. Moreover, by the learning dynamics in Table 2.1, x_e^i depends only on the estimates $\hat{a}^{i-1}$ and these estimates depend only of the noise up to step $i-1$. Thus, ϵ_e^i is independent of x_e^j for all $j < i$.

To prove the second statement we use a similar argument as in [72, Lemma 3]. Specifically, we first show that s_e^k is a martingale. Then we argue that we can use martingale convergence theorem and we have that the result follows [69, Chapter 7, Section 4].

i) To prove that $\mathbb{E}[s_e^k|s_e^{k-1},...,s_e^1] = s_e^{k-1}$ note that for all $k \in \mathbb{N}$

$$\begin{aligned}
\mathbb{E}[s_e^k|s_e^{k-1},...,s_e^1] &= \mathbb{E}\left[\sum_{i=1}^{k} \frac{x_e^i \epsilon_e^i}{\sum_{j=1}^{i}(x_e^j)^2}\Big| s_e^{k-1},...,s_e^1\right] \\
&= \mathbb{E}\left[\frac{x_e^k \epsilon_e^k}{\sum_{j=1}^{k}(x_e^j)^2}\Big| s_e^{k-1},...,s_e^1\right] + \mathbb{E}\left[\sum_{i=1}^{k-1} \frac{x_e^i \epsilon_e^i}{\sum_{j=1}^{i}(x_e^j)^2}\Big| s_e^{k-1},...,s_e^1\right] \\
&= \mathbb{E}\left[\epsilon_e^k\right] \mathbb{E}\left[\frac{x_e^k}{\sum_{j=1}^{k}(x_e^j)^2}\Big| s_e^{k-1},...,s_e^1\right] + s_e^{k-1} \\
&= 0 + s_e^{k-1} = s_e^{k-1}.
\end{aligned}$$

ii) We now show that $\mathbb{E}\left[|s_e^k|\right]$ is bounded. To this end, note that

$$\mathbb{E}\left[|s_e^k|\right] = \mathbb{E}\left[\sqrt{(s_e^k)^2}\right] \leq \sqrt{\mathbb{E}\left[(s_e^k)^2\right]},$$

where we used Jensen's inequality [69, pg. 192]. So it su□ces to show that $\mathbb{E}[(s_e^k)^2]$ is bounded. Note that

$$\begin{aligned}
\mathbb{E}\left[(s_e^k)^2\right] &= \mathbb{E}\left[\left(\sum_{i=1}^{k} \frac{x_e^i \epsilon_e^i}{\sum_{j=1}^{i}(x_e^j)^2}\right)^2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{k}\left(\frac{x_e^i \epsilon_e^i}{\sum_{j=1}^{i}(x_e^j)^2}\right)^2\right] \\
&= \sigma_e^2 \mathbb{E}\left[\sum_{i=1}^{k}\left(\frac{x_e^i}{\sum_{j=1}^{i}(x_e^j)^2}\right)^2\right] \leq \sigma_e^2 \frac{2}{(x_e^1)^2},
\end{aligned} \tag{2.7}$$

where the second equality comes from the fact that for any i, ϵ_e^i is independent from x_e^i, ϵ_e^j, and x_e^j for all $j < i$. Thus, for any $i \neq j$, by assuming without loss of generality

that $j < i$, it holds that

$$\mathbb{E}\left[\frac{x_e^i \epsilon_e^i}{\sum_{m=1}^{i}(x_e^m)^2}\frac{x_e^j \epsilon_e^j}{\sum_{m=1}^{j}(x_e^m)^2}\right] = \mathbb{E}\left[\epsilon_e^i\right]\mathbb{E}\left[\frac{x_e^i}{\sum_{m=1}^{i}(x_e^m)^2}\frac{x_e^j \epsilon_e^j}{\sum_{m=1}^{j}(x_e^m)^2}\right] = 0.$$

The last inequality in (2.7) comes from [72, Lemma 1]. Note that $x_e^1 > 0$ is a deterministic quantity, as detailed in Table 2.1. Thus, it holds that

$$\mathbb{E}\left[|s_e^k|\right] \leq \sqrt{\sigma_e^2 \frac{2}{(x_e^1)^2}} < \infty. \tag{2.8}$$

Now, to utilize martingale convergence theorem we need that $\sup_k \mathbb{E}[|s_e^k|] < \infty$. This fact follows from (2.8), as we have that for every value of k $\mathbb{E}[|s_e^k|]$is less than the deterministic value$\sqrt{\sigma_e^2 \frac{2}{(x_e^1)^2}}$. Therefore, we can apply the martingale convergence theorem [69, Chapter 7, Section 4] and we obtain that $\{s_e^k\}_{k=1}^{\infty}$ converges almost surely to a finite value. ∎

We now state two lemmas that hold for deterministic sequences and will be useful in examining the behavior of the deterministic sample paths associated to specific realizations of the noise.

Lemma 3 (Kronecker's lemma pg. 390 in [69]). *If $\{h_k\}_{k=1}^{\infty}$ and $\{g_k\}_{k=1}^{\infty}$ are two real-valued sequences for which $\{g_k\}_{k=1}^{\infty}$ is non-negative and non-decreasing to infinity then the existence of a finite-valued s such that*

$$\lim_{k\to\infty} s_k := \lim_{k\to\infty} \sum_{i=1}^{k} \frac{h_i}{g_i} = s$$

implies that

$$\lim_{k\to\infty} \frac{1}{g_k} \sum_{i=1}^{k} h_i = 0.$$

The next result is similar to the above one, except that it assumes convergence of g_k to a finite value instead of an infinite value. Consequently, one gets convergence to a finite value instead of zero.

Lemma 4 (Lemma 2(ii) in [72]). *If $\{h_k\}_{k=1}^{\infty}$ and $\{g_k\}_{k=1}^{\infty}$ are two real-valued sequences for which $\{g_k\}_{k=1}^{\infty}$ is non-negative, non-decreasing, and converges to a value $M < \infty$ then the existence of a finite-valued s such that*

$$\lim_{k\to\infty} s_k := \lim_{k\to\infty} \sum_{i=1}^{k} \frac{h_i}{g_i} = s$$

implies that

$$\lim_{k\to\infty} \frac{1}{g_k} \sum_{i=1}^{k} h_i$$

exists and is finite.

Proof. This proof is similar to [72, Lemma 2(ii)] with the extended conclusion that the limit is finite. Let $s_0 = 0$ and $s_k = \sum_{i=1}^{k} \frac{h_i}{g_i}$ for all $k \in \mathbb{N}$. Thus, we have $h_k = g_k(s_k - s_{k-1})$ and

$$\frac{1}{g_k} \sum_{i=1}^{k} h_i = \frac{1}{g_k} \sum_{i=1}^{k} g_i(s_i - s_{i-1}) = s_k - \frac{1}{g_k} \sum_{i=1}^{k-1} (g_{i+1} - g_i)s_i.$$

Now, by assumption $s_k \to s$ where s is finite. The result is thus proven if we show that also the second term $\frac{1}{g_k} \sum_{i=1}^{k-1} (g_{i+1} - g_i)s_i$ converges to a finite value. Set an arbitrary value $\delta > 0$ and choose k_0 such that $|s_k - s| < \delta$ for all $k > k_0$. Now, we have

$$\begin{aligned}
\frac{1}{g_k} \sum_{i=1}^{k-1} (g_{i+1} - g_i)s_i &= \frac{1}{g_k}(g_k - g_1)s + \frac{1}{g_k} \sum_{i=1}^{k-1} (g_{i+1} - g_i)(s_i - s) \\
&= \frac{1}{g_k}(g_k - g_1)s + \frac{1}{g_k} \sum_{i=1}^{k_0-1} (g_{i+1} - g_i)(s_i - s) + \frac{1}{g_k} \sum_{i=k_0}^{k-1} (g_{i+1} - g_i)(s_i - s).
\end{aligned}$$

Taking the limit as $k \to \infty$, we have that the first term converges to $\frac{1}{M}(M - g_1)s$ and the second term converges to $\frac{1}{M} \sum_{i=1}^{k_0-1} (g_{i+1} - g_i)(s_i - s)$, which are both finite. Now,

the absolute value of the third term is

$$\begin{aligned}\left|\frac{1}{g_k}\sum_{i=k_0}^{k-1}(g_{i+1}-g_i)(s_i-s)\right| &\leq \frac{1}{g_k}\sum_{i=k_0}^{k-1}(g_{i+1}-g_i)|s_i-s| \\ &\leq \frac{1}{g_k}\sum_{i=k_0}^{k-1}(g_{i+1}-g_i)\delta \\ &= \frac{1}{g_k}(g_k-g_{k_0})\delta < \delta,\end{aligned}$$

where we used that $\{g_k\}_{k=1}^{\infty}$ is non-negative and non-decreasing. Overall, we have proven that the limit exists and is finite. ∎

Finally, in proving our main result we use Theorem 3.1 from Dafermos and Nagurney on sensitivity analysis of the Wardrop equilibrium under a change of (expected) latency functions [29]. This result is rewritten for our scenario and proven below.

Lemma 5 (Theorem 3.1 in [29]). *Suppose that Assumption 2 holds. Let x^k be a Wardrop equilibrium of the partial information game $\mathcal{G}(\mathcal{N},\mathcal{L}^k)$ and $\bar{x}$ be the full information Wardrop equilibrium, that is, the Wardrop equilibrium of the game $\mathcal{G}(\mathcal{N},\mathcal{L})$. Also, let $l_e(x_e;a_e) := a_e x_e + b_e$, where we made explicit the dependence of the expected latency function on the congestion coe cient. Then*

$$\|x^k-\bar{x}\|^2 \leq \frac{1}{\alpha^2}\sum_{e=1}^{\mathcal{E}}(l_e(x_e^k;a_e)-l_e(x_e^k;\hat{a}_e^k))^2$$

where $\alpha := \min\{a_e\}_{e\in\mathcal{E}} > 0$.

Proof. Dafermos and Nagurney prove the result for general latency functions. We apply their result to changes in latency functions from $a_e x_e + b_e$ to $\hat{a}_e^k x_e + b_e$. To this end, we briefly recall that the (unique) Wardrop equilibrium of the game $\mathcal{G}(\mathcal{N},\mathcal{L})$ can be equivalently characterized as the (unique) solution to the variational inequality (VI) VI$(\bar{F},\mathcal{X})$, where the operator $\bar{F}:\mathbb{R}^{|\mathcal{E}|}\to\mathbb{R}^{|\mathcal{E}|}$ is defined as

$$\bar{F}(x) := [a_e x_e + b_e]_{e\in\mathcal{E}}$$

and $\mathcal{X}$ is the feasible set for the total edge flow vector x. For a definition of variational inequality and a proof of the equivalence mentioned above we refer to [29]. Similarly, let $F^k(\cdot)$ be the operator of the VI associated with the coe□cients $\hat{a}_e^k$, that is $F^k(x) := [\hat{a}_e^k x_e + b_e]_{e\in\mathcal{E}}$. Since $\bar{x}$ solves the VI in $\bar{F}(\cdot)$ and x^k solves the VI in $F^k(\cdot)$ by definition

$$\begin{aligned}\bar{F}(\bar{x})^\top(x^k - \bar{x}) &\geq 0\\ F^k(x^k)^\top(\bar{x} - x^k) &\geq 0.\end{aligned}$$

Subtracting the second inequality from the first one yields

$$\begin{aligned}&[\bar{F}(\bar{x}) - F^k(x^k)]^\top(x^k - \bar{x}) \geq 0,\\ &[\bar{F}(\bar{x}) - \bar{F}(x^k) + \bar{F}(x^k) - F^k(x^k)]^\top(x^k - \bar{x}) \geq 0,\\ &[\bar{F}(\bar{x}) - \bar{F}(x^k)]^\top(x^k - \bar{x}) + [\bar{F}(x^k) - F^k(x^k)]^\top(x^k - \bar{x}) \geq 0,\\ &[\bar{F}(x^k) - F^k(x^k)]^\top(x^k - \bar{x}) \geq [\bar{F}(\bar{x}) - \bar{F}(x^k)]^\top(\bar{x} - x^k).\end{aligned}$$

Plugging in the definition of $\bar{F}(\cdot)$ we get

$$\begin{aligned}[\bar{F}(\bar{x}) - \bar{F}(x^k)]^\top(\bar{x} - x^k) &= \sum_{e\in\mathcal{E}} a_e(\bar{x}_e - x_e^k)^2\\ &\geq \alpha\|\bar{x} - x^k\|^2\end{aligned}$$

and by Cauchy-Schwartz

$$\begin{aligned}[\bar{F}(x^k) - F^k(x^k)]^\top(x^k - \bar{x}) &\leq |[\bar{F}(x^k) - F^k(x^k)]^\top(x^k - \bar{x})|\\ &\leq \|\bar{F}(x^k) - F^k(x^k)\|\|x^k - \bar{x}\|.\end{aligned}$$

The last three inequalities give us

$$\|\bar{F}(x^k) - F^k(x^k)\|\|x^k - \bar{x}\| \geq \alpha\|\bar{x} - x^k\|^2$$

and thus

$$\frac{1}{\alpha}\|\bar{F}(x^k) - F^k(x^k)\| \geq \|\bar{x} - x^k\|$$

which concludes the proof. ∎

2.3 Convergence to the full information Wardrop equilibrium

In this section we combine the previously stated lemmas to show our main result, which is that the learning dynamics in Table 2.1 converges to the full information Wardrop equilibrium almost surely. Additionally, we show that all the congestion coe□cients corresponding to edges that are used in the full information Wardrop equilibrium are learned almost surely. To this end, let us denote by ϵ a specific noise realization of $\{\epsilon_e^k\}_{e\in\mathcal{E},k\geq 0}$ and by $s_e^k(\epsilon), x_e^k(\epsilon), \text{err}_e^k(\epsilon)$ the corresponding deterministic realizations of the martingale s_e^k, the total edge flow x_e^k, and the error term err_e^k at step k, respectively. Lemma 2 guarantees that the set of noise realizations

$$\Sigma := \{\epsilon \mid \lim_{k\to\infty} s_e^k(\epsilon) \text{ exists and is finite for all } e \in \mathcal{E}\} \tag{2.9}$$

has probability one. To prove our main result, we consider each noise realization $\epsilon \in \Sigma$ separately and we partition the edges into two sets, $S^{\infty}(\epsilon)$ and $S^{\text{finite}}(\epsilon)$, according to a specific property of the sample paths $x_e^k(\epsilon)$. For any edge in $S^{\infty}(\epsilon)$ we show the error term $\text{err}_e^k(\epsilon)$ on the estimator goes to 0 and for any edge in $S^{\text{finite}}(\epsilon)$ we show that $\text{err}_e^k(\epsilon)$ is bounded. We use these two facts to show that the flow $x^k(\epsilon)$ converges to the full information Wardrop equilibrium under the considered noise realization $\epsilon \in \Sigma$. Since the set of noise realizations Σ has probability one we thus have that the learning dynamics in Table 2.1 converge almost surely to the full information Wardrop equilibrium.

We start by studying the behaviour of the error term $\text{err}_e^k(\epsilon)$ in (2.5) for a fixed noise realization $\epsilon \in \Sigma$.

Lemma 6. *Suppose that Assumptions 1, 2, and 3 hold. Under any realization of the error ϵ the edges can be split into two sets*

1. $S^{\infty}(\epsilon) = \left\{e \in \mathcal{E} \mid \sum_{k=1}^{\infty}(x_e^k(\epsilon))^2 = \infty\right\}$
2. $S^{\text{finite}}(\epsilon) = \left\{e \in \mathcal{E} \mid \sum_{k=1}^{\infty}(x_e^k(\epsilon))^2 < \infty\right\}$.

If $\epsilon \in \Sigma$ then

1. *For any edge $e \in S^{\infty}(\epsilon)$ the error term in the ordinary least squares estimate*

$$\text{err}_e^k(\epsilon) = \frac{\sum_{i=1}^k \epsilon_e^i x_e^i(\epsilon)}{\sum_{i=1}^k (x_e^i(\epsilon))^2} \to 0$$

   as $k \to \infty$.

2. *For any edge $e \in S^{\text{finite}}(\epsilon)$ the error term in the ordinary least squares estimate*

$$\text{err}_e^k(\epsilon) = \frac{\sum_{i=1}^k \epsilon_e^i x_e^i(\epsilon)}{\sum_{i=1}^k (x_e^i(\epsilon))^2}$$

   is bounded.

Proof. Construct the two deterministic sequences $h_k = x_e^k(\epsilon)\epsilon_e^k$, $g_k = \sum_{i=1}^k (x_e^i(\epsilon))^2$ so that $\{g_k\}_{k=1}^{\infty}$ is non-negative and non-decreasing and $s_e^k(\epsilon) = \sum_{i=1}^k \frac{h_i}{g_i}$. Note that if $\epsilon \in \Sigma$ then by definition $\lim_{k\to\infty} s_e^k(\epsilon) = \lim_{k\to\infty} \sum_{i=1}^k \frac{h_i}{g_i}$ exists and is finite.

1. By definition of $S^{\infty}(\epsilon)$, $g_k \to \infty$. Thus, we can apply Kronecker's lemma (Lemma 3) and we have

$$\frac{1}{g_k}\sum_{i=1}^k h_i = \frac{1}{\sum_{i=1}^k (x_e^i(\epsilon))^2}\sum_{i=1}^k x_e^i(\epsilon)\epsilon_e^i = \text{err}_e^k(\epsilon) \to 0.$$

2. By definition of $S^{\text{finite}}(\epsilon)$ the sequence g_k converges to a finite value. Thus, we can apply Lemma 4 and we obtain that $\frac{\sum_{i=1}^k \epsilon_e^i x_e^i(\epsilon)}{\sum_{i=1}^k (x_e^i(\epsilon))^2} \to p_e(\epsilon)$ for some finite $p_e(\epsilon)$. Thus, for any $\delta > 0$ there exists a $\hat{k} > 0$ such that for any $k > \hat{k}$ we have

$$\left|\frac{\sum_{i=1}^k \epsilon_e^i x_e^i(\epsilon)}{\sum_{i=1}^k (x_e^i(\epsilon))^2} - p_e(\epsilon)\right| < \delta$$

and the error $\text{err}_e^k(\epsilon)$ is bounded.

∎

Theorem 1. *Suppose that Assumptions 1, 2, and 3 hold. Then the learning dynamics in Table 2.1 converge almost surely to the unique Wardrop equilibrium $\bar{x}$ of the full information game $\mathcal{G}(\mathcal{N}, \mathcal{L})$, that is,*

$$x^k \to \bar{x}, \quad \textit{almost surely.}$$

Proof. Let us consider an arbitrary, fixed realization of the error $\epsilon \in \Sigma$ as defined in (2.9). Now from Lemma 6 we have that

1. if $e \in S^\infty(\epsilon)$ then $\hat{a}_e^k(\epsilon) \to a_e$;
2. if $e \in S^{\text{finite}}(\epsilon)$ then $x_e^k(\epsilon) \to 0$ and there exists $\hat{k} > 0$, $M_e(\epsilon) > 0$ such that $|a_e^k(\epsilon) - a_e| < M_e(\epsilon)$ for all $k > \hat{k}$.

These two statements conclude the proof because then by Lemmas 5 and 6 we obtain that for any $k > \hat{k}$

$$\begin{aligned}
\|x^k(\epsilon) - \bar{x}\|^2 &\leq \frac{1}{\alpha^2} \sum_{e \in \mathcal{E}} (l_e(x_e^k(\epsilon); a_e) - l_e(x_e^k(\epsilon); \hat{a}_e^k(\epsilon)))^2 \\
&= \frac{1}{\alpha^2} \sum_{e \in \mathcal{E}} (x_e^k(\epsilon))^2 (a_e - \hat{a}_e^k(\epsilon))^2 \\
&= \frac{1}{\alpha^2} \sum_{e \in \mathcal{E}} (x_e^k(\epsilon))^2 (\text{err}_e^k(\epsilon))^2 \\
&= \frac{1}{\alpha^2} \sum_{e \in S^\infty(\epsilon)} (x_e^k(\epsilon))^2 (\text{err}_e^k(\epsilon))^2 + \frac{1}{\alpha^2} \sum_{e \in S^{\text{finite}}(\epsilon)} (x_e^k(\epsilon))^2 (\text{err}_e^k(\epsilon))^2 \\
&\leq \frac{1}{\alpha^2} \sum_{e \in S^\infty(\epsilon)} d^2 (\text{err}_e^k(\epsilon))^2 + \frac{1}{\alpha^2} \sum_{e \in S^{\text{finite}}(\epsilon)} (x_e^k(\epsilon))^2 (M_e(\epsilon))^2 \to 0,
\end{aligned}$$

where we used that $x_e^k(\epsilon) \leq d$ for all $e \in \mathcal{E}$ and $k \in \mathbb{N}$, where $d = \sum_{w \in \mathcal{W}} d_w$ is the total travel demand and the fact that, in the limit as $k \to \infty$, the first term in the last step goes to 0 since $\text{err}_e^k(\epsilon) \to 0$ for all $e \in S^\infty(\epsilon)$ and the second term goes to 0

since $(x_e^k(\epsilon))^2 \to 0$ for any $e \in S^{\text{finite}}(\epsilon)$. Hence for any realization $\epsilon \in \Sigma$, $x^k(\epsilon) \to \bar{x}$. Since, by Lemma 2, Σ has probability one, $x^k \to \bar{x}$ a.s. ∎

Corollary 1. *Suppose that Assumptions 1, 2, and 3 hold. For any edge e such that $\bar{x}_e > 0$, where $\bar{x}$ is the full information Wardrop equilibrium, $\hat{a}_e^k \to a_e$ a.s.*

Proof. By Theorem 1, $x^k(\epsilon) \to \bar{x}$ for all $\epsilon \in \Sigma$. Consequently, for any edge $e \in \mathcal{E}$ such that $\bar{x}_e > 0$ and for any $\epsilon \in \Sigma$ we have that $x_e^k(\epsilon) \to \bar{x}_e > 0$. Consequently, it must be that $e \in S^{\infty}(\epsilon)$. By Lemma 6, we then have that $\text{err}_e^k(\epsilon) \to 0$. Since Σ has probability one, we have proven that for any edge $e \in \mathcal{E}$ such that $\bar{x}_e > 0$ it holds $\text{err}_e^k \to 0$ a.s. ∎

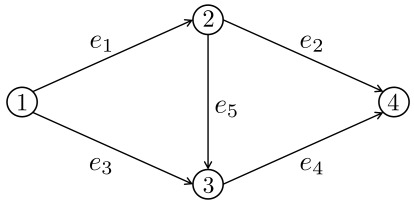
2.4 Simulation

To illustrate our theoretical results we consider a routing game over the five road Wheatstone network illustrated in Figure 2-1-A). For simplicity, we assume that $d = 1$ unit of tra□c needs to be routed from vertex 1 to vertex 4 and that the expected latency for each edge is $l_e(x_e) = x_e$. With these settings, the Wardrop equilibrium of the full information game is $\bar{x} = [0.5, 0.5, 0.5, 0.5, 0]^\top$, that is, 0.5 flow is sent in all edges except for edge 5 which is not used.

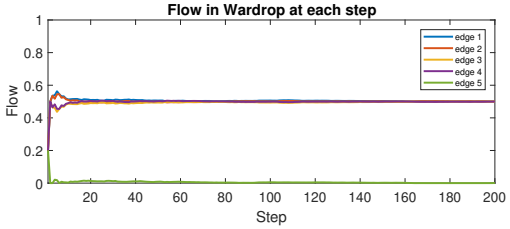
Figure 2-1-B) shows the evolution of x_e^k for each edge $e \in \mathcal{E} = \{1, 2, 3, 4, 5\}$ as a function of the step $k \in [1, 200]$ according to the learning dynamics illustrated in Table 2.1 (for one realization of the noise). The partial information Wardrop equilibrium x^k at each step k was computed by using the projection algorithm [31, Algorithm 12.1.1] applied to the $\text{VI}(F^k, \mathcal{X})$ (see the proof of Lemma 5). Note that $x^k \to \bar{x}$.

Figure 2-1-C) shows the evolution of the estimates $\hat{a}_e^k$ as a function of the step $k \in [1, 200]$. Note that the estimates of the congestion coe□cients of edges 1-4 converge to the true value $a_e = 1$. On the other hand, $\hat{a}_5^k$ does not converge to $a_5 = 1$. This is consistent with Corollary 1 since edge 5 is not used in the full information Wardrop equilibrium.

A)



B)



C)

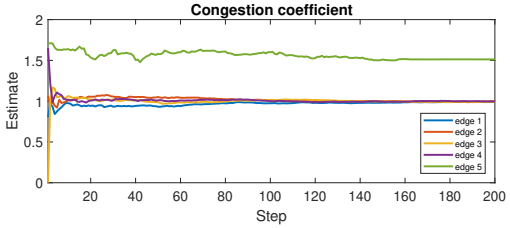


Figure 2-1: A) Wheatstone network with one unit of travel demand from node 1 to node 4, the full information Wardrop equilibrium is $\bar{x} = [0.5, 0.5, 0.5, 0.5, 0]^\top$. B) Plot of one realization of x_e^k as a function of the step $k \in [1, 200]$. C) Plot of one realization of the estimates $\hat{a}_e^k$ as a function of the step $k \in [1, 200]$.

2.5 Conclusion

In this chapter, we examine the learning dynamics in a nonatomic routing game on a network with affine stochastic edge latency functions with unknown congestion coefficients. We show that by using an ordinary least squares estimator to update the coefficient estimates the learning dynamics converges, in an almost sure sense, to the Wardrop equilibrium of the full information game and that the coefficients of the edges used in such an equilibrium are learned.

We view our work as a first step since we here assumed that all users have access to the same information. A natural but challenging extension is to assume that each user applies a similar least-squares type estimation to a user-specific information set, thus leading to a situation of heterogeneous information and routing decisions across users.

Chapter 3

Learning Dynamics in Network Games

In many strategic environments, agents' interactions are heterogeneous and an agent is directly a□ected by the decisions of its neighbors in a network of interactions. Examples range from firm competition [20, 4], adoption of innovation [11] and local public good provision [18, 6] to telecommunications [7] or network security [2]. Equilibrium behavior and dynamics in these settings have been extensively investigated under the fundamental assumption that each agent has perfect information about its utility function, see e.g. [42, 19].

In this chapter we aim at extending these results to cases where agents are uncertain about the weight with which their neighbors' actions are a□ecting their utility function and aim at learning this *network weight parameter* over repetitions of the game. We give two motivating examples of local public goods and advertising campaigns (see Examples 1 and 2).

We consider a scenario where the same set of risk-neutral agents take part in sequential repetitions of the same game and learn over time. We study simple learning dynamics whereby at each game repetition every agent: i) selects his strategy to maximize his expected utility given his current estimate of his network weight parameter, ii) observes the actions played by his neighbors and a corresponding realization of his stochastic utility, iii) uses such observations to update his network weight pa-

rameter estimate. The main challenge is that the parameter estimation process is intrinsically coupled with the strategy update process. In fact, on the one hand, the chosen strategy determines the instantaneous payo and thus the data used in the estimation procedure; on the other hand, the next strategy is computed by using the most current estimate.

Nonetheless, we show that if agents use a simple least squares procedure, then the strategies converge almost surely to the Nash equilibrium of the full information game (i.e. the Nash equilibrium that is obtained when each agent knows its network weight parameter), which under our assumptions is unique. We also prove that agents whose neighbors have a nonzero aggregate strategy in this equilibrium almost surely learn their network weight parameter. We start by deriving a su cient condition in terms of the network and of the utility function to guarantee that the best response dynamics (under full information) converge to a unique Nash equilibrium. For the case of partial information, we build on our previous results in [59] to relate the error term of the estimate to a martingale and we prove almost sure convergence of the error. We then use sensitivity results of the best response mapping to prove that, because we have convergence of the errors, the learning dynamics under the estimated parameters asymptotically converge towards the full information best response dynamics and therefore to the full information Nash equilibrium.

Literature review

In contrast to Bayesian games, where agents have incomplete information about payo s and form beliefs about their and other agents' utility functions and strategies, see e.g. [42, chapter 5.2], here, we focus on parametric learning dynamics using a simple least squares estimator.

Similar parametric learning dynamics were considered in the context of routing games in our previous work [59]. With respect to that work, we here consider a network game (where the network enters in the utility function instead of the constraint set) and we assume that at each repetition agents play a best response, given their current parameter estimate, instead of allocating flow according to a Wardrop

equilibrium. Moreover, we consider generic concave utility functions.

Learning for generic games with misspecified parameters has been recently addressed in [44]. While in the first part of that work agents update strategies through a gradient method, in the second part the authors consider learning via best response dynamics, similarly to our work, but focus on Cournot games. Instead, we consider network games where agents' payo□s are a□ected by a subset of the population (i.e. the neighbors). "Demand function" learning in Cournot games is also discussed in [43] and [14], where gradient methods are used to learn a linear demand function where the intercept or the slope is unknown. We remark that [44, 43, 14] assume other agents' strategies are not observed, but the cost functions and strategy sets are common knowledge. We instead assume agents observe the aggregate of their neighbors' strategies, which in network games is a local quantity, but we do not require agents to be aware of the payo□ functions or local settings of their neighbors. Moreover, we assume that agents update their strategies according to a best response and assume that the update of the estimates is done through ordinary least squares.

Notation: $\|x\|$ is the 2-norm of x. $\rho(A)$ is the spectral radius of the matrix A. diag(a^i) is a diagonal matrix with entry a^i in position (i,i) for each i.

3.1 Motivating examples

We start by considering two motivating examples.

Example 1 (Local public good). *Consider a game where N agents need to decide on their level of contribution to a public good (e.g. how much e ort to devote to team work) and the payo received by each player i for contribution $x^i \geq 0$ is stochastic and given by*

$$\tilde{v}^i(x^i, x^{-i}, \epsilon^i) = f\left(x^i + a^i \sum_{j \neq i} P_{ij} x^j + \epsilon^i\right) - k^i x^i,$$

where $x^{-i} = \{x^j\}_{j \neq i}$, f is a strictly increasing, concave function representing the benefit agent i obtains from the local public good, $P_{ij} = 1$ if agent j contributes to the local good of agent i and 0 otherwise, k^i is agent i's marginal cost and the noise

ϵ^i *models stochasticity in the payo realization [20]. The network weight parameter* $a^i > 0$ *is the substitutability of the local good: when* $a^i = 1$ *contributions of other agents are equivalent to contributions from agent* i*. Agents do not know* a^i *and need to learn it from stochastic payo observations. We denote the weighted contribution to the public good as perceived by agent* i *by* $y^i(x^{-i}, \epsilon^i) = a^i \sum_{j \neq i} P_{ij} x^j + \epsilon^i$*, so that the utility function can be rewritten in aggregate form as* $v^i(x^i, y^i(x^{-i}, \epsilon^i))$*.*

Example 2 (Competition in advertisement)**.** *Consider* N *firms, where each firm* i *decides a level* $x^i \geq 0$ *of investment in advertising and earns a stochastic payo*

$$\tilde{v}^i(x^i, x^{-i}, \epsilon^i) = f\left(b^i + a^i \sum_{j \neq i} P_{ij} x^j + \epsilon^i\right) x^i - k^i (x^i)^2,$$

where the cost for the advertising e ort is $k^i(x^i)^2$ *and the marginal benefit is a function of other firms' advertising e ort. Specifically, if firm* i *is the only one advertising its product the return per unit of advertisement is* $f(b^i + \epsilon^i)$*. However, if competing firms advertise the unit return may increase or decrease depending on whether the other firms sell goods that are complements or substitutes for firm* i*'s good. This is captured in* P_{ij}*; we assume* $P_{ij} = -1$ *if* i *and* j *sell competing products,* $P_{ij} = 1$ *if* i *and* j *sell complementary products, and* $P_{ij} = 0$ *if there is no interaction between the two firms' products. The level of interference of advertisements is modeled by the network weight parameter* a^i*, which may be unknown to the firm. We denote the competition on advertisement as perceived by agent* i *by* $y^i(x^{-i}, \epsilon^i) = a^i \sum_{j \neq i} P_{ij} x^j + \epsilon^i$*. Again the utility function is of the aggregate form* $v^i(x^i, y^i(x^{-i}, \epsilon^i))$*.*

The examples above are cases of stochastic network games, which have been studied under the assumption that agents know their expected utility functions. This is not always true. In this chapter, we study how agents learn such unknown weight parameters (a^i in our examples) from repeated play. Specifically, we consider a set of agents that repeat the same game (e.g. employees interacting over sequential job assignments or firms interacting in di erent ad campaigns). Each time the agents observe the aggregate e ort of the other players as well as a realization of their payo

and use these observations to learn their a^i.

3.2 The model

Consider a repeated game with N agents that interact over a network with adjacency matrix $P \in \{-1, 0, 1\}^{N \times N}$; $P_{ij} = 1$ if agent j positively influences agent i, $P_{ij} = -1$ if agent j negatively influences agent i and $P_{ij} = 0$ if agent j does not a□ect agent i. We assume $P_{ii} = 0$. Each risk neutral player $i \in \mathbb{N}[1, N]$ aims to select a strategy x^i from its feasible set $\mathcal{X}^i \subseteq \mathbb{R}$ to *maximize* the expected utility function

$$\mathbb{E}_{\epsilon^i}[v^i(x^i, y^i(x^{-i}, \epsilon^i))] \tag{3.1}$$

which depends on its own strategy x^i and on other agents' strategies via a possibly agent-dependent function y^i. In this work we assume y^i is a linear function of the neighbors strategies according to the coe□cients of the network and the network weight parameter a^i, that is,

$$y^i(x^{-i}, \epsilon^i) := a^i \sum_{j=1}^{N} P_{ij} x^j + \epsilon^i. \tag{3.2}$$

We define the aggregate of the strategies of the neighbors of agent i as $z^i(x^{-i}) := \sum_{j=1}^{N} P_{ij} x^j$ and the expected weighted aggregate as perceived by agent i as $p^i(x^{-i}) := a^i z^i(x^{-i})$. We can thus write the expected utility function as

$$\begin{aligned} \mathbb{E}_{\epsilon^i}[v^i(x^i, y^i(x^{-i}, \epsilon^i))] &= \mathbb{E}_{\epsilon^i}[v^i(x^i, p^i(x^{-i}) + \epsilon^i)] \\ &=: u^i(x^i, p^i(x^{-i})) \end{aligned} \tag{3.3}$$

We make the following regularity assumptions.

Assumption 4 (Strategy sets). *The strategy sets $\mathcal{X}^i$ are convex and compact.*

Assumption 5 (Utility functions). *For every agent i, the expected utility function $u^i(x^i, p^i)$ is continuously di erentiable and strongly concave in x^i uniformly in p^i with*

parameter $-\alpha$, $\alpha > 0$. *Moreover, the gradient* $\nabla_{x^i} u^i(x^i, p^i)$ *is Lipschitz continuous in* p^i *uniformly in* x^i *with parameter* $L > 0$. *The parameters* a^i *are such that*

$$\max_i(|a^i|)\frac{L}{\alpha}\rho(P) < 1. \tag{3.4}$$

3.2.1 Nash equilibrium and best response dynamics

The best response of an agent i to the weighted neighbors aggregate p^i is given by

$$B^i(p^i) := \arg\max_{x^i \in \mathcal{X}^i} \mathbb{E}[v^i(x^i, p^i + \epsilon^i)] = \arg\max_{x^i \in \mathcal{X}^i} u^i(x^i, p^i). \tag{3.5}$$

Under Assumption 5 the max in (3.5) is unique. A set of strategies where each player plays a best response to the other agents' strategies is a Nash equilibrium.

Definition 3 (Nash equilibrium). *A set of strategies* $\{\bar{x}^i \in \mathcal{X}^i\}_{i=1}^N$ *is a Nash equilibrium if for all agents* $i \in \{1, 2, ..., N\}$ *and for all* $x^i \in \mathcal{X}^i$,

$$u^i(\bar{x}^i, p^i(\bar{x}^{-i})) \geq u^i(x^i, p^i(\bar{x}^{-i})).$$

Equivalently, a set of strategies is a Nash equilibrium if and only if it is a fixed point of the best response mapping

$$B(x) := [B^i(p^i(x^{-i}))]_{i=1}^N, \tag{3.6}$$

where $x := [x^i]_{i=1}^N \in \mathbb{R}^N$.

The following proposition shows in the games we consider such a best response mapping is a contraction. This implies existence and uniqueness of the Nash equilibrium and that the best response dynamics converge to the Nash equilibrium.

Proposition 1. *Under Assumptions 4 and 5,* $B : \mathbb{R}^N \to \mathbb{R}^N$ *defined in* (3.6) *is a contraction with constant* $\gamma := \max_i(|a^i|)\frac{L}{\alpha}\rho(P)$. *Consequently, the Nash equilibrium exists and is unique and the best response dynamics converge to it. That is, for any* $\bar{x}_0 \in \mathbb{R}^N$ *the sequence* $\bar{x}_{k+1} = B(\bar{x}_k)$ *converges to the unique Nash equilibrium* $\bar{x}$.

The proof of Proposition 1 is an immediate generalization of [20] to games that are not linear quadratic. We include the proof in the appendix for completeness.

3.3 Learning model

In the previous section we assumed each agent i knows his network weight parameter a^i and the update consisted of responding to other agents' strategies. We next investigate what happens if each agent i does not know his parameter a^i and needs to learn it while playing.

Assumption 6 (Observation model). *At each round k each agent i observes the aggregate strategies played by its neighbors, that is $z_k^i := \sum_j P_{ij} x_k^j$ and a realization of its stochastic payo , that is $v_k^i = v^i(x_k^i, y_k^i)$ where $y_k^i := y^i(x_k^{-i}, \epsilon_k^i) = a^i z_k^i + \epsilon_k^i$. The terms ϵ_k^i are independent and identically distributed realizations of the random variable ϵ^i. We assume that $\mathbb{E}[\epsilon^i] = 0$ and $var(\epsilon^i)$ is finite.*

Assumption 7. *Each agent i knows the function v^i and for each value of x^i the function $y^i \mapsto v^i(x^i, y^i)$ is invertible.*

Remark 2. *Assumption 7 allows each agent to recover the noisy value y^i as defined in* (3.2)*. This assumption holds in both our examples if f is strictly monotone, e.g. in Example 1 if utility is increasing in total e ort by others.*

Assume that each agent i has initial estimates z_0^i and $\hat{a}_0^i$. At each iteration k each agent

1. computes his strategy x_k^i through best response to his neighbors' previous aggregate z_{k-1}^i and his current parameter estimate $\hat{a}_{k-1}^i$, that is,

$$x_k^i = \arg\max_{x^i \in \mathcal{X}^i} u^i(x^i, \hat{a}_{k-1}^i z_{k-1}^i);$$

2. observes neighbors' aggregate z_k^i and stochastic payo

$$v_k^i = v^i(x_k^i, a^i z_k^i + \epsilon_k^i);$$

3. uses the observations of z_k^i and v_k^i to update his parameter estimate through ordinary least squares.

These dynamics are summarized in Table 3.1.

Table 3.1: The learning dynamics

Initialize: Each agent i has an initial state z_0^i and initial estimate $\hat{a}_0^i$. Set $k = 1$.
Iterate until convergence; each agent:

1) Calculates the best response

$$x_k^i = \arg\max_{x^i \in \mathcal{X}^i} u^i(x^i, \hat{a}_{k-1}^i z_{k-1}^i) \tag{3.7a}$$

2) Observes

$$z_k^i = \sum_j P_{ij} x_k^j \text{ and } v_k^i = v^i(x_k^i, a^i z_k^i + \epsilon_k^i) \tag{3.7b}$$

3) Updates the parameter estimate

$$\hat{a}_k^i = \frac{1}{\sum_{t=1}^k (z_t^i)^2} \sum_{t=1}^k z_t^i y_t^i \text{ where } y_k^i = a^i z_k^i + \epsilon_k^i \tag{3.7c}$$

$k \leftarrow k + 1$

We assume that agents use ordinary least squares regression to estimate a^i. In particular, given k observations $\{v_t^i\}_{t=1}^k$, by Assumption 7 agent i can recover k observations of the form

$$\{y_t^i = a^i z_t^i + \epsilon_t^i\}_{t=1}^k \tag{3.8}$$

where $\{z_t^i\}_{t=1}^k$ are known and ϵ_t^i are i.i.d. samples. The least squares estimate of a^i

at step k is then

$$\hat{a}_{\mathsf{LS}}(\{z_t^i, y_t^i\}_{t=1}^k) := \arg\min_{a^i \in \mathbb{R}} \sum_{t=1}^{k} (y_t^i - a^i z_t^i)^2$$
$$= \frac{1}{\sum_{t=1}^{k} (z_t^i)^2} \sum_{t=1}^{k} z_t^i y_t^i. \tag{3.9}$$

Note the $\{z_t^i\}_{t=1}^{\infty}$ are not i.i.d. as z_k^i depends on $\hat{a}_{k-1}$ which depends on all $\{z_t\}_{t=1}^{k-1}$. Standard consistency results of least squares cannot be applied. However the following holds.

Proposition 2. *Under Assumptions 6 and 7, there exists a set of noise realizations Σ that has probability one and is such that for any $\epsilon \in \Sigma$ if we partition the agents into the two sets*

$$S^{\infty}(\epsilon) := \{i \in \mathbb{Z}[1, N] \mid \sum_{t=1}^{\infty} (z_t^i(\epsilon))^2 = \infty\} \tag{3.10}$$

$$S^{\mathsf{finite}}(\epsilon) := \{i \in \mathbb{Z}[1, N] \mid \sum_{t=1}^{\infty} (z_t^i(\epsilon))^2 < \infty\} \tag{3.11}$$

then it holds that

1. *if $i \in S^{\infty}(\epsilon)$ then $\lim_{k\to\infty} |\hat{a}_k^i(\epsilon) - a^i| = 0$;*

2. *if $i \in S^{\mathsf{finite}}(\epsilon)$ then there exists $M^i(\epsilon) > 0$ such that $|\hat{a}_k^i(\epsilon) - a^i| \leq M^i(\epsilon)$ for all k.*

The proof of this proposition is similar to that of Lemma 6 in [59] and is reported in the Appendix for completeness. The above proposition states that for almost all noise realizations agents can be partitioned into two groups based on the sum of neighbors' aggregate strategies: those that diverge and those that are square summable over the infinite horizon. Agents in the first group learn a^i and agents in the second group have an estimate of a^i with a finite bounded error.

3.4 Convergence

Using Propositions 1 and 2 we show that the learning dynamics summarized in Table 3.1 converge almost surely to the unique Nash equilibrium of the full information game.

Theorem 2. *Under Assumptions 4, 5, 6, and 7, the strategy vector x_k corresponding to the dynamics in Table 3.1 converges to the full information Nash equilibrium $\bar{x}$ almost surely.*

Proof. To prove this statement we prove convergence for any noise realization in Σ, as defined in Proposition 2. Fix $\epsilon \in \Sigma$ and let $x_k(\epsilon)$ be the corresponding learning dynamics, as detailed in Table 3.1. Moreover, let $\bar{x}_k$ be the best response dynamics that one obtains under full information starting from the same initial condition. Let $A = \text{diag}(a^i)$, $A_k(\epsilon) = \text{diag}(a_k^i(\epsilon))$ and $L_B = \frac{L}{\alpha}$. Then

$$\begin{aligned}
\|x_k(\epsilon) - \bar{x}_k\| &= \sqrt{\sum_{i=1}^{N} \left(B^i(\hat{a}_{k-1}^i(\epsilon) z_{k-1}^i(\epsilon)) - B^i(a^i \bar{z}_{k-1}^i)\right)^2} \\
&\leq L_B \sqrt{\sum_{i=1}^{N} \left(\hat{a}_{k-1}^i(\epsilon) z_{k-1}^i(\epsilon) - a^i \bar{z}_{k-1}^i\right)^2} \\
&= L_B \|\hat{A}_{k-1}(\epsilon) z_{k-1}(\epsilon) - A\bar{z}_{k-1}\| \\
&\leq L_B \|\hat{A}_{k-1}(\epsilon) z_{k-1}(\epsilon) - A z_{k-1}(\epsilon)\| + L_B \|A z_{k-1}(\epsilon) - A\bar{z}_{k-1}\| \\
&\leq L_B \|\hat{A}_{k-1}(\epsilon) z_{k-1}(\epsilon) - A z_{k-1}(\epsilon)\| + L_B \|A\| \|z_{k-1}(\epsilon) - \bar{z}_{k-1}\| \\
&\leq L_B \|\hat{A}_{k-1}(\epsilon) z_{k-1}(\epsilon) - A z_{k-1}(\epsilon)\| + L_B \|A\| \|P\| \|x_{k-1}(\epsilon) - \bar{x}_{k-1}\|
\end{aligned} \tag{3.12}$$

where we used that B^i is Lipschitz continuous (shown in proof of Proposition 1) for the first inequality. Note that

$$\begin{aligned}
w_k &:= L_B \|\hat{A}_k(\epsilon) z_k(\epsilon) - A z_k(\epsilon)\| \\
&= L_B \sqrt{\sum_{i=1}^{N} (\hat{a}_k^i(\epsilon) - a^i)^2 (z_k^i(\epsilon))^2} \to 0
\end{aligned} \tag{3.13}$$

by Proposition 2. In fact, if $i \in S^{\infty}(\epsilon)$ then $|\hat{a}_k^i(\epsilon) - a^i| \to 0$ and $(z_k^i(\epsilon))^2$ is bounded (as $\mathcal{X}^i$ is compact). On the other hand, if $i \in S^{\text{finite}}(\epsilon)$ then $|\hat{a}_k^i(\epsilon) - a^i|$ is bounded but $(z_k^i(\epsilon))^2 \to 0$.

By letting $\xi_k := \|x_k(\epsilon) - \bar{x}_k\|$ the system in (3.12) can be written

$$\xi_k \leq \gamma \xi_{k-1} + w_{k-1}$$

where $\gamma := L_B \|A\| \|P\|$ and $0 < \gamma < 1$ by assumption. Note that ξ_k and w_k are non-negative by definition. Consequently, all the assumptions of Lemma 11 (in the Appendix) are satisfied and we conclude that $\xi_k = \|x_k(\epsilon) - \bar{x}_k\| \to 0$.

Overall we have proven that for any noise realization $\epsilon \in \Sigma$, $\|x_k(\epsilon) - \bar{x}_k\| \to 0$. From Proposition 1 we have that $\|\bar{x}_k - \bar{x}\| \to 0$, therefore

$$\|x_k(\epsilon) - \bar{x}\| \leq \|x_k(\epsilon) - \bar{x}_k\| + \|\bar{x}_k - \bar{x}\| \to 0.$$

Since Σ has measure one, $\|x_k - \bar{x}\| \to 0$ almost surely. ∎

A corollary of the above theorem is that any agent who has a non-zero aggregate of his neighbors strategies in the full information Nash equilibrium learns a^i almost surely.

Corollary 2. *Under Assumptions 4, 5, 6, and 7, any agent i for which $z^i(\bar{x}) \neq 0$ (where $\bar{x}$ is the full information Nash equilibrium) learns the parameter a^i almost surely.*

Proof. For any $\epsilon \in \Sigma$ we have that $x_k(\epsilon) \to \bar{x}$ and thus $z_k^i(\epsilon) \to \bar{z}^i$. Consequently, if $\bar{z}^i \neq 0$ it must be $\sum_k z_k^i(\epsilon)^2 \to \infty$ and $i \in S^{\infty}(\epsilon)$ as defined in (3.10). The conclusion follows by Proposition 2. ∎

3.5 Simulations

To illustrate our results we simulate a game in the setting of Example 1 with

$$f(\ell) = \log(\ell + b) \tag{3.14}$$

for a positive constant $b > 0$. We first derive a su□cient condition to guarantee that Assumption 5 is met.

Lemma 7. *Consider a game with the structure given in Example 1 and suppose that f is as in* (3.14)*, $a^i > 0$ for all i, $\epsilon^i \sim U(-\bar{\epsilon}, \bar{\epsilon})$ for some finite $\bar{\epsilon} \in (0, b)$, $\mathcal{X}^i = [0, 1]$ and $P \in \{0, 1\}^{N \times N}$ with $P_{ii} = 0$. If $\{a^i\}_{i=1}^N$ and b are such that*

$$\max_i(\mid a^i \mid)\frac{(1 + \max(\mid a^i \mid)N + b)^2 + \frac{1}{3}\bar{\epsilon}^2}{b^2}\rho(P) < 1, \tag{3.15}$$

then Assumption 5 is met.

Proof. In this setting

$$u^i(x^i, p^i) = \mathbb{E}_{\epsilon^i}[\log(x^i + p^i + \epsilon^i + b) - k^i x^i].$$

By Leibniz's rule

$$\begin{aligned}
\nabla_{x^i} u^i(x^i, p^i) &= \nabla_{x^i}(\mathbb{E}_{\epsilon^i}[\log(x^i + p^i + \epsilon^i + b)] - k^i x^i) \\
&= \mathbb{E}_{\epsilon^i}[\nabla_{x^i} \log(x^i + p^i + \epsilon^i + b)] - k^i \\
&= \mathbb{E}_{\epsilon^i}\left[\frac{1}{x^i + p^i + \epsilon^i + b}\right] - k^i.
\end{aligned}$$

1. $u^i(x^i, p^i)$ *is strongly concave in* x^i:

$$\begin{aligned}\nabla_{x^i x^i} u^i(x^i, p^i) &= \nabla_{x^i} \mathbb{E}_{\epsilon^i}\left[\frac{1}{x^i + p^i + \epsilon^i + b}\right] \\ &= \mathbb{E}_{\epsilon^i}\left[-\frac{1}{(x^i + p^i + \epsilon^i + b)^2}\right] \\ &\leq -\frac{1}{\mathbb{E}_{\epsilon^i}[(x^i + p^i + \epsilon^i + b)^2]} \\ &= -\frac{1}{(x^i + p^i + b)^2 + \mathsf{var}(\epsilon^i)} < 0,\end{aligned}$$

the second equality follows from Leibniz's rule and the inequality follows from Jensen's inequality since $-\frac{1}{s}$ is concave in s. To obtain the uniform concavity constant $-\alpha$ note for each agent i

$$\frac{1}{(x^i + p^i + b)^2 + \mathsf{var}(\epsilon^i)} \geq \frac{1}{(1 + \max(\mid a^i \mid) N + b)^2 + \frac{1}{3}\bar{\epsilon}^2} =: \alpha \tag{3.16}$$

The last inequality holds as the maximum degree is N.

2. $\nabla_{x^i} u^i(x^i p^i)$ *is Lipschitz in* p^i: Taking the derivative with respect to p^i

$$\nabla_{p^i} \mathbb{E}_{\epsilon^i}\left[\frac{1}{x^i + p^i + \epsilon^i + b}\right] = \mathbb{E}_{\epsilon^i}\left[\frac{-1}{(x^i + p^i + \epsilon^i + b)^2}\right].$$

For any random variable ξ,

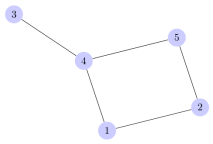
$$|\mathbb{E}[\xi]| \leq \mathbb{E}[|\xi|] \leq \max_{\mathsf{supp}(\xi)} |\xi|.$$

Hence, $|\nabla_{x^i, p^i} u^i(x^i, p^i)| \leq \max \frac{1}{(x^i + p^i + \epsilon^i + b)^2} = \frac{1}{(b - \bar{\epsilon})^2} = L$.

Plugging the values for L and the lower bound on α in to (3.4) we recover the su□cient condition given in (3.15). ∎

In our simulations we set $N = 5$, $a^i = \left[\frac{3}{20}, \frac{1}{5}, \frac{1}{5}, \frac{1}{4}, \frac{1}{20}\right]$ and $b = 1$, so that (3.15) holds. It is immediate that Assumptions 1, 3 and 4 hold as well (in particular the function $v^i(x^i, y^i) = \log(x^i + y^i + b) - k^i x^i$ is invertible in y^i). In Figure 3-1 we show

A)



B)

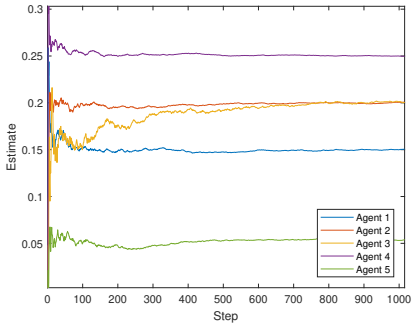


Figure 3-1: A) The network: $P_{ij} = P_{ji} \in \{0, 1\}$ and a link between agents i and j is present if and only if $P_{ij} = P_{ji} = 1$. B) One realization of the estimates $\hat{a}_k^i$ over steps $k \in [1, 1000]$.

a symmetric network and the learning dynamics starting from $\hat{a}_0^i = z_0^i = 0$ for each agent i for one noise realization.

3.6 Conclusion

We examined simple learning dynamics in a repeated network aggregative game where each agent learns an unknown network weight parameter from observations of its stochastic payo and neighbors' actions. We showed that ordinary least squares estimates coupled with best response dynamics converge to the unique Nash equilibrium, under appropriate assumptions on the network and the utility functions. The extension of our results to multi-dimensional strategies is immediate. In the future, we plan to study settings where agents do not observe other agents' actions or have more than one unknown parameter.

Chapter 4

Optimal Dynamic Information Provision in Traffic Routing

As discussed in Chapter 2, traffic congestion is a major issue for billions of commuters all around the world. Although in Chapter 2 we saw that if the latency functions are fixed and unknown the Wardrop equilibrium will be reached. Instead, in this chapter we consider a setting where the road conditions are unpredictably changing due to accidents, traffic jams, construction, and weather. In such a setting, users rely not only on their past experiences, but also on information provided by a central planner (for example through a realtime routing system) to make optimal routing decision. The fundamental question that we investigate is whether differentiated provision of information from the central planner (CP) can be used as an effective tool to coordinate the traffic demand and reduce overall traffic congestion. This is a new approach to easing congestion, whereas in the past strategies such as road tolling ([24], [25], [3], [35]) and reward-based incentive mechanisms ([32], [8]) have been focused on. Thus, in this chapter, we investigate this alternative approach to traffic regulation which leverages GPS-based navigation systems and aims to change user behavior by providing different information about road states.

We consider a repeated routing game, where the CP at each round provides personalized and private road recommendations to the drivers with the goal of minimizing overall travel time. We assume drivers are each interested in minimizing their indi-

vidual travel times. Therefore drivers will follow the recommendation from the CP only if it is in their best interest given their current belief on the state of the road. In technical terms, agents' self interest constrains the feasible set of the CP to recommendations that are *incentive compatible* (IC), see [36] or the survey on incentive design [67].

Crucially, we assume that the CP himself does not know the state of the road and needs to learn from users' observations of road conditions. This introduces an exploration-exploitation trade-off, as the CP needs to send some drivers to roads with unknown conditions to obtain up-to-date information.

We investigate these issues by developing a repeated two-road *dynamic routing game* with a finite number of (atomic) forward-looking agents (drivers). We assume that one road, the *safe* road, has a travel time that depends on the flow, that is, the number of agents on the road, via a known affine function. The second road, the *risky* road, has a travel time that is a linear function of the flow with an unknown stochastic congestion coefficient, θ. We assume each agent aims to minimize his total discounted travel time, while the CP aims to minimize the sum of all discounted travel times. When there is experimentation, meaning that some agent is using the risky road, the CP learns the state of the road at that time, and then makes recommendations to all agents using this information in the next round. Agents themselves learn the state of the risky road if they take it; otherwise they rely on the recommendations sent by the CP and their observations of the flows to form beliefs.

We first fully characterize the optimal IC recommendation system in a *two-stage setting* where the state of the risky road θ is low (L) or high (H), and remains constant over time. This state θ is unknown to both the CP and all the agents at the beginning of the first stage. We consider three cases: i) full information where all drivers are informed of the state of the risky road before the second round, ii) no information where only drivers that take the risky road in the first round know the state at the second round, and iii) partial information, where the CP will give personalized recommendations to agents that did not take the risky road before the second round. Under full information we show that, since all agents learn the state

of the risky road, under favorable road conditions significant congestion is induced at the second stage thus reducing incentives of agents to experiment in the first round. Private information may, for certain values of priors on θ, perform better than full information since the experimenter will experience very low congestion in the second round (thus reaping the entire reward of experimentation), but still leads to high average travel times as most users will be uninformed and thus restricted to the safe road. As first main results, we show that the optimal recommendation system is a compromise between these two scenario. The CP provides information to some of the uninformed agents if the risky road is low but not all, as a way to minimize average travel time while maintaining incentives for exploration. We characterize the optimal number of recommendations as solution to a quadratically constrained quadratic program.

We then extend our model to an *infinite-horizon setting*. In this case we assume the state of the risky road changes according to an underlying Markov chain, where $\theta_t \in \{L, H\}$ at each time t. This framework could model construction projects that persist over time or roads that are unfavorable in persistent weather conditions. We show that our general insights from the two-stage model generalize to this more complex setting. A new challenge in this case is however that the CP must account for the information that users receive from observations of other agents' past actions. For example, an agent on safe that sees many agents switching to risky can infer that that risky road was low at the previous round and will therefore likely be low again if the transition probability from L to H is sufficiently small. The CP must consider the agents' ability to infer information with one step delay and mitigate the possible deviations agents may take to improve their cost. We characterize an IC recommendation scheme that leads to better social cost than full information and we study optimality of such scheme under different assumptions. Under the derived scheme the CP always sends at least one agent to the risky road to have up-to-date information of the state of the world. If this agent, the experimenter, observes θ change to L, then in the next round the CP sends the myopic optimal flow, the flow that minimizes overall travel time for that round. If after this first

round the road remains favorable, then the CP potentially increases the number of agents sent to the risky road in the next round to maintain incentive compatibility among all agents. If the CP did not do this, then agents on the safe road would be incentivized to switch roads after observing the higher flow on the risky road. Overall, we again find that the CP must balance the amount of information it shares with agents to account for both decreasing the total travel time and maintaining incentive compatibility. The major challenge in analyzing this model comes from both the CP and the agents being forward looking and learning; but with di□erent and opposing objectives. The CP wants to minimize overall travel time, whereas the agents are concerned only with their own travel times. We show that such dichotomy can be solved by carefully including the IC constraints in the derivation of the optimal recommendation schemes. Our findings show that di□erentiated information provision through personalized routing recommendations can be e□ectively used as a control tool to minimize tra□c congestion.

4.0.1 Related Literature

Classic approaches to congestion control include tools such as charging tolls [24], coordinating of tra□c lights [65], or provision of subsidies to drivers for taking certain routes [32]. More recently, the idea of using information as an additional/alternative control lever was suggested in [1] and [54]. Therein the authors introduced the concept of informational Braess' paradox, formalizing the idea that providing certain subsets of drivers with more information can actually harm such drivers. These papers considered a static setting, where agents have a fixed set of information and the state of the roads is constant. In this chapter we take a step further and investigate how one can exploit this phenomenon for dynamic tra□c control under varying road conditions.

We note that a related question has been investigated by using tools of Bayesian persuasion, see [47] and [13], for *static tra c problems* in [30, 71, 55, 75]. The closest works to ours are [71], [51], and [76]. [71] consider a two-stage, two-road routing problem. Therein however the CP has perfect information about the state of the

risky road and does not need to rely on agents experimentation. As a consequence, [71] focus on the case of non-atomic agents, while we consider atomic (a finite number of) agents, which implies that agents take into account the information they generate for their own and others' future use. In addition, observations of tra□c flow in [71] are fully revealing, while in our setting there is a one step delay due to the fact that the CP's suggestions are based on previous observations and not on the current state of the risky road (which is unknown to the CP). As already noted, we also extend our analysis beyond two stages by looking at an infinite horizon model where the state of the road changes in time. [51] consider a repeated game, again in a non-atomic setting where the CP does learn from agents' actions, but agents themselves are not learning – agents at each time step only use the public message that is sent by the CP to make a myopic routing decision. [76] also consider a repeated non-atomic routing game where the CP has perfect information; agents themselves are not learning, but use an aggregated rating of the CP at each point in time. This trust rating, based on the CP's past honesty, is used by the agents to decide whether or not to follow the given recommendation. [56] study exploration in general repeated Bayesian games, which include routing games, where the CP is attempting to steer agents to certain actions to learn, but the agents themselves play the game only once and are not forward looking or learning, as we instead study here. Lastly, [52] and [70] study the dynamic setting where the game is repeated and the platform learns from agents, but with public information sharing and where the agents themselves do not learn. The key novelty in our work with respect to all the references above is that we consider a *dynamic problem with private information* where: i) both the agents and the CP are forward looking, ii) agents influence others payo□ functions, iii) the parameters change over time, requiring constant experimentation and iv) the CP depends on agents to gain information through experimentation.

On a more general note, our results contribute to a growing literature on social learning. For example, [22], [23], [66], and [49] study how a recommender system or platform may incentivize users to learn collaboratively about a product, [68] study how correlated preferences between agents may e□ect learning, [45] study a matching

problem between heterogeneous jobs and workers, with an aim to learn worker types, [41] study learning in repeated auctions, and [15] study product adoption. A survey of the literature at the interface of learning, experimentation, and information design can be found in [39]. The main feature distinguishing the routing problem addressed in our work and the applications considered in the works above is again congestion e□ects, which fundamentally modify the results since agents do not only a□ect others in their learning process but also in payo□s.

Finally, our work has connections with the classic exploration-exploitation trade o□ setting studied using the multi-armed bandit (MAB) problem, see e.g. [37]. A large literature has been devoted to extend the MAB model to di□erent settings. The most closely related to ours are [17] and [48], where multiple experimenters can learn from one another. These papers show that because of free riding by agents there will be less experimentation than in the standard MAB model. Our setting is di□erent because of three features: first, congestion creates dependent payo□s across agents; second, we focus on the IC experimentation scheme for a central planner; and third, information in our setting is neither public (as in [17] and [48]) nor fully private.

The rest of the chapter is organized as follows. In Section 4.1, we introduce the routing model. In Section 4.2 we consider a two-stage setting, compare three information schemes and explain how to characterize the optimal recommendation scheme. Finally, in Section 4.3 we detail the infinite time horizon setting and study the optimal, IC recommendation scheme. All proofs are given in the Appendix.

4.1 Model

We consider a dynamic mechanism design problem in which a central planner (CP) aims to minimize total travel time in a repeated routing game with N (atomic) agents on two roads. Agents decide their own routing to minimize their own travel time, and the CP can try to influence their choices by providing information.

Congestion model

We consider a network with two roads. One of the roads, the *safe road*, has a congestion-dependent but non-stochastic a□ne cost $S_0 + S_1(N - x_R)$ where $S_0 > 0$, $S_1 \geq 0$ and x_R is the number of agents on the risky road. The other road, the *risky road*, has a linear cost $\theta_t x_R$ where $\theta_t \in \{L, H\}$ (where L, H are scalars with $L, H > 0$) is an unknown congestion parameter that changes over time according to an underlying Markov chain with switching probabilities $\gamma_L := \mathbb{P}(\theta_t = H | \theta_{t-1} = L)$ and $\gamma_H := \mathbb{P}(\theta_t = L | \theta_{t-1} = H)$. (Our analysis can be easily generalized to a risky road with a□ne cost $\theta_t x_R + R_0$ for known $R_0 \geq 0$. We omit this for simplicity of exposition. The parameter $\theta_t = L$ represents cases where the congestion parameter is "low" which means the risky road is favorable, alternatively $\theta_t = H$ means the parameter is "high" and the safe road is preferable.

Assumption 8 (Frequency of switching)**.**

$$0 \leq \gamma_L, \gamma_H \leq \frac{1}{2}.$$

Assumption 8 imposes an upper limit on the probability that the road condition changes. Intuitively this condition suggests that it might be worthwhile for an agent to experiment, since the road is likely to remain in the same condition for more than one stage.

Agent actions and stage cost function

At each time t, each agent i chooses an action $\alpha_t^i \in \{S, R\}$ corresponding to either taking the safe ($\alpha_t^i = S$) or the risky road ($\alpha_t^i = R$). An agent's realized stage cost is then given by the travel time he experiences at stage t

$$\tilde{g}(\alpha_t^i, \alpha_t^{-i}, \theta_t) = \begin{cases} S_0 + S_1(N - x_t^R(\alpha_t)) & \text{if} \quad \alpha_t^i = S, \\ \theta_t x_t^R(\alpha_t) & \text{if} \quad \alpha_t^i = R, \end{cases}$$

where $x_t^R(\alpha_t) := \sum_{i=1}^N \mathbb{1}\{\alpha_t^i = R\}$ is the total flow on the risky road at time t.

Information structure

We assume that if an agent experiments by using the risky road at time t, he observes the true value of θ_t and this is also directly observed by the CP (because he communicates it truthfully to the CP or because of direct observation of his experience by the CP via GPS). Consequently the CP knows θ_t if and only if at least one agent takes the risky road at time t. Before $t = 0$, the CP commits to a signaling scheme to disseminate information to agents that are on the safe road and are thus uninformed. From here on we restrict our attention to *recommendation schemes*, which are a specific type of signaling scheme where the signal sent to each agent is a recommendation to take either the safe (r_S) or the risky road (r_R) and we refer to such a recommendation scheme as π (a formal definition of recommendation schemes for the two stage model is given in Definition 5 and for the infinite horizon model is given in Definition 6).

Cost function

Drivers are homogeneous and risk-neutral. Each minimizes his expected sum of travel times over T repetitions of the game discounted in time by a factor $\delta \in [0, 1)$. Let h^i_{t-1} be agent i's history after round $t-1$ and before time t (based on his observations and on the signals sent by the CP). This includes the past actions of the agent, the flows the agent experienced, the recommendations the agent received, and the state of the risky road for those times the agent took it. An agent's strategy $\xi^i_t(h^i_{t-1})$ maps any history to an action $\{S, R\}$. The agent's expected cost from time t to time T is given by

$$u^i_t(h^i_{t-1}) = \mathbb{E}\left[\sum_{k=t}^{T} \delta^{k-t} \tilde{g}(\xi^i_k(h^i_{k-1}), \xi^{-i}_k(h^{-i}_{k-1}), \theta_k)\right].$$

We focus on pure strategies; hence the expectation here is on the state of the risky road θ and on the information received from the CP.

Each agent chooses $\xi^i_t(\cdot)$ to minimize his total expected cost given the strategies of others and the recommendation scheme the CP uses. Note that the agent's strategy is chosen *after* the CP has committed to a recommendation scheme.

The CP's objective is to select a recommendation scheme π to minimize the ex-

pected total discounted travel times of agents over T stages. Let $\xi_t^{i|\pi}$ be the strategy agent i uses in equilibrium under the recommendation scheme π.

Definition 4. *A recommendation scheme π is incentive compatible if $\xi_t^{i|\pi} = R$ for any agent i that receives a recommendation r_R for stage t and $\xi_t^{i|\pi} = S$ for any agent i that receives a recommendation r_S for stage t.*

Let Π be the class of incentive compatible recommendation schemes and $g(x^R, \theta)$ be the total cost of a stage when there are x^R agents on the risky road and the road condition is θ, i.e.

$$g(x^R, \theta) = \theta(x^R)^2 + (S_0 + S_1(N - x^R))(N - x^R).$$

If no agent is using the risky road, we define $g(0) = (S_0 + S_1 N)N$ as the cost does not depend on θ. Before the game begins the CP and all agents share a common prior on the state of the risky road which we denote by $\beta = \mathbb{P}(\theta_0 = L)$. The CP wants to choose a scheme $\pi \in \Pi$ to minimize

$$V_T^\pi(\beta) := \mathbb{E}\left[\sum_{k=1}^{T} \delta^k g(x_k^R, \theta_k) \mid \mathbb{P}[\theta_0 = L] = \beta\right], \text{ where } x_k^R = \sum_{i=1}^{N} \mathbb{1}\{\xi_k^{i|\pi}(h_{k-1}^i) = R\}. \tag{4.1}$$

4.2 The two-stage model

We start by considering a two-stage model ($T = 2$) and, for simplicity, we assume that the road condition does not change between the two stages, that is $\theta_0 = \theta_1 = \theta_2 =: \theta$ (i.e. $\gamma_H = \gamma_L = 0$). We also assume no discounting ($\delta = 1$) to simplify the derived bounds. Define the expected value of θ at the beginning of the first stage as $\mu_\beta := \mathbb{E}[\theta] = \beta L + (1 - \beta)H$. In this context the objective of the CP is to minimize

$$V_2^\pi(\beta) := \mathbb{E}\left[g(x_1^R, \theta) + g(x_2^R, \theta) \mid \mathbb{P}(\theta = L) = \beta\right].$$

We adopt the following assumption.

Assumption 9 (Two stage model). *The parameters are such that*

1. $L < S_0 + S_1$,
2. $S_0 + S_1 N < \mu_\beta$.

Assumption 9.1 states that if the congestion parameter is L, the risky road is preferable (i.e. the cost of one agent on risky if $\theta = L$ is less than the cost of one agent on safe). Assumption 9.2 states that the expected cost of experimentation for one agent μ_β is greater than a fully congested safe road. Note that agents may nonetheless select the risky road in the first round since, if $\theta = L$, they can exploit this information in the second round. We let x_L^{eq} be the myopic equilibrium flow on risky if all agents know that $\theta = L$. (That is $x_L^{eq} \approx \min\left\{\frac{S_0+S_1 N}{L+S_1}, N\right\}$. The approximation comes from the fact that x_L^{eq} must be an integer as we work with an atomic model (finite number of agents).)

4.2.1 Full and private information

We start with two extreme and simple informational scenarios:

- *Full information* - if any agent takes the risky road in round one, then all agents learn θ before round two.
- *Private information* - any agent that takes the risky road in round one knows the value of θ before round two, but any agent that chose to play safe in the first round has no new information before the beginning of round two.

We first show that in any pure strategy Nash equilibrium, at most one agent experiments in the first round. We then use this result to characterize the equilibrium under both full and private information. While the result that only one agent experiments under full information is very general, the fact that only one agent experiments under private information is a consequence of some of the special features of this example. In particular, in a two-period model, there is only a limited time during which an experimenter can exploit his information. Since we assume a linear

cost function, an additional experimenter increases the travel time sufficiently such that it is not worthwhile for two agents to experiment. This result does not apply, for example, in our infinite-horizon model, studied in the next section. Nevertheless, we will see that, in that setting too, in the incentive compatible optimal mechanism, the CP will induce only one agent to experiment. Note uniqueness here refers to the total number of agents in each road.

Lemma 8. *Under Assumption 9 and any information scheme, in any pure strategy Nash equilibrium at most one agent experiments in the first round.*

Theorem 3. *Under Assumption 9 the unique pure strategy Nash equilibrium is as follows:*

- *Full information:*
  - *all agents play safe in both rounds, if*

$$\beta < \frac{H - (S_0 + S_1 N)}{H - L + (S_0 + S_1 N) - \frac{g(x_L^{eq}, L)}{N}} =: \beta_f$$

  - *otherwise, one agent experiments in the first round, and x_L^{eq} agents use the risky road in the second round if $\theta = L$, and all play safe if $\theta = H$.*

  The expected cost under equilibrium is

$$V_2^{full}(\beta) := \begin{cases} 2g(0) & \text{if} \quad \beta < \beta_f \\ g(1, \mu_\beta) + \beta g(x_L^{eq}, L) + (1 - \beta) g(0) & \text{if} \quad \beta_f \leq \beta. \end{cases}$$

- *Private information:*
  - *all agents play safe in both rounds if*

$$\beta < \frac{H - (S_0 + S_1 N)}{H + (S_0 + S_1 N) - 2L} =: \beta_p \leq \beta_f$$

– *otherwise, one agent experiments in the first round and uses the risky road if $\theta = L$ and the safe road if $\theta = H$ in the second round. All other agents play safe in both rounds.*

The expected cost under equilibrium is

$$V_2^{private}(\beta) := \begin{cases} 2g(0) & \text{if} \quad \beta < \beta_p \\ g(1, \mu_\beta) + \beta g(1, L) + (1 - \beta)g(0) & \text{if} \quad \beta_p \leq \beta. \end{cases}$$

Corollary 3. *If the prior belief β is such that*

$$\beta_p \leq \beta < \beta_f, \tag{4.2}$$

then in the pure strategy equilibrium there is experimentation under private information, but not under full information. Consequently private information has a lower expected cost ($V_2^{private}(\beta) < V_2^{full}(\beta)$).

According to the above corollary, there may exist a range of priors where it is better for the CP to provide no information rather than full information. Intuitively, this happens because providing the information that $\theta = L$ to all the agents induces congestion in the second round, thus reducing the value of information. This decreases the incentive of an agent to experiment in the first round. In other words, full information allows more agents to free-ride o on one agent's experimentation, reducing the payo of the experimenter due to congestion e ects. The next example illustrates the costs as a function of the prior belief.

Example 3. *Suppose $N = 40$, $S_0 = 10, S_1 = 1$, $L = 0.9$, and $H = 150$. The comparison of the equilibrium cost for all beliefs satisfying Assumption 9 is shown in Figure 4-1. For $\beta \in [0.50, 0.57]$, there is experimentation under private information, but no experimentation under full information.*

The fact that full information, where the conditions of the risky road are communicated to all agents, is not socially optimal motivates the rest of our analysis. We

will show that some amount of information sharing by the CP is preferable to private information and characterize the optimal recommendation scheme.

4.2.2 Unconstrained social optimum

Define x_L^{SO}, x_H^{SO} as the social optimum integer myopic flows on the risky road when it is known that $\theta = L$ or $\theta = H$ respectively, that is,

$$x_L^{SO} := \underset{x \in \{0,1,\dots,N\}}{\operatorname{argmin}} \, g(x, L), \quad x_H^{SO} := \underset{x \in \{0,1,\dots,N\}}{\operatorname{argmin}} \, g(x, H).$$

Theorem 4. *Under Assumption 9 the social optimum is given by*

- *All agents playing safe in both rounds, if*

$$\beta \leq \frac{H + g(x_H^{SO}, H) - (S_0(N+1) - S_1((2+N)N - 1))}{H - L + g(x_H^{SO}, H) - g(x_L^{SO}, L)} := \beta_{SO} \leq \beta_p$$

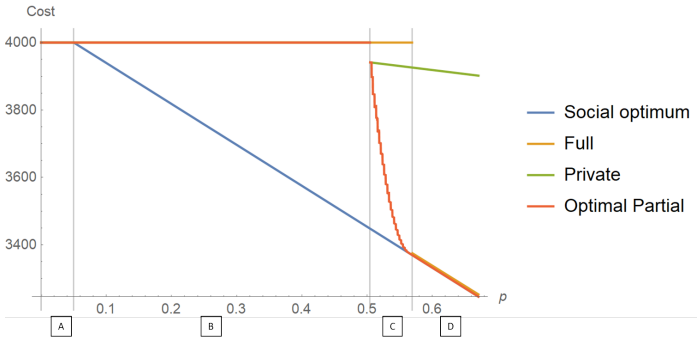
- *One agent experimenting in the first round and x_L^{SO} (x_H^{SO}) agents taking the risky road in the second round if $\theta = L$ ($\theta = H$), otherwise.*

The expected cost under the social optimum is then

$$V_2^* := \begin{cases} 2g(0) & \text{if} \quad \beta < \beta_{SO}, \\ g(1, \mu_\beta) + \beta g(x_L^{SO}, L) + (1-\beta) g(x_H^{SO}, H) & \text{if} \quad \beta \geq \beta_{SO}. \end{cases}$$

Remark 3. *Two remarks are in order. First, note that while when $\theta = H$ it is never myopically a best response for an agent to take the risky road, the previous lemma shows that the CP may still want to send some agents to the risky road in the second round (if $x_H^{SO} \geq 1$) to reduce congestion on safe for all other agents. Second, note that at least for any belief $\beta \in [\beta_{SO}, \beta_p)$ the social optimum scheme is not incentive compatible. In fact, since $\beta < \beta_p$ it is not incentive compatible for an agent to experiment in the first round (under private information the experimenter has the highest possible gain from experimentation hence if experimentation doesn't*

Figure 4-1: Example 3. We distinguish four cases based on the prior β: A) no experimentation, B) experimentation under social optimum, C) experimentation under private and optimal information, D) experimentation under all schemes.



happen under private information it cannot happen under any information scheme). Nonetheless, for $\beta > \beta_{SO}$ the CP would like to experiment by sending one agent to the risky road (because knowing the state of the road is collectively beneficial). In Example 3, $\beta_{SO} = 0.05$ is significantly lower than $\beta_p = 0.50$ suggesting that the social optimum may not be incentive compatible for a large range of beliefs.

4.2.3 Partial information

The CP can alleviate the problems of full and private information and achieve a cost that is closer to social optimum by providing recommendations in a coordinated way. The objective here is to find a balance between

- providing information to a large enough number of agents in the second stage, so that the total cost is low when $\theta = L$;
- providing information to a small enough number of agents in the second stage to avoid a high level of congestion on the risky road when $\theta = L$ to encourage experimentation in the first round.

We refine the CP's recommendation scheme for the two stage model as follows.

Definition 5 (Two stage recommendation scheme). *In the two stage model, a deterministic recommendation scheme is a pair of mappings (π_1, π_2) where $\pi_t : \{\beta\} \to \{0, 1, ..., N\}$ maps the CP's belief on the state of the risky road at time t to the number*

$\pi_t(\beta)$ *of uninformed agents to whom the CP sends a recommendation of risky before time* t*. (We assume that the* $\pi_t(\beta)$ *agents to which* r_R *is sent are chosen uniformly at random from the set of uninformed agents.) With a slight abuse of notation we let* $\pi_t(L) := \pi_t(1)$ *and* $\pi_t(H) := \pi_t(0)$*.*

Note that this definition restricts attention to recommendation systems that are *anonymous*, in the sense that the recommendations for all agents with the same beliefs (i.e. agents that took the safe road) are drawn from the same distribution. (This could also be replicated with a recommendation system that is fully anonymous, meaning that all agents receive recommendations from the same distribution, but those with beliefs determined from their experience of the risky road (the experimenters) will not follow these recommendations.) Nevertheless, the recommendation system is potentially "interim asymmetric" — meaning that some of these agents may receive di□erent recommendations. (An alternative is to impose additionally that the scheme is interim symmetric, but mixed. In this case, all agents would receive the same stochastic recommendation. Because we have a finite number of agents, this would induce additional noise in tra□c flows, hence we do not focus on this case.)

Because of Lemma 8, in any incentive compatible scheme it must be $\pi_1(\beta) \leq 1$. We already argued in Remark 3 that in any incentive compatible scheme there cannot be experimentation if $\beta < \beta_p$, hence in this range it must be $\pi_1(\beta) = \pi_2(\beta) = 0$. If instead $\beta \geq \beta_p$, we show that the optimal incentive compatible scheme selects $\pi_1(\beta) = 1$ and values of $\pi_2(L), \pi_2(H)$ obtained by solving the following quadratic integer program with quadratic constraints (corresponding to the incentive compatibility constraints).

Theorem 5. *If* $\beta \geq \beta_p$ *the optimal incentive compatible recommendation scheme is*

a solution to the following minimization problem

$$\min_{\pi_2(L),\pi_2(H)} \quad g(1,\mu_\beta) + \beta g(x_2^{R|L}, L) + (1-\beta) g(x_2^{R|H}, H) \tag{4.3a}$$

$$\text{s.t.} \quad \underbrace{\mathbb{E}_\theta[S_0 + S_1(N - x_2^{R|\theta}) \mid \text{rec. safe}]}_{\text{follow rec. of safe}} \leq \underbrace{\mathbb{E}_\theta[\theta(x_2^{R|\theta}+1) \mid \text{rec. safe}]}_{\text{deviate to risky}}, \tag{4.3b}$$

$$\underbrace{\mathbb{E}_\theta[\theta x_2^{R|\theta} \mid \text{rec. risky}]}_{\text{follow rec. of risky}} \leq \underbrace{\mathbb{E}_\theta[S_0 + S_1(N - x_2^{R|\theta}+1) \mid \text{rec. risky}]}_{\text{deviate to safe}}, \tag{4.3c}$$

$$\underbrace{\mathbb{E}[\theta]}_{\substack{\text{exp.'s cost}\\ \text{in round 1}}} + \underbrace{\beta(L x_2^{R|L}) + (1-\beta)(S_0 + S_1(N - x_2^{R|H}))}_{\substack{\text{experimenter's cost}\\ \text{in round 2}}} \leq \underbrace{2(S_0 + S_1 N)}_{\substack{\text{all playing safe}\\ \text{in both rounds}}} \tag{4.3d}$$

$$x_2^{R|L} = \pi_2(L) + 1, \; x_2^{R|H} = \pi_2(H)$$

$$\pi_2(L), \pi_2(H) \in \{0, 1, ..., N\}$$

Equations (4.3b)*,* (4.3c) *and* (4.3d) *are given in implicit form for readability, the explicit form is provided within the proof.*

Note, that $\pi_2(L) = \pi_2(H) = 0$ is a feasible solution when $\beta \geq \beta_p$ (and corresponds to private information). Moreover, if $\beta > \beta_f$, $\pi_2(L) = x_L^{eq} - 1, \pi_2(H) = 0$ is a feasible solution and has the same social cost as full information. Hence private and full information have always at least weakly higher cost than the optimal incentive compatible scheme. We show in Example 3 that the optimal incentive compatible scheme can be strictly better than private and full information (see region C in Figure 4-1).

4.3 The infinite-horizon model

We now extend our analysis to an infinite-horizon setting. For simplicity, we restrict our attention to the case when the safe road has a fixed cost S_0 (i.e. we set $S_1 = 0$), so that the cost under full information is simply $S_0/(1-\delta)$. (If $S_1 > 0$ a similar argument can be followed to derive an incentive compatible recommendation scheme. Proving optimality of such a scheme is more complicated, however. The main tech-

nical di□culty in that scenario is computing the optimal punishment for deviations. We show below that, in the case where $S_1 = 0$, the optimal punishment is providing full information.) Finally, note that even though the travel time on the safe road does not depend on congestion levels, we still assume that if an agent takes safe at time t, he observes x_t^S. We first characterize the social optimum scheme and give an example to illustrate why it may not be incentive compatible. We then introduce an incentive compatible recommendation system, prove it achieves better cost than full information and derive conditions for optimality.

4.3.1 Unconstrained social optimum

We first derive the "unconstrained" social optimal, meaning that we ignore the incentive compatibility constraints of the agents.

Suppose that the CP has a belief, $\beta_{t-1} \in [0, 1]$ about the probability that state of the road at time $t-1$ was L (we use the convention $\beta_{t-1} = 0$ if H was observed and $\beta_{t-1} = 1$ if L was observed). If the CP had complete control of the agents the myopically optimal flow to send at time t under a generic belief $\beta_{t-1} = \beta$ would be

$$x_\beta^{SO} := \underset{x \in \{0,\dots,N\}}{\operatorname{argmin}} \ \mathbb{E}_{\theta_t} \left[g(x, \theta_t) \mid \beta_{t-1} = \beta \right]. \tag{4.4}$$

Note that the myopic flow does not depend on t given the properties of Markov processes. With a slight abuse of notation we set $x_L^{SO} := x_1^{SO}$ and $x_H^{SO} := x_0^{SO}$.

In the following, we focus on cases where the cost of experimentation is high from a myopic standpoint. Specifically, we consider cases in which $x_H^{SO} = 0$, so that myopically the CP has no incentive to send agents to the risky road after observing H in the last period. Our interest is determining conditions under which experimentation happens when the CP is forward looking.

Assumption 10 (Cost of experimentation)**.** *Define the expected value of the conges-*

tion parameter θ_t, following an observation of θ_{t-1} as

$$\mu_L := \mathbb{E}[\theta_t|\theta_{t-1} = L] = (1-\gamma_L)L + \gamma_L H,$$
$$\mu_H := \mathbb{E}[\theta_t|\theta_{t-1} = H] = \gamma_H L + (1-\gamma_H)H.$$

We assume that $S_0 > 3L$ and

$$\mu_L \in [L, (1/3)S_0)$$
$$\mu_H \in \left[S_0, S_0 + \delta\gamma_H\left(\frac{S_0}{3} - \mu_L\right)\right].$$

Intuitively, Assumption 10 imposes that the expected congestion parameter μ_L following an observation of L is small enough such that the CP would myopically send two or more agents after observing L (that is, $x_L^{SO} \geq 2$). If this were not the case, then the CP would send the same flow after L and H making the problem uninteresting. The assumption also imposes that $\mu_H \geq S_0$, which implies that the CP would myopically send no agent after seeing H, thus making experimentation beneficial only because of forward-looking incentives—there would be no experimentation with myopic agents. Finally, the upper bound on μ_H ensures that the forward-looking CP always find experimentation after seeing H beneficial (rather than sending all agents to safe for one or more rounds).

Proposition 3 (Social optimum). *Under Assumptions 8 and 10, $x_H^{SO} = 0$ and $x_L^{SO} \geq 2$. Let us define the social optimum recommendation scheme as a function π^{SO} that maps the belief β that the CP has about the state of the road at time $t-1$ to the number of agents to send to the risky road at time t to minimize total discounted travel time, that is,*

$$\pi^{SO}(\beta) := \underset{x\in\{0,\dots,N\}}{\operatorname{argmin}} \mathbb{E}\left[g(x,\theta_t) + \sum_{k\geq 1}\delta^k g\left(\pi^{SO}(\beta_{t-1+k}), \theta_{t+k}\right) \mid \mathbb{P}[\theta_{t-1} = L] = \beta\right]. \tag{4.5}$$

Then $\pi^{SO}(\beta) = \max\{1, x_\beta^{SO}\}$, with x_β^{SO} as defined in (4.4).

Under the social optimum recommendation scheme derived above the CP sends

one agent (the experimenter) if the state of the risky road was H at the previous step (to explore) and $x_L^{SO} \geq 2$ if it was L (to exploit). Hence, under π^{SO} the CP always knows the state of the risky road. The next example, however, shows that this scheme is not necessarily incentive compatible. In particular, when agents make their own routing decisions, the CP may not be able to send x_L^{SO} agents when the state is L.

Example 4. *Take the extreme case where $\gamma_L = 0$. Then $x_L^{SO} \approx \frac{S_0}{2L}$ and $x_L^{eq} \approx \frac{S_0}{L}$. (Again, the approximation comes from the integer constraint of our atomic model.) Suppose that at time $t-1$ the risky road state changes from H to L. Since $\gamma_L = 0$ this will be the state of the risky road from that point forward. According to π^{SO}, at time t, the CP sends x_L^{SO} drivers to risky to exploit the low state. After time t, under π^{SO}, agents that were on risky at time t should remain on risky forever and agents that were on safe at time t should remain on safe forever.*

However, consider an agent on safe at time t. After observing the flow x_L^{SO} at time t, this agent can infer that θ has changed to L. Hence at time $t+1$ he knows that

- *if he remains on safe, as prescribed by π^{SO}, he will experience a cost of S_0 for all future times;*
- *if he switches to risky, he will experience a cost of $\approx L(\frac{S_0}{2L} + 1) = \frac{S_0}{2} + L$ for all future times;*

Under Assumption 10, $L < \frac{S_0}{3} < \frac{S_0}{2}$. Hence following π^{SO} is not incentive compatible for the agent.

In the next sections, our objective is to derive incentive compatible recommendation schemes that achieve lower cost than providing full information.

4.3.2 Partial information: Incentive compatibility

Example 4 shows that the social optimum scheme π^{SO} may not be incentive compatible because it does not take into account the fact that agents that are on the safe

road can infer the state θ_{t-2} from the flow observed at time $t-1$. For this reason, from here on we consider recommendation schemes where the CP conditions his recommendations not only on θ_{t-1} but also on θ_{t-2}. In principle, the CP could even condition on further past values of the state θ. Though we are not able to rule out formally that conditioning on $(\theta_{t-2}, \theta_{t-1})$ is optimal without loss of generality, in what follows we simplify the analysis of incentive compatible recommendation schemes by assuming that the CP will condition only on $(\theta_{t-2}, \theta_{t-1})$ and thus the relevant state can be summarized by equilibrium path beliefs $(\beta_{t-2}, \beta_{t-1})$. Based on Proposition 3, we also restrict attention to schemes which do involve experimentation for all sample paths (meaning that the CP always prefers to send one agent on the risky road).

Definition 6 (Infinite horizon recommendation scheme). *A recommendation scheme is defined as a map $\pi : [0,1] \times [0,1] \to \{1, 2, ..., N\}$ which maps the belief of the CP on the state of the risky road at time $t-2$ and $t-1$ (i.e. β_{t-2}, β_{t-1}) to the number of agents to whom the CP sends a recommendation to take the risky road, r_R. If $\theta_{t-1} = L$, we assume that all agents that were on the risky road at time $t-1$ receive a recommendation to remain on the risky road at time t; the remaining recommendations (i.e. $\pi(\cdot, L) - x^R_{t-1}$) are sent to a random subset of the agents on safe. If instead $\theta_{t-1} = H$ then recommendations are sent to a random subset of all the agents. In both cases agents that do not receive a recommendation of risky receive a recommendation of safe, r_S. Finally, if any agent deviates, the CP provides full information to all agents from then on. (Intuitively, full information is the worst incentive compatible punishment, for deviation, that the CP can impose. In fact under full information, the expected cost per round of each agent is S_0. No punishment can lead to higher cost and be incentive compatible because agents can always switch to play safe and achieve a cost of S_0.) We denote the set of all recommendation schemes of this form as $\hat{\Pi}$.*

We want to stress that the assumption of restricted history applies only to the CP, not to agents. Consequently, we are in no way restricting the optimal behavior of the agents, who can condition their actions on all of their past information.

Any scheme $\pi \in \hat{\Pi}$ can be parametrized as follows

$$\pi(\beta_{t-2}, \beta_{t-1}) = \begin{cases} a & if\ [\beta_{t-2}, \beta_{t-1}] = [L, H] \\ b & if\ [\beta_{t-2}, \beta_{t-1}] = [H, H] \\ c & if\ [\beta_{t-2}, \beta_{t-1}] = [H, L] \\ d & if\ [\beta_{t-2}, \beta_{t-1}] = [L, L]. \end{cases}$$

Note that one should also specify the values of $\pi(\beta_{t-2}, \beta_{t-1})$ for values of $\beta_1, \beta_2 \in (0, 1)$ but in the schemes we consider this is not relevant in view of the fact that the CP always sends at least one agent to experiment and thus knows the state of the risky road in the previous period. For simplicity we denote a generic scheme π of this form as $\pi_{a,b,c,d}$ and the associated social cost as $V_{a,b,c,d}$ (we consider costs starting from $\theta_0 = H$ since this choice induces the lowest possible belief and is thus the most difficult scenario for experimentation).

Example 4 showed that π^{SO} may not be incentive compatible because if agents on the safe road see the flow $N - x_L^{SO}$ they can infer that the road switched to L at time $t-2$ and may have an incentive to deviate. This intuition motivates us to focus in particular on a subclass of $\hat{\Pi}$ consisting of schemes obtained by the following modification of the social optimum policy (see also Table 4.1). If $\theta_{t-1} = H$ the CP sends one agent (to experiment) exactly as in the social optimum (i.e. we set $a = b = 1$). If instead $\theta_{t-1} = L$ then the CP sends two possibly different flows c and d depending on whether the road has just switched to L or whether it was L also in the previous period (in which case the agents on safe can infer the road changed at time $t-2$ from the flow observed at $t-1$).

Definition 7. *Consider a scheme $\pi_{a,b,c,d} \in \hat{\Pi}$ with $a = b = 1$ and $1 < c \leq d$ and denote this for simplicity as $\pi_{c,d}$.*

Since $c > 1$, under any scheme $\pi_{c,d}$, agents can learn the state of the risky road at $t-2$ by observing the flow at $t-1$. Exploiting this fact, we show that agents can summarize their history h_{t-1} with a smaller state z_t^i, as detailed next.

Lemma 9. *Under $\pi_{c,d}$ and given that other agents follow their recommendations, an agent can evaluate if a recommendation is incentive compatible using only the information $z_t^i := [x_{t-1}, \beta_{t-1}^i, r_{t-1}^i]$ where*

- *$x_{t-1} \in X := \{0, 1, ..., N\}$ is the flow observed on the risky road at the previous time (even if an agent is on safe he can infer x_{t-1} as N minus the flow on safe, hence this is common information);*

- *$\beta_{t-1}^i \in \mathcal{B} := \{L, H, U\}$ encodes the information that agent i has about the state of the road at $t-1$. If the agent was on the risky road at $t-1$, then he knows the true realization (L or H), while if he was on the safe road we denote the fact that he does not know the state with the symbol U (Unobserved); (We simply use the symbol U instead of specifying the belief with a number in $(0, 1)$.)*

- *$r_{t-1}^i \in \Lambda := \{r_S, r_R\}$ is the recommendation an agent receives between round $t-1$ and t.*

Specifically, let h_{t-1}^i be the entire history of the agent up to and including time $t-1$. For any $\alpha_t^i \in \{S, R\}$ it holds

$$\begin{aligned}
&\mathbb{E}\left[\tilde{g}(\alpha_t^i, \pi_{c,d}^{-i,t}, \theta_t) + \sum_{k=t+1}^{\infty} \delta^{k-t} \tilde{g}(\pi_{c,d}^{i,k}, \pi_{c,d}^{-i,k}, \theta_k) \mid h_{t-1}^i\right] \\
&\quad = \mathbb{E}\left[\tilde{g}(\alpha_t^i, \pi_{c,d}^{-i,t}, \theta_t) + \sum_{k=t+1}^{\infty} \delta^{k-t} \tilde{g}(\pi_{c,d}^{i,k}, \pi_{c,d}^{-i,k}, \theta_k) \mid z_t^i\right]
\end{aligned}$$

where $\pi_{c,d}^{i,k}$ is the recommendation sent at time $k > t$ by the CP if agent i takes action α_t^i at time t and follows the recommendations from there on, while $\pi_{c,d}^{-i,k}$ denotes the recommendations sent to all other agents.

Intuitively, the flow x_{t-1} is a summary of all that happened up to θ_{t-2} (this is common information) and the combination of β_{t-1}^i and r_{t-1}^i adds personalized information about an agent's knowledge of θ_{t-1} before time t. Note that if road congestion were unobserved (U), under $\pi_{c,d}$ the combination of x_{t-1} and r_{t-1}^i would

be enough to provide a unique belief on the state of the risky road. That is any agents with the same state $z_t^j = z_t^i$ have the same belief on the state of the risky road.

Our first main result is to derive sufficient conditions on c, d so that $\pi_{c,d}$ is incentive compatible.

Proposition 4 (Symmetric equilibrium). *Suppose that Assumptions 8 and 10 hold. Additionally, assume that c, d are such that*

1. $x_L^{SO} \leq c \leq d \leq x_L^{eq}$
2. $g(c, \mu_L) \leq g(2, \mu_L)$
3. *the pair (c, d) is such that agents that are on safe and receive a recommendation of r_S after observing flow d on risky will follow the recommendation, that is,*

$$\underbrace{u([d, U, r_S])}_{\text{cost of following}} \leq \underbrace{p_{d,S}\mu_L(d+1) + (1 - p_{d,S})2\mu_H}_{\substack{\text{expected stage cost of deviating} \\ \text{to risky}}} + \underbrace{\frac{\delta}{1-\delta}S_0}_{\substack{\text{continuation cost} \\ \text{of deviating}}} \tag{4.6}$$

where $p_{d,S} = \mathbb{P}(\theta_{t-1} = L \mid z_t^i = [d, U, r_S])$. The constraint (4.6) *is written implicitly for readability and an explicit formula is provided in* (B.20) *in the Appendix.*

Then, the recommendation scheme $\pi_{c,d}$ induces the symmetric equilibrium

$$\xi^i_{\pi_{c,d}}(z_t^i) = \xi^i_{\pi_{c,d}}([x_{t-1}, \beta_{t-1}^i, r_{t-1}^i]) = \begin{cases} R & \text{if } r_{t-1}^i = r_R, \\ S & \text{otherwise}, \end{cases} \tag{4.7}$$

and is thus incentive compatible.

The intuition behind the conditions derived in the Proposition 4 are given next:

1. after the road switches to L for the first time the CP sends at least the social optimum number of agents and he possibly increases the flow after that, but no more than the myopic equilibrium flow;

$$\ldots LH|\underbrace{H\ldots HL\ldots LH}_{period}|\underbrace{H\ldots HL\ldots LH}_{period}|\underbrace{H\ldots HL\ldots LH}_{period}|\underbrace{H\ldots HL\ldots LH}_{period}|H\ldots$$

State of the risky road	H	H	...	H	L	L	L	...	L	H
π^{SO} (Social optimum)	-	1	...	1	1	x_L^{SO}	x_L^{SO}	...	x_L^{SO}	x_L^{SO}
$\pi_{c,d}$ (Proposition 4)	-	1	...	1	1	c	d	...	d	d

Table 4.1: Comparison between the flows on the risky road under schemes π^{SO} and $\pi_{c,d}$ for one period.

2. the flow sent by the CP on the risky road after the road switches to L for the first time leads to a no worse stage cost than sending just two agents;

3. d is large enough so that agents that are on safe and infer $\theta_{t-2} = L$ follow the recommendation to remain on safe (thus addressing the issue identified in Example 4).

The proof of Proposition 4 is presented in the Appendix. Here we provide some intuition. The first fundamental observation that we make is that, since the experimenter after $\theta_{t-1} = H$ is chosen at random among all the agents, each agent has the same continuation cost (which we term $\bar{v}$) after he observes the risky road switching from L to H. Because of this we can divide the infinite horizon into periods (by defining the beginning of a new period as the time immediately after the risky road switches from L to H) and study incentive compatibility only until the end of the current period. The division in periods and the number of agents taking the risky road under the schemes π^{SO} and $\pi_{c,d}$ for each period are illustrated in Table 4.1.

To prove incentive compatibility of $\pi_{c,d}$ we then need to show that no agent can improve his cost by a unilateral deviation. To this end, we divide the agents into four types:

1. Agents that took the risky road at time $t-1$ and saw L: under $\pi_{c,d}$ these agents receive a recommendation of risky. Since the probability that the road changes from L to H in one step is $\gamma_L \leq \frac{1}{2}$, one should expect that following such a recommendation is incentive compatible. In particular, we show that the stage cost

obtained by following the recommendation is less than S_0 (as proven in Lemma 15 in the Appendix), hence any deviation will increase both current cost and also continuation cost (because it leads to lower information than using the risky road and is thus not profitable).

2. Agents that took the risky road at time $t-1$ and saw H: there are two cases, either the agent receives a recommendation to take the safe road (which is intuitively incentive compatible since the probability that the road changes from H to L in one step is $\gamma_H \leq \frac{1}{2}$) or the agent receives a recommendation to take the risky road. The only case when the latter happens is if the agent is selected to be the next experimenter. In this case we show that, even though experimentation is costly in terms of current payo□s, the continuation cost is lower from experimenting than from deviating (recall that after any deviation the CP provides full information in all future periods). This makes being the experimenter incentive compatible.

3. Agents that took the safe road at time $t-1$ and received a recommendation to remain on the safe road: as noted in Example 4 from observing $x_{t-1} = c > 1$ or $x_{t-1} = d > 1$ these agents can infer $\theta_{t-2} = L$. Condition (4.6) guarantees it is incentive compatible for an agent that observed $x_{t-1} = d$ to follow a recommendation of using the safe road. We show that this condition implies incentive compatibility also for the case $x_{t-1} = c$. The only remaining possibility is when $x_{t-1} = 1$, in this case the agent can infer $\theta_{t-2} = H$ and incentive compatibility is immediate.

4. Agents that took the safe road at time $t-1$ and received a recommendation to take the risky road: incentive compatibility in this case follows with the same argument as in cases 1 and 2. Indeed, the agent has either been chosen to benefit from using the risky road when the state is L (which is incentive compatible by the discussion for case 1) or he has been chosen as an experimenter (which is incentive compatible by the discussion for case 2).

4.3.3 Partial information: Optimality

Motivated by Example 4, we consider a specific scheme among those that are incentive compatible according to Proposition 4. Specifically, for the period immediately after the road switched from H to L we assume that the CP sends the flow $c = x_L^{SO}$ exactly as in the social optimum (intuitively this is possible, because agents on safe are unaware that the road condition changed). For all subsequent periods the CP sends the minimum number of agents to maintain incentive compatibility (i.e. to satisfy (4.6) for $c = x_L^{SO}$). We denote this flow by x_{LL}.

Definition 8. *We define the scheme $\pi^* \in \hat{\Pi}$ as follows*

$$\pi^*(\beta_{t-2}, \beta_{t-1}) = \begin{cases} 1 & \text{if } \beta_{t-1} = H, \\ x_L^{SO} & \text{if } \beta_{t-2} = H, \beta_{t-1} = L, \\ x_{LL} & \text{if } \beta_{t-2} = L, \beta_{t-1} = L, \end{cases} \tag{4.8}$$

where $x_{LL} := \max\{x_L^{SO}, \bar{x}_{LL}\}$ with $\bar{x}_{LL}$ being the smallest integer such that $d = \bar{x}_{LL}$ satisfies (4.6) *for $c = x_L^{SO}$. In other words, $\pi^* = \pi_{(x_L^{SO}, x_{LL})}$.*

Corollary 4. *Suppose that Assumptions 8 and 10 hold. The scheme $\pi^* \in \hat{\Pi}$ is incentive compatible and achieves strictly lower social cost than full information.*

This corollary follows immediately from Proposition 4 upon noting that the pair $c = x_L^{SO}$ and $d = x_{LL}$ satisfy the assumptions of that proposition (we prove in Lemma 19 in the Appendix that $\bar{x}_{LL} \leq x_L^{eq}$). The fact that the social cost is strictly less than full information is proven in point 1 of Lemma 16 (in the Appendix).

We next derive sufficient conditions for the scheme π^* to be not only incentive compatible, but also optimal. We consider two different regimes depending on the discount factor δ used by the agents to weight future travel times.

Large δ

We show that as $\delta \to 1$, $x_{LL} \to x_L^{SO}$. In other words, the cost under π^* converges to the cost of the social optimum as $\delta \to 1$. Define the social cost starting from belief

$\beta = H$ under the social optimum and π^* as V_H^{SO} and $V_H^{\pi^*}$, respectively.

Proposition 5. *Suppose that Assumptions 8 and 10 hold and assume $\gamma_H, \gamma_L > 0$. Then the cost under π^* approaches the social optimum as $\delta \to 1$. Formally,*

$$\lim_{\delta \to 1} \frac{V_H^{\pi^*}}{V_H^{SO}} = 1.$$

This full optimality result obtains because for large δ the policy $\pi_{(c,d)} = \pi_{(x_L^{SO}, x_L^{SO})}$ satisfies (4.6) hence π^* coincides with π^{SO}. This result is to be expected. In fact given any time t let t^H be the first time the road switches to $\theta = H$ after t (this event happens in finite time since $\gamma_L > 0$). Then under any policy $\pi_{(c,d)}$, the cost of any agent is

$$\sum_{\tau=t+1}^{t_H} \delta^{\tau - t} \text{cost}_\tau + \delta^{\tau - t_H} \frac{V_H^{\pi_{(c,d)}}}{N}.$$

For $\delta \to 1$ the first term is negligible; hence any scheme for which $V_H^{\pi} < S_0 N/(1-\delta)$ (i.e. the cost under full information) is incentive compatible. Clearly the social optimum meets this condition and hence it must be incentive compatible.

Small δ

Before stating our main result we show that for δ small, for any scheme $\pi \in \hat{\Pi}$ to be incentive compatible it must be $a = b = 1$ (thus justifying our interest in the class of schemes given in Definition 7).

Lemma 10. *Suppose that $\delta \leq \frac{1}{2}$ and that Assumptions 8 and 10 hold. Then if a scheme $\pi \in \hat{\Pi}$ is incentive compatible it must be $a = b = 1$.*

The intuition for this result is straightforward. Since, after seeing H, the CP selects the next experimenter randomly, in any scheme π there is a positive probability that the selected experimenter knows that the risky road was H in the previous period. If either a or b are greater than one, then the expected cost of the experimenter would be greater than $2\mu_H$ (which is the expected stage cost). On the other hand, if the experimenter deviates, the CP provides full information and can guarantee an

expected cost of $\frac{S_0}{1-\delta}$ from then on. Under Assumption 10, and if $\delta \leq \frac{1}{2}$,

$$\frac{S_0}{1-\delta} \leq 2S_0 \leq 2\mu_H.$$

Hence having more than one experimenter cannot be incentive compatible. (We note that instead if $\delta > \frac{1}{2}$, a scheme with a or b greater than one, might be incentive compatible. Although, sending more than one agent to experiment always gives a higher stage cost for the CP, it is unclear under higher δ whether it may benefit the CP to send a higher flow after H to drive down d either through raising the cost of deviation in this setting or through obfuscation of information by making the flow the same after H and L. Overall, when $\delta > \frac{1}{2}$ we are unable to rule out that a scheme sending more than one agent to experiment could be incentive compatible and give a lower overall cost.) Having fixed a, b we now turn to the optimal choice of c, d.

Proposition 6. *Suppose that Assumptions 8 and 10 hold and $N \geq 5$. Then*

1. *for δ su ciently small, π^* achieves the minimum social cost among all the incentive compatible schemes belonging to $\hat{\Pi}$;*
2. *for $\delta \leq \frac{1}{2}$, the scheme that minimizes the social cost among all the incentive compatible schemes belonging to $\hat{\Pi}$ is either π^* or $\tilde{\pi}^* := \pi_{x_L^{SO}+1, x_{LL}-1}$.*

To understand the previous result recall that the social optimum choice would be $c = d = x_L^{SO}$. Unfortunately, in most cases this choice is not incentive compatible because of constraint (4.6) (guaranteeing that agents that are on safe follow a recommendation of safe). Our first step in the proof of Proposition 6 is to show that the incentive compatibility constraint (4.6) can be rewritten as $f(c) \leq g(d)$ where $f(c)$ is convex in c and is minimized at a value between x_L^{SO} and $x_L^{SO} + 1$. By the integer nature of our problem, this immediately implies that at optimality c should take one of this two values (for having a larger value of c would make the constraint (4.6) harder to satisfy (leading to $d \geq x_{LL}$) and would thus lead to a scheme with higher social cost than π^*; recall that $c = x_L^{SO}$ would be the optimal choice to minimize the social cost).

If the minimizer is $c = x_L^{SO}$, then by definition it must be $d \geq x_{LL}$. If, instead, $c = x_L^{SO} + 1$ is the minimizer then there exist parameters under which $d = x_{LL} - 1$ can be incentive compatible (but we show that no smaller values of d could be).(In fact, in some cases increasing the value of c leads to a smaller continuation cost for agents that follow the recommendation of safe (since they have higher probability to be sent to the risky road when the road changes to L in the next period). When that happens $\tilde{\pi}^*$ is incentive compatible by Proposition 4.) The pair $c = x_L^{SO} + 1$ and $d = x_{LL} - 1$, defining $\tilde{\pi}^*$, may give a lower social cost than π^* for certain parameter values (because there are more future rounds with flow d in expectation than rounds with c, hence increasing c slightly to decrease d may be beneficial). (While $\tilde{\pi}^*$ has a higher cost that π^* immediately after the road switches to L (since $g(x_L^{SO}+1, \mu_L) > g(x_L^{SO}, \mu_L)$), it has lower stage cost for all the subsequent times (since $g(x_{LL} - 1, \mu_L) < g(x_{LL}, \mu_L)$). For su□ciently large values of δ this may reduce the overall cost.) The second statement of Proposition 6 follows immediately from these observations. The first statement follows from the observation that, for δ small enough, the scheme π^* must have smaller social cost than $\tilde{\pi}^*$ (since it leads to smaller cost for the stage immediately after the state of the road switches to L, and for δ small enough, this dominates the potential future gain of using $d = x_{LL} - 1$ instead of $d = x_{LL}$).

4.4 Conclusion

New GPS technologies and tra□c recommendation systems critically depend on real-time information about road conditions and delays on a large number of routes. This information mainly comes from the experiences of drivers. Consequently, enough drivers have to be induced to experiment with di□erent roads (even if this involves worse expected travel times for them). This situation creates a classic experimentation-exploitation trade-o□, but critically one in which the party interested in acquiring new information cannot directly choose to experiment but has to convince selfish, autonomous agents to do it. This is the problem we investigate in the current chapter.

There is by now a large literature on experimentation in economics and operations

research. The main focus is on the optimal amount of experimentation by trying new or less well-known options in order to acquire information at the expense of foregoing current high payo□s. The game theoretic experimentation literature, investigating situations where there are multiple agents who can generate information for themselves and others, studies issues of collective learning, free-riding and underexperimentation. Missing from the previous literature is the main focus of our chapter: a setting in which exploitation of relevant information creates payo□ dependence (for example, via congestion in the context of our routing model) and the central entity or planner has the incentives for experimentation, but has to confront the incentive compatibility of the agents, especially in view of the aforementioned payo□ dependence.

We develop a simple model to study these issues, and characterize optimal recommendation systems first in a two-stage setting and then in an infinite-horizon environment. Key aspects of our model are congestion externalities on roads (introducing payo□ dependence); a finite number of agents (so that agents take into account their impact on information as well as congestion); forward-looking behavior by agents (so that they can be incentivized by future rewards); and a central planner who can observe results from experimentation and can make recommendations but has to respect incentive compatibility (introducing the feature that this is not a direct model of experimentation). We simplify our analysis by assuming that there are only two roads and one of them is "safe", meaning that the travel time is known, non-stochastic, and does not depend on the state of nature. This contrasts with the other, "risky" road, where travel times depend on the state of nature (on which the central planner is acquiring information).

We first show that full information, whereby the central planner shares all the information he acquires with all agents, is generally not optimal. The reason is instructive about the forces in our model: full information will make all agents exploit information about favorable conditions on the risky road, and this will in turn cause congestion on this risky road, reducing the rewards the experimenter would need to reap in order to encourage his experimentation. As a result, full information may lead to insu□cient or no experimentation, which is socially costly.

We then proceed to characterizing optimal incentive compatible recommendation schemes. These typically do not induce full information, but still share some of the information obtained from the experimentation of few experimenters (in our model only one experimenter is sufficient because there is no uncertainty about the state conditional on experimentation).

In the case of infinite-horizon, the underlying state of the risky road changes according to a Markov chain. An additional issue in this case is that the incentive compatibility of non-informed agents has to be ensured as well, since they may decide to disregard the recommendation of the central planner and choose the risky road when they think travel times are lower there. This makes the characterization of the optimal recommendation scheme more challenging. We propose a relatively simple incentive compatible dynamic scheme and then establish its optimality when the discount factor is small (in particular less than 1/2) and large enough (limiting to 1).

This chapter highlights the importance of understanding how modern routing technologies (and perhaps more generally) need to induce sufficient experimentation and how they can balance the benefits from exploiting new information and ensuring incentive compatibility of experimentation as well as incentive compatibility of all non-experimenters. Investigating how these issues can be navigated in more general settings (for example, with a more realistic road network and richer dynamic and stochastic elements or in models of payoff dependence resulting from other considerations) is an important area for future work.

Chapter 5

(Sub)Optimality of Bundling in Streaming Platforms

5.1 Introduction

Since Netflix introduced a video on demand service in 2007, many other competing streaming platforms have been established. Each o□ers its own set of movies and television options to users usually as a bundle, for some fixed monthly subscription fee. In this chapter, we aim to understand when bundling is the most profit maximizing strategy for a platform and when instead platforms should begin to unbundle their goods.

In this chapter, we introduce and study a model for the interaction between users and multiple streaming platforms that provide digital content. We use this to model streaming services such as Netflix and Hulu. In our model, the platforms compete in two types of products, we take for example comedies and original content. We view these products as being on a spectrum and each platform sitting on either end of this spectrum. For example the original content is di□erent on the two platforms and each consumer has some a preferred mix of the two contents. For the comedies, we can view this again as each platform having some set of comedies and users having a preference over the mix of the two platforms. We model this setting as a 2D Hotelling problem, wherein each platform must decide whether to bundle the two products together and

sell them for a single price or separate the two and sell them for individual prices. The users then must decide which products they would like to purchase. Note each user has a preference for the mix of the products from the two platforms which is modeled as their position on the spectrum.

We first study what happens when there is only one platform in the market. This market leader enters the market with a new product and is able to capture some, but not all, of the market. Our first result establishes that in this monopoly setting there exist conditions such that the platform obtains a higher profit from bundling the goods than from selling them separately. This is the case as consumers can only buy the goods from one platform and they may have a strong preference for one good and a weak preference for another, so are willing to purchase the bundle. We then move to the setting where another platform enters the market. Our second result establishes that in our model in the duopoly setting, it is always optimal for both platforms to sell the goods separately, or "unbundle" as we use to refer to this action. We claim this is in line with what is happening in practice as though one entity may own multiple streaming services, for example Disney has a majority share of Hulu they are not combining or bundling these services together. This also supports the fact that we see more and more of these platforms arising and no merging of the platforms is occurring. We even continue to see more specialized streaming services starting, for example Crunchyroll which is for anime and there is a different streaming service for each individual sport, e.g. NBA League Pass, NHL.TV, WWE Network. This is in line with the different axis representing different genres or types of content and unbundling being optimal.

5.1.1 Related Literature

This chapter relates to the works on pricing for bundled services. Many of the works study settings where a single user wants to purchase either one or multiple products. These works include [5], [58], [9], and [61], where the authors characterize when there are two products and a single seller under what conditions bundling is more profitable than unbundling. The above works all focus on a single seller, whereas

we study bundling in the monopoly and in the competitive setting. One paper that also studies competition in bundling is [10], where they consider bundling of a large number of information goods and show how bundling may be profitable for the seller. In [10] they are considering bundling a large number of small information goods, where the customers' realized values are unknown. Here instead, we consider bundling of two *types* of products, where each type is already a bundle in itself. The platforms also know the values of the users and thus have no uncertainty in demand. Another, more recent work [53], studies platforms that are connecting sellers and buyers and how the platform should price and bundle the goods in competition.

Our work also relates to the standard Hotelling model which was introduced in [40] to study competition among two entities in one dimension. We here extend the competition to two dimensions which we model as categories or genres, for example comedies and action films. Another work that considers a two-dimensional Hotelling model is [74]. Their work only considers two platforms selling single goods, but where they differ in one direction of taste and one direction of quality, where higher quality is always preferred by the users. We instead consider two directions of taste for two goods that can be sold together or separately.

Our work is different from these papers as we compare the optimal strategy under a single seller to the duopoly case. The users in our setting not only have a value of the products that they gain from buying, but they have preferences over which platform they purchase from. This models potentially different tastes in mixtures of content that is generated by these online platforms.

The rest of the paper is organized as follows. Section 5.2 presents the model. Section 5.3 studies the pricing and bundling decisions of a single platform in the market. Section 5.4 studies pricing and bundling when two platforms are now competing for customers. Section 5.5 concludes and the Appendix contains the omitted proofs.

5.2 Model

We consider two competing media service provider platforms, for example Netflix and Hulu that o□er a selection of television shows and movies to subscribers for one subscription fee. Users have heterogeneous preferences over the movie types and the two platforms.

We use a 2D Hotelling model, where each dimension is a type of product, for example action movies or comedies. One of the platforms sits at location $(0, 0)$; the other sits at location $(1, 1)$. We model the users as a unit of demand that is distributed over the unit square. The (x, y) position of the user in the unit square determines their tastes for each of the x and y goods. If a user purchases a good, she pays not only the price that the platform sets, but also a transport cost based on the distance from the platform's position. These transport costs represent heterogeneity in users' valuations for the two goods. The transport costs are t_x and t_y. Note the fixed value of the x good is denoted V_x and similarly the value of the y good for is denoted V_y. We assume agents do not multihome, so each agent wants only one x good and one y good.

Platform i's decision involves first choosing whether to o□er the two goods as a bundle and charge one price p_i or sell the two goods separately at a price p_i^x for the x good and p_i^y for the y good.

Customer's decision involves choosing which goods (if any) to buy from which platforms. The customer makes this decision based on her taste which is represented by her location, (x, y), and the platforms' decisions.

5.3 Monopoly

We start with only one firm in the market; we choose Platform 0 without loss of generality. This firm is the market leader and enters the market earlier then the other. An agent either buys from Platform 0 or does not buy. Platform 0 still must decide whether to bundle the products and o□er a single price p_0 or unbundle the

products and o□er product x at p_0^x and product y at p_0^y.

Throughout Section 5.3 we adopt the following assumption on the transport costs.

Assumption 11. $t_x \in \left[\frac{V_x}{2}, V_x\right), t_y \in \left[\frac{V_y}{2}, V_y\right)$.

This assumption guarantees that neither the x nor the y market is entirely covered by Platform 0. This means there are some customers that choose not to purchase.

5.3.1 Platform's Problem

If the platform bundles the goods it sets one price p_0. If it does not bundle it sets a price for the x good, p_0^x, and a price for the y good, p_0^y. The goal of the platform is to maximize total revenue.

5.3.2 Customer's Problem

Each user must decide whether to purchase or not. If Platform 0 bundles the goods, then the utility of an agent at position (x, y) who subscribes to Platform 0 is given by

$$\begin{aligned} u_0(x,y) &= V_x + V_y - t_x|x-0| - t_y|y-0| - p_0 \\ &= V_x + V_y - t_x x - t_y y - p_0. \end{aligned} \tag{5.1}$$

If the agent does not subscribe, she gets utility 0.

If the Platform 0 does not bundle and sells the x and y goods separately, then the utility for a single product $j \in \{x, y\}$ from Platform 0 is

$$u_0^j(x,y) = V_j - t_j j - p_j^0. \tag{5.2}$$

When the goods are sold separately each agent can buy both goods, one good, or no goods at all.

5.3.3 When is Bundling Optimal?

In the following proposition we characterize a sufficient setting where the monopoly platform chooses to bundle the goods.

Proposition 7. *Suppose Assumption 11 holds, there exists* $\alpha > 0, \beta > 0$ *such that when* $|V_x - V_y| \leq \alpha$ *and* $|t_x - t_y| \leq \beta$*, bundling is more profitable than unbundling.*

Proposition 7 shows that when the values of the goods are close to one another and the transport costs are close to one another then bundling is more profitable than unbundling. If one of the value parameters dominates it becomes better to unbundle the goods as the platform can extract more profit from separating the goods and charging higher prices for each good individually, as many people want the good with the dominating parameter value, but may have little to no value for the other good. Similarly, if one of the transport costs dominates, then the platform can extract more by price discriminating and not forcing customers to buy both goods which may not be optimal if the transport cost for one of the goods is too high.

To illustrate this point, we show in Figure 5-1 the ratio of profit from bundling to unbundling in many cases. Specifically, we take $V_x = 1$ and $t_x = 1$ and show how this ratio changes with different values of V_y and t_y. As shown in Proposition 7, we see that when V_y and t_y are also close to 1 the profit from bundling is greater than the profit from unbundling, but as we move away from this bundling becomes much more profitable. Note in the figure we plot cases when Assumption 11 does not hold and we see how in these cases again unbundling is more profitable.

5.4 Duopoly

In the previous section we considered what happened when only one platform was in the market, Platform 0. Now, assume that Platform 1 has entered the market. How does having two platforms change the optimal bundling and pricing strategies of the platforms?

In the monopoly case we wanted an uncovered market, but we now want the

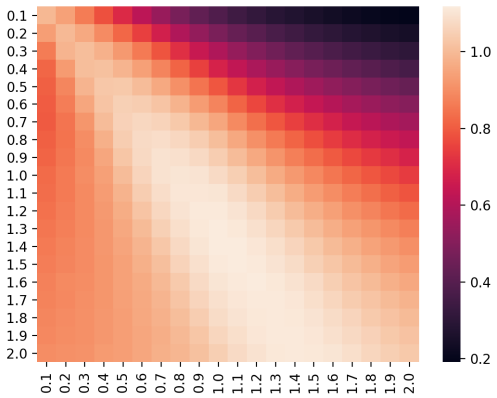


Figure 5-1: $V_x^0 = 1, t_x = 1$. Possible V_y^0 values are on the y-axis and possible t_y values are on the x-axis. The color represents the ratio of profit from bundling to profit from unbundling.

market to be covered by the two platforms. Thus, we adopt the following assumption to guarantee that no agent abstains from purchasing.

Assumption 12. $V_x \geq \frac{3}{2}t_x$, $V_y \geq \frac{3}{2}t_y$.

If Assumption 12 does not hold then there is not necessarily competition between the two platforms. Basically, each platform is a monopoly on its own side of the market. We are only interested in when there exist customers that get positive value from either platform. Otherwise, we are back in the monopoly setting.

Again, platforms must first decide whether or not to bundle. Then they must decide on their prices. Note that if one platform decides to bundle the other de facto bundles as the agents that don't buy from the bundled platform must buy both products from the unbundled platform, as under Assumption 12 the market is covered and all agents buy both goods, and thus pay a fixed fee which is the sum of the prices of the two separate items. So we just consider the two cases: i) both platforms bundle, ii) both platforms do not bundle.

In the next Proposition we show that it is the equilibrium for both platforms to sell the goods separately. This is in contrast to Proposition 7 where we saw that in the monopoly setting where there exist cases when bundling the goods was superior for Platform 0.

Proposition 8. *Suppose Assumption 12 holds, then unbundling is more profitable than bundling.*

The proof of Proposition 8 is straightforward and actually shows that when the platforms bundle they only extract the profit in a single direction of the competition, whereas when the platforms unbundle they are able to extract the profit from both directions of competition. This leads to higher profits under unbundling.

In the monopoly case it was better to bundle the goods in some cases, as it allowed the platform to get more customers while sacrificing the ability to price discriminate for the two goods. Here instead as there is competition, the price discrimination becomes more important and allows the platforms to compete on two axes instead of a single one. This allows them to split consumers that buy one good from each platform and charge more overall for these customers.

5.5 Conclusion

We introduced a model to study bundling and pricing decisions in online platforms. We start by showing that when there is only one platform o□ering two separate products, where agents have di□erent payo□s from these products based on their taste, there exist cases where it is optimal for the single platform to bundle the goods. We then move to the competition setting, where now the two platforms o□er both goods, but where preferences over the platforms depends again on a users taste. We see in this setting that unbundling is the unique optimal.

This model supports the fact that many specialized streaming services are arising and that we do not see companies that own multiple streaming services merging these into one product.

Appendix A

Proofs of Learning Dynamics in Network Games

A) Proof of Proposition 1. Under Assumptions 4 and 5, the mapping $B^i(p^i)$ as defined in (3.5) is Lipschitz continuous with constant $\frac{L}{\alpha}$. In fact, an agent's best response can be characterized as the solution of the variational inequality VI$(\mathcal{X}^i, F(\cdot\ ; p^i))$ where $F(x^i; p^i) := -\nabla_{x^i} u^i(x^i, p^i)$ and by Assumption 5 is strongly monotone with constant α uniformly in p^i. Take any two values p^i_A, p^i_B it follows by [63, Theorem 1.14] that

$$
\begin{aligned}
&\mid B^i(p^i_A) - B^i(p^i_B) \mid \\
&\qquad \leq \frac{1}{\alpha} \mid F(B^i(p^i_A); p^i_A) - F(B^i(p^i_A); p^i_B) \mid \\
&\qquad \leq \frac{L}{\alpha} \mid p^i_A - p^i_B \mid .
\end{aligned}
$$

Now, take any two vectors x_A and x_B then

$$\begin{aligned}
\|B(x_A) - B(x_B)\|^2 &= \sum_{i=1}^{N} \left(B^i(p^i(x_A^{-i})) - B^i(p^i(x_B^{-i}))\right)^2 \\
&\leq \left(\frac{L}{\alpha}\right)^2 \sum_{i=1}^{N} \left(p^i(x_A^{-i}) - p^i(x_B^{-i})\right)^2 \\
&\leq \left(\frac{L}{\alpha}\right)^2 \sum_{i=1}^{N} \left(a^i(z^i(x_A^{-i}) - z^i(x_B^{-i}))\right)^2 \\
&\leq \left(\frac{L}{\alpha} \max_i(|a^i|)\right)^2 \|P(x_A - x_B)\|^2 \\
&\leq \left(\frac{L}{\alpha} \max_i(|a^i|)\rho(P)\right)^2 \|x_A - x_B\|^2 \\
&= \gamma^2 \|x_A - x_B\|^2,
\end{aligned}$$

where we used Lipschitz continuity for the first inequality. Since $\gamma < 1$ by assumption, B is a contraction. Since $\mathcal{X}$ is closed and convex by Banach fixed point theorem the Nash equilibrium exists, is unique, and the best response dynamics $\bar{x}_{k+1} = B(\bar{x}_k)$ converge to it [31, Theorem 2.1.21].

B) Auxiliary results

Lemma 11. *Consider two non-negative sequences* $\{\xi_k \geq 0\}_{k=1}^{\infty}$, $\{w_k \geq 0\}_{k=1}^{\infty}$ *satisfying the system*

$$\xi_{k+1} \leq \gamma \xi_k + w_k,$$

for some $0 < \gamma < 1$. *Then* $w_k \to 0$ *implies* $\xi_k \to 0$.

Proof. Let us define the auxiliary sequence

$$s_k^L := \sup\{w_k, w_{k+1}, \dots\}.$$

Note that $s_k^L \geq 0$ for all k and, since $w_k \to 0$, $\lim_{k\to\infty} s_k^L = 0$. Fix a step k. Since

$\gamma > 0$, for any $p > 0$ the state in $k + p$ can be written as

$$\begin{aligned}
\xi_{k+p} &\leq \gamma^p \xi_k + \sum_{i=k}^{k+p-1} \gamma^{k+p-i-1} w_i \leq \gamma^p \xi_k + s_k^L \sum_{i=0}^{p-1} \gamma^i \\
&= \gamma^p \xi_k + s_k^L \frac{1-\gamma^p}{1-\gamma} \\
&\leq \gamma^p \left(\gamma^k \xi_0 + s_0^L \frac{1-\gamma^k}{1-\gamma} \right) + s_k^L \frac{1-\gamma^p}{1-\gamma} \\
&\leq \gamma^k \xi_0 + \gamma^p K_1 + s_k^L K_2
\end{aligned}$$

where $K_2 := 1/(1-\gamma)$ and $K_1 := K_2 s_0^L$. Now $\xi_k \to 0$ as $k \to \infty$ by using the definition of the limit, that is, by proving that for all $\epsilon > 0$, there exists an $\bar{h} > 0$ such that $\xi_{\bar{h}+h} \leq \epsilon$, for all $h \geq 0$. For any fixed $\epsilon > 0$ we can choose a $\bar{k} > 0$ such that $\gamma^{\bar{k}} \xi_0 < \frac{\epsilon}{3}$ and $s_{\bar{k}}^L K_2 < \frac{\epsilon}{3}$ and a $\bar{p} > 0$ such that $\gamma^{\bar{p}} K_1 < \frac{\epsilon}{3}$. Let $\bar{h} = \bar{k} + \bar{p}$, then

$$\begin{aligned}
\xi_{\bar{h}+h} = \xi_{\bar{k}+\bar{p}+h} &\leq \gamma^{\bar{k}} \xi_0 + \gamma^{\bar{p}+h} K_1 + s_{\bar{k}}^L K_2 \\
&\leq \gamma^{\bar{k}} \xi_0 + \gamma^{\bar{p}} K_1 + s_{\bar{k}}^L K_2 \\
&\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon \qquad \forall h \geq 0.
\end{aligned}$$

∎

C) ***Proof of Proposition 2.*** This proof is similar to the proof of Lemma 6 in [59]. We reference the reader to [59] for a more detailed exposition.

Note that the estimator in (3.9) can be rewritten by plugging in the value for y_k^i as

$$\hat{a}_k^i = a^i + \frac{\sum_{t=1}^k \epsilon_t^i z_t^i}{\sum_{t=1}^k (z_t^i)^2}.$$

The error of the parameter estimate of a^i at step k is therefore

$$\text{err}_k^i := \hat{a}_k^i - a^i = \frac{\sum_{t=1}^k \epsilon_t^i z_t^i}{\sum_{t=1}^k (z_t^i)^2} =: \frac{1}{g_k} \sum_{t=1}^k h_t,$$

where $h_k = \epsilon_k^i z_k^i$ and $g_k = \sum_{t=1}^k (z_t^i)^2$. By [59, Lemma 2] the stochastic process $s_k^i = \sum_{t=1}^k \frac{h_t}{g_t}$ is a martingale. Consequently, by the martingale convergence theorem

[69, Chapter 7, Section 4] there exists a set Σ of measure one such that for any noise realization $\epsilon \in \Sigma$ the deterministic sequence $s_k^i(\epsilon)$ converges to a finite value as $k \to \infty$. Consider a fixed noise realization $\epsilon \in \Sigma$ and note that $\{g_k(\epsilon)\}_{k=1}^{\infty}$ is non-negative and non-decreasing. Depending on the asymptotic behavior of $g_k(\epsilon)$ we can distinguish two cases:

1. if $g_k(\epsilon) \to \infty$ (i.e. $i \in S^{\infty}(\epsilon)$) then by Kronecker's lemma (stated e.g. as Lemma 3 in [59])
$$\text{err}_k^i(\epsilon) = \frac{1}{g_k(\epsilon)} \sum_{t=1}^{k} h_t(\epsilon) \to 0.$$

2. if $g_k(\epsilon)$ converges to a finite value (i.e. $i \in S^{\text{finite}}(\epsilon)$) then we can apply Lemma 4 in [59] and
$$\text{err}_k^i(\epsilon) = \frac{1}{g_k(\epsilon)} \sum_{t=1}^{k} h_t(\epsilon) \to \tilde{M}^i(\epsilon)$$
for some finite $\tilde{M}^i(\epsilon)$. Thus, for any $\delta > 0$ there exists a $\hat{k} > 0$ such that for any $k > \hat{k}$, $\left|\text{err}_k^i(\epsilon) - \tilde{M}^i(\epsilon)\right| < \delta$ and thus $\text{err}_k^i(\epsilon)$ is bounded. Thus there exists a finite value $M^i(\epsilon)$ such that $|\hat{a}_k^i(\epsilon) - a^i| \leq M^i(\epsilon)$ for all k.

Appendix B

Proofs of Optimal Dynamic Information Provision in Traffic Routing

B.1 Proofs of Section 3: Two stage example

Proof. **of Lemma 8** Suppose that $k \geq 2$ agents experiment. Then the cost of any of these agents is

$$[\text{expected cost of risky}] = k\mu_\beta + \eta_R \geq k\mu_\beta \geq 2\mu_\beta > 2(S_0 + S_1 N),$$

where we denoted by η_R the expected cost in the second round, which is for sure non-negative and we used Assumption 9. If instead the agent switches to safe he will have an expected cost of

$$[\text{expected cost of safe}] = S_0 + S_1(N - k + 1) + \eta_S \leq S_0 + S_1(N-1) + S_0 + S_1 N,$$

where we denoted by η_S the expected cost in the second round, which is at most $S_0 + S_1 N$ as the agent can always play safe in the second round and in the worst case every other agent is also playing safe. Since $S_0 + S_1(N-1) + S_0 + S_1 N < 2(S_0 + S_1 N)$,

it follows that the cost of experimenting is greater than the overall cost of taking safe. Thus, it is never a pure strategy equilibrium for more than one person to experiment – under any information scheme. ∎

Proof. **of Theorem 3** Recall we are considering pure strategy equilibria. Using Lemma 8 we can characterize when it is an agent's best response to experiment under the two different information schemes.

Full information: If no agent experiments the cost for each agent is $2(S_0 + S_1 N)$. If any agent experiments all agents learn θ before round two and there are two possibilities.

1) If $\theta = H$, everyone takes the safe road in the second round as $S_0 + S_1 N < H$.

2) If $\theta = L$, the agents play a pure strategy Nash equilibrium and split the flow across the two roads, i.e. x_L^{eq} take risky. The expected cost of equilibrium in the second round when $\theta = L$ is therefore $g(x_L^{\text{eq}}, L)/N = (x_L^{\text{eq}}/N)(L x_L^{\text{eq}}) + (N - x_L^{\text{eq}})/N(S_0 + S_1(N - x_L^{\text{eq}}))$.

All agents playing safe is an equilibrium if and only if the cost of switching to experimenting is worse than $2(S_0 + S_1 N)$. The expected cost of one agent switching is

$$\mu_\beta + \beta g(x_L^{\text{eq}}, L)/N + (1 - \beta)(S_0 + S_1 N)$$

where the first term is the cost of experimenting in the first round, the second term is the cost in the second round if $\theta = L$ weighted by $\mathbb{P}(\theta = L)$, and similarly the last term is the cost if $\theta = H$ weighted by $\mathbb{P}(\theta = H)$.

Overall, if

$$2(S_0 + S_1 N) < \mu_\beta + \beta g(x_L^{\text{eq}}, L)/N + (1 - \beta)(S_0 + S_1 N), \tag{B.1}$$

no one experiments. Otherwise, one agent experimenting in the first round is the unique pure strategy Nash equilibrium (recall by Lemma 8 that it is never incentive compatible for more than one agent to experiment). The total cost of the first round

under this equilibrium is $g(1, \mu_\beta)$ and if $\theta = H$ the total cost of the second round is $g(0)$, while if $\theta = L$ the total cost of the second round is $g(x_L^{\text{eq}}, L)$. The thresholds β_f can be obtained by imposing equality in (B.1) and solving for β.

Private information: No experimentation is an equilibrium if and only if it is not individually optimal for an agent to play risky. Similarly to the previous case, this occurs when the expected cost of switching to playing risky is worse than all agents playing safe, that is, when

$$2(S_0 + S_1 N) < \mu_\beta + \beta L + (1 - \beta)(S_0 + S_1 N), \tag{B.2}$$

where we used the fact that, under private information, it is a best response for all the agents that were on safe at time 1 to remain on safe at time 2, while for the experimenter it is a best response to take risky at time 2 if he observed L at time 1 and safe otherwise.

If the above does not hold, then there is an incentive to deviate from all playing safe and there exists an asymmetric pure strategy Nash equilibrium, where one agent experiments in the first round. By Lemma 8, this is the unique pure strategy Nash equilibrium. The expected cost of the first round is $g(1, \mu_\beta)$ and the expected cost of the second round is $\beta g(1, L) + (1 - \beta) g(0)$. The thresholds β_p can be obtained by imposing equality in (B.2) and solving for β.

Finally, note that $\beta_p \leq \beta_f$ if and only if $\frac{g(x_L^{\text{eq}}, L)}{N} \geq L$. Note that by Assumption 9 $L < S_0 + S_1$, thus $x_L^{\text{eq}} \geq 1$. If $x_L^{\text{eq}} = N$ then $g(x_L^{\text{eq}}, L) = LN^2$ and the inequality holds. Instead, if $1 \leq x_L^{\text{eq}} \leq N - 1$ then

$$\begin{aligned}
g(x_L^{\text{eq}}, L) &= L(x_L^{\text{eq}})^2 + (S_0 + S_1(N - x_L^{\text{eq}}))(N - x_L^{\text{eq}}) \\
&= L(x_L^{\text{eq}})^2 + S_0(N - x_L^{\text{eq}}) + S_1(N - x_L^{\text{eq}})(N - x_L^{\text{eq}}) \\
&\geq L x_L^{\text{eq}} + (S_0 + S_1)(N - x_L^{\text{eq}}) \\
&> L x_L^{eq} + L(N - x_L^{\text{eq}}) \\
&\geq LN
\end{aligned}$$

where the first inequality follows $x_L^{\text{eq}}, N - x_L^{\text{eq}} \geq 1$, while the second follows from $L < S_0 + S_1$. ∎

Proof. **of Corollary 3** In this interval of beliefs:

- under full information there is no experimentation and the cost is $2g(0)$.
- under private information there is experimentation. The experimenter has a total cost for the two rounds that is less than $2(S_0 + S_1 N)$, otherwise he would switch to safe. All the other agents have cost $2(S_0 + S_1(N-1)) < 2(S_0 + S_1 N)$. Therefore the total cost is strictly less than $2N(S_0 + S_1 N) = 2g(0)$.

∎

Proof. **of Theorem 4** The social optimum is the minimum total cost for the two rounds. If all agents use the safe road in both rounds then the total cost is $2g(0)$. If the CP sends *at least one agent* on the risky road in the first round the CP learns θ and can make a decision on how many agents to send on the risky road in the second round based on the value of θ. Denote these flows by $x_2^{R|L}$ and $x_2^{R|H}$. Thus, if the CP experiments with $x_1^R > 0$ agents in the first round, he is facing the following optimization problem

$$\min_{x_1^R, x_2^{R|L}, x_2^{R|H}} g(x_1^R, \mu_\beta) + \beta g(x_2^{R|L}, L) + (1-\beta) g(x_2^{R|H}, H)$$

$$\begin{aligned} \text{s.t.} \quad & 1 \leq x_1^R \leq N \\ & 0 \leq x_2^{R|L} \leq N \\ & 0 \leq x_2^{R|H} \leq N \\ & x_1^R, x_2^{R|L}, x_2^{R|H} \in \mathbb{Z}_{\geq 0}. \end{aligned}$$

This minimization can be separated into three optimization problems, one for each of the flows $x_1^R, x_2^{R|L}, x_2^{R|H}$.

- For the first round x_1^R can be obtained by solving

$$\begin{aligned} \min_{x_1^R} \quad & \mu_\beta (x_1^R)^2 + (S_0 + S_1(N - x_1^R))(N - x_1^R) \\ \text{s.t.} \quad & 1 \leq x_1^R \leq N \\ & x_1^R \in \mathbb{Z}_{\geq 0}. \end{aligned}$$

  Since $S_0 + S_1 N < \mu_\beta$, the cost is strictly increasing for $x_1^R \geq 1$ and thus the optimal solution is $x_1^R = 1$.

- For the second round, the two values $x_2^{R|L}$ and $x_2^{R|H}$ can be found separately and are just the values that minimize $g(x, L)$ and $g(x, H)$ respectively, which are x_L^{SO} and x_H^{SO} by definition.

The minimum of all agents playing safe and the objective of the above minimization problem gives the social optimum cost and the threshold β_{SO} is the belief under which the CP is indi□erent between experimentation and all agents playing safe. Note that $\beta_{SO} < \beta_p$ because if experimentation is an equilibrium under private information it implies it is optimal for the CP to experiment. Specifically if experimentation is an equilibrium under private information then β is such that

$$2(S_0 + S_1 N) \geq \mu_\beta + \beta L + (1 - \beta)(S_0 + S_1 N).$$

This implies

$$\begin{aligned} & 2(S_0 + S_1 N) + 2(S_0 + S_1 N)(N - 1) \\ & \qquad \geq \mu_\beta + (S_0 + S_1 N)(N - 1) + \beta(L + (S_0 + S_1 N)(N - 1)) \\ & \qquad\qquad + (1 - \beta)((S_0 + S_1 N)N) \\ & \qquad \geq \mu_\beta + (S_0 + S_1(N - 1))(N - 1) + \beta g(x_L^{\text{SO}}, L) + (1 - \beta) g(x_H^{\text{SO}}, H) \\ \iff & 2(S_0 + S_1 N)N \geq \mu_\beta + (S_0 + S_1(N - 1))(N - 1) + \beta g(x_L^{\text{SO}}, L) + (1 - \beta) g(x_H^{\text{SO}}, H) \end{aligned}$$

which is the condition for experimentation to be optimal for the CP. ∎

Proof. **of Theorem 5**

Recall by Lemma 8 that in any equilibrium there is at most one experimenter in the first round. Hence we can distinguish two cases:

1. **No experimentation:** if no agent experiments the social cost is $2g(0)$;

2. **One experimenter:** For a recommendation scheme π to be incentive compatible, it must be that $\pi(L)$, $\pi(H)$ are such that an agent follows the recommendation he is given. We study incentive compatibility starting from the **second round**. In this case there are three type of agents:

   - **Type 1 (experimenter in round 2)**: The experimenter knows the value of θ since he observed it in the first round. We next show that without loss of optimality we can restrict our attention to recommendation schemes where it is a best response for the experimenter to take risky in the second round if $\theta = L$ and safe if $\theta = H$.

     – $\theta = H$: The experimenter's cost on safe is $S_0 + S_1(N - \pi(H))$, the cost on risky is $H(\pi(H) + 1)$. The conclusion follows since,

     $$S_0 + S_1(N - \pi(H)) \leq S_0 + S_1 N < H \leq H(\pi(H) + 1),$$

     where we used the fact that $H > S_0 + S_1 N$ by assumption.

     – $\theta = L$: Let $x_2^{R|L}$ be the equilibrium flow on the risky road in the second round (this a priori may or may not include the experimenter). For incentive compatibility it must be $x_2^{R|L} \leq x_L^{\mathsf{eq}}$. In fact if that was not the case, consider an agent that was on safe in the first round and receives a recommendation of risky. This agent doesn't know θ, but he knows that in both cases ($\theta = L$ or $\theta = H$) switching to safe would give a better cost. Hence a scheme that leads to $x_2^{R|L} > x_L^{\mathsf{eq}}$ is not incentive compatible. We then distinguish two cases for the experimenter:

       if the experimenter belongs to the flow $x_2^{R|L}$ then deviating to safe is not convenient because $x_2^{R|L} \leq x_L^{\mathsf{eq}}$;

- if the experimenter is on safe, then either $x_2^{R|L} < x_L^{\text{eq}}$ in which cases it is convenient for the experimenter to deviate to risky or $x_2^{R|L} = x_L^{\text{eq}}$.

Overall the only case when it might be convenient for the experimenter to take safe after observing $\theta = L$ is for recommendation schemes such that $x_2^{R|L} = x_L^{\text{eq}}$. Note that the social cost of such a scheme is the same as full information. We are going to show at the end of this proof that the optimal solution of (4.3) is weakly less than full information.

Overall, the previous discussion shows that we can assume $x_2^{R|L} = \pi(L)+1$ and $x_2^{R|H} = \pi(H)$ without loss of optimality. For simplicity we denote these flows by x_L and x_H in the rest of this proof.

- **Type 2 (recommended safe)**: An agent of this type took safe in the first round and received a recommendation to take safe, signal r_S, before the second round. His expected cost of following the recommendation is

$$S_0 + S_1(\mathbb{P}(\theta = L \mid r_S)(N - x_L) + \mathbb{P}(\theta = H \mid r_S)(N - x_H)),$$

as the flow he will experience depends on how many agents are being sent to risky. Deviating gives an expected cost of

$$\mathbb{P}(\theta = L \mid r_S)L(x_L + 1) + \mathbb{P}(\theta = H \mid r_S)H(x_H + 1).$$

By Bayes rule

$$\begin{aligned}
\mathbb{P}(\theta = L \mid r_S) &= \frac{\beta\mathbb{P}(r_S \mid \theta = L)}{\beta\mathbb{P}(r_S \mid \theta = L) + (1-\beta)\mathbb{P}(r_S \mid \theta = H)} \\
&= \frac{\beta\frac{N-x_L}{N-1}}{\beta\frac{N-x_L}{N-1} + (1-\beta)\frac{N-x_H-1}{N-1}} \\
&= \frac{\beta(N-x_L)}{\beta(N-x_L) + (1-\beta)(N-x_H-1)} \\
\mathbb{P}(\theta = H \mid r_S) &= 1 - \mathbb{P}(\theta = L \mid r_S).
\end{aligned}$$

Thus, the full constraint is

$$S_0 + S_1 \left(\frac{\beta(N - x_L)^2}{\beta(N - x_L) + (1-\beta)(N - x_H - 1)} + \frac{(1-\beta)(N - x_H)(N - x_H - 1)}{\beta(N - x_L) + (1-\beta)(N - x_H - 1)} \right) \tag{B.3}$$

$$\leq \frac{\beta(N - x_L)L(x_L + 1)}{\beta(N - x_L) + (1-\beta)(N - x_H - 1)} + \frac{(1-\beta)(N - x_H - 1)H(x_H + 1)}{\beta(N - x_L) + (1-\beta)(N - x_H - 1)}.$$

- **Type 3 (recommended risky)** An agent of this type took safe in the first round and received a recommendation to take risky, signal r_R, before the second round. His expected cost of following the recommendation is

$$\mathbb{P}(\theta = L \mid r_R) L x_L + \mathbb{P}(\theta = H \mid r_R) H x_H$$

and deviating gives an expected cost of

$$S_0 + S_1(\mathbb{P}(\theta = L \mid r_R)(N - x_L + 1) + \mathbb{P}(\theta = H \mid r_R)(N - x_H + 1)).$$

By Bayes rule

$$\begin{aligned}
\mathbb{P}(\theta = L \mid r_R) &= \frac{\beta \mathbb{P}(r_R \mid \theta = L)}{\beta \mathbb{P}(r_R \mid \theta = L) + (1-\beta)\mathbb{P}(r_R \mid \theta = H)} \\
&= \frac{\beta \frac{x_L - 1}{N-1}}{\beta \frac{x_L - 1}{N-1} + (1-\beta)\frac{x_H}{N-1}} \\
&= \frac{\beta(x_L - 1)}{\beta(x_L - 1) + (1-\beta)x_H} \\
\mathbb{P}(\theta = H \mid r_R) &= 1 - \mathbb{P}(\theta = L \mid r_R).
\end{aligned}$$

Thus, the full constraint is

$$\frac{\beta(x_L - 1)}{\beta(x_L - 1) + (1-\beta)x_H} L x_L + \frac{(1-\beta)x_H}{\beta(x_L - 1) + (1-\beta)x_H} H x_H \tag{B.4}$$

$$\leq S_0 + S_1 \left(\frac{\beta(x_L - 1)}{\beta(x_L - 1) + (1-\beta)x_H}(N - x_L + 1) \right. \tag{B.5}$$

$$\left. + \frac{(1-\beta)x_H}{\beta(x_L - 1) + (1-\beta)x_H}(N - x_H + 1) \right)$$

The constraints above are for the second round, we next consider the **first round**. We already know by Lemma 8 that it is not convenient for any agent on safe to join the experimenter. Hence we only need to ensure that it is incentive compatible for the experimenter to experiment in the first round. Equivalently, we need to show that experimenting gives a weakly lower cost than all agents playing safe for two rounds ($2(S_0 + S_1 N)$), which leads to the constraint

$$\beta L + (1-\beta)H + \beta L x_L + (1-\beta)(S_0 + S_1(N - x_H)) \leq 2(S_0 + S_1 N). \quad \text{(B.6)}$$

The CP then solves the constrained optimization problem given in (4.3), where the objective function is the total travel time summed over the two periods and the IC constraints (4.3b), (4.3c) and (4.3d) can be explicitly rewritten as detailed in (B.3), (B.4), and (B.6) respectively.

Finally it is easy to show that the choices $\pi_2(L) = \pi_2(H) = 0$ and, for $\beta \geq \beta_f$, $\pi_2(L) = x_L^{\text{eq}} - 1, \pi_2(H) = 0$ are feasible (i.e. satisfy (B.3), (B.4), and (B.6)) and lead to the same social cost as private and full information respectively, thus proving that partial information is weakly better than private and full information.

∎

B.2 Proofs of Section 4: Infinite horizon

B.2.1 Proof of Proposition 3

We start by showing that under the given assumptions $x_H^{\text{SO}} = 0$ and $x_L^{\text{SO}} \geq 2$.

- $\boldsymbol{x_H^{\text{SO}} = 0}$: Note that

$$\mathbb{E}_{\theta_t}\left[g(x, \theta_t) \mid \beta_{t-1} = 0\right] = \mu_H x^2 + S_0(N - x) = g(x, \mu_H).$$

We next show that under the given assumptions $g(0, \mu_H) \leq g(1, \mu_H)$. Together with the fact that $g(x, \mu_H)$ is strongly convex in x, this proves the desired statement. Note that

$$g(0, \mu_H) \leq g(1, \mu_H) \Leftrightarrow S_0 N \leq \mu_H + S_0(N-1) \Leftrightarrow S_0 \leq \mu_H,$$

and the latter inequality holds by Assumption 10.

- $\boldsymbol{x_L^{\mathrm{SO}} \geq 2}$: x_L^{SO} is the minimizer of $g(x, \mu_L)$, which is strongly convex in x. For the minimizer to be ≥ 2, the cost at $x = 2$ must be strictly less than at $x = 1$, that is,

$$g(2, \mu_L) = 4\mu_L + S_0(N-2) < \mu_L + S_0(N-1) = g(1, \mu_L)$$

rearranging gives

$$\mu_L < \frac{1}{3} S_0,$$

which holds by Assumption 10.

Since the CP has full control, (4.5) is an optimal control problem and we can apply the one-step deviation principle to prove optimality. To this end, we distinguish two cases

- Consider any β such that $\boldsymbol{x_\beta^{\mathrm{SO}} \geq 1}$:
  Sending a number of agents di□erent from x_β^{SO} doesn't lead to a profitable deviation. In fact, the stage cost would be higher (since x_β^{SO} is the minimizer of $g(x, \mu_\beta)$) and the continuation cost would be the same (as no more information can be gained by sending more agents to the risky road).

- Consider any β such that $\boldsymbol{x_\beta^{\mathrm{SO}} = 0}$:
  In this case sending $\pi^{\mathrm{SO}}(\beta) = 1$ agent is not myopically optimal and the stage cost could be reduced by sending no agent. Nonetheless, we show that because sending one agent provides information about the state of the risky road,

$\pi^{\text{SO}}(\beta) = 1$ is the best strategy if the CP is forward looking. There are two possible deviations:

1. The CP sends more than one agent: similarly to the previous case the stage cost increases and the continuation cost stays the same. Hence this deviation is not profitable.

2. The CP does not send any agents: to analyze this case, let β' be the belief that $\theta_{t+1} = L$ when an agent has belief β that $\theta_t = L$ (i.e. $\beta' = \beta(1-\gamma_L) + (1-\beta)\gamma_H$). We start by noting that $x^{\text{SO}}_\beta = 0 \Rightarrow x^{\text{SO}}_{\beta'} \leq 1 \Rightarrow \pi^{\text{SO}}(\beta') = 1$ (see Lemma 12 below). The cost that the CP encounters by sending one agent at time t is

$$V(\beta) = \underbrace{\mu_\beta + S_0(N-1)}_{\text{stage cost time t}} + \delta \underbrace{\left(\beta'(\mu_L(x^{\text{SO}}_L)^2 + S_0(N - x^{\text{SO}}_L)) + (1-\beta')(\mu_H + S_0(N-1))\right)}_{\text{stage cost time t+1}} + \delta^2 \dots$$

while if he deviates and sends no agent at time t the cost is

$$\begin{aligned}\tilde{V}(\beta) &= \underbrace{S_0 N}_{\text{stage cost time t}} + \delta \underbrace{(\mu_{\beta'}\pi^{\text{SO}}(\beta')^2 + S_0(N - \pi^{\text{SO}}(\beta')))}_{\text{stage cost time t+1}} + \delta^2 \dots \\ &= \underbrace{S_0 N}_{\text{stage cost time t}} + \delta \underbrace{(\beta'(\mu_L + S_0(N-1)) + (1-\beta')(\mu_H + S_0(N-1)))}_{\text{stage cost time t+1}} + \delta^2 \dots\end{aligned}$$

where we used $\mu_{\beta'} = \beta'\mu_L + (1-\beta')\mu_H$ and $\pi^{\text{SO}}(\beta') = 1$. Note that we did not report the stage costs from time $t+2$ on as they are equal under both schemes. Overall, $\pi^{\text{SO}}(\beta) = 1$ is optimal if $V(\beta) \leq \tilde{V}(\beta)$ or equivalently,

$$\begin{aligned}&\mu_\beta + S_0(N-1) + \delta\beta'(\mu_L(x^{\text{SO}}_L)^2 + S_0(N - x^{\text{SO}}_L)) \leq S_0 N + \delta\beta'(\mu_L + S_0(N-1)) \\ &\Leftrightarrow \mu_\beta \leq S_0 + \delta\beta'[g(1,\mu_L) - g(x^{\text{SO}}_L, \mu_L)]. \qquad \text{(B.7)}\end{aligned}$$

Note that $g(1,\mu_L) - g(x^{\text{SO}}_L, \mu_L) \geq 0$ since $x^{\text{SO}}_L = \arg\min_x g(c, \mu_L)$. Moreover when β increases μ_β decreases. Hence it suffices to prove that (B.7)

holds for the smallest possible value of β which is 0.

We note that $g(1,\mu_L) - g(x_L^{\mathsf{SO}},\mu_L) \geq S_0 - 3\mu_L$ since

$$\begin{aligned} g(1,\mu_L) - g(x_L^{\mathsf{SO}},\mu_L) &\geq g(1,\mu_L) - g(2,\mu_L) \\ &= \mu_L + S_0(N-1) - 4\mu_L - S_0(N-2) = S_0 - 3\mu_L. \end{aligned}$$

Note that $S_0 - 3\mu_L \geq \frac{S_0}{3} - \mu_L \geq 0$ by Assumption 10 and thus a su□cient condition for (B.7) to hold when $\beta = 0$ (and $\beta' = \gamma_H$) is

$$\mu_H \leq S_0 + \delta\gamma_H \left[\frac{S_0}{3} - \mu_L\right],$$

which holds by Assumption 10.

Lemma 12. $x_\beta^{\mathsf{SO}} = 0 \Rightarrow x_{\beta'}^{\mathsf{SO}} \leq 1$.

Proof. A su□cient condition for $x_{\beta'}^{\mathsf{SO}} \leq 1$ is

$$g(1,\mu_{\beta'}) = \mu_{\beta'} + S_0(N-1) < 4\mu_{\beta'} + S_0(N-2) = g(2,\mu_{\beta'}) \Leftrightarrow \mu_{\beta'} > \frac{S_0}{3}.$$

We next show that $\mu_{\beta'} > \frac{\mu_\beta}{3}$. The conclusion then follows since $x_\beta^{\mathsf{SO}} = 0$ implies

$$g(0,\mu_\beta) = S_0 N < \mu_\beta + S_0(N-1) = g(1,\mu_\beta) \Leftrightarrow \mu_\beta > S_0.$$

To show $3\mu_{\beta'} > \mu_\beta$ recall that $\mu_{\beta'} = \beta\mu_L + (1-\beta)\mu_H \geq \beta L + (1-\beta)\mu_H$ and $\mu_H = (1-\gamma_H)H + \gamma_H L \geq \frac{1}{2}H$. Hence

$$3\mu_{\beta'} \geq 3\beta L + 3(1-\beta)\mu_H \geq \frac{3}{2}\beta L + \frac{3}{2}\beta L + \frac{3}{2}(1-\beta)H \geq \frac{3}{2}\mu_\beta > \mu_\beta.$$

∎

B.2.2 Preliminary statements in support of the proof of Proposition 4

To prove our main Proposition 4 we start with some additional statements. We first prove that the agent's state can be simplified as detailed in Lemma 9 in the main text.

Proof of Lemma 9: If all agents are following the scheme $\pi_{c,d}$ then the flow on the risky road at time $t-1$ is distinct depending on whether $\theta_{t-2} = H$ or $\theta_{t-2} = L$. Thus, either an agent was on the risky road at time $t-1$ and observed θ_{t-1} or the agent was on safe and can infer x_{t-1} and (from that) θ_{t-2}. By the Markov property of θ and the stationarity of the recommendation policy no information before θ_{t-2} is useful to the agents. Thus, the state z_t^i is a su□cient summary for any agent to determine his expected ongoing cost, as well as the information that other agents have. If agents do not follow the scheme, then all agents receive full information and this state is still su□cient as every agent and the CP will have symmetric information. ■

From here on we consider the values of c, d fixed (satisfying the assumptions of Proposition 4) and we denote by $u^*(z_t^i)$ the expected cost under $\pi_{c,d}$ of an agent whose state is z_t^i. We note that the expected cost for any agent that knows that the risky road was high at the previous step ($\theta_{t-1} = H$) and before receiving a recommendation for time t is the same, no matter his state. Intuitively, this is true because according to the recommendation scheme $\pi_{c,d}$ if $\theta_{t-1} = H$ then at the next step the CP sends the recommendation r_R to one and only one agent (the experimenter) selected at random among all the agents independent of previous actions or knowledge.

Lemma 13. *The expected cost*

$$\bar{v} := \mathbb{E}_{\pi_{c,d}}[u^*([x_{t-1}, \theta_{t-1}^i, r_{t-1}^i]) \mid x_{t-1} = x, \theta_{t-1}^i = H]$$

is the same for all $x \geq 1$.

Proof. Whenever $\theta_{t-1} = H$ is observed, a new experimenter is chosen among all

Table B.1: Expected stage costs if agents follow the recommendation scheme given in Proposition 4

State of the risky road	H	H	...	L	L	L	...	H
Number of agents on risky	-	1	...	1	c	d	...	d
Expected stage cost on risky	-	μ_H	...	μ_H	$\mu_L c$	$\mu_L d$	...	$\mu_L d$
Expected stage cost on safe	-	S_0	...	S_0	S_0	S_0	...	S_0

agents. Thus every agent, no matter which road he was on, has an identical likelihood of being chosen as the experimenter at time t. Note that $\bar{v}$ conditions on the knowledge that $\theta_{t-1} = H$, hence there is no need for distinguishing states where agents do not know the state of the road. In other words, the expectation is only over the recommendation scheme $\pi_{c,d}$ and thus

$$\begin{aligned}\bar{v} &= \frac{1}{N} \underbrace{u^*([x_{t-1}, H, r_R])}_{\text{ongoing cost as the experimenter}} + \frac{N-1}{N} \underbrace{u^*([x_{t-1}, H, r_S])}_{\text{ongoing cost on safe}} \\ &= \frac{1}{N} u^*([1, H, r_R]) + \frac{N-1}{N} u^*([1, H, r_S]).\end{aligned}$$

In the second line we substitute the observed flow with 1 since it is unimportant (i.e. these costs are the same for any $x_{t-1} \geq 1$); given the state $\theta_{t-1} = H$, the flow in the previous round will not e□ect the ongoing cost: as at time t one agent will be on risky and $N-1$ will be on safe. ∎

Lemma 13 simplifies our analysis because it implies that we can partition the infinite horizon into consecutive periods by defining the beginning of a new period as the time immediately after the risky road switches from L to H, see Table 4.1 in the main text. Conditioned on the agents knowing that a new period has begun, their ongoing cost from that point on is the same (i.e. $\bar{v}$) independent of their history. This observation simplifies the analysis of incentive compatibility and optimality. Table 4.1 in the main text illustrates the flow in the risky road under the social optimum and the scheme described in Proposition 4 within one period. Table B.1 illustrates the expected stage cost for agents taking the risky road or the safe road under the scheme described in Proposition 4.

We start our analysis by deriving closed form expressions and relations for the cost $u^*(z)$ of different states z reached under $\pi_{c,d}$.

Lemma 14 (Closed form expression of auxiliary cost)**.**

1. $u^*([c, L, r_R]) = u^*([d, L, r_R]) = \frac{1}{1-\delta(1-\gamma_L)} (\mu_L d + \delta\gamma_L \bar{v})$
2. $u^*([c, L, r_S]) = u^*([d, L, r_S]) = \frac{1}{1-\delta(1-\gamma_L)}(S_0 + \delta\gamma_L \bar{v})$

Proof. 1. The first equality follows from the fact that the flow on the risky road after $\theta = L$ and the flow c or d is d. The second equality follows from $u^*([d, L, r_R]) = \mu_L d + \delta((1-\gamma_L)u^*([d, L, r_R]) + \gamma_L \bar{v})$ (see (B.10)).

2. The first equality follows similarly to the above point and the second follows from $u^*([d, L, r_S]) = S_0 + \delta((1-\gamma_L)u^*([d, L, r_S]) + \gamma_L \bar{v})$ (see (B.10)).

∎

Lemma 15. $\mu_L c \leq \mu_L d \leq S_0$ *and* $c \geq \frac{S_0}{2\mu_L} - \frac{1}{2}$.

Proof. The first chain of inequalities follows immediately from the assumption $c \leq d \leq x_L^{\text{eq}}$ and $\mu_L x_L^{\text{eq}} \leq S_0$ by the equilibrium condition. Finally, since x_L^{SO} is either the closest integer that to $\frac{S_0}{2\mu_L}$, one gets $c \geq x_L^{\text{SO}} \geq \frac{S_0}{2\mu_L} - \frac{1}{2}$. ∎

Remark 4. *According to Definition 6, if any agent deviates the punishment is full information. Specifically, we assume that the CP sends a recommendation of risky to each agent with probability* $\frac{S_0}{N\mu_\beta}$*, so that the expected cost on risky is exactly equal to the fixed cost* S_0 *of the safe road. The continuation cost after any deviation is therefore*

$$u^*_{dev} = S_0 + \delta S_0 + \delta^2 S_0 + \ldots = \frac{1}{1-\delta} S_0,$$

Independent of the belief β*.*

Lemma 16. *The following statements hold:*

1. $\bar{v} < \frac{1}{1-\delta} S_0$
2. $u^*([d, L, r_R]) \leq \frac{1}{1-\delta} S_0$

3. $u^*(1,L) := \left(\frac{c-1}{N-1}u^*([1,L,r_R]) + \frac{N-c}{N-1}u^*([1,L,r_S])\right) \leq u^*([d,L,r_S]) \leq \frac{1}{1-\delta}S_0$

4. $u^*(c,L) := \left(\frac{d-c}{N-c}u^*([c,L,r_R]) + \frac{N-d}{N-c}u^*([c,L,r_S])\right) \leq u^*([d,L,r_S]) \leq \frac{1}{1-\delta}S_0$

Proof. 1.

$$\bar{v} = \frac{1}{N}\mu_H + \frac{N-1}{N}S_0 + \delta\left(\gamma_H\left(\frac{c}{N}\mu_L c + \frac{N-c}{N}S_0 \right.\right.$$
$$\left.\left. +\delta\left(\gamma_L\bar{v} + (1-\gamma_L)\left(\frac{d}{N}u^*([d,L,r_R]) + \frac{N-d}{N}u^*([d,L,r_S])\right)\right)\right) + (1-\gamma_H)\bar{v}\right)$$

where $u^*([d,L,r_R])$ and $u^*([d,L,r_S])$ are as defined in Lemma 14. To simplify exposition recall that

$$g(1,\mu_H) := \mu_H + (N-1)S_0$$
$$g(c,\mu_L) := \mu_L c^2 + (N-c)S_0$$
$$g(d,\mu_L) := \mu_L d^2 + (N-d)S_0$$

then $\bar{v}$ can be written as

$$\bar{v} = \frac{1-\delta(1-\gamma_L)}{N(1-\delta)(1-\delta(1-\gamma_H-\gamma_L))}\left(g(1,\mu_H) + \delta\gamma_H\left(g(c,\mu_L) + \frac{\delta(1-\gamma_L)}{1-\delta(1-\gamma_L)}g(d,\mu_L)\right)\right) \tag{B.8}$$

To show the inequality holds we first multiply both sides by $(1-\delta)N$

$$\frac{1-\delta(1-\gamma_L)}{1-\delta(1-\gamma_L-\gamma_H)}\left(g(1,\mu_H) + \delta\gamma_H\left(g(c,\mu_L) + \frac{\delta(1-\gamma_L)}{1-\delta(1-\gamma_L)}g(d,\mu_L)\right)\right) < S_0 N$$
$$\iff (1-\delta(1-\gamma_L))(g(1,\mu_H) + \delta\gamma_H g(c,\mu_L)) + \delta^2\gamma_H(1-\gamma_L)g(d,\mu_L)$$
$$< (1-\delta(1-\gamma_L-\gamma_H))S_0 N$$
$$\iff (1-\delta(1-\gamma_L))(\mu_H + S_0(N-1) + \delta\gamma_H g(c,\mu_L)) + \delta^2\gamma_H(1-\gamma_L)g(d,\mu_L)$$
$$< (1-\delta(1-\gamma_L))S_0 N + \delta\gamma_H S_0 N - \delta^2\gamma_H(1-\gamma_L)S_0 N + \delta^2\gamma_H(1-\gamma_L)S_0 N$$
$$\iff (1-\delta(1-\gamma_L))(\mu_H + \delta\gamma_H g(c,\mu_L)) + \delta^2\gamma_H(1-\gamma_L)g(d,\mu_L)$$
$$< (1-\delta(1-\gamma_L))(S_0 + \delta\gamma_H S_0 N) + \delta^2\gamma_H(1-\gamma_L)S_0 N.$$

By Lemma 15, $g(d, \mu_L) \leq S_0 N$, so it is su□cient to show

$$\mu_H + \delta\gamma_H g(c, \mu_L) < S_0 + \delta\gamma_H S_0 N \iff \mu_H < S_0 + \delta\gamma_H (S_0 N - g(c, \mu_L)).$$

Note that, by assumption, $g(c, \mu_L) \leq g(2, \mu_L) < g(1, \mu_L)$ hence

$$S_0 N - g(c, \mu_L) > S_0 N - (S_0(N-1) + \mu_L) = S_0 - \mu_L > \frac{S_0}{3} - \mu_L.$$

Thus, it is su□cient if

$$\mu_H \leq S_0 + \delta\gamma_H \left(\frac{S_0}{3} - \mu_L \right)$$

which holds by Assumption 10.

2. The cost $u^*([d, L, r_R])$ is defined in Lemma 14. Then

$$\begin{aligned} u^*([d, L, r_R]) =& \frac{1}{1 - \delta(1-\gamma_L)} (\mu_L d + \delta\gamma_L \bar{v}) \leq \frac{1}{1-\delta} S_0 \\ \iff & (1-\delta)(\mu_L d + \delta\gamma_L \bar{v}) \leq (1 - \delta(1-\gamma_L)) S_0 \\ \iff & (1-\delta)\mu_L d + (1-\delta)\delta\gamma_L \bar{v} \leq (1-\delta) S_0 + \delta\gamma_L S_0 \end{aligned}$$

and the result holds by part 1 of this lemma and by Lemma 15.

3. Expanding the cost $u^*(1, L)$

$$\begin{aligned} u^*(1, L) = & \frac{c-1}{N-1}\mu_L c + \frac{N-c}{N-1} S_0 \\ & + \delta \left((1-\gamma_L) \left(\frac{d-1}{N-1} u^*([d, L, r_R]) + \frac{N-d}{N-1} u^*([d, L, r_S]) \right) + \gamma_L \bar{v} \right). \end{aligned}$$

Note that $u^*([d, L, r_S]) = S_0 + \delta(1-\gamma_L) u^*([d, L, r_S]) + \delta\gamma_L \bar{v}$. Then $u^*(1, L) \leq u^*([d, L, r_S])$ holds if

$$\frac{c-1}{N-1}\mu_L c + \frac{N-c}{N-1} S_0 + \frac{\delta(1-\gamma_L)}{1-\delta(1-\gamma_L)} \left(\frac{d-1}{N-1}\mu_L d + \frac{N-d}{N-1} S_0 \right) \leq S_0 + \frac{\delta(1-\gamma_L)}{1-\delta(1-\gamma_L)} S_0$$

which is true by Lemma 15.

For the second inequality, we need

$$\frac{1}{1-\delta(1-\gamma_L)}(S_0+\delta\gamma_L\bar{v}) \leq \frac{1}{1-\delta}S_0 \tag{B.9}$$
$$\iff (1-\delta)(S_0+\delta\gamma_L\bar{v}) \leq (1-\delta+\delta\gamma_L)S_0$$

and the result follows from the first point of the lemma.

4. By using Lemma 14, we can expand $u^*(c,L)$ as

$$\begin{aligned} u^*(c,L) &= \frac{d-c}{N-c}\left(\frac{1}{1-\delta(1-\gamma_L)}\left(\mu_L d+\delta\gamma_L\bar{v}\right)\right)+\frac{N-d}{N-c}\left(\frac{1}{1-\delta(1-\gamma_L)}\left(S_0+\delta\gamma_L\bar{v}\right)\right) \\ &= \frac{1}{1-\delta(1-\gamma_L)}\left(\frac{d-c}{N-c}\mu_L d+\frac{N-d}{N-c}S_0+\delta\gamma_L\bar{v}\right), \end{aligned}$$

It follows by Lemma 15, that

$$u^*(c,L) \leq \frac{1}{1-\delta(1-\gamma_L)}(S_0+\delta\gamma_L\bar{v}) = u^*([d,L,r_S])$$

and the result follows from part 3 of this lemma.

∎

B.2.3 Proof of Proposition 4

As in the two stage case, we break the agents into **types based on individual's information.** We detail the possible states agents reach under $\pi_{c,d}$ and show that $\xi_{\pi_{c,d}}$ is an equilibrium by showing that at each state any deviation would lead to a higher expected cost. Each equilibrium constraint in this setting is dynamic (i.e. consists of both stage and continuation cost).

Type 1 ($\beta^i_{t-1}=L$, $r^i_{t-1}=r_R$)

Agents of this type took the risky road at time $t-1$ and observed L. Under $\pi_{c,d}$ these agents receive a recommendation to remain on risky, so $r^i_{t-1}=r_R$ for all cases

below. We further distinguish different states based on the observed flow on risky at time $t-1$. We present our results in decreasing order of number of other agents that took risky at $t-1$. We show that in each case an agent that observes L follows the recommendation and stays on risky.

- $[\mathbf{x_{t-1}}, \boldsymbol{\beta}^{\mathbf{i}}_{\mathbf{t-1}}, \mathbf{r}^{\mathbf{i}}_{\mathbf{t-1}}] = [\mathbf{d}, \mathbf{L}, \mathbf{r_R}]$: this agent is part of what we call the "incentive compatible" flow, which is the flow d. If everybody follows the recommendation then $x_t = d$. Agent i's expected costs under $\pi_{c,d}$ (if he follows or if he deviates) are therefore

  - Following:

$$u^*([d,L,r_R]) = \underbrace{\mu_L d}_{\substack{\text{stage cost} \\ \text{for } t}} + \delta\big((1-\gamma_L)\underbrace{u^*([d,L,r_R])}_{\substack{\text{ongoing cost} \\ \text{if } \theta_t = L}} + \gamma_L \underbrace{\mathbb{E}_{\pi_{c,d}}[u^*([d,H,r^i])]}_{\text{ongoing cost if } \theta_t = H}\big),$$

$$= \mu_L d + \delta\left((1-\gamma_L)\, u^*([d,L,r_R]) + \gamma_L \bar{v}\right), \tag{B.10}$$

  - Deviating to safe:

$$\underbrace{S_0}_{\text{stage cost for } t} + \underbrace{\delta u^*_{dev}}_{\text{ongoing cost}} = S_0 + \frac{\delta}{1-\delta} S_0 \tag{B.11}$$

  Thus, taking the risky road is an equilibrium if

$$\mu_L d + \delta\left((1-\gamma_L)\, u^*([d,L,r_R]) + \gamma_L \bar{v}\right) \leq S_0 + \frac{\delta}{1-\delta} S_0. \tag{B.12}$$

  The inequality holds by Lemmas 15, 16-1, and 16-2.

- $[\mathbf{x_{t-1}}, \boldsymbol{\beta}^{\mathbf{i}}_{\mathbf{t-1}}, \mathbf{r}^{\mathbf{i}}_{\mathbf{t-1}}] = [\mathbf{c}, \mathbf{L}, \mathbf{r_R}]$ this agent is part of the "exploiters" (agents that are sent to the risky road the first period after the experimenter saw L). If everybody follows the recommendation the flow on risky will be $x_t = d$.

  By Lemma 14, the costs are the same as in state $[d, L, r_R]$ (as given in (B.10) and (B.11)). The equilibrium constraint is therefore identical to (B.12) and is satisfied.

- $[\mathbf{x_{t-1}}, \boldsymbol{\beta}^{\mathbf{i}}_{\mathbf{t-1}}, \mathbf{r}^{\mathbf{i}}_{\mathbf{t-1}}] = [\mathbf{1}, \mathbf{L}, \mathbf{r_R}]$ this agent is the current experimenter and just saw the road change to "low". If everybody follows the recommendation the next flow on risky would be $x_t = c$. His expected costs are

  - Following:

$$u^*([1, L, r_R]) = \mu_L c + \delta((1-\gamma_L)u^*([c, L, r_R]) + \gamma_L \bar{v}). \tag{B.13}$$

  - Deviating to safe:

$$S_0 + \delta u^*_{\text{dev}} = S_0 + \frac{\delta}{1-\delta} S_0$$

  The agent follows the recommendation if

$$\mu_L c + \delta((1-\gamma_L)u^*([c, L, r_R]) + \gamma_L \bar{v}) \leq S_0 + \frac{\delta}{1-\delta} S_0.$$

  This inequality holds by Lemmas 14, 15, 16-1, and 16-2.

Type 2 ($\beta^i_{t-1} = H$)

These agents took risky and saw H, this means a new experimenter will be chosen. They each will receive a recommendation of either r_R or r_S. Since the state is H the flow on risky in the next round is $x_t = 1$ no matter the previous flow and we can divide the states by recommendations

- $r^i_{t-1} = r_S$. He is not the experimenter.

  - Following the recommendation and taking safe

$$u^*([-, H, r_S]) = S_0 + \delta(\gamma_H u^*(1, L) + (1-\gamma_H)\bar{v}) \tag{B.14}$$

- Deviating

$$2\mu_H + \frac{\delta}{1-\delta}S_0$$

Thus, the incentive constraint holds by Lemma 16-1, 16-3, and Assumption 10.

- $r_{t-1}^i = r_R$. This agent has been selected to be the experimenter. Recall that if the agent deviates then full information is provided from then on and the cost is S_0 at every round (see Remark 4). The agent's expected costs are

  - Following the recommendation and taking risky

$$u^*([x_{t-1}, H, r_R]) = \mu_H + \delta(\gamma_H u^*([1, L, r_R]) + (1-\gamma_H)\bar{v}) \qquad \text{(B.15)}$$

  - Deviating

$$S_0 + \frac{\delta}{1-\delta}S_0$$

The incentive constraint can be written

$$\mu_H + \delta(\gamma_H u^*([1, L, r_R]) + (1-\gamma_H)\bar{v}) \leq S_0 + \gamma_H \frac{\delta}{1-\delta}S_0 + (1-\gamma_H)\frac{\delta}{1-\delta}S_0$$

and by Lemma 16-1 it su□ces to show

$$\mu_H + \delta\gamma_H u^*([1, L, r_R]) \leq S_0 + \gamma_H \frac{\delta}{1-\delta}S_0.$$

Note that

$$u^*([1, L, r_R]) = \mu_L c + \delta((1-\gamma_L)u^*([c, L, r_R]) + \gamma_L\bar{v}) \leq \mu_L c + \frac{\delta}{1-\delta}S_0$$

where the upper bound follows from Lemma 14-1, 16-1 and 16-2. Thus, it is

su□cient if

$$\begin{aligned}\mu_H &\leq S_0 + \delta\gamma_H(S_0 - \mu_L c) \\ &\leq S_0 + \delta\gamma_H\left(S_0 - \mu_L\left(\frac{S_0}{2\mu_L} - \frac{1}{2}\right)\right) \\ &= S_0 + \delta\gamma_H\frac{1}{2}(S_0 + \mu_L)\end{aligned}$$

where the second inequality follows by Lemma 15. The result holds by Assumption 10.

Type 3 ($\beta^i_{t-1} = U, r^i_{t-1} = r_S$)

These agents took safe at time $t-1$ and received a recommendation to remain on safe, r_S. Note that since these agents took safe at time $t-1$ they do not know θ_{t-1}. Receiving a recommendation of safe either means that $\theta_{t-1} = L$ (and the agent continues to be part of the flow on the safe road) or that $\theta_{t-1} = H$ (a new cycle has begun but the agent is not the new experimenter). For simplicity we denote the probability of the first event by $p_{x,S} = \mathbb{P}(\theta_{t-1} = L \mid x_{t-1} = x, r^i_{t-1} = r_S)$. Note that this probability depends on the flow x observed at $t-1$. The following lemma relates these probabilities for the cases $x_{t-1} = d$ and $x_{t-1} = c$.

Lemma 17. *The following statements hold*

1. $p_{d,S} = \frac{1-\gamma_L}{1-\gamma_L+\gamma_L\frac{N-1}{N}}$
2. $p_{c,S} = \frac{(1-\gamma_L)\frac{N-d}{N-c}}{(1-\gamma_L)\frac{N-d}{N-c}+\gamma_L\frac{N-1}{N}}$
3. $p_{1,S} = \frac{\gamma_H\frac{N-c}{N-1}}{(1-\gamma_H)\frac{N-1}{N}+\gamma_H\frac{N-c}{N-1}}$
4. $p_{c,S} \leq p_{d,S}$

Proof. 1. Note that the agent can infer $\theta_{t-2} = L$ from the flow being d on the risky

road at time $t-1$. Therefore, by Bayes rule

$$p_{d,S} := \mathbb{P}(\theta_{t-1} = L \mid [d, U, r_S]) \tag{B.16}$$
$$= \frac{(1-\gamma_L)\mathbb{P}(r_S \mid \theta_{t-1} = L, x_{t-1} = d)}{(1-\gamma_L)\mathbb{P}(r_S \mid \theta_{t-1} = L, x_{t-1} = d) + \gamma_L \mathbb{P}(r_S \mid \theta_{t-1} = H, x_{t-1} = d)}$$
$$= \frac{1-\gamma_L}{1-\gamma_L+\gamma_L\frac{N-1}{N}}. \tag{B.17}$$

2. Note that the agent can infer $\theta_{t-2} = L$ from the flow being c on the risky road at time $t-1$. By Bayes rule

$$p_{c,S} := \mathbb{P}(\theta_{t-1} = L \mid [c, U, r_S]) \tag{B.18}$$
$$= \frac{(1-\gamma_L)\mathbb{P}(r_S \mid \theta_{t-1} = L, x_{t-1} = c)}{(1-\gamma_L)\mathbb{P}(r_S \mid \theta_{t-1} = L, x_{t-1} = c) + \gamma_L \mathbb{P}(r_S \mid \theta_{t-1} = H, x_{t-1} = c)}$$
$$= \frac{(1-\gamma_L)\frac{N-d}{N-c}}{(1-\gamma_L)\frac{N-d}{N-c}+\gamma_L\frac{N-1}{N}}.$$

3. Note that the agent can infer $\theta_{t-2} = H$ from the flow being 1 on the risky road at time $t-1$. By Bayes rule

$$p_{1,S} := \mathbb{P}(\theta_{t-1} = L \mid [1, U, r_S])$$
$$= \frac{\gamma_H \mathbb{P}(r_S \mid \theta_{t-1} = L, x_{t-1} = 1)}{(1-\gamma_H)\mathbb{P}(r_S \mid \theta_{t-1} = H, x_{t-1} = 1) + \gamma_H \mathbb{P}(r_S \mid \theta_{t-1} = L, x_{t-1} = 1)}$$
$$= \frac{\gamma_H\frac{N-c}{N-1}}{\gamma_H\frac{N-c}{N-1}+(1-\gamma_H)\frac{N-1}{N}}.$$

4. For positive α, β, η

$$\frac{\alpha}{\alpha+\beta} \geq \frac{\eta}{\eta+\beta} \iff \alpha \geq \eta.$$

Let $\alpha := 1-\gamma_L$, $\beta := \gamma_L\frac{N-1}{N}$, and $\eta := (1-\gamma_L)\frac{N-d}{N-c}$. Then $p_{d,S} = \frac{\alpha}{\alpha+\beta}$ and $p_{c,S} = \frac{\eta}{\eta+\beta}$. The fact that $d \geq c$ implies $\alpha \geq \eta$ and therefore $p_{d,S} \geq p_{c,S}$.

∎

The possible states for agents of Type 2 are

- $[\mathbf{x_{t-1}}, \boldsymbol{\beta}^{\mathbf{i}}_{\mathbf{t-1}}, \mathbf{r}^{\mathbf{i}}_{\mathbf{t-1}}] = [\mathbf{d}, \mathbf{U}, \mathbf{r_S}]$: these agents were on safe at time $t-1$, observed

flow d and received a recommendation to remain on safe. Their expected costs are

- Following the recommendation and taking safe

$$u^*([d,U,r_S]) = p_{d,S}\underbrace{u^*([d,L,r_S])}_{\text{cost if } \theta_{t-1}=L} + (1-p_{d,S})\underbrace{u^*([-,H,r_S])}_{\text{cost if } \theta_{t-1}=H}.$$

- Deviating to risky

$$p_{d,S}\mu_L(d+1) + (1-p_{d,S})2\mu_H + \frac{\delta}{1-\delta}S_0$$

This inequality holds by assumption (4.6).

- $[\mathbf{x_{t-1}}, \boldsymbol{\beta}^{\mathbf{i}}_{\mathbf{t-1}}, \mathbf{r}^{\mathbf{i}}_{\mathbf{t-1}}] = [\mathbf{c}, \mathbf{U}, \mathbf{r_S}]$: The agent's expected costs are

- Following the recommendation and taking safe

$$u^*([c,U,r_S]) = p_{c,S}\underbrace{u^*([d,L,r_S])}_{F_1} + (1-p_{c,S})\underbrace{u^*([-,H,r_S])}_{F_2}.$$

- Deviating

$$p_{c,S}\underbrace{\left(\mu_L(d+1) + \frac{\delta}{1-\delta}S_0\right)}_{D_1} + (1-p_{c,S})\underbrace{\left(2\mu_H + \frac{\delta}{1-\delta}S_0\right)}_{D_2}$$

By inspection, this case is similar to the previous case ($[d,U,r_S]$). The only difference is the beliefs ($p_{d,S}$ for the previous case, and $p_{c,S}$ for this case). The incentive compatibility constraint (4.6) for the previous case can be written compactly as

$$p_{d,S}F_1 + (1-p_{d,S})F_2 \leq p_{d,S}D_1 + (1-p_{d,S})D_2$$

or equivalently

$$p_{d,S}(F_1 - D_1) + (1 - p_{d,S})(F_2 - D_2) \leq 0.$$

We next show that $(F_2 - D_2) \leq (F_1 - D_1)$. Since, by Lemma 17 $p_{c,S} \leq p_{d,S}$, by properties of convex combinations, this suffices to show that

$$p_{c,S}(F_1 - D_1) + (1 - p_{c,S})(F_2 - D_2) \leq p_{d,S}(F_1 - D_1) + (1 - p_{d,S})(F_2 - D_2) \leq 0,$$

as desired. To show $(F_2 - D_2) \leq (F_1 - D_1)$ we equivalently show $(F_2 + D_1) \leq (F_1 + D_2)$. Note

$$\begin{aligned} F_2 &= u^*([-, H, r_S]) = S_0 + \delta\left(\gamma_H u^*(1, L) + (1 - \gamma_H)\bar{v}\right) \\ F_1 &= u^*([d, L, r_S]) = S_0 + \delta((1 - \gamma_L)u^*([d, L, r_S]) + \gamma_L \bar{v}). \end{aligned}$$

Hence $(F_2 + D_1) \leq (F_1 + D_2)$ can be rewritten as

$$\begin{aligned} & S_0 + \delta((1 - \gamma_H)\bar{v} + \gamma_H u^*(1, L)) + \mu_L(d + 1) + \frac{\delta}{1 - \delta} S_0 \\ & \leq S_0 + \delta(\gamma_L \bar{v} + (1 - \gamma_L)u^*([d, L, r_S])) + 2\mu_H + \frac{\delta}{1 - \delta} S_0 \\ \iff & \delta(1 - \gamma_L - \gamma_H)\bar{v} + \mu_L(d + 1) \leq \delta((1 - \gamma_L)u^*([d, L, r_S]) - \gamma_H u^*(1, L)) + 2\mu_H. \end{aligned}$$

By Lemma 15 and by Assumption 10, $\mu_L d + \mu_L \leq S_0 + S_0 \leq 2\mu_H$ and it is sufficient to show

$$(1 - \gamma_L - \gamma_H)\bar{v} \leq (1 - \gamma_L)u^*([d, L, r_S]) - \gamma_H u^*(1, L).$$

The right hand side can be lower bounded using Lemma 16-3 by

$$(1 - \gamma_L - \gamma_H)u^*([d, L, r_S])$$

and incentive compatibility holds if

$$\begin{aligned} & \bar{v} \leq u^*([d, L, r_S]) \\ \iff & \bar{v} \leq \frac{1}{1-\delta(1-\gamma_L)}(S_0 + \delta\gamma_L\bar{v}) \\ \iff & (1-\delta)\bar{v} \leq S_0 \end{aligned}$$

which holds by Lemma 16-1.

- $[\mathbf{x_{t-1}}, \boldsymbol{\beta}^{\mathbf{i}}_{\mathbf{t-1}}, \mathbf{r}^{\mathbf{i}}_{\mathbf{t-1}}] = [\mathbf{1}, \mathbf{U}, \mathbf{r_S}]$ These agents were on safe at time $t-1$ and observed $x_{t-1} = 1$, thus they infer that $\theta_{t-2} = H$. The agent's expected costs are

  - Following the recommendation of safe

$$u^*([1, U, r_S]) = p_{1,S}u^*([1, L, r_S]) + (1 - p_{1,S})u^*([-, H, r_S]).$$

  - Deviating

$$p_{1,S}\left(\mu_L(c+1)\right) + (1-p_{1,S})2\mu_H + \frac{\delta}{1-\delta}S_0.$$

The incentive compatibility constraint can thus be written as

$$\begin{aligned} & p_{1,S}\left(S_0 + \delta((1-\gamma_L)u^*(c, L) + \gamma_L\bar{v})\right) + (1-p_{1,S})(S_0 + \delta(\gamma_H u^*(1, L) + (1-\gamma_H)\bar{v})) \\ & \quad \leq p_{1,S}\left(\mu_L(c+1)\right) + (1-p_{1,S})\left(2\mu_H\right) + \frac{\delta}{1-\delta}S_0. \end{aligned}$$

where $u^*(c, L)$ and $u^*(1, L)$ are as defined in Lemma 16. It follows from Lemma 16-1, 16-3, and 16-4 that

$$\delta\left(p_{1,S}((1-\gamma_L)u^*(c, L) + \gamma_L\bar{v}) + (1-p_{1,S})(\gamma_H u^*(1, L) + (1-\gamma_H)\bar{v})\right) \leq \frac{\delta}{1-\delta}S_0.$$

The result is then proven by the following lemma.

Lemma 18. $S_0 \leq p_{1,S}(\mu_L(c+1)) + (1-p_{1,S})(2\mu_H)$.

Proof. Plugging in the expression of $p_{1,S}$ derived in Lemma 17 and rearranging

$$\left((1-\gamma_H)\frac{N-1}{N} + \gamma_H\frac{N-c}{N-1}\right) S_0 \leq \gamma_H\frac{N-c}{N-1}(\mu_L(c+1)) + (1-\gamma_H)\frac{N-1}{N}2\mu_H.$$

Since $c+1 \geq \frac{S_0}{2\mu_L}$ (by Lemma 15) the inequality above holds if

$$\begin{aligned} &\left((1-\gamma_H)\frac{N-1}{N} + \gamma_H\frac{N-c}{N-1}\right) S_0 \leq \gamma_H\frac{N-c}{N-1}\frac{S_0}{2} + (1-\gamma_H)\frac{N-1}{N}2\mu_H \\ \iff &\left((1-\gamma_H)\frac{N-1}{N}\right) S_0 + \gamma_H\frac{N-c}{N-1}\frac{S_0}{2} \leq (1-\gamma_H)\frac{N-1}{N}2\mu_H. \end{aligned}$$

By Assumption 10 $S_0 \leq \mu_H$, so the result holds if

$$\gamma_H\frac{N-c}{N-1}\frac{1}{2} \leq (1-\gamma_H)\frac{N-1}{N}.$$

The last inequality holds since, for all $N \geq 2$,

$$\gamma_H\frac{N-c}{N-1}\frac{1}{2} < \frac{\gamma_H}{2} \leq \frac{(1-\gamma_H)}{2} \leq (1-\gamma_H)\frac{N-1}{N},$$

where we used $c > 1$ and $\gamma_H \leq (1-\gamma_H)$ (since $\gamma_H \leq \frac{1}{2}$). ∎

Type 4 ($\beta_t^i = U, r_t^i = r_R$)

These agents took safe at time $t-1$ and received a recommendation to take risky, r_R. We show that each of these cases is comparable to one that has already been detailed. Thus, following the recommendation is optimal. The possible states of an agent of type 4 are:

- $[\mathbf{x_{t-1}}, \boldsymbol{\beta}^{\mathbf{i}}_{\mathbf{t-1}}, \mathbf{r}^{\mathbf{i}}_{\mathbf{t-1}}] = [\mathbf{d}, \mathbf{U}, \mathbf{r_R}]$: receiving a recommendation of risky means the agent is an experimenter. This is equivalent to the case $[-, H, r_R]$ as this agent can infer that $\theta_{t-1} = H$ since, under $\pi_{c,d}$, the CP does not send a r_R if the flow

is d and $\theta = L$. Specifically, by Bayes rule

$$\mathbb{P}(\theta_{t-1} = L \mid [d, U, r_R])$$
$$= \frac{(1-\gamma_L)\mathbb{P}(r_R \mid \theta_{t-1} = L, x_{t-1} = d)}{(1-\gamma_L)\mathbb{P}(r_R \mid \theta_{t-1} = L, x_{t-1} = d) + \gamma_L \mathbb{P}(r_R \mid \theta_{t-1} = H, x_{t-1} = d)} = 0.$$

Since it is a best response to take risky when the state is $[-, H, r_R]$, it is also a best response to take risky when the state is $[d, U, r_R]$.

- $[\mathbf{x_{t-1}}, \boldsymbol{\beta}^{\mathbf{i}}_{\mathbf{t-1}}, \mathbf{r}^{\mathbf{i}}_{\mathbf{t-1}}] = [\mathbf{c}, \mathbf{U}, \mathbf{r_R}]$: receiving a recommendation of risky could mean an agent is part of d (if $\theta_{t-1} = L$) **or** is the new experimenter (if $\theta_{t-1} = H$). Denote by $p_{c,R}$ agent's i belief that $\theta_{t-1} = L$. His expected cost of following the recommendation is therefore

$$u^*([c, U, r_R]) = p_{c,R} u^*([d, L, r_R]) + (1 - p_{c,R}) u^*([1, H, r_R]).$$

The cost of deviating is the convex combination of deviating under the two possible states to safe. In both of these cases the best response is to take risky (see the states $[d, L, r_R]$ and $[1, H, r_R]$ discussed earlier). Thus, it is a best response for the agent in this state to take risky.

- $[\mathbf{x_{t-1}}, \boldsymbol{\beta}^{\mathbf{i}}_{\mathbf{t-1}}, \mathbf{r}^{\mathbf{i}}_{\mathbf{t-1}}] = [\mathbf{1}, \mathbf{U}, \mathbf{r_R}]$: receiving a recommendation of risky means that the agent is either a part of the exploiter flow c (if $\theta_{t-1} = L$) **or** has been selected to be the next experimenter (if $\theta_{t-1} = H$). Denote by $p_{1,R}$ agent's i belief that $\theta_{t-1} = L$. His expected cost of following the recommendation and taking risky can then be written as

$$u^*([1, U, r_R]) = p_{1,R} u^*([1, L, r_R]) + (1 - p_{1,R}) u^*([1, H, r_R]).$$

Similarly, the cost of deviating is a convex combination of deviating to safe from state $[1, L, r_R]$ and deviating to safe from state $[1, H, r_R]$. In both of these cases the best response is to take risky. Thus, it is a best response for the agent in this state to take risky.

Type 5 (off the equilibrium path)

O the equilibrium path the CP provides full information, that is, sends recommendations of risky to each agent with probability $\frac{S_0}{\mu_\beta N}$. Note that in the full information regime every agent has the same belief β on the state of the risky road (equal to the belief of the CP). Moreover the continuation cost for every agent is the same, no matter his action. Since following the received recommendation is myopically optimal no agent has an incentive to deviate. At each point every agent's total expected cost is $1/(1-\delta)S_0$.

B.2.4 Proof of Corollary 4

The corollary follows from the following lemma.

Lemma 19. *Let $\bar{x}_{LL}$ be the smallest integer d that satisfies* (4.6) *when $c = x_L^{\mathsf{SO}}$, then*

$$\bar{x}_{LL} \leq x_L^{\mathsf{eq}}. \tag{B.19}$$

Proof. Note that

$$\begin{aligned} u^*([d, U, r_S]) &= p_{d,S}\frac{S_0 + \delta\gamma_L\bar{v}}{1-\delta(1-\gamma_L)} + (1-p_{d,S})\left(S_0 + \delta\left(\gamma_H u^*(1,L) + (1-\gamma_H)\bar{v}\right)\right) \\ &= p_{d,S}\left(S_0 + \frac{\delta(S_0(1-\gamma_L)+\gamma_L\bar{v})}{1-\delta(1-\gamma_L)}\right) + (1-p_{d,S})\left(S_0 + \delta\left(\gamma_H u^*(1,L) + (1-\gamma_H)\bar{v}\right)\right). \end{aligned}$$

Hence by Lemma 16

$$u^*([d, U, r_S]) \leq p_{d,S}\left(S_0 + \frac{\delta S_0}{1-\delta}\right) + (1-p_{d,S})\left(S_0 + \frac{\delta S_0}{1-\delta}\right) = S_0 + \frac{\delta S_0}{1-\delta}.$$

A sufficient condition for (4.6) to hold is therefore,

$$S_0 < p_{d,S}\mu_L(d+1) + (1-p_{d,S})2\mu_H.$$

We next show that $d = x_L^{\mathsf{eq}}$ satisfies this condition and thus (4.6). To this end, recall

that $\mu_L(x_L^{\text{eq}}+1) > S_0$ and $\mu_H \geq S_0$ hence

$$p_{d,S}\mu_L(d+1) + (1-p_{d,S})2\mu_H > p_{d,S}S_0 + (1-p_{d,S})S_0 = S_0,$$

as desired. ∎

B.2.5 Proof of Proposition 5

The proposition follows from the following lemma.

Lemma 20. *As* $\delta \to 1$, $x_{LL} \to x_L^{\text{SO}}$.

Proof. Recall x_{LL} is the smallest integer d (weakly greater than x_L^{SO}) that satisfies (4.6) for $c = x_L^{\text{SO}}$. Specifically, plugging in values from Lemma 14, taking the safe road (i.e. following the recommendation) is the best response if

$$\begin{aligned} u^*([d, U, r_S]) &= p_{d,S}\frac{S_0 + \delta\gamma_L\bar{v}}{1-\delta(1-\gamma_L)} + (1-p_{d,S})\left(S_0 + \delta\left(\gamma_H u^*(1,L) + (1-\gamma_H)\bar{v}\right)\right) \\ &\leq p_{d,S}\mu_L(d+1) + (1-p_{d,S})2\mu_H + \frac{\delta}{1-\delta}S_0. \end{aligned} \tag{B.20}$$

We let $c = d = x_L^{\text{SO}}$ in (B.20) and show as $\delta \to 1$ the constraint holds. By Lemma 16.3 and Lemma 14 it holds

$$\begin{aligned} u^*(1,L) &= \frac{x_L^{\text{SO}}-1}{N-1}\left(\frac{1}{1-\delta(1-\gamma_L)}(\mu_L x_L^{\text{SO}} + \delta\gamma_L\bar{v})\right) + \frac{N-x_L^{\text{SO}}}{N-1}\left(\frac{1}{1-\delta(1-\gamma_L)}(S_0 + \delta\gamma_L\bar{v})\right) \\ &= \frac{1}{1-\delta(1-\gamma_L)}\left(\frac{x_L^{\text{SO}}-1}{N-1}\mu_L x_L^{\text{SO}} + \frac{N-x_L^{\text{SO}}}{N-1}S_0\right) + \bar{v}\left(\frac{\delta\gamma_L}{1-\delta(1-\gamma_L)}\right). \end{aligned}$$

Hence (B.20) becomes

$$\underbrace{p_{d,S}\left(\frac{1}{1-\delta(1-\gamma_L)}S_0\right)+(1-p_{d,S})\left(S_0+\frac{\delta\gamma_H}{1-\delta(1-\gamma_L)}\left(\frac{x_L^{\mathsf{SO}}-1}{N-1}\mu_L x_L^{\mathsf{SO}}+\frac{N-x_L^{\mathsf{SO}}}{N-1}S_0\right)\right)}_{T_1}$$
$$+\bar{v}\underbrace{\left(p_{d,S}\frac{\delta\gamma_L}{1-\delta(1-\gamma_L)}+(1-p_{d,S})\left(\frac{\delta^2\gamma_H\gamma_L}{1-\delta(1-\gamma_L)}+\delta(1-\gamma_H)\right)\right)}_{T_2}$$
$$\leq \underbrace{p_{d,S}\left(\mu_L(x_L^{\mathsf{SO}}+1)\right)+(1-p_{d,S})\left(2\mu_H\right)}_{T_3}+\underbrace{\frac{\delta}{1-\delta}S_0}_{T_4}. \tag{B.21}$$

It follows from (B.8) with $g(c,\mu_L)=g(d,\mu_L)=g(x_L^{\mathsf{SO}},\mu_L)$ that

$$\bar{v}=\frac{1}{(1-\delta)}\underbrace{\frac{1-\delta(1-\gamma_L)}{N(1-\delta(1-\gamma_H-\gamma_L)}\left(\mu_H+(N-1)S_0+\frac{\delta\gamma_H}{1-\delta(1-\gamma_L)}\left(\mu_L(x_L^{\mathsf{SO}})^2+(N-x_L^{\mathsf{SO}})S_0\right)\right)}_{T_5}.$$

Substituting in (B.21) and multiplying both sides by $(1-\delta)$ yields

$$T_1(1-\delta)+T_5T_2\leq T_3(1-\delta)+\delta S_0.$$

Since T_1 and T_3 are finite and $\lim_{\delta\to1}T_2=1$, when $\delta\to1$ a su□cient condition for (B.21) to hold is

$$\begin{aligned}\lim_{\delta\to1}T_5 :=& \frac{\gamma_L}{(\gamma_H+\gamma_L)N}\left(\mu_H+(N-1)S_0+\frac{\gamma_H}{\gamma_L}\left(\mu_L(x_L^{\mathsf{SO}})^2+(N-x_L^{\mathsf{SO}})S_0\right)\right)<S_0\\ \iff& \gamma_L\mu_H+\gamma_LS_0(N-1)+\gamma_H(\mu_L(x_L^{\mathsf{SO}})^2+S_0(N-x_L^{\mathsf{SO}}))<(\gamma_H+\gamma_L)S_0N\\ \iff& \mu_H<S_0+\frac{\gamma_H}{\gamma_L}(-\mu_L(x_L^{\mathsf{SO}})^2+S_0x_L^{\mathsf{SO}})\\ &=S_0+\frac{\gamma_H}{\gamma_L}x_L^{\mathsf{SO}}(S_0-\mu_Lx_L^{\mathsf{SO}}).\end{aligned}$$

The inequality holds by Assumption 10 if

$$\frac{\gamma_H}{\gamma_L}x_L^{\mathsf{SO}}(S_0-\mu_Lx_L^{\mathsf{SO}})>\gamma_H\left(\frac{S_0}{3}-\mu_L\right)$$

this follows from $1/\gamma_L > 1$ and

$$x_L^{\mathsf{SO}}(S_0 - \mu_L x_L^{\mathsf{SO}}) \geq x_L^{\mathsf{SO}}\left(S_0 - \frac{S_0}{2} - \frac{\mu_L}{2}\right) = \frac{x_L^{\mathsf{SO}}}{2}(S_0 - \mu_L) \geq S_0 - \mu_L \geq \frac{S_0}{3} - \mu_L$$

where the first inequality follows from $x_L^{\mathsf{SO}} \leq \frac{S_0}{2\mu_L} + \frac{1}{2}$ (can be proven similarly as in Lemma 14) and the second inequality follows from $x_L^{\mathsf{SO}} \geq 2$. ∎

B.2.6 Proof of Proposition 6

We have $\pi^* = \pi_{1,1,x_L^{\mathrm{SO}},x_{LL}}$ and $V^* = V_{1,1,x_L^{\mathrm{SO}},x_{LL}}$. We denote an arbitrary optimal incentive compatible scheme (OICS) in $\hat{\Pi}$ as $\bar{\pi} := \pi_{\bar{a},\bar{b},\bar{c},\bar{d}}$ with social cost $\bar{V}$. We proceed in steps.

1. **Proof of Lemma 10: Any OICS $\bar{\pi}$ must satisfy $\bar{a}, \bar{b} \leq 1$**
   For any recommendation in $\hat{\Pi}$ the experimenters, i.e. the agents chosen to take risky at time t when $\theta_{t-1} = H$ are selected at random from all agents. Thus, with positive probability the agent or agents chosen as part of a or b know that $\theta_{t-1} = H$. We next show that if a or b is greater than one, then it is not incentive compatible for an agent to follow this recommendation when $\delta \leq 1/2$. Specifically, following gives cost of at least

   $$2\mu_H + \delta((1-\gamma_H)\hat{v}_L + \gamma_H \hat{v}_H)$$

   where $\hat{v}_L$ and $\hat{v}_H$ are some positive continuation costs. Deviating to safe has cost

   $$\frac{1}{1-\delta} S_0 \leq 2S_0 \quad \text{for} \quad \delta \leq 1/2.$$

   Since $\mu_H \geq S_0$ and the continuation costs $\hat{v}_L$ and $\hat{v}_H$ are positive it is not incentive compatible for the agent to follow and take risky when $a, b \geq 2$. Thus, for incentive compatibility to hold at most one agent can be sent to risky.

2. **Any OICS satisfies $\bar{c} = x_L^{\mathbf{SO}}$ or $\bar{c} = x_L^{\mathbf{SO}} + 1$.** In this proof we assume $x_{LL} > x_L^{\mathsf{SO}}$ (if not π^* coincides with the social optimum and is thus already an

OICS) hence x_{LL} is the minimum integer satisfying (B.20) for $c = x_L^{\mathsf{SO}}$. The incentive compatibility constraint (B.20) can be equivalently rewritten as

$$\begin{aligned}
&p_{d,S}u^*([d,L,r_S]) + (1-p_{d,S})\left(S_0 + \delta\left((1-\gamma_H)\bar{v} + \gamma_H\left(\frac{c-1}{N-1}\mu_L c + \frac{N-c}{N-1}S_0\right.\right.\right.\\
&\qquad\left.\left.\left. +\delta\left(\gamma_L\bar{v} + (1-\gamma_L)\left(\frac{d-1}{N-1}u^*([d,L,r_R]) + \frac{N-d}{N-1}u^*([d,L,r_S])\right)\right)\right)\right)\\
&\leq p_{d,S}\mu_L(d+1) + (1-p_{d,S})2\mu_H + \frac{\delta}{1-\delta}S_0.
\end{aligned} \tag{B.22}$$

Recall from Lemma 17 that

$$p_{d,S} = \frac{1-\gamma_L}{1-\gamma_L+\gamma_L\frac{N-1}{N}} \tag{B.23}$$

does not depend on d, hence within this proof to avoid confusion we denote this by p_S. By substituting the expressions of $u^*([d,L,r_S])$ and $u^*([d,L,r_R])$ computed in Lemma 14 and the expression of $\bar{v}$ given in (B.8) the LHS of the IC constraint (B.22) can be rewritten as

$$\begin{aligned}
&p_S u^*([d,L,r_S]) + (1-p_S)\left(S_0 + \delta\left((1-\gamma_H)\bar{v} + \gamma_H\left(\frac{c-1}{N-1}\mu_L c + \frac{N-c}{N-1}S_0\right.\right.\right.\\
&\qquad\left.\left.\left. +\delta\left(\gamma_L\bar{v} + \frac{(1-\gamma_L)}{1-\delta(1-\gamma_L)}\left(\frac{d-1}{N-1}\mu_L d + \frac{N-d}{N-1}S_0 + \delta\gamma_L\bar{v}\right)\right)\right)\right)\\
&= p_S\frac{1}{1-\delta(1-\gamma_L)}\left(S_0 + \delta\gamma_L\bar{v}\right) + (1-p_S)\left(S_0 + \delta\left((1-\gamma_H)\bar{v} + \gamma_H\left(\frac{c-1}{N-1}\mu_L c\right.\right.\right.\\
&\qquad\left.\left.\left. +\frac{N-c}{N-1}S_0 + \frac{\delta}{1-\delta(1-\gamma_L)}\left(\gamma_L\bar{v} + (1-\gamma_L)\left(\frac{d-1}{N-1}\mu_L d + \frac{N-d}{N-1}S_0\right)\right)\right)\right)\right)\\
&= \frac{p_S S_0}{1-\delta(1-\gamma_L)} + (1-p_S)\left(S_0 + \delta\gamma_H\left(\frac{c-1}{N-1}\mu_L c + \frac{N-c}{N-1}S_0\right.\right.\\
&\qquad\left.\left. +\frac{\delta(1-\gamma_L)}{1-\delta(1-\gamma_L)}\left(\frac{d-1}{N-1}\mu_L d + \frac{N-d}{N-1}S_0\right)\right)\right)\\
&\qquad + \bar{v}\left(\underbrace{p_S\frac{\delta\gamma_L}{1-\delta(1-\gamma_L)} + (1-p_S)\delta\left((1-\gamma_H) + \gamma_H\frac{\delta\gamma_L}{1-\delta(1-\gamma_L)}\right)}_{\tau}\right)
\end{aligned}$$

$$= \frac{p_S S_0}{1-\delta(1-\gamma_L)} + (1-p_S)\left(S_0 + \delta\gamma_H\left(\frac{c-1}{N-1}\mu_L c + \frac{N-c}{N-1}S_0 + \right.\right.$$
$$\left.\left.\frac{\delta(1-\gamma_L)}{1-\delta(1-\gamma_L)}\left(\frac{d-1}{N-1}\mu_L d + \frac{N-d}{N-1}S_0\right)\right)\right)$$
$$+ \frac{\tau\tilde{\tau}}{N}\left[g(1,\mu_H) + \delta\gamma_H g(c,\mu_L) + \delta^2\frac{(1-\gamma_L)}{1-\delta(1-\gamma_L)}\gamma_H g(d,\mu_L)\right],$$

for

$$\begin{aligned}\tilde{\tau} :=& \frac{1-\delta(1-\gamma_L)}{(1-\delta)(1-\delta(1-\gamma_H-\gamma_L))} \\ \tau :=& p_S\frac{\delta\gamma_L}{1-\delta(1-\gamma_L)} + (1-p_S)\delta\left((1-\gamma_H) + \gamma_H\frac{\delta\gamma_L}{1-\delta(1-\gamma_L)}\right)\end{aligned} \tag{B.24}$$

independent of a, b, c, d. Note that this is separable in c and d. Specifically, by bringing all the terms depending on c on the LHS and all the terms depending on d on the RHS, we can rewrite the IC constraint (B.22) as $f(c) \leq g(d)$ with

$$\begin{aligned}f(c) :=& \frac{\delta(1-p_S)\gamma_H}{N-1}\underbrace{((c-1)\mu_L c + (N-c)S_0)}_{=:f_1(c)} + \frac{\tau\tilde{\tau}}{N}\delta\gamma_H\underbrace{g(c,\mu_L)}_{=:f_2(c)} + k \\ g(d) :=& p_S\mu_L(d+1) - \delta(1-p_S)\gamma_H\frac{\delta(1-\gamma_L)}{1-\delta(1-\gamma_L)}\left(\frac{d-1}{N-1}\mu_L d + \frac{N-d}{N-1}S_0\right) \\ & - \frac{\tau\tilde{\tau}}{N}\delta^2\frac{(1-\gamma_L)}{1-\delta(1-\gamma_L)}\gamma_H g(d,\mu_L) \\ k :=& \frac{p_S S_0}{1-\delta(1-\gamma_L)} + (1-p_S)S_0 + \frac{\tau\tilde{\tau}}{N}g(1,\mu_H) - (1-p_S)2\mu_H - \frac{\delta}{1-\delta}S_0.\end{aligned} \tag{B.25}$$

Note that k is a constant independent of c and d. The functions $f_1(c)$ and $f_2(c)$ are quadratic, convex and, disregarding the integer constraint, are minimized when $c = \frac{\mu_L+S_0}{2\mu_L} = \frac{S_0}{2\mu_L} + \frac{1}{2}$ and when $c = \frac{S_0}{2\mu_L}$, respectively. This means that, disregarding the integer constraint, $f(c)$ is minimized for some $\tilde{c}^* \in \left[\frac{S_0}{2\mu_L}, \frac{S_0}{2\mu_L} + \frac{1}{2}\right]$. Let c^* be the integer minimizer of $f(c)$ (i.e. the integer closest to $\tilde{c}^*$). Recall that x_L^{SO} is the integer minimizer of $g(c,\mu_L)$ (i.e. the integer closest to $\frac{S_0}{2\mu_L}$).

There are two possible cases:

(a) *if there exists an integer $\hat{c}$ in the interval $\left[\frac{S_0}{2\mu_L}, \frac{S_0}{2\mu_L} + \frac{1}{2}\right]$ then $c^* = \hat{c} = x_L^{\mathsf{SO}}$.* In fact, $|\hat{c} - \frac{S_0}{2\mu_L}| \leq \frac{1}{2} \Rightarrow \hat{c} = x_L^{\mathsf{SO}}$ and $|\hat{c} - \tilde{c}^*| \leq \frac{1}{2} \Rightarrow \hat{c} = c^*$;

(b) *if there is no integer in the interval* $\left[\frac{S_0}{2\mu_L}, \frac{S_0}{2\mu_L} + \frac{1}{2}\right]$ *then either* $c^* = x_L^{\text{SO}}$ *or* $c^* = x_L^{\text{SO}} + 1$.

In fact, let $\hat{c}^-$ be the largest integer smaller than $\frac{S_0}{2\mu_L}$ and $\hat{c}^+$ be the smallest integer larger than $\frac{S_0}{2\mu_L} + \frac{1}{2}$ (i.e. $\hat{c}^- + 1$). Since there is no integer in $\left[\frac{S_0}{2\mu_L}, \frac{S_0}{2\mu_L} + \frac{1}{2}\right]$ it must be $|\hat{c}^- - \frac{S_0}{2\mu_L}| \leq \frac{1}{2} \Rightarrow \hat{c}^- = x_L^{\text{SO}}$. On the other hand c^* is equal to either $\hat{c}^-$ or $\hat{c}^+$ depending on whether $\tilde{c}^*$ is smaller or larger than $\frac{\hat{c}^- + \hat{c}^+}{2}$.

Thus, $f(c)$ is minimized at $c^* = x_L^{\text{SO}}$ or $c^* = x_L^{\text{SO}} + 1$ and for any $\tilde{c} \neq \{x_L^{\text{SO}}, x_L^{\text{SO}} + 1\}$, we get $f(\tilde{c}) \geq f(x_L^{\text{SO}})$.

We now prove that if $(\tilde{c}, \tilde{d})$ satisfies (B.20) with $c \neq \{x_L^{\text{SO}}, x_L^{\text{SO}} + 1\}$ then $\tilde{d} \geq x_{LL}$. In fact it must be $g(\tilde{d}) \geq f(\tilde{c}) \geq f(x_L^{\text{SO}})$. Since x_{LL} is the minimum integer satisfying $g(d) \geq f(x_L^{\text{SO}})$ it must be $\tilde{d} \geq x_{LL}$.

Hence any IC scheme with $\tilde{c} \neq \{x_L^{\text{SO}}, x_L^{\text{SO}} + 1\}$ has higher cost then π^* (since $\tilde{c} \geq x_L^{\text{SO}}$ and $\tilde{d} \geq x_{LL}$) and cannot be optimal. (Recall that by (B.8) the cost is $\tilde{\tau}\left[g(1, \mu_H) + \delta\gamma_H g(c, \mu_L) + \delta^2 \frac{(1-\gamma_L)}{1-\delta(1-\gamma_L)} \gamma_H g(d, \mu_L)\right]$ and $g(c, \mu_L)/g(d, \mu_L)$ are minimized for $c = x_L^{\text{SO}}/d = x_L^{\text{SO}}$ and strictly increasing functions for larger values of c/d.)

3. **Any OICS satisfies** $\bar{d} \geq x^{LL} - 1$

From the previous point we know, that in any OICS it must be $\bar{c} \in \{x_L^{\text{SO}}, x_L^{\text{SO}} + 1\}$.

- If $\bar{c} = x_L^{\text{SO}}$ then by definition x_{LL} is the minimum value of $\bar{d}$ to maintain incentive compatibility.
- If instead $\bar{c} = x_L^{\text{SO}} + 1$ we show that to maintain incentive compatibility it must be $\bar{d} \geq x_{LL} - 1$. In fact suppose by contradiction that $f(x_L^{\text{SO}} + 1) \leq g(x_{LL} - 2)$, then we show that under our assumptions $f(x_L^{\text{SO}}) \leq g(x_{LL} - 1)$, which is absurd since x_{LL} is the minimum integer satisfying the IC

constraint. To this end it su□ces to show

$$f(x_L^{\mathsf{SO}}) - f(x_L^{\mathsf{SO}} + 1) \leq g(x_{LL} - 1) - g(x_{LL} - 2).$$

Note that

$$\begin{aligned}
& f(x_L^{\mathsf{SO}}) - f(x_L^{\mathsf{SO}} + 1) \leq g(x_{LL} - 1) - g(x_{LL} - 2) \\
\iff & \frac{\delta(1-p_S)\gamma_H}{N-1}\left(S_0 - 2x_L^{\mathsf{SO}}\mu_L + \frac{\delta(1-\gamma_L)}{1-\delta(1-\gamma_L)}(S_0 - 2(x_{LL}-2)\mu_L)\right) \\
& + \frac{\tau\tilde{\tau}}{N}\delta\gamma_H\left(S_0 - (2x_L^{\mathsf{SO}}+1)\mu_L + \frac{\delta(1-\gamma_L)}{1-\delta(1-\gamma_L)}(S_0 - (2(x_{LL}-2)+1)\mu_L)\right) \\
& \leq p_S\mu_L.
\end{aligned} \tag{B.26}$$

Now, by $x_{LL} \geq x_L^{\mathsf{SO}} + 1$ (Recall that if $x_{LL} = x_L^{\mathsf{SO}}$ then π^* is equivalent to the social optimum and it is therefore the OICS.) and $x_L^{\mathsf{SO}} \geq \frac{S_0}{2\mu_L} - \frac{1}{2}$ (see Lemma 15) the following two inequalities hold

$$\begin{aligned}
S_0 - (2(x_{LL}-2)+1)\mu_L \leq S_0 - 2(x_{LL}-2)\mu_L &\leq S_0 - 2\mu_L(x_L^{\mathsf{SO}} - 1) \\
&\leq S_0 - S_0 + \mu_L + 2\mu_L = 3\mu_L \\
S_0 - (2x_L^{\mathsf{SO}}+1)\mu_L \leq S_0 - 2\mu_L x_L^{\mathsf{SO}} &\leq \mu_L
\end{aligned}$$

and by $\delta \leq 1/2$

$$\frac{\delta(1-\gamma_L)}{1-\delta(1-\gamma_L)} \leq 1.$$

Thus, the left hand side of (B.26) can be upper bounded by

$$4\mu_L\left(\frac{\delta(1-p_S)\gamma_H}{N-1} + \frac{\tau\tilde{\tau}}{N}\delta\gamma_H\right).$$

So it is su□cient if

$$
\begin{aligned}
& 4\left(\frac{\delta(1-p_S)\gamma_H}{N-1}+\frac{\tau\tilde{\tau}}{N}\delta\gamma_H\right)\leq p_S \\
\iff & 4\delta\gamma_H\left((1-p_S)+\frac{N-1}{N}\tau\tilde{\tau}\right)\leq (N-1)p_S \\
\overset{\text{(B.24)}}{\iff} & 4\delta\gamma_H\left((1-p_S)+\frac{N-1}{N}\tilde{\tau}\left(p_S\frac{\delta\gamma_L}{1-\delta(1-\gamma_L)}+\right.\right. \\
& \qquad \left.\left.(1-p_S)\delta\left((1-\gamma_H)+\gamma_H\frac{\delta\gamma_L}{1-\delta(1-\gamma_L)}\right)\right)\right)\leq (N-1)p_S \\
\overset{\text{(B.23)}}{\iff} & 4\delta\gamma_H\left(\gamma_L\frac{N-1}{N}+\frac{N-1}{N}\tilde{\tau}\left(\frac{\delta\gamma_L(1-\gamma_L)}{1-\delta(1-\gamma_L)}\right.\right. \\
& \qquad \left.\left.+\gamma_L\frac{N-1}{N}\delta\left((1-\gamma_H)+\gamma_H\frac{\delta\gamma_L}{1-\delta(1-\gamma_L)}\right)\right)\right)\leq (N-1)(1-\gamma_L) \\
\iff & 4\delta\gamma_H\left(\gamma_L+\tilde{\tau}\underbrace{\left(\frac{\delta\gamma_L(1-\gamma_L)}{1-\delta(1-\gamma_L)}+\gamma_L\frac{N-1}{N}\delta\left((1-\gamma_H)+\gamma_H\frac{\delta\gamma_L}{1-\delta(1-\gamma_L)}\right)\right)}_{\hat{\tau}}\right) \\
& \qquad \leq (1-\gamma_L)N \\
\iff & 4\delta\gamma_H\left(\gamma_L+\tilde{\tau}\hat{\tau}\right)\leq (1-\gamma_L)N. \qquad \text{(B.27)}
\end{aligned}
$$

where the second equivalence comes from plugging in τ and the third from plugging in p_S and multiplying by $\left(1-\gamma_L+\gamma_L\frac{N-1}{N}\right)$.

Now, we show $\tilde{\tau}\leq 2$. By (B.24)

$$
\begin{aligned}
\tilde{\tau} &= \frac{1-\delta(1-\gamma_L)}{(1-\delta)(1-\delta(1-\gamma_H-\gamma_L))}=\frac{1-\delta(1-\gamma_L)}{(1-\delta(1-\gamma_H))(1-\delta(1-\gamma_L))-\delta^2\gamma_H\gamma_L}\leq 2 \\
&\iff 1-\delta(1-\gamma_L)\leq 2((1-\delta(1-\gamma_H))(1-\delta(1-\gamma_L))-\delta^2\gamma_H\gamma_L) \\
&\iff 1\leq 2\left((1-\delta(1-\gamma_H))-\frac{\delta^2\gamma_H\gamma_L}{1-\delta(1-\gamma_L)}\right) \\
&\iff 1\leq 2\left(1-\delta+\delta\gamma_H\left(1-\frac{\delta\gamma_L}{1-\delta(1-\gamma_L)}\right)\right).
\end{aligned}
$$

By $\frac{\delta\gamma_L}{1-\delta(1-\gamma_L)}\leq 1$ the right hand side can be lower bounded by $2(1-\delta)$.

Then by $\delta \leq 1/2$ the inequality holds. Now we bound $\hat{\tau}$ by 1

$$\begin{aligned}
&\frac{\delta\gamma_L(1-\gamma_L)}{1-\delta(1-\gamma_L)} + \gamma_L \frac{N-1}{N}\delta\left((1-\gamma_H) + \gamma_H \frac{\delta\gamma_L}{1-\delta(1-\gamma_L)}\right) \\
&= \frac{\delta\gamma_L(1-\gamma_L) + \delta\gamma_L\frac{N-1}{N}((1-\gamma_H)(1-\delta(1-\gamma_L)) + \delta\gamma_H\gamma_L)}{1-\delta(1-\gamma_L)} \\
&\leq \frac{\delta(1-\gamma_L)+\delta\gamma_L}{1-\delta(1-\gamma_L)} = \frac{\delta}{1-\delta(1-\gamma_L)} \leq 1.
\end{aligned}$$

Plugging in the bounds for $\tilde{\tau}$ and $\hat{\tau}$ in (B.27) leads to the sufficient condition

$$4\delta\gamma_H(\gamma_L+2) \leq (1-\gamma_L)N.$$

Then by $\delta \leq 1/2, \gamma_H \leq 1/2, \gamma_L \leq 1/2$ it is sufficient if $N \geq 5$, which is true by assumption.

We are now ready to prove the two main statements:

Proof of statement 2: Overall, we know that π^* is incentive compatible and achieves minimum cost among the IC schemes with $\bar{c} = x_L^{\mathsf{SO}}$. From the points above, we know that the only other possibility is $\bar{c} = x_L^{\mathsf{SO}} + 1$ in which case $\bar{d} \geq x_{LL} - 1$. Any choice of $\bar{d} \geq x_{LL}$ leads to higher cost than V^* hence the only possibility left is $\tilde{\pi}^*$. If $\tilde{\pi}^*$ is IC and has cost $\tilde{V}^*$ lower than V^* then that is the OICS otherwise π^* is. Whether that happens or not depends on the chosen parameters.

Proof of statement 1: We next show that for $\delta \to 0$, $\tilde{V}^* > V^*$. To this end, recall the expression for $\bar{v}$ in (B.8) and $\tilde{\tau}$ in (B.24). Then

$$\begin{aligned}
V^* &:= \tilde{\tau}\left[g(1,\mu_H) + \delta\gamma_H g(x_L^{\mathsf{SO}},\mu_L) + \delta^2\frac{(1-\gamma_L)}{1-\delta(1-\gamma_L)}\gamma_H g(x_{LL},\mu_L)\right], \\
\tilde{V}^* &:= \tilde{\tau}\left[g(1,\mu_H) + \delta\gamma_H g(x_L^{\mathsf{SO}}+1,\mu_L) + \delta^2\frac{(1-\gamma_L)}{1-\delta(1-\gamma_L)}\gamma_H g(x_{LL}-1,\mu_L)\right],
\end{aligned}$$

For $\delta \to 0$ the terms in δ^2 are negligible and the conclusion follows by $g(x_L^{\mathsf{SO}},\mu_L) < g(x_L^{\mathsf{SO}}+1,\mu_L)$.

B.3 Monopoly

If Platform 0 does not bundle the products and sells them separately then the profit is given by

$$\pi_0^{UB} := \frac{1}{4t_x}V_x^2 + \frac{1}{4t_y}V_y^2.$$

If instead, the platform chooses to bundle, there is a more complicated optimization problem

If a platform sets the price p_0 then any agent (x, y) such that

$$V_x + V_y - t_x x - t_y y - p_0 \geq 0 \tag{B.28}$$

will buy the bundle. This gives us any agent on the line

$$y(x) = \frac{1}{t_y}(V_x + V_y - t_x x - p_0) \tag{B.29}$$

is indi□erent between purchasing and not purchasing. Thus, anyone below this line will purchase and anyone above this line will not purchase. Now, there are several cases of demand based on where the two values $y(0)$ and $y(1)$ fall. Specifically, the demand could be any of the four forms shown in Figure B-1.

For ease we define $\bar{x}$ and $\hat{x}$ as follows:

$$\bar{x} : y(\bar{x}) = 0; \quad \bar{x} = \frac{V_x + V_y - p_0}{t_x} \tag{B.30}$$

$$\hat{x} : y(\hat{x}) = 1; \quad \hat{x} = \frac{V_x + V_y - t_y - p_0}{t_x}. \tag{B.31}$$

In Case 1 we have the maximization problem

$$\pi_1^{B2} := \max_{p_0, y(0)\in[0,1], y(1)\in[0,1]} p_0 \frac{1}{2}\left(y(0) + y(1)\right) \tag{B.32}$$

$$= \max_{p_0, y(0)\in[0,1], y(1)\in[0,1]} p_0 \frac{1}{2}(2V_x^0 + 2V_y^0 - t_x - 2p_0). \tag{B.33}$$

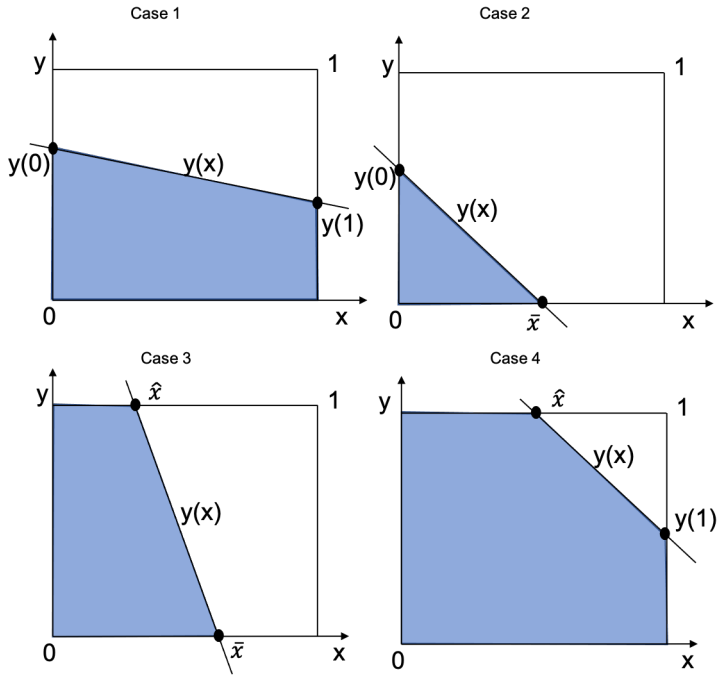


Figure B-1: The four possible cases of demand in the monopoly setting.

Note, the interior optimal p is $p = \frac{1}{4}(2(V_x^0 + V_y^0) - t_x)$ and the constraints hold if the parameters are such that

$$y(0) = \frac{1}{4t_y}(2(V_x + V_y) + t_x) \in [0, 1] \tag{B.34}$$

$$y(1) = \frac{1}{2t_y}(V_x + V_y) \in [0, 1] \tag{B.35}$$

The revenue is then given by

$$\pi_0^{B1} = \frac{1}{16t_y}(2(V_x + V_y) - t_x)^2. \tag{B.36}$$

In Case 2 we have the maximization problem

$$\pi_0^{B2} := \max_{p_0, y(0) \in [0,1], y(1) \leq 0} p_0 \frac{1}{2} y(0) \bar{x} \tag{B.37}$$

$$= \max_{p_0, y(0) \in [0,1], y(1) \leq 0} p_0 \frac{1}{2t_x t_y}(V_x + V_y - p)^2. \tag{B.38}$$

The optimal revenue in this case is given by

$$\pi_0^{B2} = \frac{t_x(V_x + V_y - t_x)}{2t_y} \tag{B.39}$$

In Case 3 we have the maximization problem

$$\pi_0^{B3} := \max_{p_0, y(0)\geq 1, y(1)\leq 0} p_0 \frac{1}{2}(\bar{x} + \hat{x}) \tag{B.40}$$

$$= p_0 \frac{1}{2t_x}(2V_x + 2V_y - t_y - 2p_0). \tag{B.41}$$

In Case 4 we then have the maximization problem

$$\pi_0^{B4} := \max_{p_0, y(0)\geq 1, y(1)\in[0,1]} p_0 \left(1 - \frac{1}{2}(1 - \hat{x})(1 - y(1))\right) \tag{B.42}$$

$$= \max_{p_0, y(0)\geq 1, y(1)\in[0,1]} p_0 \left(1 - \frac{1}{2t_x t_y}(t_x + t_y - V_x - V_y + p_0)^2\right). \tag{B.43}$$

Thus, the overall optimization problem for Platform 0 in the monopoly setting is

$$\pi_0^B := \max\{\pi_0^{B1}, \pi_0^{B2}, \pi_0^{B3}, \pi_0^{B4}\}. \tag{B.44}$$

Proof of Proposition 7. If $\frac{V_x}{2t_x} \leq 1$ and $\frac{V_y}{2t_y} \leq 1$, which holds under Assumption 11 then the profit for unbundling is given by

$$\pi_0^{UB} = \frac{1}{4t_x}V_x^2 + \frac{1}{4t_y}V_y^2 \tag{B.45}$$

For Case 1 of bundling we get the profit for the platform is

$$\pi_0^B = \frac{1}{16t_y}(2(V_x + V_y) - t_x)^2. \tag{B.46}$$

Note that the two constraints that must hold to be in Case 1 are that

$$y(0) = \frac{1}{4t_y}(t_x + 2(V_x + V_y)) \in [0, 1] \tag{B.47}$$

$$y(1) = \frac{1}{2t_y}(V_x + V_y) \in [0, 1] \tag{B.48}$$

which both hold if $t_y \leq \frac{V_x+V_y}{2}$.

If, instead $t_y > \frac{V_x+V_y}{2}$, we can consider Case 2 and a similar argument follows.

If in Case 1, to have $\pi_0^{UB} > \pi_0^B$ it must be the case that

$$\frac{1}{4t_x}V_x^2 + \frac{1}{4t_y}V_y^2 < \frac{1}{16t_y}(2(V_x + V_y) - t_x)^2 \tag{B.49}$$

which is equivalent to

$$0 < 4\left(1 - \frac{t_y}{t_x}\right)(V_x)^2 + 8V_xV_y - 4t_x(V_x + V_y) + t_x^2. \tag{B.50}$$

The right hand side can be lower bounded using $t_x \in \left[\frac{V_x^0}{2}, V_x^0\right]$

$$4\left(1 - \frac{t_y}{t_x}\right)(V_x)^2 + 8V_xV_y - 4V_x(V_x + V_y) + \frac{(V_x)^2}{2} \tag{B.51}$$

which simplifies to

$$4\left(1 - \frac{t_y}{t_x}\right)(V_x)^2 + 4V_x(V_y - V_x) + \frac{(V_x)^2}{2} \tag{B.52}$$

There exist positive α and β such that this is positive. ∎

B.4 Duopoly

Note in the unbundling case this just becomes two separate standard Hotelling problems.

- Unbundling profit:

$$\pi_0^{UB} = \frac{1}{2}t_x + \frac{1}{2}t_y$$
$$\pi_1^{UB} = \frac{1}{2}t_x + \frac{1}{2}t_y$$

- Bundling profit:

$$\pi_0^{B} = \frac{1}{2}t_y$$
$$\pi_1^{B} = \frac{1}{2}t_y$$

Proof of Proposition 8. We have $\pi_0^{UB} > \pi_0^{B}$ always holds, as $t_x > 0$. ∎

Bibliography

[1] Daron Acemoglu, Ali Makhdoumi, Azarakhsh Malekian, and Asuman Ozdaglar. Informational braess paradox: The e□ect of information on tra□c congestion. *Operations Research*, 66(4):893–917, 2018.

[2] Daron Acemoglu, Azarakhsh Malekian, and Asu Ozdaglar. Network security and contagion. *Journal of Economic Theory*, 166:536–585, 2016.

[3] Daron Acemoglu and Asuman Ozdaglar. Competition and e□ciency in congested markets. *Mathematics of operations research*, 32(1):1–31, 2007.

[4] Daron Acemoglu, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. Networks, shocks, and systemic risk. In Yann Bramoullé, Andrea Galeotti, and Brian Rogers, editors, *The Oxford Handbook of the Economics of Networks*. Oxford University Press, 2015.

[5] William James Adams and Janet L Yellen. Commodity bundling and the burden of monopoly. *The quarterly journal of economics*, pages 475–498, 1976.

[6] Nizar Allouch. On the private provision of public goods on networks. *Journal of Economic Theory*, 157:527–552, 2015.

[7] Eitan Altman, Thomas Boulogne, Rachid El-Azouzi, Tania Jiménez, and Laura Wynter. A survey on networking games in telecommunications. *Computers & Operations Research*, 33(2):286–311, 2006.

[8] Itai Arieli. Transfer implementation in congestion games. *Dynamic Games and Applications*, 5(2):228–238, 2015.

[9] Mark Armstrong. A more general theory of commodity bundling. *Journal of Economic Theory*, 148(2):448–472, 2013.

[10] Yannis Bakos and Erik Brynjolfsson. Bundling and competition on the internet. *Marketing science*, 19(1):63–82, 2000.

[11] Coralio Ballester, Antoni Calvó-Armengol, and Yves Zenou. Who's who in networks. Wanted: The key player. *Econometrica*, 74(5):1403–1417, 2006.

[12] Martin Beckmann, Charles B. McGuire, and Christopher B. Winsten. Studies in the Economics of Transportation. Technical report, 1956.

[13] Dirk Bergemann and Stephen Morris. Information design, bayesian persuasion, and bayes correlated equilibrium. *American Economic Review*, 106(5):586–91, 2016.

[14] Gian-Italo Bischi, Lucia Sbragia, and Ferenc Szidarovszky. Learning the demand function in a repeated cournot oligopoly game. *International Journal of Systems Science*, 39(4):403–419, 2008.

[15] Ilai Bistritz and Achilleas Anastasopoulos. Characterizing non-myopic information cascades in bayesian learning. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 2716–2721. IEEE, 2018.

[16] Avrim Blum, Eyal Even-Dar, and Katrina Ligett. Routing without regret: On convergence to Nash equilibria of regret-minimizing algorithms in routing games. In *Proceedings of the 25th Annual ACM Symposium on Principles of Distributed Computing*, pages 45–52. ACM, 2006.

[17] Patrick Bolton and Christopher Harris. Strategic experimentation. *Econometrica*, 67(2):349–374, 1999.

[18] Yann Bramoullé and Rachel Kranton. Public goods in networks. *Journal of Economic Theory*, 135(1):478–494, 2007.

[19] Yann Bramoullé and Rachel Kranton. Games played on networks. In Yann Bramoullé, Andrea Galeotti, and Brian Rogers, editors, *The Oxford Handbook of the Economics of Networks*, chapter 5. Oxford University Press, Oxford, 2015.

[20] Yann Bramoullé, Rachel Kranton, and Martin D'Amours. Strategic interaction and networks. *The American Economic Review*, 104(3):898–930, 2014.

[21] Jean-Philippe Chancelier, Michel De Lara, and Andre De Palma. Risk aversion, road choice, and the one-armed bandit problem. *Transportation Science*, 41(1):1–14, 2007.

[22] Yeon-Koo Che and Johannes Hörner. Optimal design for social learning. 2015.

[23] Yeon-Koo Che and Johannes Hörner. Recommender systems as incentives for social learning. Technical report, Working paper, 2017.

[24] Richard Cole, Yevgeniy Dodis, and Tim Roughgarden. Pricing network edges for heterogeneous selfish users. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 521–530, 2003.

[25] Richard Cole, Yevgeniy Dodis, and Tim Roughgarden. How much can taxes help selfish routing? *Journal of Computer and System Sciences*, 72(3):444–467, 2006.

[26] Roberto Cominetti. Equilibrium routing under uncertainty. *Mathematical Programming*, 151(1):117, 2015.

[27] Giacomo Como, Ketan Savla, Daron Acemoglu, Munther A. Dahleh, and Emilio Frazzoli. Stability analysis of transportation networks with multiscale driver decisions. *SIAM Journal on Control and Optimization*, 51(1):230–252, 2013.

[28] José R. Correa, Andreas S. Schulz, and Nicolás E. Stier-Moses. A geometric approach to the price of anarchy in nonatomic congestion games. *Games and Economic Behavior*, 64(2):457–469, 2008.

[29] Stella Dafermos and Anna Nagurney. Sensitivity analysis for the asymmetric network equilibrium problem. *Mathematical Programming*, 28(2):174–184, 1984.

[30] Sanmay Das, Emir Kamenica, and Renee Mirka. Reducing congestion through information design. In *Communication, Control, and Computing (Allerton), 2017 55th Annual Allerton Conference on*, pages 1279–1284. IEEE, 2017.

[31] Francisco Facchinei and Jong-Shi Pang. *Finite-dimensional variational inequalities and complementarity problems*. Springer Science & Business Media, 2007.

[32] Bryce L Ferguson, Philip N Brown, and Jason R Marden. Carrots or sticks? the e□ectiveness of subsidies and tolls in congestion games. In *2020 American Control Conference (ACC)*, pages 1853–1858. IEEE, 2020.

[33] Simon Fischer, Harald Räcke, and Berthold Vöcking. Fast convergence to Wardrop equilibria by adaptive sampling methods. *SIAM Journal on Computing*, 39(8):3700–3735, 2010.

[34] Simon Fischer and Berthold Vöcking. On the evolution of selfish routing. In *ESA*, volume 4, pages 323–334. Springer, 2004.

[35] Lisa Fleischer, Kamal Jain, and Mohammad Mahdian. Tolls for heterogeneous selfish users in multicommodity networks and generalized congestion games. In *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 277–285. IEEE, 2004.

[36] Drew Fudenberg and Jean Tirole. *Game Theory*. MIT Press, 1991.

[37] John C Gittins. Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(2):148–164, 1979.

[38] Alain Haurie and Patrice Marcotte. On the relationship between Nash-Cournot and Wardrop equilibria. *Networks*, 15(3):295–308, 1985.

[39] Johannes Hörner and Andrzej Skrzypacz. Learning, experimentation and information design. In *Survey Prepared for the 2015 econometric summer meetings in Montreal*, 2016.

[40] Harold Hotelling. Stability in competition. In *The collected economics articles of Harold Hotelling*, pages 50–63. Springer, 1990.

[41] Krishnamurthy Iyer, Ramesh Johari, and Mukund Sundararajan. Mean field equilibria of dynamic auctions with learning. *Management Science*, 60(12):2949–2970, 2014.

[42] Matthew O Jackson and Yves Zenou. Games on networks. In P Young and S Zamir, editors, *Handbook of game theory*, volume 4. 2014.

[43] Hao Jiang, Uday V Shanbhag, and Sean P Meyn. Learning equilibria in constrained nash-cournot games with misspecified demand functions. In *Decision and Control and European Control Conference (Conference on Decision and Control-ECC), 2011 50th IEEE Conference on*, pages 1018–1023. IEEE, 2011.

[44] Hao Jiang, Uday V Shanbhag, and Sean P Meyn. Distributed computation of equilibria in misspecified convex stochastic nash games. *IEEE Transactions on Automatic Control*, 63(2):360–371, 2018.

[45] Ramesh Johari, Vijay Kamble, and Yash Kanoria. Matching while learning. *arXiv preprint arXiv:1603.04549*, 2016.

[46] Adam Kalai and Santosh Vempala. E□cient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71(3):291–307, 2005.

[47] Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101(6):2590–2615, 2011.

[48] Godfrey Keller, Sven Rady, and Martin Cripps. Strategic experimentation with exponential bandits. *Econometrica*, 73(1):39–68, 2005.

[49] Ilan Kremer, Yishay Mansour, and Motty Perry. Implementing the "wisdom of the crowd". *Journal of Political Economy*, 122(5):988–1012, 2014.

[50] Walid Krichene, Benjamin Drighes, and Alexandre Bayen. On the convergence of no-regret learning in selfish routing. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 163–171, 2014.

[51] Lichun Li, Olivier Massicot, and Cedric Langbort. Sequential public signaling in routing games with feedback information. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 2735–2740. IEEE, 2018.

[52] Yunpeng Li, Costas A Courcoubetis, and Lingjie Duan. Dynamic routing for social information sharing. *IEEE Journal on Selected Areas in Communications*, 35(3):571–585, 2017.

[53] Xiaogang Lin, Yong-Wu Zhou, Wei Xie, Yuanguang Zhong, and Bin Cao. Pricing and product-bundling strategies for e-commerce platforms with competition. *European Journal of Operational Research*, 283(3):1026–1039, 2020.

[54] Je□rey Liu, Saurabh Amin, and Galina Schwartz. E□ects of information heterogeneity in bayesian routing games. *arXiv preprint arXiv:1603.08853*, 2016.

[55] Mohammad Hassan Lotfi, Richard J La, and Nuno C Martins. Bayesian congestion game with tra□c manager: Binary signal case. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 327–333. IEEE, 2018.

[56] Yishay Mansour, Aleksandrs Slivkins, Vasilis Syrgkanis, and Zhiwei Steven Wu. Bayesian exploration: Incentivizing exploration in bayesian games. *arXiv preprint arXiv:1602.07570*, 2016.

[57] Jason R. Marden, Gürdal Arslan, and Je□ S. Shamma. Joint strategy fictitious play with inertia for potential games. *IEEE Transactions on Automatic Control*, 54(2):208–220, 2009.

[58] R Preston McAfee, John McMillan, and Michael D Whinston. Multiproduct monopoly, commodity bundling, and correlation of values. *The Quarterly Journal of Economics*, 104(2):371–383, 1989.

[59] Emily Meigs, Francesca Parise, and Asuman Ozdaglar. Learning dynamics in stochastic routing games. In *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 259–266. IEEE, 2017.

[60] Ishai Menache and Asuman Ozdaglar. Network games: Theory, models, and dynamics. *Synthesis Lectures on Communication Networks*, 4(1):75, 2011.

[61] Domenico Menicucci, Sjaak Hurkens, and Doh-Shin Jeon. On the optimality of pure bundling for a monopolist. *Journal of Mathematical Economics*, 60:33–42, 2015.

[62] Dov Monderer and Lloyd S. Shapley. Potential games. *Games and Economic Behavior*, 14(1):124–143, 1996.

[63] Anna Nagurney. *Network economics: A variational inequality approach*, volume 10. Springer Science & Business Media, 2013.

[64] Evdokia Nikolova and Nicolas E. Stier-Moses. *Stochastic selfish routing*, pages 314–325. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[65] Markos Papageorgiou, Christina Diakaki, Vaya Dinopoulou, Apostolos Kotsialos, and Yibing Wang. Review of road tra□c control strategies. *Proceedings of the IEEE*, 91(12):2043–2067, 2003.

[66] Yiangos Papanastasiou, Kostas Bimpikis, and Nicos Savva. Crowdsourcing exploration. *Management Science*, 64(4):1727–1746, 2017.

[67] Lillian J Ratli□, Roy Dong, Shreyas Sekar, and Tanner Fiez. A perspective on incentive design: Challenges and opportunities. *Annual Review of Control, Robotics, and Autonomous Systems*, 2:305–338, 2019.

[68] Virag Shah, Jose Blanchet, and Ramesh Johari. Bandit learning with positive externalities. In *Advances in Neural Information Processing Systems*, pages 4918–4928, 2018.

[69] Albert N. Shiryaev. *Probability, volume 95 of Graduate Texts in Mathematics*. Springer-Verlag, New York„ 1996.

[70] Hamidreza Tavafoghi, Akhil Shetty, Kameshwar Poola, and Pravin Variaiya. Sequential public signaling in routing games with feedback information. 2019.

[71] Hamidreza Tavafoghi and Demosthenis Teneketzis. Informational incentives for congestion games. In *Communication, Control, and Computing (Allerton), 2017 55th Annual Allerton Conference on*, pages 1285–1292. IEEE, 2017.

[72] John B. Taylor. Asymptotic properties of multiperiod control rules in the linear regression model. *International Economic Review*, 15(2):472–484, 1974.

[73] John Glen Wardrop. Some theoretical aspects of road tra□c research. *Proceedings of the Institution of Civil Engineers*, 1(3):325–362, 1952.

[74] Sunil Wattal, Rahul Telang, and Tridas Mukhopadhyay. Information personalization in a two-dimensional product di□erentiation model. *Journal of Management Information Systems*, 26(2):69–95, 2009.

[75] Manxi Wu and Saurabh Amin. Information design for regulating tra□c flows under uncertain network state. In *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 671–678. IEEE, 2019.

[76] Yixian Zhu and Ketan Savla. On routing drivers through persuasion in the long run. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 4091–4096. IEEE, 2019.