ABSTRACT

Title of dissertation: ESSAYS IN MARKET DESIGN

Hector A Lopez Carbajal, Doctor of Philosophy, 2016

Dissertation directed by: Professor Peter Cramton

Department of Economics

In this dissertation I study three problems in market design: the allocation of resources to schools using deferred acceptance algorithms, the demand reduction of employees on centralized labor markets, and the alleviation of traffic congestion. I show how institutional and behavioral considerations specific to each problem can alleviate several practical limitations faced by current solutions. For the case of traffic congestion, I show experimentally that the proposed solution is effective.

In Chapter 1, I investigate how school districts could assign resources to schools when it is desirable to provide stable assignments. An assignment is stable if there is no student currently assigned to a school that would prefer to be assigned to a different school that would admit him if it had the resources. Current assignment algorithms assume resources are fixed. I show how simple modifications to these algorithms produce stable allocations of resources and students to schools.

In Chapter 2, I show how the negotiation of salaries within centralized labor markets using deferred acceptance algorithms eliminates the incentives of the hiring firms to strategically reduce their demand. It is well-known that it is impossible to eliminate these incentives for the hiring firms in markets without negotiation of salaries.

Chapter 3 investigates how to achieve an efficient distribution of traffic congestion on a road network. Traffic congestion is the product of an externality: drivers do

not consider the cost they impose on other drivers by entering a road. In theory, Pigouvian prices would solve the problem. In practice, however, these prices face two important limitations: i) the information required to calculate these prices is unavailable to policy makers and ii) these prices would effectively be new taxes that would transfer resources from the public to the government. I show how to construct congestion prices that retrieve the required information from the drivers and do not transfer resources to the government. I circumvent the limitations of Pigouvian prices by assuming that individuals make some mistakes when selecting routes and have a tendency towards truth-telling. Both assumptions are very robust observations in experimental economics.

ESSAYS IN MARKET DESIGN

by

Hector A Lopez Carbajal

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2016

Advisory Committee:
Professor Peter Cramton, Chair
Professor Erkut Ozbay
Professor Daniel Vincent
Professor Soohyung Lee
Professor Robert Windle

Dedication

I dedicate this work to my wife, Rocio Teresa Garcia Balcazar, and my son, Joaquin Eduardo Lopez Balcazar. You, my family, have been my strength and motivation every day of my life. I love you.

Acknowledgments

I am grateful to many people. I am most grateful to my parents — Arturo Lopez Joses and Maria Cruz Carbajal Garcia — who have supported me in every way and in every moment of my life. They are an incredible example of perseverance and love. I was also blessed with my sister, Helga Maricruz Lopez Carbajal, who made my early years full of laughs and happiness.

I am grateful to my wife, Rocio Teresa Garcia Balcazar, who left family, friends, church and country to pursue this dream of mine. This endeavor was infinitely easier with her on my side.

I am grateful to my advisor and committee. I thank Erkut Ozbay who continuously encouraged me to address important questions. I specially thank my advisor, Peter Cramton, who opened the door to countless opportunities to grow as an economist and whose ideas deeply shaped my view of economics as a science and as a practical tool for policy making. I thank Soohyung Lee for her support as a professor and as a colleague. I must say it is a pleasure to work with her. I thank Daniel Vincent for the many conversations we had in all these years. His careful eye only made my theory better.

I am grateful to all my professors. I thank Levent Ülkü who recommended my admission to the University of Maryland; without his kind words I would not be here. I thank John Shea and the departments of economics for their support.

Many other people made this journey very enjoyable. I enjoyed many nice hours and conversations with Özlem, Kristian, Naomi, Nona, Jingting, Ricardo, Marisol, Camilo, Alfredo, Santiago, Mateo, Daisuke and Wakana, Thomas and Fred, Enrique and Sara, Andrés and Vico, Matías and Vale, Jorge and Claudia, Rafael and Juliana, Juan and Javi. I also enjoyed many hours of play with Andy G and Jonathan Song. I shall specially mention Norma and Byron Silvera, whose kindness was warmly felt by my family and many others.

Contents

List of Tables

List of Figures

Chapter 1: Resources and Constraints in Matching Markets

Abstract. *A matching model is developed in which several matching markets can be analyzed simultaneously as a single matching market with constraints. This model is capable of dealing with complex allocation problems such as deciding funding in school choice problems or regional caps in residents matching. In this model a stable allocation must resist deviations within markets and across markets. We offer a strategy-proof-for-doctors mechanism capable of simultaneously selecting a market and an allocation in that market. The allocation selected is shown to be stable (within market and across markets) and efficient.*

In the United States, public school systems are increasingly adopting a new funding strategy where funding "follows the child" to the school he or she attends. A weighted student formula (WSF) at the school district level is its most common manifestation. Under WSF resources are allocated to students, rather than to schools and programs, based on their specific needs. The WSF ensures more funding is allocated to students with more expensive educational needs. Specific weights are tailored to meet student needs in every district. In Boston, for example, weights are distributed across eight categories: grade, poverty, disabilities, continued disabilities, interrupted formal education, English learners, risk students and vocational students. WSF's are a modern solution to a difficult allocation problem in schooling: funding to schools and programs must depend not only the number, but the identity and needs of every student. The WSF strategy is a reminder that all school districts, using WSF's or not, face two complex allocation decisions: resources to schools and students to schools.

The second allocation problem has been studied successfully in recent years. Following the seminal paper of Abdulkadirouglu and Sönmez (2003), a lot of attention has been devoted to several aspects of the admission process; including manipulability,[1] efficiency[2] and diversity.[3] These efforts have provided school districts across the nation with tools to improve their admission systems. For example, the NY and Boston public school systems now use a version of the deferred acceptance algorithm proposed in the literature.[4] However, in all previous models, it is assumed that resources assigned to schools are fixed, leaving the first allocation problem unattended. Current designs are efficient only when there is absolutely no freedom in allocating resources to schools.

In this paper, we make progress towards a unified solution to both allocation problems when the stability and efficiency of the final allocation are important. In order to gain insight about the new economic phenomena present in this extended model, consider the prototypical school choice problem. In a particular school district, there are several school programs; some of them may share buildings, laboratories, dining rooms and buses, but most importantly all of them share the same financial budget. There is a common pool of applicants. However, not all applicants have the same needs. For instance, some may require special facilities or instructors. Two competing procedures suggest themselves as possible solutions to these

[1] Abdulkadiroğlu et al. (2011); Abdulkadiroglu et al. (2009); Ergin and Sönmez (2006); Hatfield and Kojima (2009); Kojima and Pathak (2009); Pathak and Sönmez (2008); Sönmez (1997)

[2] Aygün and Sönmez (2013); Balinski and Sönmez (1999); Dur et al. (2013); Erdil and Ergin (2008); Ergin (2002)

[3] Abdulkadiroğlu (2005); Kominers and Sönmez (2013)

[4] See for example, Abdulkadiroğlu et al. (2005); Abdulkadiroglu et al. (2006); Abdulkadirouglu et al. (2005)

allocation problems.

First, one could allocate and fix resources to schools in advance, i.e. before matching students to schools, in order to determine the number and kind of seats in every school. Denote as market f a market with fixed resources. Assume some stable allocation is chosen in market f. Hence in market f, the final allocation is individually rational and no blocking pair exists. However, when potentially transferable resources are introduced to the model, the stability of the chosen allocation is artificial. It is possible that a student would like to move to a school different from the one he or she has been assigned and that school would like to admit the student, but currently has no resources to do so. Furthermore, the student is using resources in the assigned school and those resources are transferable, in some degree, to the other school. If resources are moved from one school to the other then, call this new market g. In market g, the chosen allocation would not be stable. One possible solution is to sequentially satisfy all would-be blocking pairs, however, it is not a suitable solutions since it is well know that even in fixed-resources environments it leads to problems.[5] Furthermore, it is not clear whether there is a market and an allocation in that market capable of eliminating all blocking pairs within markets and across markets. We extend the concept of stability to allow for inter-market blocking pairs and show that, under suitable conditions, a stable allocation exists. In principle, it would be possible to calculate a stable allocation in every market and then compare them all to choose one that is stable across markets. Such procedure is computationally very expensive. Moreover, it is highly unlikely for such a procedure to have nice incentive properties. Hence we focus our attention on finding a stable allocation and its market simultaneously.

Second, suppose the allocation of resources across schools is not fixed , and every school decides independently, subject to a matching process, the number and iden-

[5] See Roth and Sotomayor (1992)

tity of admitted students. In this admission process a new economic force would be in play: the admission of one student by a school directly affects the admission capacity of all other schools by reducing the common pool of resources, i.e. there are externalities in the resources dimension. Hence some coordination between schools is required in the admission process. We develop a new algorithm capable of solving the externalities produced in the admission process using property rights as the coordination device. Suppose all feasible allocations in this market are denoted f,g, etc. Every feasible allocation implicitly assigns resources to every school. Suppose f assigns students $d_1, ..., d_5$ to school h, then it is possible to assign school h with property rights over five seats and inform all other schools to admit students consistent with h having five seats. We show that the assignment of property rights can be done simultaneously with the matching process and show that, under suitable conditions, it produces a stable and strategy-proof mechanism.

As hinted by the terminology used above, studying several markets and a single market with constraints is essentially the same. Consider for example a collection of markets f, g, h,... and then produce a single market with constraints where f, g, h,... represent the constraints that must be satisfied. The process can be done in the opposite direction. We exploit this insight and the ideas from both approaches outlined above to produce a unified framework and offer a mechanism that is stable (within and across markets), efficient and strategy-proof for students.

Our design is based on the model of matching with contracts of Hatfield and Milgrom (2005). There are three essential components: i) a set of feasible sets of contracts (a set of markets), ii) a central authority called the matchmaker and iii) the relevant concept of stability under constraints (stability within and across markets). We follow the conventional terminology of referring to market participants as "doctors" and "hospitals".

When constraints are present, the (tentative) admission decisions of one hospital affect other hospitals through the constraints binding them together. This externality-like effect can only be efficiently resolved by the matchmaker. Even in the simplest economy with externalities there is a need for a government to restore efficiency, usually achieved using: taxes, quotas and property rights. The first two require the government to know the efficient level of activity in the market, while the third does not. We introduce the matchmaker and provide it with a way to establish “property rights” over feasible allocations in order to achieve efficiency. The matchmaker will guarantee both the feasibility and the efficiency of the final match using property rights. Matchmakers exist in real matching markets with constraints (markets where there is flexibility in the allocation of resources). Consider the school choice example above: an authority actually exists and it has been making decisions that allocate property rights over resources. We include this important real market participant in the model.

The third piece of our construction is the relevant concept of stability. In a model with no constraints a stable allocation is an individually rational set of contracts with no blocking sets, i.e. no hospital and group of doctors is willing to reject their current contracts and implement contracts by themselves. We extend this definition and deem a set of contracts stable if it is i) conditionally stable (given property rights) and ii) non-extendable in property rights (no resources are wasted). Our definition is a natural extension of stability without constraints and reduces to it as constraints are removed. Other models with constraints have developed their own notions of stability with constraints but they usually do not reduce to the standard definition as constraints are removed.[6]

Models of matching with constrains have been developed since Abdulkadirouglu and Sönmez (2003) introduced the school choice problem with affirmative action. They modified the standard Gale-Shapley algorithm to allow for maximum quan-

[6] See for example, Kamada and Kojima (2013) .

tities of certain types of students. Kamada and Kojima (2013) recently developed a model of matching to find efficient matching of medical graduates to residency programs in Japan. In their model, there are regional constraints limiting the maximum number of doctors hirable in certain regions. They introduced a matchmaker to assign property rights, which in their model specify the maximum number of residents a hospital can hire. Furthermore, they showed that the matchmaker preferences have to satisfy a substitutability condition for a stable allocation to exist and to achieve strategy-proofness for doctors. Our model extends these models in two important ways: i) we allow for complex constraints and ii) we extend the domain of admissible preferences from responsive to substitutable preferences. In order to consider complex constraints we introduce the concept of **property rights** (contracts from which the doctor dimension has been eliminated) and extend the domain of hospitals choice functions to include not only a set of contracts, but also a set of property rights. We introduce the **contract replacement property** into the model to manage substitutions of property rights. When the contract replacement property is satisfied, hospitals are only willing to substitute contracts with the same property rights, substitution of contracts with different property rights is possible, but must be induced by constraints and not the unconstrained preferences. Finally, we introduce the **consistency of choice property** to extend admissible preference from responsive to substitutable preferences. When a hospital satisfies the consistency of choice property, it never rejects a contract that was previously held because its property rights have been increased.

In a recent paper, Kominers and Sönmez (2013) studied a model of matching for diversity with the objective of achieving a certain distribution of students, effectively introducing constraints and externalities in the way seats can choose students (seats being the relevant decision units since each seat could have different preferences). However, in their model, externalities not only affect resources endowed to seats; they affect the identity of admissible students. They dealt with this problem by directly constructing schools choice functions, a suitable solution in their environment, but not in a market where hospitals represent agents with preferences. However, the mechanics of both models are similar. In both models, there is no stable allocation that is preferred by all members of one side of the market. Hence, an exogenous device is introduced to the model. In their case, a doctors precedence and in our case a matchmaker. In our language, precedence grants property rights to doctors while the matchmaker grants property rights to hospitals. Both devices make decisions whenever agents' preferences do not agree.

1.1 An example

Suppose a school district wants to assign 4 students d_1, d_2, d_3, d_4 to 2 schools h_1, h_2. The school district has some flexibility regarding the funding it provides to each school. Schools produce seats with the funding their receive. For simplicity, suppose the school district can choose between $(q_1, q_2) = (2, 0)$ and $(q_1, q_2) = (1, 2)$, where q_i is the maximum number of students h_i can admit. This could happen if it is more expensive to produce a seat in h_1 than in h_2 because of differentiated investments already in place. For example, h_1 could need to build a classroom for the second seat whereas other seats only require to cover variable costs.

Suppose this district uses the Gale-Shapley algorithm to allocate its students and preferences (schools' priorities) are as follows:

$$\begin{array}{ll} & d_1 : h_2, h_1 \\ h_1 : d_3, d_4 & d_2 : h_2, h_1 \\ h_2 : d_1, d_2, d_3, d_4 & d_3 : h_1, h_2 \\ & d_4 : h_2 \end{array}$$

The unique stable matching in each scenario, $(q_1, q_2) = (2, 0)$ and $(q_1, q_2) = (1, 2)$, are as follow:

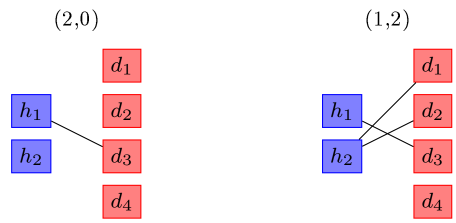


Figure 1.1: Unique stable matches

Notice that $(q_1, q_2) = (1, 2)$ Pareto dominates $(q_1, q_2) = (2, 0)$ both when both schools and students' preference matter (as if this example represented a labor market) and when only students's preferences matter (as in a school choice problem). Suppose the school district cares about efficiency and would like to select a Pareto efficient stable matching. In this example one-sided and two-sided Pareto efficiency coincide for simplicity. How can the school district decide which of the two assignments to select?

Suppose the school district decides to fund schools according to their tentative admission. If schools are free to admit as many students they want, the first round of the student-proposing Gale-Shapley algorithm will result in the following tentative admissions.

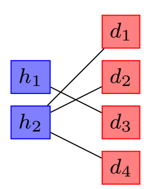


Figure 1.2: First tentative assignment

Notice that the number of seats required in each school to support this assignment is $(q_1, q_2) = (1, 3)$. At this stage, when some preferences have been revealed, it seems reasonable to select $(q_1, q_2) = (1, 2)$ rather than $(q_1, q_2) = (2, 0)$ since setting $(1, 2)$ would produce less rejections (hence keeping more students in their first choice) than $(0, 2)$. Suppose the school district does so and lets the Gale-Shapley algorithm to continue. Then the final assignment would be the optimal one.

This paper extends the previous ideas by generalizing the keeping-more-students-in-their-first-choice idea in the following way. Consider the set of tentative admissions $\{(h_1, d_3), (h_2, d_1), (h_2, d_2), (h_2, d_4)\}$ and drop the doctor dimension to produce the following multiset[7] $\{(h_1), (h_2), (h_2), (h_2)\}$. Notice that the assignment $(q_1, q_2) = (1, 2)$ can be represented by $\{(h_1), (h_2), (h_2)\}$ and $(q_1, q_2) = (2, 0)$ by $\{(h_1), (h_1)\}$. Notice that $\{(h_1), (h_2), (h_2)\}$ is a sub-multiset of the multiset of tentative assignments while $\{(h_1), (h_1)\}$ is not. Thus $\{(h_1), (h_2), (h_2)\}$ should be chosen. If assignments (or tentative assignments) had specific transaction characteristics, they could be represented by a set of contracts of the form $\{(h_1, d_3, s_{13}), (h_2, d_1, s_{21}), (h_2, d_2, s_{22}), (h_2, d_4, s_{24})\}$, where s_{ij} represent the transactional characteristics of the particular match in question. In this case, removing the doctor dimension would produce a multiset of the form $\{(h_1, s_{13}), (h_2, s_{21}), (h_2, s_{22}), (h_2, s_{24})\}$. In this paper, elements of the form (h_i, s) are called property rights.

1.2 Model

Matching models have been successfully used in many applications where transfers and prices cannot be used to signal market participants the relative value and cost of different allocations. Classical examples are matching doctors to hospitals for residency programs, students to colleges and workers to firms.

1.2.1 Feasible Sets of Contracts

Throughout the paper, capital letters will represent both sets and their cardinality. If Y is a set and w is an element, we denote $Y \cup \{w\}$ by Yw. The set of doctors is denoted by D and the set of hospitals by H with typical elements d and h, respec-

[7] A multiset is a list of elements in which repetition of elements is possible, but order does not matter.

tively. We assume there is a finite set of contracts $\mathcal{X}$. A contract $x \in \mathcal{X}$ specifies all payoff relevant conditions between a doctor and a hospital. In general, if S is the set of conditions that can be described in the market, every element of S could specify a salary, a workload, a series of tasks, etc, then $\mathcal{X} \subset D \times H \times S$. Throughout the paper we do not exploit the product structure of the set of contracts and we only use the fact that every contract specifies a doctor and a hospital. It will be assumed that doctors can sign at most one contract and hospitals can sign multiple contracts, but at most one with a particular doctor.

It is usually assumed that any contract in $\mathcal{X}$ can be signed independently (the only restriction being that doctors can sign at most one contract), i.e. if contract x_1 names doctor d_1 and hospital h_1 they can sign it if both agree on its terms, regardless of the contracts signed between say hospital h_2 and doctor d_2. However, when there are constraints, the signature of a contract by a hospital h_1 may prevent the signature of other contracts naming *different* hospitals or doctors. As mentioned in the introduction, when constraints are present there is an externality-like effect when agents sign contracts. For example, consider a Mathematics department and Computer Science department that share a building, classrooms and budget. Without a pre-allocation of resources, the admission decisions made by the math department directly affect the ability to admit students of the computer science department. In the resident matching environment, if two hospitals share a common regional cap, a hospital can prevent other from signing contracts that might violate the cap.

A set of contracts $f \subset \mathcal{X}$ is feasible if all contracts in f can be signed simultaneously. Let $\mathcal{F}$ be the set of all such sets. It is assumed that if a particular set of contracts is feasible, then all of its subsets are also feasible, i.e. $f \in \mathcal{F}$ and $g \subset f$ implies $g \in \mathcal{F}$. It will also be assumed that no contract is trivial $\mathcal{X} = \bigcup_{\mathcal{F}} f$. Both conditions are very weak, the first requires free disposal of contracts and the second only establishes that the model should consider contracts that could be feasible in at least one instance. For notational convenience $\mathcal{F}$ will be described only by its maximal members. In addition, we impose the following condition called doctor independence. Let $f \in \mathcal{F}$ and let g be a set of contracts identical to f where the doctors names have been changed and no doctor is named more than once. $\mathcal{F}$ is doctor independent if and only if $f \in \mathcal{F}$ and $g \subset \mathcal{X}$ implies $g \in \mathcal{F}$. Doctor independence precludes restrictions from excluding certain doctors from matchings allowed to some other doctors. The following examples show how feasibility constraints are implemented in the model.

Example 1.1. Consider a market with two hospitals h_1 and h_2 and two doctors d_1 and d_2 with the constraint that only one contract can be signed. Then set of contracts would be:

$$\mathcal{X} = \{(h_1, d_1), (h_1, d_2), (h_2, d_1), (h_2, d_2)\}$$

and the set of feasible sets would be:

$$f_1 = \{(h_1, d_1)\} \qquad f_2 = \{(h_1, d_2)\}$$

$$f_3 = \{(h_2, d_1)\} \qquad f_4 = \{(h_2, d_2)\}$$

$$\mathcal{F} = \{f_1, f_2, f_3, f_4\}$$

The following example shows that constraints could include contractual characteristics.

Example 1.2. A school district has two schools h_1 and h_2 and three students d_1, d_2 and d_3. The district can finance either two seats in h_1 and one in h_2 or three seats in h_2 and offer students a stipend of s.

$$\mathcal{X} = \{(h_1, d_1), (h_1, d_2), (h_1, d_3), (h_2, d_1, s), (h_2, d_2, s), (h_2, d_3, s), (h_2, d_1), (h_2, d_2), (h_2, d_3)\}$$

and the set of feasible sets would be:

$$f_1 = \{(h_2, d_1, s), (h_2, d_2, s), (h_2, d_3, s)\}$$
$$f_2 = \{(h_1, d_1), (h_1, d_2), (h_2, d_3)\}$$
$$f_3 = \{(h_1, d_1), (h_2, d_2), (h_1, d_3)\}$$
$$f_4 = \{(h_2, d_1), (h_1, d_2), (h_1, d_3)\}$$
$$\mathcal{F} = \{f_1, f_2, f_3, f_4\}$$

In order to define property rights we need to introduce the concept of multisets. Multisets are generalized sets and are intermediate objects between sets and ordered tuples. In a multiset the order does not matter, but there could be many copies of the same element. For our purposes the following examples display the meaning of simple operations using multisets.

$$\{1, 1, 1, 3\} \cap \{1, 1, 2\} = \{1, 1\}$$
$$\{1, 1\} \cup \{1, 2\} = \{1, 1, 2\}$$
$$\{1, 1\} \subset \{1, 1, 1, 2\}$$
$$|\{1, 1\}| = 2$$
$$\{1, 1\} \uplus \{1, 2\} = \{1, 1, 1, 2\}$$

Every $f \in \mathcal{F}$ induces property rights f^* for hospitals named in f. For every $f \subset \mathcal{X}$ let f^* be the multiset formed by all elements of f after dropping the doctors dimension. For example, if $f = \{(h_1, d_1, s), (h_1, d_2), (h_1, d_4), (h_2, d_3, p)\}$ then $f^* = \{(h_1, s), (h_1), (h_1), (h_2, p)\}$. Let $\mathcal{F}^*$ be the set of all sets of property rights. We introduce some standard notation used throughout the paper.

For any agent $a \in D \cup H$ and set of contracts $Y \subset \mathcal{X}$, set of feasible contracts $f \in \mathcal{F}$, or set of rights $f^* \in \mathcal{F}^*$, let Y_a, f_a, f^*_a be the set (multiset) of elements naming agent a, respectively.

For any set of contracts $Y \subset \mathcal{X}$ (including singletons), set of feasible contracts $f \in \mathcal{F}$, or set of rights $f^* \in \mathcal{F}^*$, let $H(Y)$, $H(f)$, $H(f^*)$ be the set of hospitals named in Y,f, f^* respectively. Let $D(Y)$, $D(f)$ be the set of doctors named in Y and f, respectively. Analogously $A(\cdot)$ is the set of agents.

Each doctor d has preferences $\succeq_d$ over $\mathcal{X}_d \cup \{\emptyset\}$ and is characterized by a single-valued choice function $C_d(X) = max_{\succeq_d}\{x \in X_d \cup \{\emptyset\}\}$. Let $R_d(X) = X - C_d(X)$ be the set of rejected contracts when the set X is offered. It will be assumed that preferences are strict.[8] Let $C_D(X) = \underset{d\in D}{\cup} C_d(X)$ and $R_D(X) = X - C_D(X)$.

Each hospital h has preferences $\succeq_h$ over $2^{\mathcal{X}_h}$. It will be assumed that hospital preferences are strict.[9] However, departing from the original model of matching with contracts, hospital preferences will be characterized by a choice function with two arguments, the first being a set of contracts and the second a multiset of property rights. Intuitively, a hospital will be offered a series of choice problems and it will choose its most preferred chosen sent among those problems. Formally, $C_h(X, f^*) = max_{\succeq_h}\{Y \subset X_h \mid x, y \in Y \Rightarrow D(x) \neq D(y),\ \ Y^* \subset f^*\}$. Let $R_h(X, f^*) = X - C_h(X, f^*)$ be the set of rejected contracts. Let $C_h(X) = C_h(X, X^*)$ and $R_h(X) = X - C_h(X)$. Let $C_H(X) = \underset{H}{\cup} C_h(X)$ and $R_H(X) = X - C_H(X)$.

In this model an allocation is a set of contracts, as it determines payoffs for all participants in the market. We study stable allocations. Intuitively, a stable allocation is such that no participant would like to unilaterally reject some of the contracts he holds or would be able to find a partner to sign a mutually agreeable contract. In matching problems without constraint stability is usually defined as follows:

Definition 1.1. The set of contracts X is **unconstrained stable** if:

i) It is individually rational: $X = \underset{h\in H}{\cup} C_h(X) = \underset{d\in D}{\cup} C_d(X)$

ii) It is unblocked: there is no hospital h and set of contracts $Y \neq C_h(X)$ such that:

[8] It could also be assumed that C_d satisfies $C_d(X) \in Y$ and $Y \subset X$ implies $C_d(X) = C_d(Y)$. This property has been called consistency by Kamada and Kojima (2013) and irrelevance of rejected contracts by Aygün and Sönmez (2013) and assumes a regular behavior when breaking indifferences among contracts.

[9] It could also be assumed that C_H is such that $C_H(X) \in Y$ and $Y \subset X$ implies $C_H(X) = C_H(Y)$

$Y = C_h(X \cup Y) = \underset{d \in D(Y)}{\cup} C_d(X \cup Y)$

When there are constraints there is an obvious problem with the previous definition, i.e. it is not guaranteed that a particular unconstrained stable set of contracts is feasible:

Example 1.3. Consider the regional cap example with feasible set given by:

$$\mathcal{F} = \{\{(h_1, d_1)\}, \{(h_1, d_2)\}, \{(h_2, d_1)\}, \{(h_2, d_2)\}\}$$

Consider the following preferences:

$$h_1 : d_1, d_2 \qquad d_1 : h_2, h_1$$

$$h_2 : d_2, d_1 \qquad d_2 : h_1, h_2$$

$X = \{(h_1, d_2), (h_2, d_1)\}$ is unconstrained stable but infeasible.

The previous example shows that the concept of unconstrained stability cannot be readily applied to the problem of matching with constraints. One might hope that adding a feasibility requirement would suffice to restore the concept. However, the following definition and example show that we need to restrict unconstrained stability even more.

Definition 1.2. The set of contracts X is **constrained stable** if:

i) It is feasible: $X \in \mathcal{F}$

ii) It is individually rational: $X = \underset{h \in H}{\cup} C_h(X) = \underset{d \in D}{\cup} C_d(X)$

iii) It is unblocked: there is no hospital h and set of contracts $Y \neq C_h(X)$ such that:

$$Y = C_h(X \cup Y) = \underset{d \in D(Y)}{\cup} C_d(X \cup Y)$$

$$Y \cup X_{A(X) \setminus A(Y)} \in \mathcal{F}$$

An allocation is constrained stable if i) it is feasible, ii) individually rational and ii) there is no blocking hospital and set of doctors that could implement a deviation by themselves (taking restrictions into account). The following example shows that adding feasibility is not enough.

Example 1.4. As in the previous example, assume that the feasible set is given by:

$$\mathcal{F} = \{\{(h_1, d_1)\}, \{(h_1, d_2)\}, \{(h_2, d_1)\}, \{(h_2, d_2)\}\}$$

and that preferences are as follows:

$$h_1 : d_1, d_2 \qquad d_1 : h_2, h_1$$

$$h_2 : d_2, d_1 \qquad d_2 : h_1, h_2$$

There are four candidate sets that satisfy feasibility: $X_1 = \{(h_1, d_1)\}$, $X_2 = \{(h_1, d_2)\}$, $X_3 = \{(h_2, d_1)\}$ and $X_4 = \{(h_2, d_2)\}$. Consider $X_1 = \{(h_1, d_1)\}$. In this allocation the pair (h_2, d_2) would like to sign a contract, but the constraint prevents them from doing so. However, the pair (h_2, d_1) would like to and **could** implement a deviation from form $X_1 = \{(h_1, d_1)\}$ to $X_3 = \{(h_2, d_1)\}$. Continuing the analysis reveals that X_1 lead to X_3, X_3 to X_4, and X_4 to X_1. X_2 is not stable because h_1 would reject d_2 in favor of d_1.

The previous example shows that there are matching problems with constraints where all feasible allocations contain at least one blocking pair that could implement a deviation by themselves. However, X_1 and X_2 are very different. In X_2, h_1 is given the right to hold one contract and he is using it on his second most preferred doctor. This allocation would not be unconstrained stable conditional on given rights. We exploit this characteristic to define stability in our model. Both X_1 and X_4 would be good candidates to be stable in this market, although a mechanism would be able to select only one of them as the final allocation. If reports are used to decide which one of them is chosen, then doctors will have an incentive to manipulate their preferences. Hence we let the matchmaker decide.

1.2.2 The Matchmaker

There are practical and theoretical reasons to introduce an authority to make decisions that market participants cannot make by themselves (in an efficient and strategy-proof way).

On a practical level, matchmakers exist in real matching with constraints models. School district authorities decide over budget and resource allocations among schools, policy makers decide over constraints in medical matching and labor matching, colleges decide over allocations of funds and physical resources among different departments, etc. These authorities make important decisions in the final allocation of their respective matching markets and need to be included in the model.

On a theoretical level, constraints bind hospitals decisions, therefore a decision by a hospital creates an externality over other hospitals by "tightening" constraints. Even in a simple economy with externalities there is a need for a government to restore efficiency, usually achieved through taxes, quotas and property rights. The first two require the government to know the efficient level of activity in the market while the third does not. We introduce a matchmaker and provide it with a way to establish "property rights" over feasible allocations in order to achieve efficiency.

We introduce the matchmaker m into the model of matching with constraints by his preferences $\succeq_m$ over $\mathcal{F}^*$. We assume $\succeq_m$ is complete, transitive and strict. We assume the matchmaker always prefers more contracts signed than less contracts i.e. $|f^*| \geq |g^*|$ implies $f^* \succeq_m g^*$.

1.2.3 Stability

As shown by previous examples, the set of (feasible) unconstrained stable and the set of constrained stable allocations could be empty. However, as hinted in example 4, property rights can be used to define stability. We define stability in a matching market with constraints as follows:

Definition 1.3. The set of contracts X is **stable** if:

1. It is feasible: $X \in \mathcal{F}$
2. It is individually rational: $\underset{h \in H}{\cup} C_h(X) = \underset{d \in D}{\cup} C_d(X)$
3. It is unblocked: Let $Y \subset \mathcal{X}$. If $H(Y) = h$ and $X_h \neq Y = C_h(X \cup Y) = C_{D(Y)}(X \cup Y)$, then:

$$X_h = C_h(X \cup Y, X_h^*)$$

$$((X \setminus X_h) \cup Y)^* \notin \mathcal{F}^*$$

The intuition behind the last two conditions is straightforward: hospitals should behave optimally conditional on the assigned property rights and no addition of property rights should be possible. This definition is a generalization of unconstrained stability and they are equivalent in a model with no constraints.
With this definition of stability in hand, the previous example displays a stable allocation.

Example 1.5. As in the previous example, assume that the feasibility set is given by:

$$\mathcal{F} = \{\{(h_1, d_1)\}, \{(h_1, d_2)\}, \{(h_2, d_1)\}, \{(h_2, d_2)\}\}$$

and that preferences are as follows,

$$h_1 : d_1, d_2 \qquad d_1 : h_2, h_1$$

$$h_2 : d_2, d_1 \qquad d_2 : h_1, h_2$$

Now there are two stable sets of contracts, namely $X_1 = \{(h_1, d_1)\}$ and $X_4 = \{(h_2, d_2)\}$.

1.3 Matching with Constraints Mechanism

Since the ultimate test of a model as the one presented here is its applicability in real situations we would like to offer a mechanism with attractive properties regarding four broad goals: efficiency, transparency, simplicity and fairness.

As is standard in matching models, we achieve fairness by constructing an anonymous mechanism. In other words, our mechanism will consider only the stated preferences of the participants and not their identity when determining an allocation. In addition, we focus on the stability of the final allocation. Stable allocations are fair in the following sense: no participant is able to provide a "reason" to be matched with a more preferred partner, i.e. having a higher priority or being preferred by a different participant than the current match.

In this environment, with ordinal preferences, the most appealing concept of efficiency is Pareto optimality of the final allocation. In order to make efficiency and fairness compatible we focus on Pareto optimal allocations among the set of stable allocations.

We consider that transparency of the mechanism is also crucial, in particular participants should be able to know the rules of the mechanism and the party implementing the mechanism must be able to tell participants, without conflict of interest, that it is in their best interest to reveal their private information. Therefore, we rely on strategy-proof implementation.

We achieve simplicity by finding a mechanism that simultaneously solves both allocation problems, resources to hospitals and doctors to hospitals, simultaneously. Since matching problems are combinatorial problems, solving the problems sequentially would be a rather computationally daunting task. We show this is the case for a special case of the seat production environment.

1.3.1 Construction

We start the construction of our mechanism by making the following observation: if hospitals decide the contracts they sign independently, the final allocation might not be feasible. Hence we need the matchmaker to allocate "property rights" to eliminate the externality-like effect present in the matching with constraints. Formally, the externality-like effect is characterized as follows: consider a set of contracts X and assume hospitals choose contracts independently without considering that their decision affects and is affected by the decisions of other hospitals through the binding constraints. Then, as shown in previous examples, it is possible that $\underset{H}{\cup} C_h(X) \notin \mathcal{F}^*$. In order to guarantee the feasibility of the chosen set, hospitals need to be aware of their feasible choices at a given decision moment. However, property rights over feasible options do not exist, i.e. all participants are entitled in common to resources. The matchmaker assigns property rights by offering appropriate feasible choice problems to hospitals.

Let X be a set of contracts and let $\mathcal{F}^*(X)$ be the set of property right sets at X, defined as follows: $\mathcal{F}^*(X) = \{f^* \in \mathcal{F}^* \mid f \subset X \text{ and } f \in \mathcal{F}\}$. Let $\mathcal{C}_m(X) = max_{\succeq_m} \mathcal{F}(X)$. Through this process the matchmaker is able to assign property rights over different conflicting allocations of resources. After a particular allocation of resources has been selected, hospitals choose contracts. Thus h_1 selects $C_{h_1}(X, \mathcal{C}_m(C_H(X)))$, h_2 selects $C_{h_2}(X, \mathcal{C}_m(C_H(X)))$ and so on. This mechanism is described in the following example.

Example 1.6. A school district has two schools h_1 and h_2 and three students d_1, d_2 and d_3. The district can finance either two seats in h_1 and one in h_2 or three seats in h_2 and offer students a stipend of s.

$\mathcal{X} = \{(h_1, d_1), (h_1, d_2), (h_1, d_3), (h_2, d_1, s), (h_2, d_2, s), (h_2, d_3, s), (h_2, d_1), (h_2, d_2), (h_2, d_3)\}$

and the set of feasible sets would be:

$$f_1 = \{(h_2, d_1, s), (h_2, d_2, s), (h_2, d_3, s)\}$$

$$f_2 = \{(h_1, d_1), (h_1, d_2), (h_2, d_3)\}$$

$$f_3 = \{(h_1, d_1), (h_2, d_2), (h_1, d_3)\}$$

$$f_4 = \{(h_2, d_1), (h_1, d_2), (h_1, d_3)\}$$

$$\mathcal{F} = \{f_1, f_2, f_3, f_4\}$$

Consider the following preferences[10] for schools. Preferences are responsive. h_2 prefers students with stipend to students without stipend:

$$h_1 : d_1, d_2, d_3$$

$$h_2 : d_1, d_2, d_3$$

Hence

$$C_{h_1}(\mathcal{X}) = \{(h_1, s_1), (h_1, s_2), (h_1, s_3)\}$$

$$C_{h_2}(\mathcal{X}) = \{(h_2, s_1, s), (h_2, s_2, s), (h_2, s_3, s)\}$$

Not only are the previous school choices not mutually feasible, the choice by the first school is not feasible. The matchmaker can assign property rights in order to guarantee the feasibility of the final allocation. Consider first preliminary constructions:

$$\mathcal{F}(C_H(\mathcal{X})) = \{\{h_1, h_1\}, \{(h_2, s), (h_2, s), (h_2, s)\}\}$$

The matchmaker prefers more contracts to less:

$$\mathcal{C}_m(C_H(\mathcal{X})) = \{(h_2, s), (h_2, s), (h_2, s)\}$$

Thus,

$$C_{h_1}(\mathcal{X}, f_1^*) = \emptyset$$

$$C_{h_2}(\mathcal{X}, f_1^*) = f_1$$

The matchmaker intervention produces the following aggregate behavior of the market:

[10]Strictly speaking preferences are over contracts and not students, but for simplicity we identify students with their unique contract.

$$\widetilde{C}_H(X) = \underset{h \in H}{\cup} C_h(X, \mathcal{C}_m(C_H(X)))$$

In order to define the Matching with Constraints Mechanism we introduce the Generalized Gale Shapley Algorithm [11], [12] as defined by Hatfield and Milgrom (2005). Define the following order in $\mathcal{X} \times \mathcal{X}$, $(X, Y) \geq (X', Y')$ if and only if $X \subset X'$ and $Y' \subset Y$.

Definition 1.4. The **Generalized Gale-Shapley** Algorithm (**GS**) is defined as the iterated application of $F(X_H, X_D) = (\mathcal{X} - R_D(X_D), \mathcal{X} - R_H(X_H))$. If the iterated process finishes in a fixed point $(\overline{X}, \overline{Y}) = F(\overline{X}, \overline{Y})$ starting from (X, Y) we define $GS(X, Y) = \overline{X} \cap \overline{Y}$. Analogously, $GS(X, Y, f^*) = \overline{X} \cap \overline{Y}$ denotes the fixed point, if any, of iterated applications of $F(X_H, X_D, f^*) = (\mathcal{X} - R_D(X_D), \mathcal{X} - R_H(X_H, f^*))$

With the aggregate behavior of the market defined as above we can define our matching with constraints mechanism as follows:

Definition 1.5. The **Matching with Constraints Mechanism** (**MC**) is defined as the iterated applications of $\widetilde{F}(X_H, X_D) = (\mathcal{X} - R_D(X_D), \mathcal{X} - \widetilde{R}_H(X_H))$. If the iterated process finishes in a fixed point $(\overline{X}, \overline{Y}) = \widetilde{F}(\overline{X}, \overline{Y})$ starting from (X, Y) we define $MC(X, Y) = \overline{X} \cap \overline{Y}$.

1.3.2 Properties

In this section we introduce five regularity conditions to guarantee the existence of a stable set of contracts. The first two are the appropriate generalizations of the

[11] GS has been central in the market design literature and its applications are wide. Applications to allocate medical students (Roth (1984a); Roth and Peranson (1999)), students to public universities (Balinski and Sönmez (1999)), and medical graduate to residency programs in Japan (Kamada and Kojima (2013)) are some examples.

[12] This algorithm was initially proposed by Gale and Shapley (2013). Subsequent developments extended the GS algorithm to more general preferences by Roth (1984b) and more general environments by Hatfield and Milgrom (2005).

substitutes condition and the law of aggregate demand as proposed by Hatfield and Milgrom (2005), which in a model without constraints are sufficient to achieve our goals. The third regularity condition is a generalization of the substitability condition introduced by Kamada and Kojima (2013), which control the level at which property rights are transferred to one hospital from another. We introduce two new conditions to allow the model to handle complex constraints; the **consistency of choice property,** to compare different allocations in different markets and the **contract replacement property**, to correlate hospital preferences and constraints.

The substitutes condition[13] has been shown to guarantee the existence of a stable allocation in a model without constraints.[14] In addition, every market without constraints can be represented by a market with only one feasible set of contracts. Hence, a matching market with constraints is a model where simultaneous unconstrained markets coexist. We assume that every unconstrained market satisfies the substitutes condition. In our notation this assumption is equivalent to the following condition we call the **strong substitutes condition.**

Definition 1.6. C_h satisfies the **strong substitutes condition** if and only if for every X and Y and $f^* \in \mathcal{F}^*$ such that $X \subset Y$, we have $R_h(X, f^*) \subset R_h(Y, f^*)$. C_h satisfies the **substitutes condition** if and only if for every X and Y such that $X \subset Y$, we have $R_h(X) \subset R_h(Y)$

The law of aggregate demand was introduced by Hatfield and Milgrom (2005) in order to obtain strategy-proofness in a model without constraints. Analogously to the substitutes condition, we assume that the law of aggregate demand is satisfied in every market. In our model, this is achieved by a condition we call the **strong**

[13] The substitutes condition was introduced to the matching literature by Kelso and Crawford (1982). Weaker conditions are known, see Hatfield and Kojima (2010).

[14] for example, see Hatfield and Milgrom (2005)

law of aggregate demand.

Definition 1.7. C_h satisfies the **strong law of aggregate demand** if and only if for every X and Y and $f^* \in \mathcal{F}^*$ such that $X \subset Y$, we have $|C(X, f^*)| \leq |C(Y, f^*)|$. C_h satisfies the **law of aggregate demand** if and only if for every X and Y such that $X \subset Y$, we have $|C(X)| \leq |C(Y)|$.

The strong substitutes condition and the strong law of aggregate demand characterize hospital preferences in this model. Kamada and Kojima (2013) identified a condition, they call substitability, analogous to the substitutes condition to characterize the matchmaker's preferences. We generalize the substitability condition to assign property rights in out model. When C_m satisfies the substitutes condition, the matchmaker never increases the property rights of any hospital after having restricted that hospital. The set of property rights of a given hospital can increase or decrease, but cannot increase after decreasing.

Definition 1.8. $\mathcal{C}_m$ satisfies the **substitutes condition** if and only if for every X^* and Y^* such that $X^* \subset Y^*$, we have $\mathcal{C}_m(Y^*) \cap X^* \subset \mathcal{C}_m(X^*)$.

In this model, in which many markets are studied as a single market with constraints, it is necessary to introduce some regularity in hospital preferences across different market. To see this, suppose X is an unconstrained stable allocation in market f; hence there is no blocking set Y in market f^*, however, there might be blocking sets in other markets, say g^*. We introduce the **consistency of choice** property in order to compare stable allocations in different markets.

Definition 1.9. C_h satisfies **consistency of choice** if and only if for every X and g^*,$f^* \in \mathcal{F}^*$ such that $g^* \subset f^*$, we have $C_h(X, g^*) \subset C_h(X, f^*)$.

One of the main characteristics of matching models, for example the original model of Gale and Shapley (2013), has been the so called tentative acceptance i.e. the ability of a hospital to hold a doctor until a better one arrives and then substitute one for the other. In a model without constraints or in a model in which all contracts are of the same type (for example distributional constraints) any hospital can substitute contracts freely. However, in a model with complex constraints and contracts this is not the case. Suppose for example that a hospital would like to substitute a contract w with a contract x, if they induce the same property rights, then the hospital can do so immediately. However, if they induce different property rights then the substitution might not be possible. We introduce a **contract replacement property** to control these cases.

Definition 1.10. $\mathcal{F}$ satisfies the **contract replacement property** if and only if for every X and $x, w \notin X$ such that $C_H(Xw) = Xw$ and $C_H(Xwx) = Xx$, we have $x^* = w^*$

When the contract replacement property is satisfied, hospitals are only willing to substitute contracts with the same property rights. Substitution of contracts with different property rights is also possible, but must be induced by constraints and not the unconstrained preferences.

With the five regularity conditions in place we proceed to show the properties of this model. We first show that $\widetilde{C}_H$ satisfies the substitutes condition, which guarantees the existence of an unconstrained stable allocation in the model with only one hospital whose preferences are represented by $\widetilde{C}_H$. The next step is to show that unconstrained stable allocations in the aggregate model are stable allocations in the model with many hospitals and constraints. We show next that the Matching with Constraints Mechanism is strategy-proof for doctors by showing that $\widetilde{C}_H$ also satisfies the law of aggregate demand. Finally we show that the Matching with Constraints Mechanism is efficient.

Theorem 1.1. *Suppose $\mathcal{F}$ satisfies the contract replacement property, C_m satisfies the substitutes condition and C_h satisfies the strong law of aggregate demand, the* *strong substitutes condition and the consistency of choice condition for all $h \in H$, then $\widetilde{C}_H$ satisfies the substitutes condition.*

Proof. We first show that $C_H(Yz)^* \subset C_H(Yzx)^*$ in order to use the substitutes property of C_m. Since C_H satisfies the substitutes condition and the law of aggregate demand, there are three cases to consider. i) If x is not chosen, $C_H(Yz)^* = C_H(Yzx)^*$; ii) if x is simply chosen, i.e $C_H(Yzx) = C_H(Yz)x$, then $C_H(Yz)^* \subset C_H(Yzx)^*$; iii) if x is chosen in favor of w, i.e. $C_H(Yzx) = (C_H(Yz) \setminus w)x$, then $C_H(Yz)^* \subset C_H(Yzx)^*$ since the contract replacement property implies that $x^* = y^*$. In order to prove that $\widetilde{C}_H$ satisfies the substitutes condition we show that $z \in \widetilde{C}_H(Yzx)$ implies $z \in \widetilde{C}_H(Yz)$ for arbitrary Y, x, z. $z \in \widetilde{C}_H(Yzx)$ implies that $z \in C_H(Yzx, f^*)$ for $f^* = C_m(C_H(Yzx))$. By the strong substitutes condition we have $z \in C_H(Yz, f^*)$ and by the definition of choice function we have $z \in C_H(Yz, f^* \cap (Yz)^*)$. Let $g^* = C_m(C_H(Yz))$, then $f^* \cap (Yz)^* \subset g^*$ by the substitutes condition of C_m. Finally, by the consistency of choice property we have $z \in C_H(Yz, g^*) = \widetilde{C}_H(Yz)$ □

If $\widetilde{C}_H$ satisfies the substitutes condition, then $F(X_H, X_D) = (\mathcal{X} - R_D(X_D), \mathcal{X} - \widetilde{R}_H(X_H))$ is a monotone function with non-empty set of fixed points. The set of fixed points is a lattice. In the following theorem we show that these fixed points are stable allocations.

Theorem 1.2. *Suppose $\mathcal{F}$ satisfies the contract replacement property, C_m satisfies the substitutes condition and C_h satisfies the strong law of aggregate demand, the strong substitutes condition and the consistency of choice condition for all $h \in H$, , then the set of stable allocations is not empty.*

Proof. Since $\widetilde{C}_H$ satisfies the substitutes condition, then $F(X_H, X_D) = (\mathcal{X} - R_D(X_D), \mathcal{X} - \widetilde{R}_H(X_H))$ has a non-empty set of fixed points, which forms a complete lattice. Every fixed point is an unconstrained stable set of contracts. Suppose X is an unconstrained stable set of contracts. We show that X is stable. By construction, X is feasible and individually rational. Let Y be a set of contracts such that $D(Y) = h$, and $X_h \neq Y = C_h(XUY) = C_{D(Y)}(XUY)$. Suppose that $X_d \neq Z = C_h(XUY, X_h^*)$, then there is Z' such that $X \neq Z' = \widetilde{C}_H(XUZ')$, which is a contradiction. Suppose now that $((X \setminus X_h) \cup Y)^* \in \mathcal{F}^*$, then $C_h(XUY)$ is feasible so $\widetilde{C}_H(XUY) = ((X \setminus X_h) \cup Y) \neq X$, a contradiction. Hence X is stable. □

The next couple theorems show that the law of aggregate demand is also satisfied; hence the Matching with Constraints Mechanism is strategy-proof for doctors.

Theorem 1.3. *Suppose $\mathcal{F}$ satisfies the contract replacement property, C_m satisfies the substitutes condition and C_h satisfies the strong law of aggregate demand, the strong substitutes condition and the consistency of choice condition for all $h \in H$, , then $\widetilde{C}_H$ satisfies the law of aggregate demand.*

Proof. It is sufficient to show that for any Y and x we have $\left|\widetilde{C}_H(Y)\right| \leq \left|\widetilde{C}_H(Yx)\right|$. Since C_H satisfies the law of aggregate demand and the substitutes condition, we have 3 cases to consider. i) If x is not chosen, then $\widetilde{C}_H(Y) = \widetilde{C}_H(Yx)$; ii) If x is simply chosen, let $f^* = C_m(C_H(Y))$ and $g^* = C_m(C_H(Yx))$. Since x was simply chosen we have that $C_H(Y)^* \subset C_H(Yx)^*$, thus by the substitutes property of C_m we have that $\left|\widetilde{C}_H(Y)\right| \leq \left|\widetilde{C}_H(Yx)\right|$; iii) If x is chosen in place of w, then by the contract replacement property we have that $\left|\widetilde{C}_H(Y)\right| \leq \left|\widetilde{C}_H(Yx)\right|$. □

Theorem 1.4. *Suppose $\mathcal{F}$ satisfies the contract replacement property, C_m satisfies the substitutes condition and C_h satisfies the strong law of aggregate demand, the strong substitutes condition and the consistency of choice condition for all $h \in H$, then the Matching with Constraints mechanism is strategy-proof for doctors.*

Proof. Since $\widetilde{C}_H$ satisfies the law of aggregate demand and the substitutes condition, the Matching with Constraints mechanism is strategy-proof for doctors by a result in Hatfield and Milgrom (2005). □

In a matching market with constraints stability is not efficient either. Consider a market with two hospitals, one only looking for doctors specialized in oncology and the other in neurology. Suppose resources consist on the appropriate equipment for each practice but are assigned incorrectly. The equipment for oncology to the neurology hospital and vice versa. Then the allocation where both hospital hire no doctor is stable, but not efficient. In order to move to an efficient allocation of resources a joint redistribution of rights would be needed. Since stability does not involve multiple hospitals, stability does not imply efficiency.

Theorem 1.5. *Suppose $\mathcal{F}$ satisfies the contract replacement property, C_m satisfies the substitutes condition and C_h satisfies the strong law of aggregate demand, the strong substitutes condition and the consistency of choice condition for all $h \in H$, , then the Matching with Constraints mechanism is efficient.*

Proof. Let X be the set of contracts assigned by the Matching with Constraints mechanism. Suppose there is a feasible set of contracts Y such that $Y_h = C_h(XY) = C_{D(Y_d)}(XY)$ for all hospitals and $Y_d = C_d(XY)$ for all doctors. Hence, there is Y such that $X \neq Y = \widetilde{C}_H(XY) = C_{D(Y)}(XUY)$. Hence X is not chosen by the Matching with Constraints Mechanism. □

1.3.3 Structure

In this section we show the structure produced by this model. Consider, for example, a market with constraints representing five individual markets $f_1^*, ..., f_5^*$. It is well-known that, in a market without constraints, the set of stable allocations forms a lattice. In the following figure the lattice for market f_4^* is represented by a rectangle, with maximum stable allocation $GS(\emptyset, \mathcal{X}, f_4^*)$ and minimum stable allocation $GS(\mathcal{X}, \emptyset, f_4^*)$.

The following theorem establishes that, in general, some of these lattices can be ordered by their maximum and minimum elements.

Figure 1.3: The set of stable allocations of market f_4

Theorem 1.6. Let C_H^1 *and* C_H^2 *be two choice functions satisfying the substitutes condition such that* $C_H^1(X) \subset C_H^2(X)$ *for all* X, *then* $GS^1(\emptyset, \mathcal{X}) \leq GS^2(\emptyset, \mathcal{X})$ and $GS^1(\mathcal{X}, \emptyset) \leq GS^2(\mathcal{X}, \emptyset)$. Furthermore, $GS^1(\emptyset, \mathcal{X}) = GS^1(GS^2(\emptyset, \mathcal{X})_1, GS^2(\emptyset, \mathcal{X})_2)$.

Proof. (Unconstrained) stable allocations are the fixed points of the following functions: $F_1(X_H, X_D) = (\mathcal{X} - R_D(X_D), \mathcal{X} - R_H^1(X_H))$ in the first market and $F_2(X_H, X_D) = (\mathcal{X} - R_D(X_D), \mathcal{X} - R_H^2(X_H))$ in the second market. Hence $F_1(X_H, X_D) \leq F_2(X_H, X_D)$. Let X_1 be the highest fixed point of F_1 and X_2 be the highest fixed point of F_2. Since $X_1 = F_1(X_1) \leq F_2(X_1)$, hence $X_1 \leq X_2 = \vee\{x \mid x \leq F_2(x)\}$. For the second part consider the following inequality and apply F_1 as many times as necessary $GS^1(\emptyset, \mathcal{X}) \leq GS^2(\emptyset, \mathcal{X}) \leq (\emptyset, \mathcal{X})$. □

The next figure depicts the lattices of markets $f_1^*, \ldots, f_5^*$ with the ordering provided by the previous theorem. Every gray rectangle represents the set of stable allocations in each market. The solid black rectangle represents the set of stable allocations of the market formed by the union of the smaller markets. The red dots are the Pareto frontier for doctors.

In some applications it might be desirable that the final selected allocation is a Pareto efficient allocation from the point of view of the doctors. The Matching with Constraints mechanism does exactly that.

Theorem 1.7. *The* ***Matching with Constraints Mechanism*** *is Pareto optimal for doctors among the set of stable sets of contracts.*

Proof. Suppose X is the set of contracts chosen by the Matching with Constraints

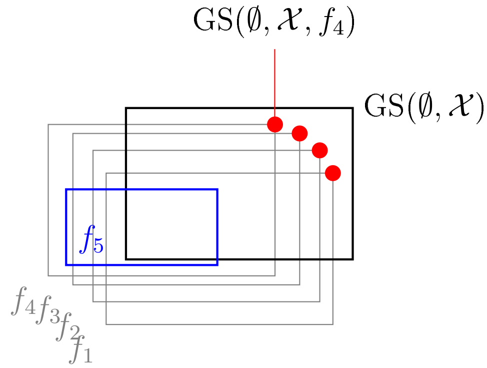


Figure 1.4: The set of stable allocations of markets $f_1, ..., f_5$

Mechanism. Suppose Y is a stable set of contracts that Pareto dominates X for doctors. Then Y is (unconstrained) stable in the market with only one hospital represented by $\widetilde{C}_H$. Thus contradicting the fact that X is Pareto optimal for doctors in the market with only one hospital. □

1.4 Conclusion

We developed a matching model in which several matching markets can be analyzed simultaneously as a single matching market with constraints. We offered a mechanism capable of finding a stable (within and across markets) and efficient allocation. Our mechanism is strategy-proof for doctors and finds the appropriate market and the stable allocation in that market simultaneously. The model is capable of implementing several design previously studied independently, including models of affirmative action (Abdulkadirouglu and Sönmez (2003)) and distributional constraints (Kamada and Kojima (2013)).

As shown in the main text, the required properties for the matching mechanism with constraints impose a limit on the complexity of constraints and the structure of preferences. However, it is possible to extend the complexity of constraints by giving up structure in some other parts of the model. One such move could be to move from substitutable preferences to additive preferences, in that case one could dispense of the contract replacement property since no hospital would "naturally" replace a doctor for another, and if they do is because a constraint.

In this paper we have focused on a simultaneous solution to the allocation problems (market and allocation) since computational simplicity is often of practical concern. Moving to a sequential solution could open the door to even more complex constraints at the expense of computational power. The simplest example would be to calculate all stable allocations in all markets and then choose one in the Pareto frontier. However, it is not clear if such a procedure could have nice incentive properties in general.

A second extension to the complexity of constraints would be achievable by specializing our preferences. In particular, in the individual rationality dimension, if some structure is provided to the individual rational allocations for hospitals the model could potentially discard the assumption of free disposal of contracts. One such specialization is the minimum quotas in schools or hospitals.

The tradeoffs of complexity in matching models are still being studied and this paper hopefully has made some of them easier to understand.

Chapter 2: Strategy-proofness for Hospitals in Matching Markets

Abstract. *Strategy-proof implementation is one of the many elements that have contributed to the successful application of matching theory in real life. However, in many-to-one matching markets without transfers (e.g., doctors to hospitals with fixed salaries) there is no stable mechanism which is strategy-proof for hospitals. Furthermore, strategy-proofness and stability cannot be achieved for both hospitals and doctors simultaneously even in one-to-one matching markets. This paper shows that in many-to-one matching markets with transfers it is possible to guarantee stability and strategy-proofness-for-hospitals whenever an opportunity cost condition is satisfied. In addition, it is shown that stability and strategy-proofness are possible for both hospitals and doctors simultaneously . Finally, it is shown that strategy-proofness can be achieved in the interior of the core.*

Many matching markets have successfully adopted centralized mechanisms as an alternative to the price system.[1] In these centralized markets, the final allocation is computed using information provided by market participants. Typically, the final allocation is chosen to be stable with respect to the provided information. However, it is desirable that stability holds with respect to the actual information. Hence,

[1]See Abdulkadiroğlu et al. (2005); Abdulkadirouglu et al. (2005); Abdulkadirouglu and Sönmez (2003) for School Choice, Roth et al. (2004) for Kidney Exchange, Sönmez (2013); Sönmez and Switzer (2013) for Branch of Choice and Roth and Peranson (1999); Crawford and Knoer (1981); Kelso and Crawford (1982); Roth and Xing (1994) for Residents Markets.

participants' incentives to report truthfully are extremely important. In practice, it is often a goal to make participants best strategy to report their actual information, regardless of other agents' reports. Whenever this is achieved it is said that the mechanism is strategy-proof.

Strategy-proof mechanisms possess several advantages over other mechanisms. First, it is possible to guarantee that efficiency or other properties hold with respect to actual and not only with respect to reported information. This is particularly important for many institutions since it has been shown experimentally that in some non-strategy-proof mechanisms, up to 80% of agents misrepresent their preferences.[2] Furthermore, submitted preferences are often used in welfare assessments, for example by computing the number of participants obtaining their first choice, second choice and so on.[3] Second, no resources are wasted by market participants in order to compute better-than-truthful reports. In an auction, which is a special kind of matching market, bidders spend a lot of time and money devising their strategies and they often hire auction consultants.[4] In school choice, parents spend time and money to obtain better outcomes from non-strategy-proof mechanisms.[5] Third, participants with more information about the market cannot take advantage of less informed participants. This is particularly important in school choice, where equal access to education is often a goal.[6]

In a series of papers, Roth studied several incentive properties of stable mechanisms in two-sided markets without transfers. In these markets there are two kinds of agents: hospitals and doctors. Hospitals want to be matched with a set of doc-

[2]See (Chen and Sönmez, 2006) for an example

[3]See Featherstone (2011) for a discussion about rankings as a welfare criterion.

[4]Some problems faced by bidders in high-stake auctions are described in Cramton and Schwartz (2000); Milgrom et al. (2009)

[5]See Pathak and Sönmez (2008, 2013) for the study of several non-strategy-proof mechanisms.

[6]See also Pathak and Sönmez (2008, 2013) for a revealing discussion about the issue.

tors and doctors want to be matched with at most one hospital. The market is said to be one-to-one whenever hospitals want to be matched with at most one doctor and many-to-one otherwise. Roth showed that (i) in one-to-one matching markets, it is possible to obtain strategy-proofness for doctors or hospitals(Roth, 1982), but not both simultaneously (Roth, 1984b) and (ii) in many-to-one matching markets, strategy-proofness can be guaranteed for doctors, but not for hospitals (Roth, 1985).

This paper studies many-to-one markets with transfers and shows that stability and strategy-proofness for both hospitals and all agents (doctors and hospitals simultaneously) are possible by characterizing the conditions under which the Vickrey-Clarke-Groves (VCG) mechanism is stable.[7] The VCG mechanism is always strategy-proof.

This paper uses a version of the assignment game proposed by Shapley and Shubik (1971) as generalized by Kelso and Crawford (1982). In this model, doctors want to be matched to at most one hospital and hospitals can be matched to multiple doctors. In this model, transfers can be made continuously, in discrete quantities, or not at all.

The first contribution of this paper is to show that a VCG mechanism, different from the pivot mechanism, is stable whenever agents' preferences satisfy an opportunity cost condition. This condition is satisfied whenever every doctor can find a hospital that offers him at least his opportunity cost. In this mechanism, hospital i receives a payoff $\pi_i^V = V(A) - V(A \setminus i)$, where V is the coalitional value function and A is the set of agents in the market. Leonard (1983) proved the result in one-to-one matching markets and his technique rests completely on the unit demand assumption. Gul and Stacchetti (1999) proved the result for replica economies.

The second contribution is to identify markets where strategy-proofness, together

[7] See Vickrey (1961)

with stability, can be achieved for all agents. The above mechanism cannot be used directly for all agents since, in general, $\sum_{i \in A} \pi_i^V > V(A)$. However, careful inspection reveals that if agent i is receiving a payoff of π_i^V he is fully capturing **two** marginal values. When a new agent enters a matching market, say a doctor, two effects take place. First, a new agent who demands hospitals enters the market; and second, a new object is available for hospitals currently in the market. In order to obtain strategy-proofness we only need agents to be able to capture the marginal value of their information i.e. the marginal value they produce as demanders not as objects. For agent i, this marginal value is captured by $\pi_i^U = U(S) - U(S \setminus i)$, where $U(T)$ is the maximum value that can be achieved with all agents present only using the information of agents in T. Thus π_i^U is the marginal contribution of agent i's private information and $\pi_i^U - \pi_i^V$ is the marginal contribution of agent i's existence as an object. If no agent's information is pivotal (the agent himself is pivotal), then the U mechanism is strategy-proof for all agents and stable whenever the private values of an efficient matching belong to the set of stable payoffs.

The third contribution is to show that strategy-proofness either for hospitals or doctors can be achieved without offering an extremal matching. In general, strategy-proofness is achieved by offering agents their most preferred stable allocation.[8] In continuous transfers models, this payoff is characterized by $\pi_i^V = V(A) - V(A \setminus i)$. However, as discussed above, strategy-proofness can also be obtained by offering agents the marginal contribution of their information, $\pi_i^U = U(A) - U(A \setminus i)$. We show that the stability of these payoffs is directly linked to the proportion of surplus generated at each side of the market. In particular, for any vector of stable payoffs $\{\pi_i\}_A$, there is a division of total surplus that makes the U mechanism stable and strategy-proof, i.e. $\pi_i = U(A) - U(A \setminus i)$ for all agents.

[8] See for example, Crawford and Knoer (1981); Kelso and Crawford (1982); Roth and Sotomayor (1992)

Finally, it is shown that strategy-proofness for hospitals can be implemented in discrete transfers. The level to which transfers can be used in a matching markets varies significantly. However, this paper shows that allowing transfers to be negotiated in the matching process not only would improve efficiency (with respect to the reports), but also makes agents more willing to report their true private information. This increases the efficiency (with respect to the actual preferences) and accountability of the market.

Incentives in matching markets have been studied systematically since Roth's contributions (1982; 1984b; 1985). A first line of research has been devoted to finding restrictions on preference profiles for mechanisms to achieve strategy-proofness. Demange and Gale (1985) showed that Roth's conclusions hold in very general one-to-one environments where agents have preferences over each other and all have possibly different valuations over money. Alcalde and Barberà (1994) showed that the Gale-Shapley algorithm is the mechanism that achieves stability and strategy-proofness in the biggest set of preference profiles where both are possible. Sönmez (1997; 1999) introduced two kinds of manipulations observed in real matching markets: capacities misrepresentations and pre-arranged matches. He showed that there is no mechanism capable of avoiding such manipulations in general. Later on, Kojima (2007) and Kesten (2012) showed that some preference domains avoid those manipulations. This paper continues this line of research and shows that hospital strategy-proofness is possible when an opportunity cost condition holds.

A second line of research has studied incentives in large markets. In particular, Roth and Peranson (1999), Immorlica and Mahdian (2005) and Kojima and Pathak (2009) show that strategy-proofness for almost all agents is possible as the number of agents increase, but the diversity of preferences decreases. Roth and Peranson (1999) showed that this property holds in some physician markets. In small markets, however, manipulability is still possible with current mechanisms. This paper continues the large market spirit by showing that as the number of possible transfers increases, strategy-proofness for all hospitals becomes possible. This fact could have great practical implications since the discreteness of transfers traded in the markets is a market design variable.

This paper is also related to the literature on VCG auctions. It is well known that the VCG payments are lower than the lowest anonymous linear Walrasian equilibrium payments. However, several deviations from anonymous linear prices can achieve VCG payments. Ausubel (2006) uses personalized linear prices. Bikhchandani and Ostroy (2002); de Vries et al. (2007); Mishra and Parkes (2007) use personalized non-linear prices. The opportunity cost condition implies that VCG payoffs can be implemented with anonymous linear prices.

2.1 Matching Markets

Throughout the paper, capital letters will represent both sets and their cardinality. Similarly, throughout the paper, when an assumption is stated, it is considered as true in all subsequent parts of the paper, including theorems. We denote the union of two sets Y and X by YX. The matching market is formed by two kinds of agents, doctors and hospitals. The set of doctors is denoted by D and the set of hospitals by H with typical elements d and h, respectively. We denote the set of all agents by $A = HD$. We assume there is a finite number of agents in the market. It will be assumed that hospitals can be matched to several doctors, but doctors can be matched to at most one hospital.

Each doctor d has preferences $u_d : H\{\emptyset\} \to \mathbb{R}$. Each hospital h has a capacity c_h and preferences $u_h : C_h \to \mathbb{R}$, where C_h is the set of subsets of D with cardinality at most c_h. We assume that there is a common utility metric, that affects agents payoffs linearly, i.e. if agent a is matched to set T and receives a transfer of t_a, then his payoff is $\pi_a = u_a(T) + t_a$. The level at which transfers can be made in a particular market varies considerably across applications. In school choice, societies have decided that wealth should not determine who gets the better schools; hence modeled as a no-transfers market. In cadet branching, cadets have the opportunity to serve longer times in order to obtain a branch they like more; hence some discrete transfers are used. Labor markets with flexible transfers would be modeled as continuous transfers matching markets. In order to simplify the distinction between models with different degrees of transfers we introduce the definition of a q-market.

Definition 2.1. A **q-market** is a market where all transfers made are multiples of $\frac{1}{q} \in \mathbb{R}_+$. The 0-market represents the no-transfers case and the ∞-market represents the continuous case.

A q-matching Mt in a q-market is a disjoint collection of sets of doctors $M = \{D_i\}_{i \in H+1}$ such that $|D_h| \leq c_h$ and a vector of transfers (multiples of $\frac{1}{q}$) $t \in \mathbb{R}^D$ such that $t_d = 0$ for every $d \in D_{H+1}$. A hospital h is matched to the set of doctors D_h and receives a transfer of $-\sum_{i \in D_h} t_i$. D_{H+1} are unassigned doctors. The set of matchings in a market with A agents is denoted by $M(A)$ and the set of matchings where all transfers are zero is denoted by $M_0(A)$. Given a particular matching Mt, the associated payoff of agent a is denoted by π_a^{Mt} and the associated private value is u_a^{Mt}.
Given a particular matching, every coalition of hospitals and doctors can arrange offers to improve their payoffs from any initial situation; they can try to improve their monetary component with their current partners, they can try to form new partnerships, or reject current ones. If such improving coalitions cannot be formed, then the matching is called stable. Hence stability is the basic notion of equilibrium in matching markets. In addition to its theoretical appeal, stability has been proved fundamental for the correct performance of real matching markets since unstable allocations typically lead to an unraveling of the market.

Definition 2.2. The q-matching Mt is **stable** in q-market if and only if

- $\pi_a^{Mt} \geq u_a(\emptyset)$ for all $a \in A$
- There is no q-matching $(Mt)^*$ and $h \in H$ such that $\pi_a^{(Mt)^*} \geq \pi_a^{Mt}$ for all $a \in hD_h^*$ with at least one strict inequality.

Notice that the second condition applies only to members of the "blocking coalition" hD_h^* and hence to find a "blocking matching" $(Mt)^*$ it would be sufficient to assign members of hD_h^* together and leave everyone else unmatched.

It is well-known that if hospitals' preferences satisfy the substitutes condition, then the set of stable matchings is not empty.[9] In discrete markets, if agents' preferences between matchings are strict, then there is a hospital-optimal matching Mt_H and a doctor-optimal matching Mt_D.[10] The associated payoffs for agent a are denoted by $\pi_a^{Mt_H}$ and $\pi_a^{Mt_D}$, respectively. In the continuous case, a hospital-optimal matching Mt_H and a doctor-optimal matching Mt_D always exit and the associated payoffs for agent a are also denoted by $\pi_a^{Mt_H}$ and $\pi_a^{Mt_D}$. In the continuous case, the optimal matchings might fail to be unique, but their associated payoffs will be. We assume hospitals' preferences satisfy the substitutes condition. We first define the demand correspondence and value function for hospitals.

Definition 2.3. For every hospital $h \in H$, $T_h(t) = \underset{S \in 2^D}{argmax}\ u_h(S) - \sum_{i \in S} t_i$ is its **demand correspondence** and $v_h(t) = \underset{S \in 2^D}{max}\ u_h(S) - \sum_{i \in S} t_i$ is its **value function**.

Definition 2.4. u_h satisfies the **substitutes condition** if and only if for every $t, t^* \in \mathbb{R}^D$ such that $t \leq t^*$, for every $T \in T_h(t)$, there is $T^* \in T_h(t^*)$ such that $T \cap S \subset T^*$; where S is the set of doctors with $t_s = t_s^*$.

In addition to the substitutes condition we assume that every hospital needs to hire at least one doctor to produce any surplus and every doctor - hospital pair produce more surplus together than the unmached doctor.

[9] See Kelso and Crawford (1982)

[10] See Hatfield and Milgrom (2005)

Definition 2.5. u_h satisfies the **marginal product** condition if and only if $u_h(\emptyset) = 0$ and for all $d \in D$ and $D_h \subset D$ such that $d \notin D_h$ and $dD_h \in C_h$ we have $u_h(dD_h) + u_d(h) - u_h(D_h) \geq u_d(\emptyset)$.

The above condition allows for cases with $u_h(dD_h) < u_h(D_h)$ with $d \notin D_h$ i.e. a hospital would need to be compensated to hire a doctor. In the context of resident matching, this assumption could feel unnatural, however, in other applications such a school choice, it is the norm. Schools and colleges usually charge students to get admitted. Unless otherwise noted, it is assumed that the substitutes and marginal product conditions are satisfied by all u_h throughout the paper.

2.2 Strategy-Proofness for Hospitals in the continuous market

We begin this section by providing some standard preliminary definitions. A **mechanism** ϕ is a function that maps preference profiles to q-matchings in a q-market. The matching at preference profile $\{u_i\}_{i \in A}$ is denoted by $\phi(\{u_i\}_{i\in A}) \in M(A)$. A mechanism ϕ is said to be **strategy-proof** for agent a if there exist no preference profile u'_a and preferences $\{u_i\}_{i\in A\setminus a}$ for all other agents such that $\pi_a^{\phi(u'_a, \{u_i\}_{i\in A\setminus a})} > \pi_a^{\phi(\{u_i\}_{i\in A})}$. A mechanism ϕ is said to be **strategy-proof for** B (set) if it is strategy-proof for all $b \in B$. That is, no agent in B has an incentive to misreport his preferences under the mechanism. The next proposition, due to Roth (1985), establishes that there is no strategy-proof for Hospitals mechanism in the 0-market. Throughout the paper, "propositions" will be used to state important known results.

Proposition 2.1. ***Roth 85***. *There is no stable and strategy-proof mechanism for Hospitals in the 0-market.*

Example 2.1. Consider a market with three hospitals and four doctors. Hospitals and Doctors have cardinal valuations over each other. The left matrix contains the values each hospital assigns to every doctor and the number of doctors they are willing to hire. The right matrix contains the values for doctors. It is assumed that both groups assign a value of zero to being unassigned.

Preferences

Hospitals

	d_1	d_2	d_3	d_4	q
h_1	4	3	2	1	2
h_2	4	3	2	1	1
h_3	3	2	4	1	1

Doctors

	d_1	d_2	d_3	d_4
h_1	2	2	3	3
h_2	1	3	1	2
h_3	2	1	2	1

Suppose transfers are prohibited and a stable allocation is to be implemented. In this case, the ordinal representation is sufficient to characterize the set of stable matchings. Under the true preference profile, there is only one stable matching, shown on the left. On the right there is an improving deviation for h_1.

True Preferences

$h_1: d_1 \; d_2 \; d_3 \; d_4$

$h_2: d_1 \; d_2 \; d_3 \; d_4$

$h_3: d_3 \; d_1 \; d_2 \; d_4$

$d_1: h_3 \; h_1 \; h_2$

$d_2: h_2 \; h_1 \; h_3$

$d_3: h_1 \; h_3 \; h_2$

$d_4: h_1 \; h_2 \; h_3$

Deviation

$h_1: d_1 \; d_4$

$h_2: d_1 \; d_2 \; d_3 \; d_4$

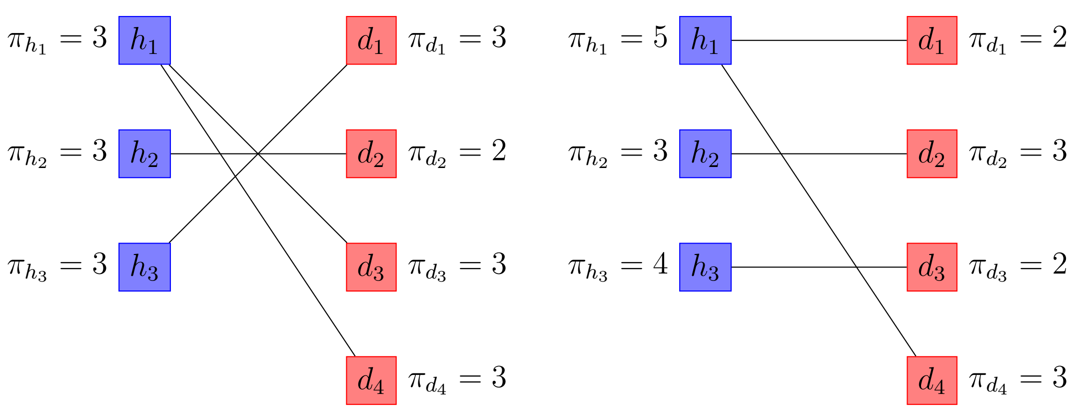
$h_3: d_3 \; d_1 \; d_2 \; d_4$

$d_1: h_3 \; h_1 \; h_2$

$d_2: h_2 \; h_1 \; h_3$

$d_3: h_1 \; h_3 \; h_2$

$d_4: h_1 \; h_2 \; h_3$



Roth's proposition applies to 0-markets. In ∞-markets, however, there is an efficient and strategy-proof mechanism: VCG. The VCG mechanism is defined below. First we define the coalitional value function.

Definition 2.6. $V(S) = \max_{Mt \in M_0(S)} \sum_{a \in S} \pi_a^{Mt}$ is the **coalitional value function.**

With the coalitional value function at hand, we define the VCG mechanism for a set of agents B.

Definition 2.7. The **VCG for agents in** B works as follows. It is assumed that the mechanism knows u_a for every $a \in A \setminus B$. Agent $b \in B$ sends u_b to the mechanism and an outcome $(Mt)^* \in \underset{Mt \in M_0(A)}{argmax} \sum_{a \in A} \pi_a^{Mt}$ is implemented. Agent b is charged a payment $p_b = -\sum_{a \in A \setminus b} \pi_a^{(Mt)^*} + W(A \setminus b)$. Where W is a coalitional value function with the reported and known preferences. Let $P = \sum_{b \in B} p_b$ be the total collected payments, let $\{t_a\}_{a \in A \setminus B}$ be a set of real numbers such that $P = \sum_{a \in A \setminus B} t_a$. Then the payoff for agent $a \in A \setminus B$ is given by $\pi_a^{(Mt)^*} + t_a$. Whenever there is a set of transfers $\{t_a\}_{a \in A \setminus B}$ such that the final payoffs are stable, those transfers are implemented.

Lemma 2.1. *The VCG mechanism for agents in B is strategy-proof for agents in B.*

If agents in B play their dominant strategy, then every agent $b \in B$ receives a payoff equal to $\pi_b = V(A) - V(A \setminus b)$ (agent b's VCG payoff). We assume agents always play their dominant strategy. Notice that VCG payoffs for members of B only depend on the value of the coalitional value function and not on a particular efficient allocation i.e. VCG payoffs and payments are well-defined even when there are multiple optimal allocations. Furthermore, notice that the only condition on payments (and payoffs) for members of $A \setminus B$ is budget balancedness. Whenever there is a set of transfers transfers $\{t_a\}_{a \in A \setminus B}$ such that the final payoffs are stable, we say that the VCG for set B is stable. Example 2.2 shows that stability for hospitals is possible in some cases while Example 2.3 shows that it is not always possible.

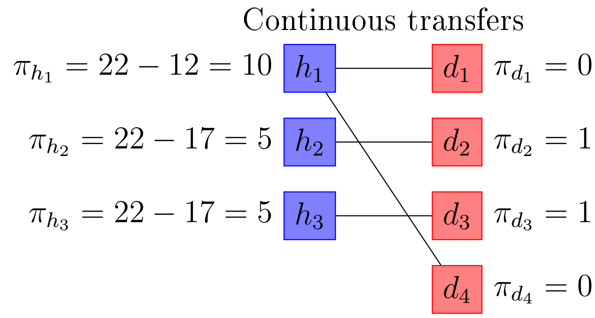
Example 2.2. Suppose we have the market of example 2.1, but transfers can be continuously adjusted. Then, VCG for hospitals is strategy-proof and stable.

Hospitals

	d_1	d_2	d_3	d_4	q
h_1	4	3	2	1	2
h_2	4	3	2	1	1
h_3	3	2	4	1	1

Doctors

	d_1	d_2	d_3	d_4
h_1	2	2	3	3
h_2	1	3	1	2
h_3	2	1	2	1



In the previous example, doctors payoffs are obtained after solving a system of equations together with some inequalities. For instance, h_2 is matched to d_2, together they generate a surplus of 6 and h_2's VCG payoff is 5, then h_2's payoff must be 1. The question is if, in general, the payments collected by the mechanism can be distributed to doctors in a stable way.

In auction theory, Ausubel and Milgrom (2002) have shown that the substitutes condition is sufficient for the VCG payoffs to be stable, i.e. all collected payments by the VCG mechanism can be paid to the auctioneer and the outcome is stable. Unfortunately, in matching markets, the substitutes condition is not sufficient to obtain the same result. Consider the following example.

Example 2.3. There are two identical hospitals and three identical doctors. $u_h(A) = 0$ if $|A| = 0$,$u_h(A) = 10$ if $|A| = 1$, $u_h(A) = 18$ if $|A| = 2$, $u_h(A) = 20$ if $|A| = 3$. $u_d(\emptyset) = u_d(h) = 0$. In this market, both hospitals receive a payoff of $\pi_h = V(A) - V(A/h) = 28 - 20 = 8$ when the VCG mechanism is used. However, in the unique stable allocation both hospitals receive a payoff $\pi_h = 2$ and all doctors receive a payoff $\pi_d = 8$.

If we want VCG payoffs for Hospitals to agree with their maximum stable payoff, $\pi_h^{Mt} = V(A) - V(A \setminus h)$ for every hospital, we need to be able to construct a payoff equivalent (for all agents different than h) allocation in the market with agents A/h. In particular, if doctor d is matched to hospital h in the market with A agents, then we need to find an allocation with $A \setminus h$ agents that provides d with π_d^{Mt}. The natural candidate is one of his blocking coalitions, i.e. one of the coalitions that would be formed if he were offered anything less than π_d^{Mt}. Unfortunately, it is possible that a hospital h' belongs to a blocking coalition with some doctor d and a different (incompatible) coalition with d'. This is illustrated in the following example.

Example 2.4. Consider a market with two hospitals and three doctors. The matrix contains the joint surpluses. It is assumed that both groups assign a value of zero to being unassigned. t and s are a real numbers.

Hospitals

	d_1	d_2	d_3	q
h_1	3	4	0	2
h_2	2	3	1	1

Hospital-optimal stable payoffs

$\pi_{h_1} = 4 \quad \pi_{d_1} = 1$

$\pi_{h_2} = 1 \quad \pi_{d_2} = 2$

$\pi_{d_3} = 0$

VCG for Hospitals payoffs

$\pi_{h_1} = 5 \quad \pi_{d_1} = 2 + s + t$

$\pi_{h_2} = 1 \quad \pi_{d_2} = -t$

$\pi_{d_3} = -s$

If $\pi_{d_1} < 1$, then d_1 and h_2 form a blocking coalition.

If $\pi_{d_2} < 2$, then d_2 and h_2 form a blocking coalition.

However, in the market without h_1, h_2 cannot honor his blocking offers with d_1 and d_2 simultaneously.

The previous example shows that the substitutes condition is not enough to guarantee the equivalence between the maximum stable payoff and the VCG payoff for hospitals. There are, however, some regularities that the substitutes condition can provide. Proposition 2.2 states well-known results that are used in our discussion and proof of Theorem 2.1.

Proposition 2.2. *In all matching markets,*

- $\pi_d^{Mt_D} = V(A) - V(A \setminus d)$ *for every* $d \in D$. *Leonard (1983)*

- $\pi_d^{Mt_H} = V(dA) - V(A)$ *for every* $d \in D$. *Gul and Stacchetti (1999). Where* $V(Ad)$ *is the value of a market where an identical doctor* d *is added.*

Proposition 2.2 characterizes doctors' payoffs at the hospital-optimal and doctor-optimal matchings. Furthermore, it implies the strategy-proofness and stability of VCG for doctors; as it shows that the VCG payoff for doctors is equal to their doctor-optimal stable payoff. In this section we provide an analogous result for hospitals. Example 2.4 shows that this cannot be achieved in general, however, we show that if preferences satisfy a joint restriction we call the **opportunity cost condition**, then the equivalence can be guaranteed.

Definition 2.8. V, the coalitional value, function satisfies the **opportunity cost condition** if and only if for all $d \in D$ and $h \in H$, $V(A \setminus h) - V(A \setminus hd) \geq V(Ad) - V(A)$.

On the right hand side of the inequality we have $V(Ad) - V(A)$, this is the opportunity cost of doctor d in the optimal assignment. Intuitively, if a new copy of doctor d were added to the market, the copy would go to the second highest value allocation, as the highest value is occupied by the original d. On the left hand side of the inequality we have $V(A \setminus h) - V(A \setminus hd)$, this is the marginal value of d in a market where h is not present. When the opportunity cost holds, every doctor can find a hospital that offers him at least his opportunity cost.

Theorem 2.1. *Suppose* V *satisfies the opportunity cost condition, then* $\pi_h^{Mt_H} = V(A) - V(A \setminus h)$ *for every* $h \in H$ *in the continuous transfers market.*

Corollary 2.1. *Under theorem 2.1 assumptions the VCG for Hospitals is stable and strategy-proof.*

Theorem 2.1 is completely analogous to that of Leonard (1983). When preferences satisfy the opportunity cost condition it is possible to use VCG for hospitals as a strategy-proof and stable mechanism. The opportunity cost condition is a joint condition on preferences, however, there are some individual preferences that imply it. For instance, linear preferences and unit demand preferences.

Definition 2.9. Hospital preferences **are linear** if and only if for all $h \in H$, $c_h = |H|$ and for all $D_h \subset D$, $u_h(D_h) = \sum_{d \in D_h} u_h(d)$. Hospitals preferences are of **unit demand** if for all $h \in H$, $c_h = 1$.

Lemma 2.2. *Unit demand and linear hospital preferences satisfy the opportunity cost condition.*

2.3 Strategy-Proofness for Hospitals And Doctors in the continuous market

We begin this section with another impossibility result due to Roth (1982).

Proposition 2.3. ***Roth 82.*** *There is no stable and strategy-proof mechanism for hospitals and doctors in the 0-market, even when hospitals have unit demands.*

Example 2.5. Consider a market with two hospitals and two doctors. Hospitals and Doctors have cardinal valuations over each other. The left matrix contains the values each hospital assigns to every doctor. All hospital are willing to hire at most one doctor. The right matrix contains the values for doctors. It is assumed that both groups assign a value of zero to being unassigned

Continuous transfers

Hospitals

	d_1	d_2
h_1	2.5	1
h_2	3	4.5

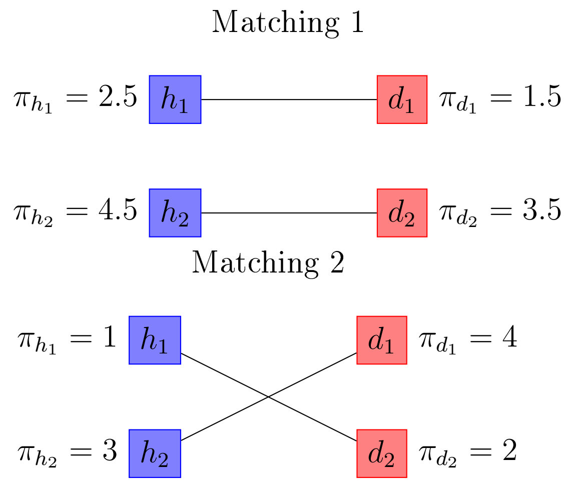
Doctors

	d_1	d_2
h_1	1.5	4
h_2	2	3.5

Suppose transfers are prohibited and a stable allocation is to be implemented. In this case, the ordinal representation is sufficient to characterize the set of stable matchings. Under the true preference profile, there are two stable matchings, shown on right. Suppose the bottom matching is chosen by the mechanism.

$h_1: \; d_1 \; d_2 \qquad d_1: \; h_2 \; h_1$

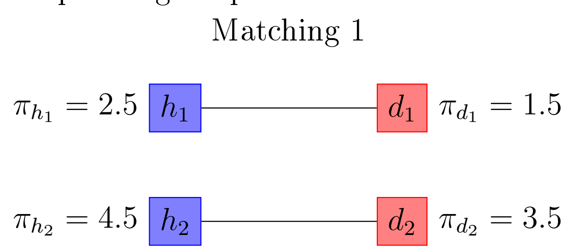
$h_2: \; d_2 \; d_1 \qquad d_2: \; h_1 \; h_2$



Then h_1 can manipulate the outcome by manipulating his preferences.

$h_1: \; d_1 \qquad d_1: \; h_2 \; h_1$

$h_2: \; d_2 \; d_1 \qquad d_2: \; h_1 \; h_2$



The mechanism provided by Theorem 2.1 cannot be used directly in this case to obtain strategy-proofness for all agents. In general, $\sum_{a \in A} V(A) - V(A \setminus a) > V(A)$. Hence, we need to reduce agents payoffs without losing strategy-proofness. Intuitively, agents needs to be able to capture the marginal value generated by their reports. In auction environments, this is precisely achieved by offering every agent a a payoff equal to $V(A) - V(A \setminus a)$. In matching environments, this is not the case. When a new agent enters a matching market, say a doctor d, two effects take place: (1) a new agent who demands hospitals enters the market and (2) a new object is available for hospitals currently in the market. $V(A) - V(A \setminus d)$ captures both marginal effects. In order to obtain strategy-proofness we only need the first one. Doctor d's information's marginal value is captured by $U(S) - U(S \setminus d)$, where $U(T)$ is the maximum value that can be achieved with all agents present only using the information of agents in T. The U function is defined below.

Definition 2.10. $U(S) = \max_{Mt \in M_0(A)} \sum_{a \in S} \pi_a^{Mt}$ is the **optimal matching function**

It is possible to define a second VCG mechanism using the U function.

Definition 2.11. The U-**VCG mechanism for agents in** B works as follows. It is assumed that the mechanism knows u_a for every $a \in A \setminus B$. Agent $b \in B$ sends u_b to the mechanism and an outcome $(Mt)^* \in \underset{Mt \in M_0(A)}{argmax} \sum_{a \in A} \pi_a^{Mt}$ is implemented. Agent b is charged a payment $p_b = -\sum_{a \in A \setminus b} \pi_a^{(Mt)^*} + W'(A \setminus b)$. Where W' is the optimal matching function with the reported and known preferences. Let $P = \sum_{b \in B} p_b$ be the total collected payments, let $\{t_a\}_{a \in A \setminus B}$ be a set of real numbers such that $P = \sum_{a \in A \setminus B} t_a$. Then the payoff for agent $a \in A \setminus B$ is given by $\pi_a^{(Mt)^*} + t_a$. Whenever there is a set of transfers $\{t_a\}_{a \in A \setminus B}$ such that the final payoffs are stable, those transfers are implemented.

Lemma 2.3. *The U-VCG mechanism for agents in B is strategy-proof.*

If agents in B play their dominant strategy, then every agent $b \in B$ receives a payoff equal to $\pi_b = U(A) - U(A \setminus b)$. We assume agents always play their dominant strategy. The difference between the VCG and the U-VCG mechanisms is the last term in their payoffs; $U(A/b) = \max_{Mt \in M_0(A)} \sum_{a \in A/b} \pi_a^{Mt}$ while $V(A/b) = \max_{Mt \in M_0(A/b)} \sum_{a \in A/b} \pi_a^{Mt}$. If we apply the U-VCG to the previous example we achieve stability and strategy-proofness for all agents in the market.

Example 2.6. Consider the market of example 2.5 and use the U-VCG mechanism.

Hospitals

	d_1	d_2
h_1	2.5	1
h_2	3	4.5

Doctors

	d_1	d_2
h_1	1.5	4
h_2	4.5	3.5

The payoff for h_1 is calculated as follows:

$\pi^U_{h_1} = U(A) - U(A \setminus h_1) = 2.5$

$U(A) = \max_{Mt \in M_0(A)} \sum_{a \in S} \pi^{Mt}_a = V(h_1 d_1) + V(h_2 d_2)$

$= 4 + 8 = 12$

$U(A \setminus h_1) = \max_{Mt \in M_0(A)} \sum_{a \in A \setminus h_1} \pi^{Mt}_a = u_{d_1}(h_1) + V(h_2 d_2)$

$= 1.5 + 8 = 9.5$

Analogously,

$\pi^U_{d_1} = 12 - 10.5 = 1.5$; $\pi^U_{h_2} = 12 - 7.5 = 4.5$;

$\pi^U_{d_2} = 12 - 8.5 = 3.5$

It is routine to check that the payoffs in example 2.6 are stable. It is instructive to compare the U-VCG payoffs with the hospital-optimal and doctor-optimal stable matchings. In particular, $\pi^{Mt_H}_{h_1} = 4$, $\pi^{Mt_H}_{h_2} = 7$, $\pi^{Mt_H}_{d_1} = 0$ and $\pi^{Mt_H}_{d_2} = 1$ for the hospital-optimal stable matching and $\pi^{Mt_D}_{h_1} = 0$, $\pi^{Mt_D}_{h_1} = 1$, $\pi^{Mt_H}_{d_1} = 4$ and $\pi^{Mt_H}_{d_2} = 7$ for the doctor-optimal stable matching. The U-VCG mechanism is achieving strategy-proofness and stability without offering either doctors or hospitals their most preferred stable matching. In the following section we study this property of the U-VCG mechanism. In the previous example, the U-VCG delivers stable payoffs for all agents. In general, this is not the case. Consider the following example.

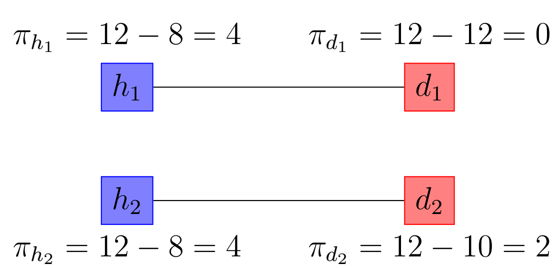
Example 2.7. The U-VCG mechanism is not stable for this market, as it removes resources from the market.

Hospitals

	d_1	d_2
h_1	4	5
h_2	4	4

Doctors

	d_1	d_2
h_1	0	0
h_2	1	4



Both markets in example 2.6 and 2.7 share the same coalitional function, hence they have the same set of stable payoffs. As it can be observed in the examples, the division of surplus between agents plays a fundamental role in the stability of the

U-VCG mechanism. The next definition formalizes the idea of a division of surplus.

Definition 2.12. Let $\{u_a\}_A$ be a matching market with hospitals H, doctors D and $A = HD$. Consider any other market with the same set of agents A, but different preferences $\{u'_a\}_A$. We say that $\{u_a\}_A$ is a **division** of $\{u'_a\}_A$ if and only if for all $D_h \subset D$ and h we have $u_h(D_h) + \sum_{d \in D_h} u_d(h) = u'_h(D_h) + \sum_{d \in D_h} u'_d(h)$ and for all $d \in D$ $u_d(\emptyset) = u'_d(\emptyset)$.

Notice that if $\{u_a\}_A$ is a division of $\{u'_a\}_A$, then the opposite is also true i.e. divisions form an equivalence relation in the set of preferences profiles. Furthermore, as lemma 2.4 establishes, divisions have no impact on the set of stable matchings, which only depend on the coalitional value function.

Lemma 2.4. *Let $\Pi \subset \mathbb{R}^A$ be the set of payoffs arising from stable allocations in market $\{u_a\}_A$. Let $\{u'_a\}_A$ be a division of $\{u_a\}_A$ and let $\Pi' \subset \mathbb{R}^A$ be its set of stable payoffs. Then $\Pi = \Pi'$.*

Whereas the VCG mechanism, and its stability for hospitals, only depends on the coalitional value function, the stability of the U-VCG mechanism depends on the particular division of surplus in the market. Consider examples 2.6 and 2.7. They share the same coalitional value function and set of stable matchings and payoffs. However, when the U-VCG mechanism is used to elicit preferences, the resulting payoffs are stable only for the division in example 2.6. The following theorem shows that the payoffs delivered by the U-VCG mechanism are stable for at least one representative of each equivalence class in the space of preference profiles.

Theorem 2.2. *Let $\pi \in \Pi$, then there is a division such that $\pi^U_a = \pi_a$, i.e. there is a division of surplus such that the U-VCG mechanism is stable and strategy-proof for all agents.*

Theorem 2.2 does not say that the U-VCG mechanism delivers stable payoffs in every market. However, for every market there is a division of surplus that would make the payoffs delivered by the U-VCG mechanism stable. In other words, for every fixed set of agents A and coalitional function V there is market characterized by utility functions on $\{u_a\}_A$ such that the U-VCG mechanism would deliver stable payoffs in that market.

Theorem 3 describes the conditions under which utility functions $\{u_a\}_A$ produce a market for which the U-VCG mechanism is produces stable payoffs. We first introduce the concept of pivotal information. We say that agent i's information is not pivotal if when his information is disregarded, but he is still considered part of the matching market, the optimal allocation does not change.

Definition 2.13. Let $(Mt)^* \in \underset{Mt \in M_0(A)}{argmax} \sum_{a \in A} \pi_a^{Mt}$. Agents i's information is **not pivotal**, with respect to $(Mt)^*$, whenever $(Mt)^* \in \underset{Mt \in M_0(A)}{argmax} \sum_{a \in A \setminus \{i\}} \pi_a^{Mt} = \emptyset$.

Theorem 2.3. *Let* $(Mt)^* \in \underset{Mt \in M_0(A)}{argmax} \sum_{a \in A} \pi_a^{Mt}$ *and let* $U_a^{(Mt)^*}$ *be agent a's private value in* $(Mt)^*$. *Suppose* V *satisfies the opportunity cost condition, no agent's information is pivotal with respect to* $(Mt)^*$ *and* $(U_a^{(Mt)^*})_A \in \Pi$, *then the U-VCG mechanism is strategy-proof for all agents and stable.*

If any agent's information is pivotal, then U-VCG will collect a positive payment from this agent. If the U-VCG is used to elicit preferences from only one side of the market, then there is the possibility (studied in the next section) of redistributing the payments to the other side to maintain all the surplus in the market. When eliciting preferences from both sides of the market, this possibility disappears. Furthermore, any agent whose information is not pivotal will have a payoff equal to his private surplus at the chosen allocation. Unfortunately, both conditions are independent. The next example shows that there are non-pivotal markets where $(U_a)_A \notin \Pi$ and pivotal markets where $(U_a)_A \in \Pi$.

Example 2.8. On the left, there is a non-pivotal market where $(U_a)_A \notin \Pi$. On the right, there is a pivotal market where $(U_a)_A \in \Pi$.

Hospitals

	d_1	d_2
h_1	3	4
h_2	0	1

Doctors

	d_1	d_2
h_1	2	4
h_2	0	4

Hospitals

	d_1	d_2
h_1	4	2
h_2	0	0

Doctors

	d_1	d_2
h_1	0	1
h_2	0	0

The U-VCG payments can be modified in a market where no agent's information is pivotal but $(U_a)_A \notin \Pi$. Specifically, let Mt be any stable matching and let t_d be the minimum salary allowed for doctor d, i.e. regardless of the hiring hospital, d will charge at least t_d. Of course, this will impact the true preferences in the market, as now, doctor d will have preferences $u'_d(h) = u_d(h) + t_d$ and hospital h will have preferences $u'_h(D_h) = u_h(D_h) - \sum_{d \in D_h} t_d$. With these new preferences, the market is non-pivotal and $(U'_a)_A \in \Pi$.

2.4 Non-extremal Strategy-Proofness.

In this section we study the one-sided U-VCG. As a first motivation, we consider example 2.6 from the previous section. It can be observed that the U-VCG mechanism achieves strategy-proofness for all agents and stability, but does **not** depend on offering any side their most preferred stable allocation. Example 2.9 shows that strategy-proofness and stability can be achieved in the interior of the set of stable payoffs.

Example 2.9. Strategy-proofness can be obtained in the interior of the set of stable payoffs.

Hospitals

	d_1	d_2
h_1	2.5	1
h_2	3	4.5

Doctors

	d_1	d_2
h_1	1.5	4
h_2	2	3.5

The U-VCG uses the following values:

$U(A) = V(h_1d_1) + V(h_2d_2) = 4 + 8 = 12$

$U(A\backslash h_1) = u_{d_1}(h_1)+V(h_2d_2) = 1.5+8 = 9.5$

$U(A\backslash h_2) = V(h_1d_1)+u_{d2}(h_2) = 4+3.5 = 7.5$

$U(A \setminus d_1) = u_{h_1}(d_1) + V(h_2d_2) = 2.5 + 8 = 10.5$

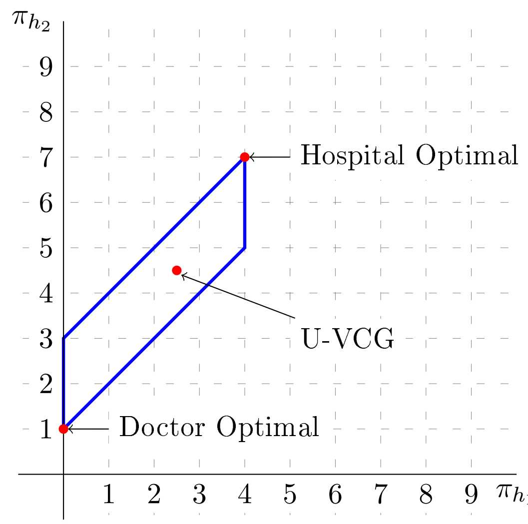
$U(A\backslash d_2) = V(h_1d_1)+u_{h_2}(d_2) = 4+4.5 = 8.5$

Hence,

$\pi^U_{h_1} = 12 - 9.5 = 2.5$; $\pi^U_{h_2} = 12 - 7.5 = 4.5$;

$\pi^U_{d_1} = 12 - 10.5 = 1.5$; $\pi^U_{d_2} = 12 - 8.5 = 3.5$

Hospital Stable Payoffs



We know that the U-VCG mechanism is not stable in the market of example 2.7. However, the following example shows that the U-VCG for doctors is stable and strategy-proof for doctors.

Example 2.10. One-sided strategy-proofness can be obtained without offering that side their most preferred stable payoff.

Hospitals

	d_1	d_2
h_1	4	5
h_2	4	4

Doctors

	d_1	d_2
h_1	0	0
h_2	1	4

The U-VCG uses the following values:

$U(A) = V(h_1d_1) + V(h_2d_2) = 4 + 8 = 12$

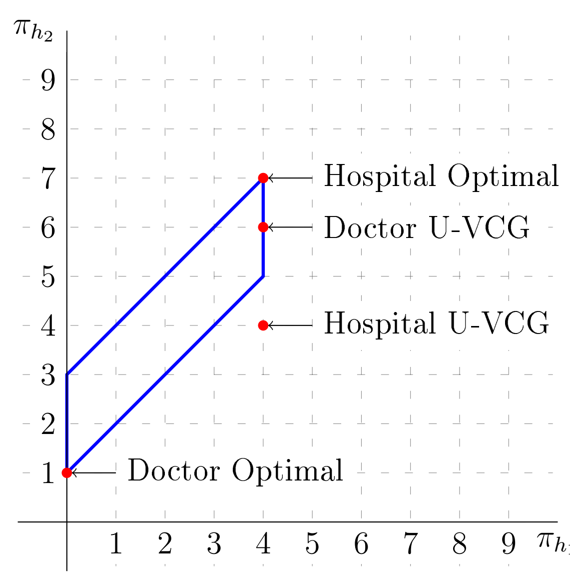
$U(A\backslash h_1) = u_{d_1}(h_1)+V(h_2d_2) = 0+8 = 8$

$U(A\backslash h_2) = V(h_1d_1)+u_{d2}(h_2) = 4+4 = 8$

$U(A\backslash d_1) = u_{h_1}(d_1)+V(h_2d_2) = 4+8 = 12$

$U(A\backslash d_2) = u_{h_1}(d_2)+V(h_2d_1) = 5+5 = 10$

Hospital Stable Payoffs



Hence the U-VCG for doctors delivers:

$\pi_{h_1} = V(h_1 d_1) - 0 = 4$;

$\pi_{h_2} = V(h_2 d_2) - 2 = 6$;

$\pi_{d_1} = 12 - 12 = 0$; $\pi_{d_2} = 12 - 10 = 2$

Hence the U-VCG for hospitals delivers:

$\pi^U_{h_1} = 12 - 8 = 4$; $\pi^U_{h_2} = 12 - 8 = 4$;

$\pi_{d_1} = V(h_1 d_1) - 4 = 0$;

$\pi_{d_2} = V(h_2 d_2) - 4 = 4$

As noted above, the performance of the U-VCG mechanism depends on the division of surplus between both sides of the market. In this section, a special class of division for which the U-VCG for hospitals is stable and strategy-proof are studied. It is assumed that preferences are linear. When preferences are linear, surplus can be moved freely from one side to the other. In particular, divisions of the form $u_h(D_h) = \alpha(u_h(D_h) + \sum_{d \in D_h} u_d(h))$ for all $\alpha \in [0,1]$ are well defined.

Definition 2.14. Let $\alpha \in [0,1]$. We say that the hospital and doctors preferences forms an α **division** if $u_h(D_h) = \alpha(u_h(D_h) + \sum_{d \in D_h} u_d(h))$ for all h and $D_h \in D$.

For a fixed alpha, let $\pi^U_h(\alpha)$ the payoff assigned to hospital h by the hospital U-VCG mechanism. The following lemma describes a few properties of π^U_h.

Lemma 2.5. *Let $\pi^U_h(\alpha)$ the payoff assigned to hospital h by the hospital U-VCG mechanism, then π^U_h is a monotone increasing piece-wise linear function satisfying:*

- $\pi^U_h(0) = 0$ *and* $\pi^U_h(1) = V(A) - V(A \setminus h)$
- $\frac{d}{d\alpha}\pi^U_h(\alpha) \geq \alpha(V(A) - V(A \setminus h))$ *for all* $\alpha \in [0,1]$
- $\frac{d}{d\alpha}\pi^U_h(0) = V(hD_h)$ *and* $\frac{d}{d\alpha}\pi^U_h(0) = 0$*, where D_h is the set of doctors assigned to h at the efficient matching.*

The following theorem establishes several conditions for the one-sided U-VCG mechanism to deliver stable payoffs in markets characterized by α divisions.

Theorem 2.4. *Suppose preferences are linear and Π has a non-empty interior. Let h_d be the hospital matched with d at a fixed efficient matching.*

- *If the lowest stable payoff for every hospital is zero and* $V(h_d d) > V(h'd)$ *for all* d *and* $h' \neq h_d$*, then there is* $\beta \in (0,1)$ *such that for any* $\alpha \in [0, \beta]$ $\pi_h^U(\alpha)$ *for all* $h \in H$ *is stable.*

- *For every doctor there are at least two hospitals* h, h' *such that* $V(hd) > 0$ *and* $V(h'd) > 0$*, then there is* $\beta \in (0,1)$ *such that for any* $\alpha \in [\beta, 1]$ $\pi_h^U(\alpha)$ *for all* $h \in H$ *is stable.*

Suppose that in a particular market all the surplus is generated by hospitals. In this case, hospitals' preferences could be elicited using the VCG for hospitals mechanism and the hospital-optimal matching would be implemented in a strategy-proof manner. In this case every hospital h would receive $\pi_h^U(1) = V(A) - V(A \setminus h)$ as a payoff and doctors would receive their minimal stable payoff. Now suppose that in that market, a small quantity of surplus is shifted from hospitals; for example by offering some wages to doctors. In this new market with wages, the set of stable payoffs has not changed as $V(hd)$ remains unchanged by the wage offered by h. In this market, using VCG for hospitals would still be possible, however, according to theorem 2.4 if the U-VCG for hospitals mechanism is used instead, then an interior solution would be achieved i.e. hospitals would receive less than their maximum stable payoffs and doctors would receive more than their minimum stable payoffs.

2.5 Hospital Strategy-proofness in the q-Market

Roth's theorems show that, in the 0-market, it is not possible to achieve strategy-proofness for hospitals or all agents simultaneously. In sharp contrast, in the ∞-market, strategy-proofness and stability for hospitals is possible. Which model describes better a particular real market depends on the institutional environment. In school choice, transfers are completely prohibited whereas, in residents matching, the monetary component of any transaction is a fundamental part. In markets where transactions are allowed, the level of discreteness of transfers is a market design variable. The following theorem shows that, if strategy-proofness for hospitals is important in a particular market, there is always a discrete level of transactions that achieves it, i.e. transfers only need to be as small as the smallest common factor between the set of possible valuations.

Theorem 2.5. *Suppose preferences are integer valued and V satisfies the opportunity cost condition, then strategy-proofness and stability are possible in any q-market for any $q \in \mathbb{Z}$ such that $q \geq 1$.*

When a stable allocation is to be implemented in a market where discrete quantities of money can be exchanged the generalized Gale-Shapley mechanism is the standard solution. For example, in the formed proposed by Kelso and Crawford (1982). One important limitation of this mechanism is that it is not strategy-proof for hospitals. Theorem 2.1 establishes that, when preferences satisfy the opportunity cost condition, and monetary transfers can be continuously adjusted, the VCG for hospitals is stable and strategy-proof. These two properties can be extended to markets with discrete quantities of money by using a lowest common denominator fraction i.e. the smallest tradable quantity of money is sufficiently small to offset the smallest change in valuations. One very simple way of expressing this idea is assuming that preferences are integer-valued. Thus in a market in which preferences are expressed in thousands of dollars, e.g. the value of a one year contract between a hospital and a doctor is in thousands of dollars, a stable allocation can be achieved using the VCG for hospitals mechanism instead of the Gale-Shapley mechanism.

2.6 Conclusion

In 2002, 16 law firms filed a class action law suit, representing 3 former residents seeking to represent all residents, arguing that the NRMP violated antitrust laws and was a conspiracy to depress resident's wages. The complaint was:

> Defendants and others have illegally contracted, combined and conspired among themselves to displace competition in the recruitment, hiring, employment and compensation of resident physicians, and to impose a scheme of restraints which have the purpose and effect of fixing, artificially depressing, standardizing and stabilizing resident physician compensation and other terms of employment.
>
> Defendants' illegal combination and conspiracy has restrained competition in the employment of resident physicians by:
>
> (a) stabilizing wages below competitive levels by exchanging competitively sensitive information regarding resident physician compensation and other terms of employment;
>
> (b) eliminating competition in the recruitment and employment of resident physicians by assigning prospective resident physician employees to positions through the National Resident Matching Program ("NRMP"); and
>
> (c) establishing and complying with anticompetitive accreditation standards and requirements through the Accreditation Council for Graduate Medical Education ("ACGME").

The suit was dismissed on August 12, 2004 in an Opinion, Order & Judgment by Judge Paul L. Friedman based on evidence regarding the structure of wages in other decentralized industries and expert opinions from several economists.

To this date no wage negotiation takes place in the NRMP, however, this paper shows that allowing transfers to be negotiated in the matching process not only would "enhance competition", it would also make agents more willing to report their true private information. As argued in the introduction, this would increase the efficiency and accountability of the whole program.

2.7 Appendix

This appendix contains the proofs for the theorems stated in the paper. In order to relate the results of this paper to the literature on VCG auctions we first introduce establish a connection between matching markets and auction markets. The connection establishes that for any stable matching there is an equivalent Walrasian equilibrium, defined as follows.

Definition 2.15. Let $\{u_a\}_A$ be a matching market with hospitals H and doctors D. Suppose that for all $d \in D$ we have $u_d(h) = 0$ for all $h \in H$. A vector $t \in \mathbb{R}^D$ is a **Walrasian Equilibrium Price** (WEP) if and only if there is $D_h \in T(t)$ such that $\bigcup_H D_h = D$ and for all $h \neq h'$ $D_h \bigcap D_{h'} = \emptyset$. $(\{D_h\}_H, t)$ is a **Walrasian Equilibrium** (WE).

We first consider a couple of lemmas.

Lemma. 2.4 *Let $\Pi \subset \mathbb{R}^A$ be the set of payoffs arising from stable allocations in market $\{u_a\}_A$. Let $\{u'_a\}_A$ be a division of $\{u_a\}_A$ and let $\Pi' \subset \mathbb{R}^A$ be its set of stable payoffs. Then, $\Pi = \Pi'$.*

Proof. By definition, $\pi \in \Pi$ if and only if $\sum_C \pi_a \geq V(C)$ for all $C \subset A$ and $\sum_A \pi_a = V(A)$. By definition of a division, $V(C) = V'(C)$ for all $C \subset A$, hence $\Pi = \Pi'$. To see this, consider for example $V(hD_h) = \underset{Mt \in M_0(hD_h)}{max} \sum_{a \in hD_h} u_a(Mt_a)$ and let $D^* \subset D_h$ such that $V(hD_h) = V(hD^*)$ i.e. D^* is the optimal subset of doctors among D_h.

Then by definition of a division, for any $D \subset D_h$ we have $u_h(D) + \sum_{d \in D^*} u_d(h) = u'_h(D) + \sum_{d \in D} u'_d(h)$, in particular we have $u_h(D^*) + \sum_{d \in D^*} u_d(h) = u'_h(D) + \sum_{d \in D^*} u'_d(h) = V(hD^*) = V'(hD^*) = V'(hD_h) = V(hD_h)$. □

Lemma 2.6. *Let $\Pi \subset \mathbb{R}^A$ be the set of payoffs arising from stable allocations in market $\{u_a\}_A$. Let $\{u'_a\}_A$ be a division of $\{u_a\}_A$ such that $u'_d(h) = 0$ for all $h \in H$. Then $(\pi_H, \pi_D) = \pi \in \Pi$ if and only if $\pi_h = v'_h(\pi_D)$ and π_D is a Walrasian Equilibrium Price in the division $\{u'_a\}_A$.*

Proof. Since $(\pi_H, \pi_D) = \pi \in \Pi$ implies that $\pi_h = V(hD_h) - \sum_{D_h} \pi_d$ if h is matched to D_h and $\pi_h \geq V(hD'_h) - \sum_{D'_h} \pi_d$ for every $D'_h \subset D$, then $\pi_h = v'_h(\pi_D)$ and π_D is a Walrasian Equilibrium Price in the division $\{u'_a\}_A$. □

For completion, as the following results are well-known, we include the proofs for Proposition 2.

Proposition. 2.2 *In all matching markets,*

- $\pi_d^{Mt_D} = V(A) - V(A \setminus d)$ *for every $d \in D$.* Leonard (1983)
- $\pi_d^{Mt_H} = V(dA) - V(A)$ *for every $d \in D$.* Gul and Stacchetti (1999). *Where $V(Ad)$ is the value of a market where an identical doctor d is added.*

Proof. We show that $\pi_d^{Mt_D} = V(A) - V(A \setminus d)$ for every $d \in D$. Let Mt_D be the doctors optimal matching. Fix d, if d is unassigned in Mt_D, then the result follows. If d is assigned to hospital h, let $M't'_D$ be a matching where everything is identical to Mt_D but t_d and d are removed. In this market, h holds whatever doctors he is assigned at the current transfers since h preferences satisfy the substitutes condition. Since the new allocation is stable, it is efficient. Hence $\sum_{A \setminus d} \pi_a^{Mt_D} = V(A \setminus d)$. If $u_h(D_h \setminus d) - \sum_{D_h \setminus d} \pi_d^{Mt_D} < u_h(D_h) - \sum_{D_h} \pi_d^{Mt_D}$, then d can increase his payoff in the original market, contradicting the maximality of $\pi_d^{Mt_D}$. Thus $\pi_d^{Mt_D} = V(A) - V(A \setminus d)$.

To show that $\pi_d^{Mt_H} = V(Ad) - V(A)$ for every $d \in D$. Consider a market with agents A plus a copy of doctor d and a new hospital h'. Let $u_{h'}(d) < V(Ad) - V(A)$ and zero otherwise. In this new market h' does not get d (or his copy) in any stable allocation and hence $\pi_d^{Mt_H} \geq V(Ad) - V(A)$ since $(\pi_D^{Mt_H})$ constitute a Walrasian Equilibrium. Suppose now that $u_{h'}(d) > V(Ad) - V(A)$ and zero otherwise. Now in every stable allocation h' gets d (or his copy) and hence $\pi_d \leq V(Ad) - V(A)$ for any stable allocation in the economy $Ah'd$. Since any stable allocation in $Ah'd$ induces a stable allocation in A we have $\pi_d^{Mt_H} \leq V(Ad) - V(A)$. Hence $\pi_d^{Mt_H} = V(Ad) - V(A)$. □

Theorem. 2.1 *Suppose V satisfies the opportunity cost condition in the continuous transfers market, then $\pi_h^{Mt_H} = V(A) - V(A \setminus h)$ for every $h \in H$.*

Proof. We first show that $\pi_h^{Mt_H} \leq V(A) - V(A \setminus h)$ for all $h \in H$. Without loss of generality, consider the division where all surplus belongs to hospitals. To show that $\pi_h^{Mt_H} \leq V(A) - V(A \setminus h)$ for all $h \in H$. Let h be any hospital and let $\pi_h^{Mt_H}$ be its maximum stable payoff, then $\pi_h^{Mt_H} = V(A) - \sum\limits_{A \setminus h} \pi_a^{Mt_H}$. Since the hospital preferred stable allocation belongs to the core, we have $\sum\limits_{A \setminus h} \pi_a^{Mt_H} \geq V(A/h)$. Hence the result. For the converse, without loss of generality, consider the division where all surplus belongs to hospitals. For notational simplicity let s^* be the set of doctors for hospital s at the **doctor** optimal stable allocation when hospital h is not present and let let s_* be the set of doctors for hospital s at the hospital optimal stable allocation when hospital h is present. Let π^* and π_* be the corresponding prices. By the optimality of s^* we have $v_s(\pi^*) \geq u_s(s_*) - \sum\limits_{s_*} \pi_d^*$, this implies that $v_s(\pi^*) \geq u_s(s_*) - \sum\limits_{s_*} \pi_{*d} + \sum\limits_{s_*} \pi_{*d} - \sum\limits_{s_*} \pi_d^* = v_s(\pi_*) + \sum\limits_{s_*} (\pi_{*d} - \pi_d^*)$ i.e $0 \geq v_s(\pi_*) - v_s(\pi^*) + \sum\limits_{s_*} (\pi_{*d} - \pi_d^*)$ for all s. We also have that $V(A) = \sum\limits_{H} v_s(\pi_*) + \sum\limits_{D} \pi_{*d}$ and $V(A/h) = \sum\limits_{H/h} v_s(\pi^*) + \sum\limits_{D} \pi_d^*$. Subtracting we have $V(A) - V(A/h) = \sum\limits_{H} v_s(\pi_*) + \sum\limits_{D} \pi_{*d} - \sum\limits_{H/h} v_s(\pi^*) - \sum\limits_{D} \pi_d^*$, reorganizing terms we have $V(A) - V(A/h) =$

$v_h(\pi_*) + \sum_{H\setminus h}[v_s(\pi_*) - v_s(\pi^*) + \sum_{s_*}(\pi_{*d} - \pi^*_d)] + \sum_{h_*}(\pi_{*d} - \pi^*_d)$. Since the second and third components are non-positive we have $V(A) - V(A/h) \leq v_h(\pi_*)$ (the third component is non-positive by the opportunity cost condition). □

Lemma. 2.1 *and* 2.3. *Both VCG and U-VCG for agents in B are strategy-proof for agents in B.*

Proof. Suppose agent $b \in B$ has preferences u_b and the reported and known preferences for other agents are $\{u_a\}_{A\setminus b}$. Then the VCG payoff for b when sending u_b is $u_b(M^*) + \sum_{a\in A\setminus b} u_a(M^*) - W(A \setminus b)$ and $u_b(M^{**}) + \sum_{a\in A\setminus b} u_a(M^{**}) - W(A \setminus b)$ when sending u'_b, where $(Mt)^* \in \underset{Mt\in M_0(A)}{argmax} \sum_{a\in A\setminus b} u_a(M) + u_b(M)$ and $(Mt)^{**} \in \underset{Mt\in M_0(A)}{argmax} \sum_{a\in A\setminus b} u_a(M) + u'_b(M)$. Since $u_b(M^*) + \sum_{a\in A\setminus b} u_a(M^*) \geq u_b(M^{**}) + \sum_{a\in A\setminus b} u_a(M^{**})$ we have that $u_b(M^*) + \sum_{a\in A\setminus b} u_a(M^*) - W(A\setminus b) \geq u_b(M^{**}) + \sum_{a\in A\setminus b} u_a(M^{**}) - W(A\setminus b)$. Hence, VCG is strategy-proof for b. The proof for U-VCG is analogous. □

Corollary. 2.1 *The VCG for Hospitals is stable and strategy-proof.*

Proof. The VCG for hospitals is always strategy-proof and and delivers playoffs equal to $V(A) - V(A \setminus h)$ for every $h \in H$. According to the previous theorem these payoffs are identical to the hospital-optimal stable payoffs. Thus, VCG is stable.

□

Corollary 2.2. *In auction markets, VCG payments coincide with the value of the assigned goods at the lowest Walrasian Equilibrium.*

Lemma. 2.2 *Unit demand and linear hospital preferences satisfy the opportunity cost condition.*

Proof. Suppose that $\pi_h^{Mt_H} = V(A) - V(A \setminus h)$ for all $h \in H$ and fix a hospital h^*, then there is a stable matching in the market without h^* in which all agents receive $\pi_a^{Mt_H}$. By construction for any $C \subset A \setminus h^*$, $\sum_C \pi_a^{Mt_H} \geq V(C)$ and $V(A \setminus$

$h^*) = \sum_{A \setminus h^*} \pi_a^{Mt_H}$. In the market without h^* doctor d has an optimal stable payoff of $V(A \setminus h^*) - V(A \setminus h^* d)$ (by Leonard's theorem) and $V(A \setminus h^*) - V(A \setminus h^* d) \geq \pi_d^{Mt_H} = V(Ad) - V(A)$, where the inequality comes from the optimality of the doctor-optimal stable payoff and the equality from Gul and Stacchetti's theorem. Leonard's theorem shows that $\pi_h^{Mt_H} = V(A) - V(A \setminus h)$ for all $h \in H$ in the unit demand case. For the linear case we show directly that $\pi_h^{Mt_H} = V(A) - V(A \setminus h)$ for all $h \in H$. Fix a hospital h^* and let B the set of doctors matched with h^*. Let every doctor in B who has a payoff equal to his outside option be unmatched. For every other $d \in B$, there is a hospital h', set of doctors A and $D \subset B$ such that $V(h'AD) = \sum_{a \in h'AB} \pi_a^{Mt_H}$. Let C be the set of doctors assigned to h'. We show that $V(h'CB) = \sum_{a \in h'CB} \pi_a^{Mt_H}$. $V(h'CB) = \sum_{a \in h'C} \pi_a^{Mt_H} + \sum_{a \in B} \pi_a^{Mt_H} + (\sum_{a \in A} \pi_a^{Mt_H} + \pi_{h'}^{Mt_H} - V(h'A))$, the term in brackets is non-negative by the stability of Mt_H and hence $V(h'CB) \geq \sum_{a \in h'C} \pi_a^{Mt_H} + \sum_{a \in B} \pi_a^{Mt_H}$ which together with the stability inequality $V(h'CB) \leq \sum_{a \in h'C} \pi_a^{Mt_H} + \sum_{a \in B} \pi_a^{Mt_H}$ imply the result. □

Theorem. 2.2 *Let $\pi \in \Pi$, then there is a division such that $\pi_a^U = \pi_a$, i.e. there is a division of surplus such that the U-VCG mechanism is stable and strategy-proof for all agents.*

Proof. Let x_a be the match of agent a in an efficient allocation. Let $u_a(x_a) = \pi_a$. For any h and set of doctors $D_h \neq x_h$ let $u_h(D_h) = V(D_h h) - \sum_{d \in D_h \cap x_h} \pi_d$. For any d and $h \neq x_d$ let $u_d(h) = 0$. With this division, the allocation is stable whenever any agent reports a zero value for every match, hence efficient. □

Theorem. 2.3 *Let U_a be agent a's private value in an efficient assignment. Suppose no agent's information is pivotal. i.e. removing one agent's information does not change the efficient assignment, and $(U_a)_A \in \Pi$, then the U-VCG mechanism is strategy-proof for all agents and stable.*

Proof. By construction, the U-VCG mechanism is strategy-proof for all agents, if no agents information is pivotal, then $U(A) - U(A \setminus a)$ is equal to the private value U_a for every agent, which by assumption belongs to Π. □

Lemma. 2.5 *Let $\pi_h^U(\alpha)$ the payoff assigned to hospital h by the hospital U-VCG mechanism, then π_h^U is a monotone increasing piecewise linear function satisfying:*

- $\pi_h^U(0) = 0$ *and* $\pi_h^U(1) = V(A) - V(A \setminus h)$
- $\frac{d}{d\alpha}\pi_h^U(\alpha) \geq \alpha(V(A) - V(A \setminus h))$ *for all* $\alpha \in [0,1]$
- $\frac{d}{d\alpha}\pi_h^U(0) = V(hD_h)$ *and* $\frac{d}{d\alpha}\pi_h^U(0) = 0$, *where* D_h *is the set of doctors assigned to* h *at the efficient matching.*

Proof. By definition $U(A/h)(\alpha) = \underset{Mt \in M_0(A)}{max} \sum_{a \in A/hD_h} u_a^{Mt} + (1-\alpha)V(hD_h)$. By the envelope theorem (see for example Milgrom and Segal (2002)) $U(A/h)(\alpha)$ is a piecewise linear function and $U(A/h)(\alpha)$ is monotone (since $V(hD_h) \geq 0$ for all h and D_h that are ever chosen at the optimum). Hence $\pi_h^U(\alpha) = U(A) - U(A \setminus h)(\alpha)$ is an increasing piecewise linear function such that $\pi_h^U(0) = 0$, $\pi_h^U(1) = V(A) - V(A \setminus h)$, $\frac{d}{d\alpha}\pi_h^U(0) = V(hD_h)$ where D_h is the set of doctors matched with h and $\frac{d}{d\alpha}\pi_h^U(1) = 0$. Since $\pi_h^U(0) = 0$ and $\frac{d}{d\alpha}\pi_h^U(0) = V(hD_h) \geq V(A) - V(A \setminus h)$ we have $\pi_h^U(\alpha) \geq \alpha(V(A) - V(A \setminus h))$ for all $\alpha \in [0,1]$. □

Theorem. 2.4 *Suppose preferences are linear and Π has a non-empty interior. Let h_d be the hospital matched with d at the efficient matching.*

- *If the lowest stable payoff for every hospital is zero and $V(h_d d) > V(h'd)$ for all d and $h' \neq h_d$, then there is $\beta \in (0,1)$ such that for any $\alpha \in [0,\beta]$ $\pi_h^U(\alpha)$ for all $h \in H$ is stable.*
- *If for every doctor there are at least two hospitals h, h' such that $V(hd) > 0$ and $V(h'd) > 0$, then there is $\beta \in (0,1)$ such that for any $\alpha \in [\beta, 1]$ $\pi_h^U(\alpha)$ for all $h \in H$ is stable.*

Proof. Since the lowest stable payoff for every hospital is zero, $\pi^U(0)$ is stable. Since $V(h_d d) > V(h'd)$ for all d and $h' \neq h_d$, it is possible to increase h payoff by $\epsilon > 0$ and have a set of stable payoffs, i.e. it is possible to reduce the payoff of all doctors matched with hospital h without them being able to form a blocking coalition. Hence $\epsilon(V(h_1 D_{h1}), \cdots, V(h_H D_H))$ is a stable payoff for hospitals for ϵ sufficiently small. Let $\beta = sup\{\epsilon > 0 | \epsilon(V(h_1 D_{h1}), \cdots, V(h_H D_H)) \, is \; a \; stable \; payoff\}$. Since Π is a convex closed set we have the result.

Suppose that for every doctor there are at least two hospitals h, h' such that $V(hd) > 0$ and $V(h'd) > 0$. Suppose there is a hospital h such that $\frac{d}{d\alpha}\pi_h^U(\alpha) > 0$ for every $\alpha \in (\epsilon, 1)$ for every $\epsilon > 0$. This implies that for all $\alpha \in (\epsilon, 1)$, $\pi_h^U(\alpha) = V(hD_h^*) > 0$ i.e. hospital h is assigned doctors in D_h^*, even when they produce a surplus $V(hD_h^*)(1-\alpha)$; however, there for every doctor in D_h^* there is another hospital h' such that $V(h'd) > V(h'd)(1-\alpha)$. Thus for every h there is ϵ_h for which $\pi_h^U(\alpha) = 0$ for all $\alpha \in [\epsilon_h, 1]$. Let $\beta = max\{\epsilon_h\}$. □

Theorem. *2.5 Suppose preferences are integer valued, then strategy-proofness and stability are possible in any q-market for any $q \in \mathbb{Z}$ such that $q \geq 1$.*

Proof. Suppose the VCG mechanism is used. If preferences are integer valued, then agent b payment is equal to $p_b = u_b(M_b) - W'(A) + W'(A \setminus b) \in \mathbb{Z}$. If in addition, $q \geq 1$ all possible payments are implementable as a matching. Hence, strategy-proofness and stability are possible in the ∞-market. □

Chapter 3: The Power of Weak Incentives

Abstract. *A social planner would like a socially optimal outcome to be chosen in an environment with externalities. The standard approach to solving the social planner's problem is to design mechanisms with desirable incentive properties such as strategy-proofness or equilibrium uniqueness. These mechanisms make the desired outcome a Nash equilibrium and rely on agents' rationality to coordinate on it. I introduce mechanisms with weak incentives to offer a different approach. These mechanisms make the desired outcome a Nash equilibrium, but rely on agents' behavioral traits - instead of rationality - to coordinate on the desired outcome. A mechanism with weak incentives is an indirect mechanism in which the payoff of agent i does not depend on his report. These mechanisms shed light on the relative importance between making the desired outcome a Nash equilibrium and offering incentives to coordinate on it. As an application, I show that in large economies, if players' reports are true on average, mechanisms with weak incentives solve the social planner's problem. I demonstrate this result using an experimental congestion game. In the lab, a mechanism with weak incentives realized 95% of the efficiency achieved by a social planner with full information. This suggest that lie-aversion, a well-established behavioral trait, can be used to design effective mechanisms.*

Ever since Hurwicz (1972) introduced the concept of incentive compatibility, the accepted wisdom has been that the minimal requirement to implement a social goal is to have a mechanism in which the social optimum is a Nash equilibrium. In practice, however, the standard approach has been to require stronger incentive properties because incentive compatible mechanisms potentially have undesired Nash equilibria, or their desired Nash equilibria might not be easy to reach. This approach has been used in kidney exchange (Roth et al. (2004)), school choice (Abdulkadirouglu and Sönmez (2003)) and military assignments (Sönmez and Switzer (2013)).

Providing strong incentive properties has been successful in practice, but it has limited the study of mechanisms in at least three ways: i) it is not applicable to problems that are incompatible with these incentive properties, ii) it fails to incorporate behavioral traits as a model of human behavior and iii) it leaves many interesting questions out of the scope. The first limitation is well-understood, but it has typically been addressed by replacing one incentive property for another. This swap is not always possible. The second limitation is more delicate. There is evidence that mechanisms with strong incentive properties sometimes work and sometimes fail. Typically, their success is attributed to their incentives; however, this interpretation is inconsistent with their failures. Furthermore, there is evidence that mechanisms without strong incentives properties sometimes succeed. These observations are consistent with the existence of behavioral traits. Finally, once behavioral traits are acknowledged, it is possible to investigate, for example, if some strategy-proof mechanisms are significantly better than others.

This paper addresses the second limitation and shows that behavioral traits can be as effective as strong incentive properties in solving social problems. Specifically, this paper i) introduces mechanisms with weak incentives – the minimal incentives for the social goal to be a rational choice, and ii) shows that these mechanism can rely on behavioral traits to solve externality problems in big economies. The objective is to achieve efficiency in an environment with externalities: each agent in a group must select an action, but the efficient profile of actions depends on the agents' private information. In this environment, a mechanism with weak incentives is an indirect mechanism in which each individual selects an action and reports his private information. The mechanism assigns prices that reflect the externalities produced by each action. These mechanisms possess the efficient profile of actions as a Nash equilibrium, but do not incentivize the truthful revelation of private information. Hence, this class of mechanisms constitutes a natural way to define the incremental value of incentives.

The main drawback of using mechanisms with weak incentives is that they generically possess many equilibria because best responses are thick, as all reports are associated with the same payoff for any given action. This does not prevent them from solving the social planner's problem. Suppose, for example, that agents have a tendency to report the truth when they cannot profit from misrepresenting their private information. In this case, a mechanism with weak incentives would be as effective as a mechanism with stronger incentive properties. This is indeed the typical assumption of strategy-proof mechanisms, as they also often possess equilibria other than truth-telling.

Of course, human beings might or might not report their private information when confronted with weak incentives. The question is for actual human behavior: What kind of problems can be effectively solved? This paper explores this question by showing that externality problems in big economies can be effectively solved by mechanisms with weak incentives for a large class of behavioral assumptions. Their effectiveness is confirmed in the experimental laboratory using a congestion problem.

Mechanisms with weak incentives are effective in solving externality problems in which average truth-telling is sufficient for achieving efficiency. For example, the efficient provision of a public good requires that the sum of net benefits is accurately signed; if some agents overstate their values while others shade by the same amount, the result would still be efficient.[1] Analogously, correcting a negative externality requires the calculation of the social marginal cost, which typically is the sum of individual marginal costs of affected parties. In these cases, the welfare function depends on the average private value, not on each individual value. Knowledge of the average type at the efficient outcome is enough to implement it. Hence, actions can be priced correctly even if some agents misrepresent their private information.

To study the coordination problem, this paper uses non-equilibrium adjustment processes. These processes characterize how agents select actions and reports, given a current profile of actions and reports. This tool is commonly used in evolutionary game theory. It is shown that a concave welfare function is sufficient for a large class of non-equilibrium adjustment processes to converge to the efficient Nash equilibrium in problems characterized by the average type. Both conditions, dependence on average values and concavity of the welfare function, are common in economic problems. This theoretical result provides reasons to believe that this class of mechanism could be effective in real life. However, the true test of the effectiveness of a mechanism is empirical.

A traffic congestion game is used to test the effectiveness of a mechanism with weak incentives in the experimental laboratory. Traffic congestion represents an ideal application. It is a big problem in which a very large number of agents play each

[1] The purchase (or funding) of a unit of public good by one agent has a positive externality on other agents.

other repeatedly.[2] Commuters have heterogeneous values of commuting and time.[3] The welfare function is concave and depends on the average value of time. In principle, a social planner could ensure efficient behavior by introducing a congestion price equal to the social marginal cost at the efficient level of traffic. In practice, however, policymakers lack the information to set such a price.[4] A mechanism design approach is still necessary.

The experimental design consists of a driving game in which 14 subjects independently decide whether to "drive" or "not drive" on a fixed road for 30 rounds of play. At the beginning of the game, every subject was randomly and privately assigned two numbers: i) a value of commuting and ii) a value of time. Neither the distribution nor the support of values was revealed to the subjects. Types were chosen to fulfill the following three functions: (i) replicate a large market, (ii) minimize the set of agents who belong to both the Nash equilibrium without congestion pricing and the social optimum, and (iii) allow for zero efficiency gains with the message mechanism.

Two main treatments were considered: *no price* and *message price*. The first treatment represents a situation with no congestion prices and the second uses a mechanism with weak incentives. The message price treatment uses agents' messages

[2]Empirical studies have found that the loss of welfare due to traffic congestion is between \$32 and \$121 billion dollars every year in the United States. According to Schrank et al. (2012), the congestion "invoice" for the cost of extra time and fuel in 498 urban areas in 2011 was (in 2011 dollars): \$121 billion. On the other hand, Litman (2014) considers that \$32 is a more appropriate value, as the former report consider a value of time "unreasonable" high. The value of time considered by the former is \$16.79 per hour and \$8.37 by the latter. These studies also have a different position on the efficient level of congestion.

[3]The value of commuting is the utility derived from getting from A to B. The value of time is the opportunity cost of every unit of time spent on the road.

[4]This lack of information is a problem that no system has been able to solve in practice. For example, both the Congestion Charge in London and Singapore's Area Licensing Scheme, which are deemed the most successful congestion systems in the world, use demand estimations and an objective level of congestion to set the congestion price to be charged to drivers. Z.F. Li (1999) describes the evolution the the Singapore's Area Licensing Scheme, which originally had a target reduction of 25% - 30%. According to the transport for London report (2003), the London's congestion charging was originally intended to reduce traffic by 10% - 15%.

about their value of time and the observed level of traffic to calculate congestion prices. Traffic observations are used to measure the marginal impact, in time, of adding an extra vehicle to the road. Messages are used to measure the cost of the marginal increase in time.

Four additional treatments were considered to provide control and robustness to the findings. The *fixed price* treatment provides a measure of the maximum observable efficiency. This treatment considers a social planner with access to all private information and imposes the optimal fixed congestion price in all rounds. The *dynamic price* treatment follows the same structure of the *message price*, but behaves as if all agents reported the truth all the time. The *balanced* treatment considers budget-balanced versions of the dynamic and message treatments. The *random* treatment considers random types instead of the constructed types used in other treatments.

The experimental results are promising. Efficiency is measured with respect to the observed efficiency achieved by the fixed price treatment, as this treatment represents the maximum possible efficiency a policymaker could achieve in a real situation. The observed efficiencies are as follows: 65.90% (13.01%) for the no price treatment and 95.00% (3.44%) for the message treatment.[5] The random treatment achieved an efficiency of 91.74% (9.3%).[6] However, when one of the six sessions is omitted, the efficiency of the random treatment becomes 95.65% (3.2%). The low efficiency, 72.21%, achieved by one of the random sessions was due to the small scale of the experiment. In the controlled sessions, types were chosen to represent a big market. In the low efficiency session, 4 out of 14 subjects had a market power

[5]The standard deviation is reported in brackets. The paired Wilcoxon signed-rank test was used to reject the null hypothesis that the treatments with congestion prices achieve the same efficiency as the treatment with no price. In all cases the null was rejected at a confidence level of 99%.

[6]This efficiency is measured with respect to the maximum theoretical efficiency associated with each draw of random types. The theoretical efficiency associated with the message price treatment is 91.46% (3.31%).

inconsistent with a big market.

This paper is related to the literature on mechanism design, the growing literature on behavioral implementation, and the well-established literature on congestion pricing.

The inconsistent performance of strong rational incentives provides evidence that human behavior - not accounted for in the rational model - plays a role in the success of many mechanisms. The most famous, but not unique, example of a mechanism that fails despite providing strong rational incentives is the second price auction, which is strategy-proof. Kagel et al. (1987) report an experiment in which bidders do not report their true value.[7] Attiyeh et al. (2000) and Kawagoe and Mori (2001) report experiments in which another strategy-proof mechanism, a version of the Vickrey-Clarke-Groves (VCG) mechanism, achieve rates of truth-telling as low as 10%. There are mechanisms that display the opposite behavior. Double auctions are the most well-known example of a mechanism that is typically not incentive compatible, but performs well most of the times. Smith (1962; 1980) shows that the double auction mechanism consistently achieves the competitive equilibrium outcome despite agents' manipulation possibilities. Budish and Kessler (2014) show that the mechanism for the fair allocation of indivisible goods without money proposed by Budish (2011) performs well in practice, despite providing opportunities for manipulation.[8]

The above inconsistencies have led to two different views towards behavioral traits. The first view considers that mechanisms should be robust to behavioral traits. Saijo et al. (2007) propose double implementation, both in Nash and weakly dominant strategies. Li (2015) proposes implementation in obviously-strategy-proof

[7] This is a prevalent phenomenon as Kawagoe and Mori (2001); Kagel and Levin (1993) report similar findings.

[8] Similarly, Che and Tercieux (2015) propose a mechanism which is neither strategy-proof, nor stable, nor efficient to obtain a matching that approximately obtains the three properties.

strategies.[9] These notions exacerbate the first limitation mentioned above, as they are harder to provide in practice. Bierbrauer et al. (2014) considers mechanisms that are robust to individuals with social preferences. Their characterization depends on payoff equivalent reports, a characteristic also present in this paper. Farhi and Gabaix (2015) implement an optimal tax scheme with behavioral agents who might not perfectly optimize their budgets, and they show that the optimal tax scheme is simple, a characteristic shared with this paper. These similarities are in spirit, not in the letter. However, they might help us understand how behavioral implementation is different or similar to rational implementation. de Clippel (2014) shows they are not entirely different, but that their connection is still not well understood. The second view leverages on behavioral traits to achieve social goals. This paper belongs to this second branch. In this branch there are several papers that explore mechanisms without strong incentives, but do not explicitly address how the desired Nash equilibrium is reached. Abdulkadiroğlu et al. (2011) and Abdulkadiroğlu et al. (2015) propose a non-truth-telling mechanism for school choice that improves upon a strategy-proof mechanism but provide no evidence that these gains could be realized in practice or how. Featherstone and Niederle (2015) shows experimentally that these non-truth-telling equilibria might me very difficult to reach in practice and propose a truth-telling-not-strategy-proof mechanism, however, their experiments only suggest a *potential* for truth-telling-not-strategy-proof mechanisms, since they do not explicitly address how their subjects reach equilibrium. There are papers that use non-equilibrium strategies as means of implementation. Fragiadakis and Troyan (2015) shows that focal, non-equilibrium, strategies can be used to improve efficiency in an assignment game. In contrast to the mentioned papers, this paper: i) deals with externalities instead of assignment games,

[9] A strategy is obviously dominant if, for any deviating strategy, starting from any earliest information set where both diverge, the best possible outcome from the latter is no better than the worst possible outcome of the former.

ii) provides a general framework for understanding equilibrium selection in terms of behavioral traits, iii) shows explicitly that average-truth-telling is sufficient to converge to the efficient outcome, and iv) designs an experiment that allows one to attribute the success of the mechanism to the aforementioned behavioral trait. Both Featherstone and Niederle (2015) and Fragiadakis and Troyan (2015) experimental results can be interpreted as leveraging on the agents' tendency to report the truth - a feature also present in this paper and well-established in the behavioral game theory literature (Gneezy (2005); Erat and Gneezy (2012); Gneezy et al. (2013)).

This paper is also related to the literature on congestion abatement systems. Externalities and externality abatement have been studied consistently at least since Pigou (1920), who proposed to charge agents the value of the marginal externality they produce at the efficient social allocation. As mentioned before, this approach requires information not available to the policymaker. Many solutions have been studied. For example, Sandholm (2002; 2005; 2010) provides a systematic treatment of the dynamics of congestion prices in continuous time. Both Li (2002) and Yang et al. (2004) provide evidence that prices can also be adjusted in discrete time. Yang and Wang (2011) study systems of tradable permits. They show that the system can achieve full efficiency when the market for permits is perfectly competitive. Continuing their work, Wang et al. (2014) showed that the system of tradable permits can be guaranteed to achieve the social optimum allocation by adjusting the quantity of permits according to the observed price in the permits market. Nie (2012) have shown that these tradable permit systems are very sensitive to transaction costs in the permits market. Guo and Yang (2010) show that, when demand is fixed, it is possible to achieve budget balancedness using an appropriate combination of taxes and subsidies. The message system can achieve budget balancedness even when demand respond to prices. Several studies have taken congestion games to the experimental lab. Schneider and Weimann (2004), Selten et al. (2007), and

Hartman (2012) study route choice behavior with and without congestion prices. Rapoport et al. (2004) and Rapoport et al. (2014) study entry games with and without congestion prices. In both the experimental and theoretical literature on congestion, it is assumed that the policymaker or mechanism knows the value of the externality i.e. knows every agents' value of time and that this value is homogeneous. The theory and experiment in this paper do not assume knowledge of private information nor its homogeneity in the population.

3.1 Mechanisms with weak incentives

This section introduces mechanisms with weak incentives in a general framework to highlight the interactions between rational incentives, information, and behavioral traits in mechanisms designed to solve the social planer's problem in an environment with externalities. The purpose of these mechanisms is to isolate behavioral traits as an equilibrium refinement. These mechanisms offer the social optimum as a Nash equilibrium, but do not incentivize agent's to coordinate on it. Furthermore, it is assumed that agents have private information, but lack common knowledge of the distribution of types. The informational assumption might hold in some real life applications.

Consider a set of agents $N = \{1, \ldots, N\}$. Agents must select an action simultaneously and independently from each other. Agent i selects actions from the finite set X_i. An action profile $x = (x_1, \ldots, x_N)$ describes an action for each agent. The set of action profiles is denoted by $X = \Pi_N X_i$. Agent i is described entirely by his type $\theta_i \in \Theta_i$. Types are private information. Let $\theta = (\theta_1, \ldots, \theta_N)$ and $\Theta = \Pi_N \Theta_i$. Individuals have quasilinear utility functions $v_i(x, \theta_i, t) = u_i(x, \theta_i) + t$, where $u_i : X \times \Theta_i \to \mathbb{R}$ depends on everyone actions and i's private information. Agent i knows his type θ_i and his set of strategies X_i, but does not know the distribution of types.

The profile of actions $x \in X$ is efficient at θ if $\sum_N u_i(x, \theta_i) \geq \sum_N u_i(y, \theta_i)$ for all $y \in X$. The efficiency level associated with an action profile x at θ is $V(x, \theta) = \sum_N u_i(x, \theta_i)$. The set of efficient profiles of actions at θ is denoted by $x^*(\theta)$. A profile of actions $x \in X$ is a Nash equilibrium at θ if $v_i(x_i, x_{-i}, \theta_i) \geq v_i(y_i, x_{-i}, \theta_i)$ for all $y_i \in X_i$ for all $i \in N$. The set of Nash equilibria at θ is denoted by $x(\theta)$. In many situations there is no efficient Nash equilibrium i.e. $x(\theta) \cap x^*(\theta) = \emptyset$. Consider the following example.

Example 3.1. Consider a situation with two agents $N = \{1, 2\}$ and actions $X_1 = \{a_1, b_1\}$ and $X_2 = \{a_2, b_2\}$. Each agent has two possible types: $\Theta_1 = \{\theta_1, \theta_1'\}$ and $\Theta_2 = \{\theta_2, \theta_2'\}$. Suppose payoffs are as follow:

(θ_1, θ_2)	a_2	b_2	(θ_1, θ_2')	a_2	b_2
a_1	$4, 3$	$2, 2$	a_1	$4, 3$	$2, 4$
b_1	$3, 5$	$1, 4$	b_1	$3, 1$	$1, 2$
(θ_1', θ_2)	a_2	b_2	(θ_1', θ_2')	a_2	b_2
a_1	$2, 3$	$4, 2$	a_1	$2, 3$	$4, 4$
b_1	$3, 5$	$5, 4$	b_1	$3, 1$	$5, 2$

The efficient profile of actions and Nash equilibria are as follow: $x^*(\theta_1, \theta_2) = (b_1, a_2)$ and $x(\theta_1, \theta_2) = (a_1, a_2)$, $x^*(\theta_1', \theta_2) = (b_1, b_2)$ and $x(\theta_1', \theta_2) = (b_1, a_2)$, $x^*(\theta_1, \theta_2') = (a_1, a_2)$ and $x(\theta_1, \theta_2') = (a_1, b_2)$, $x^*(\theta_1', \theta_2') = (a_1, b_2)$ and $x(\theta_1', \theta_2') = (b_1, b_2)$.

A social planner would like to ensure that a member of $x^*(\theta)$ is chosen by the agents for all $\theta \in \Theta$ by introducing a mechanism. A mechanism is a pair M, g, with $M = \Pi_N M_i$ and $g : M \to O$, where M_i is player's i message space and $O = X \times \mathbb{R}^N$ is the outcome space. A mechanism assigns a profile of actions $g_x(m)$ and transfers $g_t(m)$ for every profile of messages $m = (m, \ldots, m_N)$. A message profile m is a Nash equilibrium at $\theta \in \Theta$ if $v_i(g_x(m_i, m_{-i}), \theta_i, g_{t,i}(m_i, m_{-i})) \geq v_i(g_x(m_i', m_{-i}), \theta_i, g_{t,i}(m_i', m_{-i}))$ for all $m_i' \in M_i$ and $i \in N$. The set of Nash equilibria in the mechanism M, g at θ is denoted by $m_g(\theta)$. The mechanism M, g is efficient whenever $x^*(\theta) \cap g_x(m_g(\theta)) \neq \emptyset$

for all $\theta \in \Theta$. In this case $m_g^*(\theta)$ is a selection of $m_g(\theta)$ such that $g_x(m_g^*(\theta)) \in x^*(\theta)$ for all $\theta \in \Theta$. A message m_i is a dominant strategy for agent i at θ_i if $v_i(g_x(m_i, m_{-i}), \theta_i, g_{t,i}(m_i, m_{-i})) \geq v_i(g_x(m_i', m_{-i}), \theta_i, g_{t,i}(m_i', m_{-i}))$ for all $m_i' \in M_i$ and $m_{-i} \in M_{-i} = \Pi_{N \setminus i} M_j$. A mechanism M, g is budget balanced at $m \in M$ if $\sum_N t_i(m) \leq 0$. It is assumed that the social planner knows Θ, but not the distribution of types.

It is widely accepted that the existence of an efficient mechanism is not sufficient to guarantee that $x^*(\theta)$ will be chosen by the agents for all $\theta \in \Theta$ because there might be multiple equilibria. This problem has been addressed in many different ways. For example, offering a unique equilibrium guarantees that the only rational choice is the desired outcome and making truth-telling a weakly dominant strategy makes it easier to coordinate in the truth-telling equilibrium even when there are other equilibria. There are many other options, but all of them share one characteristic: they limit the set of problems that can be solved and demand a level of rationality that might not be available in practice. This paper offers an alternative approach for dealing with multiple equilibria: rely on agents' behavior to coordinate on the desired outcome. This is done by providing a mechanism that has the desired outcome $x^*(\theta)$ as a Nash equilibrium, but does not incentivize agents to select it. This class of mechanism posses weak incentives.

A mechanism M, g is a mechanism with weak incentives if $M_i = X_i \times \Theta_i$, $g_x(x, \theta) = x$ and $g_t(x, \theta) = p : X \times \Theta \to \mathbb{R}^N$ is such that is such that $v_i(x, \theta_i, p_i(x, \theta_i', \hat{\theta}_{-i})) = v_i(x, \theta_i, p_i(x, \theta_i'', \hat{\theta}_{-i}))$ for all $x \in X$, $\theta_i', \theta_i'' \in \Theta_i$ and $\hat{\theta}_{-i} \in \Theta_{-i}$. Agents select an action and send a report about their type, but their payoff does not depend on the particular report they send. Hence, this class of mechanisms do not incentivize the revelation of private information. As in the case of direct mechanisms, it is possible to choose p such that $(x^*(\theta), \theta)$ becomes a Nash equilibrium for all $\theta \in \Theta$. For each agent, p_i is a list of prices for each action in X_i. The construction of an efficient set of prices p relies on the celebrated Vickrey - Clarke - Groves mechanism (VCG).[10] The VCG mechanism is an efficient direct mechanism with $M = \Theta$, $g_x(m) \in \underset{y}{argmax} V(y,m)$ and $g_{t,i}(m) = \sum_{N \setminus i} u_j(x(m), m_j) - h_i(m_{-i})$, where $h_i : \Theta_{-i} \to \mathbb{R}$. Truth-telling is a dominant strategy in the VCG mechanism. To obtain an efficient mechanism with weak incentives, let $p_i(x,\theta) = \sum_{N \setminus i} u_j(x,\theta_j) - h_i^w(x_{-i},\theta_{-i})$ be the price associated with x_i when other agents select x_{-i}, where $h_i : X_{-i} \times \Theta_{-i} \to \mathbb{R}$. These prices define the weak VCG mechanism (wVCG). Both transfers and prices can be set to represent the marginal impact of the introduction of an agent, in the case of the VCG, or the selection of an action, for the wVCG. This is achieved by setting $h_i(m_{-i}) = max \sum_{N \setminus i} u_j(x, m_j)$ for all $i \in N$ for VCG and selecting a default profile of actions $x^0 \in X$ and letting $h_i^w(x_{-i},\theta_{-i}) = \sum_{N \setminus i} u_j((x_i^0, x_{-i}), \theta_j)$ for all $i \in N$ for wVCG. Unless otherwise noted, these transfers and prices will be used in all examples. The differences between the VCG and the wVCG are illustrated in the following example.

Example 3.2. Consider the problem from example 1. The games induced by VCG and wVCG with default $x^0 = (b_1, b_2)$ when the true state of the world is (θ_1, θ_2) are shown below. Transfers and prices are added to (or subtracted from) the payoff associated with each profile of messages.

[10] Vickrey (1961); Clarke (1971); Groves (1973)

VCG

(θ_1, θ_2)	θ_2	θ_2'
θ_1	$3+0, 5-1$	$4-1, 3+0$
θ_1'	$1-1, 4+0$	$2+0, 2-1$

wVCG

(θ_1, θ_2)	a_2, θ_2	a_2, θ_2'	b_2, θ_2	b_2, θ_2'
a_1, θ_1	$4-2, 3+2$	$4+2, 3+2$	$2-2, 2+0$	$2+2, 2+0$
a_1, θ_1'	$4-2, 3-2$	$4+2, 3-2$	$2-2, 2+0$	$2+2, 2+0$
b_1, θ_1	$3+0, 5+2$	$3+0, 5+2$	$1+0, 4+0$	$1+0, 4+0$
b_1, θ_1'	$3+0, 5-2$	$3+0, 5-2$	$1+0, 4+0$	$1+0, 4+0$

In wVCG, there are 4 Nash equilibria

$$m(\theta) = \{(b_1, \theta_1, a_2, \theta_2), (a_1, \theta_1, a_2, \theta_2'), (b_1, \theta_1', a_2, \theta_2), (b_1, \theta_1', b_2, \theta_2)\}$$

in VCG $m(\theta_1, \theta_2) = (\theta_1, \theta_2)$ is the unique equilibrium in dominant strategies. Both mechanism are efficient.

The following propositions show some properties of mechanisms with weak incentives. All proofs are in the appendix.

Proposition 3.1. *There is an efficient mechanism with weak incentives, namely the wVCG.*

A mechanism with weak incentives makes the efficient allocation a rational choice i.e. any $x \in x^*(\theta)$ can be supported as a Nash equilibrium, however, agents are not incentivized to reveal their private information. This weakening in solution concept, with respect to strategy-proofness, allows for some new possibilities. In particular, budget balancedness is always possible to obtain.

Proposition 3.2. *There is a budget balanced mechanism with weak incentives for any profile of actions. In particular, any efficient profile of actions can be supported as a budget balanced Nash equilibrium.*

In some applications sending a report and selecting an action could be difficult for the agents or for the agency collecting the prices. In these cases, decisions could be preferably made sequentially. The next proposition shows that efficiency can also be achieved in this manner.

Proposition 3.3. *Any efficient profile of actions can be supported as a subgame perfect Nash equilibrium of a sequential mechanism with weak incentives.*

The above propositions and example show crucial differences between VCG and wVCG. VCG induces the efficient profile of actions by incentivizing the revelation of private information while wVCG allows for efficiency without incentivizing agents to select the socially desirable outcome. The wVCG mechanism depends completely on agents' behavioral traits to coordinate on the desired outcome. The next section develops the idea of behavioral traits as an equilibrium refinement.

3.2 Mechanisms with weak incentives in large average economies

This section develops a model in which behavioral traits are used as an equilibrium refinement for a mechanism with weak incentives. In this model, agents can adjust their strategies over time, allowing the emergence of the desired Nash equilibrium as a social convention. The model is developed in continuous time and agents for technical convenience.

Agents have a common and finite set of actions $S = \{1, \ldots, S\}$ with typical element s.[11] The common set of types Θ is finite with typical element $\theta = (\theta_1, \theta_2)$,

[11] This can be done without loss of generality by letting $S = \cup_N X_i$

$\theta_j \in \mathbb{R}^S$.[12] There is a positive mass of agents μ_θ of each type θ. The mass of agents of type θ doing s is denoted by $x_{\theta s} \geq 0$. Profiles of actions are replaced by distributions of actions $x \in X = \{x \in \mathbb{R}_+^{|\Theta| \times |S|} | \sum_s x_{\theta s} = \mu_\theta\}$. The mass of agents, of any type, doing s is denoted by $x_s \geq 0$. The anonymous distribution of actions $X' = \{x \in \mathbb{R}_+^{|S|} | \sum_s x_s = \sum_\theta \mu_\theta\}$ describes what actions are being taken without specifying which type is doing them. For every $x \in X$, let x' be such that $x'_s = \sum_\theta x_{\theta s}$.

An agent with type θ doing s has utility function $u_{\theta s}(x) = F_s(x')\theta_{1s} + \theta_{2s}$, where $F : X' \to \mathbb{R}^S$, $F \in C^2$ is an observable externality function.[13] To simplify notation, $F(x')$ will be denoted by $F(x)$. Types are scaled so that for every $\theta \in \Theta$ there is an action s_θ such that $u_{\theta s}(x) = 0$ for all $x \in X$. Social welfare is captured by $W(x) = \sum_\theta \sum_s x_{\theta s} u_{\theta s} = \sum_s F_s(x) x_s \bar{\theta}_{1s}(x) + \sum_s x_s \bar{\theta}_{2s}(x)$, where $\bar{\theta}_{1s}(x) = \frac{1}{x_s} \sum_\theta x_{\theta s}\theta_{1s}$ and $\bar{\theta}_{2s}(x) = \frac{1}{x_s} \sum_\theta x_{\theta s}\theta_{2s}$ represent the average type doing action $s \in S$. It is assumed that W is strictly concave. The efficient distribution of actions x^* is characterized by the first order conditions of the Kuhn-Tucker problem:[14]

$$\begin{gathered} F_s(x^*)\theta_{1s} + \theta_{2s} + \sum_j \frac{\partial F_j}{\partial x_s} \sum_\theta x^*_{\theta j}\theta_{1j} = \lambda_\theta - \lambda_{\theta s} \text{ for all } \theta \in \Theta,\ s \in S \\ \lambda_\theta \geq 0,\ \lambda_\theta[\sum x^*_{\theta j} - \mu_\theta] = 0 \text{ for all } \theta \in \Theta \\ \lambda_{\theta s} \geq 0,\ \lambda_\theta[x^*_{\theta s}] = 0 \text{ for all } \theta \in \Theta,\ s \in S \end{gathered} \tag{3.1}$$

A distribution of actions x constitutes a Nash equilibrium if $v_{\theta s}(x) = \underset{j \in S}{max}\ v_{\theta j}(x)$ whenever $x_{\theta s} > 0$. Equivalently, x is a Nash equilibrium if there is $k_\theta \geq 0$ such that $v_{\theta s}(x) = k$ whenever $x_{\theta s} > 0$ and $v_{\theta s}(x) \leq k$ whenever $x_{\theta s} = 0$.

---

[12]This can be done without loss of generality by letting $\Theta = \cup_N \Theta_i$

[13]If there are no externalities, there is no need for a mechanism as each agent could select his favorite action without hurting others. Both positive and negative externalities are considered.

[14]The Lagrangian function is $L(x, \lambda) = W(x) - \sum_\theta \lambda_\theta(\sum_j x^*_{\theta j} - \mu_\theta) + \sum_\theta \sum_s \lambda_{\theta s} x_{\theta s}$

Pigou (1920) realized that efficiency can be achieved in the presence of externalities if agents internalize them through prices. In this case, a price equal to $p_s(x^*) = \sum_j \frac{\partial F_j}{\partial x_s} \sum_\theta x^*_{\theta j}\theta_{1j}$ for doing action $s \in S$ would make the condition for optimality and Nash equilibrium identical. To see this observe that the first order conditions imply the conditions for a Nash equilibrium with $k = \lambda_\theta$, $x_{\theta s} > 0$ implies that $F_s(x^*)\theta_{1s} + \theta_{2s} + \sum_j \frac{\partial F_j}{\partial x_s} \sum_\theta x^*_{\theta j}\theta_{1j} = \lambda_\theta = k$ and $x_{\theta s} = 0$ implies that $F_s(x^*)\theta_{1s} + \theta_{2s} + \sum_j \frac{\partial F_j}{\partial x_s} \sum_\theta x^*_{\theta j}\theta_{1j} = \lambda_\theta - \lambda_{\theta s} \leq k$.

The main problem with the above approach is that the efficient average type of action s, $\bar{\theta}^*_{1s} = \bar{\theta}_{1s}(x^*)$, is unknown to the social planner. However, pricing an action based on reported types and observed actions is feasible i.e. $p_s(x, \tilde{x}) = \sum_j \frac{\partial F_j}{\partial x_s} \sum_\theta \tilde{x}_{\theta j}\theta_{1j}$ where $\tilde{x}_{\theta j}$ is the mass of agents reporting being of type θ. This pricing mechanism is a mechanism with weak incentives. When these prices are used, the efficient distribution of actions can be supported as a Nash equilibrium.

As in the discrete case, the mechanism with weak incentives with prices $p_s(x, \tilde{x}) = \sum_j \frac{\partial F_j}{\partial x_s} \sum_\theta \tilde{x}_{\theta j}\theta_{1j}$ for all $s \in S$ has multiple equilibria. In particular, for any fixed distribution of type reports $\tilde{x}$, there is a Nash equilibrium $x(\tilde{x})$ that satisfies $F_s(x(\hat{x}))\theta_{1s} + \theta_{2s} + p_s(x(\hat{x}), \hat{x}) = k_\theta$ whenever $x(\hat{x})_{\theta s} > 0$ and $F_s(x(\hat{x}))\theta_{1s} + \theta_{2s} + p_s(x(\hat{x}), \hat{x}) \leq k_\theta$ whenever $x(\hat{x})_{\theta s} = 0$.

To understand if agents have any chance of coordinating in the efficient profile of actions first assume that agents always reveal their private information truthfully. In this case, prices $p_s(x) = \sum_j \frac{\partial F_j}{x_s} \sum_\theta x_{\theta j}\theta_{1j}$ would only depend on the current profile of actions and the multiplicity of Nash equilibria disappears.[15] In standard game theory, it is almost always assumed that the existence of a single Nash equilibrium is sufficient for agents to coordinate on it. This section uses a different tool: evolutionary game theory. This theory replaces the strong rational and informational

[15] When the identity of each individual in a continuum is considered, there is still a continuum of equilibria as agents of a particular type could distribute themselves differently and still respect the aggregate distribution of types and actions.

assumptions in standard game theory with assumptions about non-equilibrium behavior.[16]

Agents' individual actions determine a particular distribution of actions x. When x is Nash equilibrium, it is in the best interest of all agents to follow it. Conversely, when a non-equilibrium distribution of actions is specified, there is a positive mass of agents who can gain by changing their action. However, it is not clear when a sequence of non-equilibrium distributions of actions and their respective deviations actually lead to a Nash equilibrium. Thus characterizing non-equilibrium behavior is essential to study the convergence properties of mechanisms with weak incentives. This approach specifies how actions associated with the same payoff are chosen, a critical element in the study of mechanism with weak incentives.

Mean dynamics and Lyapunov functions are introduced to characterize non-equilibrium behavior. A mean dynamic $V : X \to \mathbb{R}^{|\Theta|\times|S|}$ is a function that defines an equation of motion $\dot{x} = V(x)$ on the space of distributions of actions. V is called admissible if:

$$\begin{aligned} V \quad & \text{is Lipschitz continuous} \\ V_{\theta s}(x) \geq 0 \quad & \text{whenever } x_{\theta s} = 0 \\ \sum_S V_{\theta s}(x) = 0 \quad & \text{for all } \theta \in \Theta \\ V(x) = 0 \quad & \text{implies } x \text{ is a Nash equilibrium} \end{aligned}$$

A function $L : X \to \mathbb{R}$ such that $\nabla L(x)' V(x) \leq 0$ for all $x \in X$ is a Lyapunov function for V. An admissible mean dynamic V with Lyapunov function L has important properties: (i) there is a unique solution trajectory $x : \mathbb{R}_+ \to X$ from any initial point $x \in X$, (ii) all solution trajectories stay in the space X, (iii) all rest points of V are Nash equilibria, and (iv) all accumulation points of solution trajectory x are critical points of $L \circ x$.[17] The following proposition shows that, when all

[16] Aumann and Brandenburger (1995), for example, have shown that reaching a Nash equilibrium instantaneously requires strong informational conditions.

[17] These are well-known results in the theory of differential equations. The first condition im-

agents report their types truthfully, agents can successfully coordinate on the efficient Nash equilibrium.

Proposition 3.4. *Let* $v_{\theta s}(x) = F_s(x)\theta_{1s} + \theta_{2s} + \sum\limits_j \frac{\partial F_j}{\partial x_s} \sum\limits_\theta x_{\theta j}\theta_{1j}$ *for all* $\theta \in \Theta$ *and* $s \in S$ *and* V *an admissible mean dynamic such that* $V(x) \cdot \nabla W(x) > 0$ *whenever* $V(x) \neq 0$*, then every solution trajectory of* V *converges to the efficient distribution of actions* x^**.*

V satisfies $V(x) \cdot \nabla W(x) > 0$ for all x such that $V(x) \neq 0$ whenever, on aggregate, agents adjust their actions by increasing their payoffs over time; this adjustment does not need to be optimal for any agent, in particular, the payoff for some individual agents might decrease as long as the aggregate welfare increases.

If agents are not guaranteed to tell the truth, the pricing mechanism becomes a function of their reports as well as their actions. In this case, prices become $p_s(x, \tilde{x}) = \sum\limits_j \frac{\partial F_j}{\partial x_s} \sum\limits_\theta \tilde{x}_{\theta j}\theta_{1j}$ where x is the observable distribution of actions and $\tilde{x}$ is the reported distribution of types. A mean dynamic $\hat{V} : \hat{X} \to \mathbb{R}^{|\Theta| \times |S \times \Theta|}$ describes both the action and reporting behavior, where $\hat{X} = \{\hat{x} \in \mathbb{R}_+^{|\Theta| \times |S \times \Theta|} | \sum_s x_{\theta s \hat{\theta}} = \mu_\theta\}$. $\hat{x}_{\theta s \hat{\theta}}$ is the mass of agents of type θ taking action s and reporting $\hat{\theta}$ as their type. Letting $x_{\theta s} = \sum_{\hat{\theta}} \hat{x}_{\theta s \hat{\theta}}$ and assuming that $\sum_{\hat{\theta}} V_{\theta s \hat{\theta}}(\hat{x}) = \sum_{\hat{\theta}} V_{\theta s \hat{\theta}}(\hat{y})$ for every $\hat{x}$ and $\hat{y}$ such that $x = y$, every mean dynamic $\hat{V}$ induces a mean dynamic V by letting $V_{\theta s}(x) = \sum_{\hat{\theta}} V_{\theta s \hat{\theta}}(\hat{x})$. Such a mean dynamic is called an average truth-telling dynamic if, in addition, its induced V is admissible and $V(x) \cdot \nabla W(x) > 0$ whenever $V(x) \neq 0$.

Proposition 3.5. *Let* $\hat{V}$ *be an average truth-telling mean dynamic, then the mechanism with weak incentives defined by* $p_s(x, \hat{x}) = \sum\limits_j \frac{\partial F_j}{\partial x_s} \sum\limits_\theta \hat{x}_{\theta j}\theta_{1j}$ *converges to the efficient* x^* *distribution of actions.*

---

plies existence of a solution to $\dot{x} = V(x)$ by the Picard–Lindelöf theorem. The second and third conditions guarantee that the solution does not leave X. The last condition follows the intuition provided by the Nash equilibrium: agents at a Nash equilibrium do not change their actions while agents in a non-equilibrium do. See Sandholm (2010) for an introduction.

In theory, agents following an average truth-telling mean dynamic would converge to the efficient distribution of actions. In practice, do agents converge to the efficient distribution of actions? The next section explores this question by using a mechanism with weak incentives to solve an externality problem in the experimental laboratory. The experiment is framed as a traffic congestion problem as real traffic involves a large number of agents who lack enough information about each other to justify convergence to equilibrium by means of the rational model.

3.3 A traffic congestion model

This sections specializes the model developed in section two to describe a traffic congestion problem and describe different interventions a social planner could implement under different informational assumptions. These interventions are latter tested in the experimental laboratory.

Real life traffic congestion occurs when thousands of drivers use a road network. During congested times, the marginal effect of each individual on the total congestion is very small, but the total effect can be large. Drivers do not know each other, and do not coordinate routes or departure times. These characteristics are better captured by the continuous agents model.

A continuum of agents want to commute using a single road during a single peak time of the day. The total time spent by each agent commuting is a function of the number of agents on the road and is characterized by a strictly increasing and strictly convex, twice differentiable function $t : \mathbb{R} \to \mathbb{R}_+$. There is a finite set of types Θ, with typical element θ and mass denoted by μ_θ. Every type is characterized by two values: θ_{1d} is the value of time and θ_{2d} is the value of commuting. All types have an outside option with value 0, staying home. All agents choose between commuting and staying home, $S = \{d, h\}$.

Outcomes are identified by a distribution of actions $x \in X = \{x \in \mathbb{R}_+^{|\Theta|\times|S|} | \sum_s x_{\theta s} =$

$\mu_\theta\}$, where $x_{\theta d}$ represents the mass of agents of type θ who drive. The utility received by an agent of type θ for driving is $u_\theta(x) = \theta_{2d} - \theta_{1d} t(\sum_\theta x_{\theta d})$. When there is no risk of confusion, x will be used to denote both the total number of drivers on the road and the strategy distribution.

3.3.1 Congestion prices

A social planner would select a strategy distribution that maximizes welfare. The aggregate welfare for a strategy distribution x is given by $W(x) = \sum_\theta \sum_s x_{\theta s} u_{\theta s}$. The efficient distribution of actions is characterized by the first order conditions in (1). In real life, there are no social planners, but policy makers facing informational and political constraints. In the following sections we analyze how a policy maker could implement or approximate the social planner's solution under different informational and political constraints. Since t is observable it is assumed that policy makers know t.

3.3.1.1 Full information

Suppose a policy maker had complete information about the commuting time function t and the mass of each type μ_θ, then he could calculate the optimal allocation x^* and impose a fixed optimal price of driving equal to $P^* = t'(x^*) \sum_\theta \theta_{1d} x^*_{\theta d}$.

3.3.1.2 Unknown demand

Assume that the policy maker has no information regarding the demand for commuting but can perfectly identify the types i.e. upon observing an agent, the policy maker can identify θ_{1d} but not θ_{2d}. This is a very strong assumption, but allows the study of the gradual loss of information from the policy maker's perspective. This lack of information prevents the policy maker from implementing the optimal fixed congestion price $P* = t'(x^*)\sum_\theta \theta_{1d} x^*_{\theta d}$. In this case, the following dynamic congestion price can be implemented: $P^D(x) = t'(x)\sum_\theta \theta_{1d} x_{\theta d}$.

3.3.1.3 Unknown demand and unknown social cost

Suppose the policy maker has no information regarding the demand or social cost. Policy makers can observe the total number of drivers on the road, but cannot distinguish their types. Thus the implementation of the dynamic tax $P^D(x) = t'(x)\sum_\theta \theta_{1d} x_{\theta d}$ becomes impossible. The policy maker, however, could ask drivers to report their value of time and observe traffic; with this information, a mechanism with weak incentives characterized by the following prices becomes a natural candidate:$P^M(x, \hat{x}) = t'(x)\sum_\theta \theta_{1d} \hat{x}_{\theta d}$.

3.3.1.4 Revenue neutrality

On top of informational constraints, policy makers usually face political constraints. In the case of externality abatement, the imposition of a congestion price is usually seen as a bad alternative, since it involves a new "tax". Hence it is important to consider revenue neutral alternatives.

In the context of this model, revenue neutrality is simple to achieve since any congestion price can be replaced by a smaller price on driving and a transfer for not driving. For example, the dynamic congestion price $P^D(x) = t'(x)\sum_\theta \theta_{1d} x_{\theta d}$ can be replaced by a smaller price $P^{BD}(x) = \frac{\mu - x}{\mu} t'(x)\sum_\theta \theta_{1d} x_{\theta d}$ and a transfer $S^{BD}(x) = \frac{x}{\mu} t'(x)\sum_\theta \theta_{1d} x_{\theta d}$, where $\mu = \sum_\theta \mu_\theta$. The analogous division can be implemented for the message congestion price.

3.4 A mechanism with weak incentives in the laboratory.

The main objective of the experiment is to test if the message system proposed above allows drivers to converge to the socially optimal traffic congestion level. The previous section provides some evidence that, under average truth-telling, the social optimum would be observed. The empirical effectiveness is tested in the experimental laboratory.

The experimental design consists of a driving game in which 14 subjects independently decide whether to "drive" or "not drive" on a fixed road for 30 rounds of play.[18] At the beginning of each game, every subject was randomly and privately assigned a type characterized by two numbers: a value of commuting and a value of time. These values are held fixed over the 30 rounds of play. Neither the distribution nor the support of values was revealed to the subjects. There is a fixed set of types.

Types were chosen to fulfill the following three functions: (i) produce at most one marginal agent, (ii) minimize the set of agents who belong to both the Nash equilibrium without congestion pricing and the social optimum, and (iii) allow for zero efficiency gains with the message congestion price.

Congestion occurs when thousands of drivers use the road at the same time. However, designing an experiment that requires thousands of subjects would be both impractical and expensive. This large numbers problem is addressed through the experimental design. When there is a large number of drivers, the impact of each individual on one another is small. In particular, the small increase in travel time produced by the introduction of one single driver to a road would change the decision of a small number of current drivers. This feature is reproduced in the experiment by carefully selecting types. In the experiment, when an agent changes his

[18] In two out of nine session the number of drivers was 16.

driving decision i.e. drives if he was not driving or the other way around, at most one other agent finds it profitable to change his behavior.

The goal of a congestion price is to change the behavior of agents. An effective system would not only produce the right level of traffic congestion, but also the right set of drivers. In this experiment, types are used to minimize the set of agents who belong to both the Nash equilibrium without congestion pricing and the social optimum. The equilibrium without congestion pricing consists of 10 drivers and the social optimum consists of 6. However, only two drivers belong to both allocations. In other words, 12 out of 14 agents have to change their behavior with the introduction of congestion pricing. This radical change in the set of drivers is a strong test for the effectiveness of the system.

Inevitably the message congestion price system will produce a continuum of equilibria. The experimental design exploits this feature by providing the social optimum and the outcome without congestion pricing as Nash equilibria. This prevents the message price treatment from producing artificial efficiency gains.

Figures 3.1 and 3.2 contain the list of types used in the experiment and illustrate their distribution. The congestion function $t(x) = \frac{x^3}{12}$ was chosen to have commute values and time values on a relatively equal scale.

Type	Value of Time	Value of Commuting	No Congestion Price	Social Optimum
1	2.40	70.00	x	x
2	3.60	80.00	x	x
3	6.00	32.00	x	
4	9.00	35.00	x	
5	12.00	38.00	x	
6	15.00	41.00	x	
7	18.00	44.00	x	
8	21.00	48.00	x	
9	24.00	51.00	x	
10	27.00	54.00	x	
11	60.96	82.65		x
12	77.02	76.35		x
13	99.00	99.50		x
14	101.00	100.99		x

Figure 3.1: Type allocations

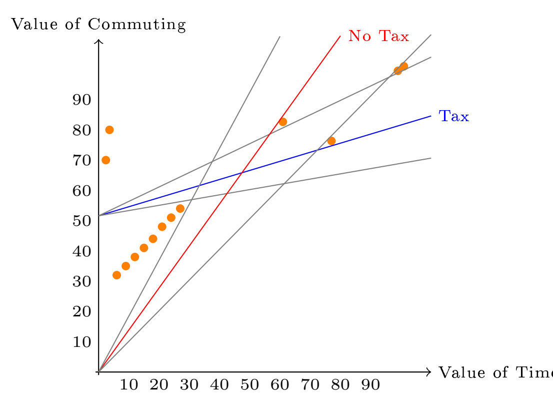


Figure 3.2: Experimental types

In figure 3.2, every dot represents a type. The red line represents the equilibrium time when there is no congestion price and the blue line represents the optimal time when the optimal fixed price is imposed. The gray lines are variations of time when a driver is added or removed. When no congestion price is in place, all agents above the red line would find profitable to drive; with the optimal fixed congestion price in place, only those above the blue line would find profitable to drive. Only two types are above both lines. Suppose there is no congestion price and all the agents above the red line are driving and consider the exit of one of the current drivers. This would reduce congestion and travel time for everyone. In particular, at the current time (the gray line below the red) only one type would find profitable to start driving (type 11) i.e. there is at most one marginal agent.

In theory, with the above types and congestion function, the Nash equilibrium without congestion pricing achieves an efficiency level of 301.3 experimental dollars whereas the social optimum achieves an efficiency of 406.3 experimental dollars, an increase of 34.8%. In practice, the efficiency level associated with no congestion price could be lower or higher than the Nash equilibrium efficiency. Hence, the benefits, if any, of the message system have to be measured against observed efficiencies.

Two main treatments were considered: *no price* and *message price*. The first treatment represents a situation with no congestion prices and the second uses a mechanism with weak incentives. The message price treatment uses agents' messages about their value of time and the observed level of traffic to calculate congestion prices. Traffic observations are used to measure the marginal impact, in time, of adding an extra vehicle to the road. Messages are used to measure the cost of the marginal increase in time.

Four additional treatments were considered to provide control and robustness to the findings. The *fixed price* treatment provides a measure of the maximum observable efficiency. This treatment considers a social planner with access to all private information and imposes the optimal fixed congestion price in all rounds. The *dynamic price* treatment follows the same structure of the *message price*, but behaves as if all agents reported the truth all the time. The *balanced* treatment considers budget-balanced versions of the dynamic and message treatments. The *random* treatment considers random types instead of the constructed types used in other treatments. Each treatment was run 6 times.

Every treatment is associated with a hypothesis derived from the theory section.

1. *The no congestion price treatment will achieve the theoretical efficiency associated with no congestion price*

2. *The fixed price treatment will achieve the theoretical optimal efficiency*

3. *The dynamic price treatment will achieve the same efficiency as the fixed price treatment*

4. There are two hypothesis associated with the message price treatment

   (a) *Subjects will play an average-truth-telling mean dynamic*

   (b) *The message treatment will achieve the same efficiency as the fixed price treatment*

5. *The balanced treatments will achieve the same efficiency as the unbalanced treatments*

6. There are two hypothesis associated with the random treatment

   (a) *Subjects will play an average-truth-telling mean dynamic*

   (b) *The random treatment will achieve the same level of efficiency as the message price treatment*

To further replicate the large economy environment, every experimental subject managed ten identical drivers. In every round, each subject decides whether to drive or not; if he decides to drive, a driver of his type is introduced to the road (up to ten); if he decides to not drive, a driver is removed from the road (up to zero). The experiment was run at the Experimental Economics Lab at the University of Maryland. There were 130 participants, all undergraduate students at the University of Maryland. There were nine sessions. No subject participated in more than one session. In every session, subjects participated in six different treatments. Treatments were played in random order. Participants were seated in isolated booths. The experiment is programmed in z-Tree (Fischbacher (2007)).

At the beginning of each treatment, each subject was randomly assigned a type, i.e. a value of commuting D and a value of time v. In addition, they were informed that in some rounds they could face a congestion price T or a transfer S and that their experimental payoffs would depend on the observed time t using the following formulas: $D - \frac{vt}{60} - T$ for driving and S for not driving. In all rounds, subjects could see on screen the current values of T and S, the history of times for all previous rounds and their private information. In addition, a table with several time scenarios ($t = 5$ to $t = 85$ in steps of 5) with the values for driving and not driving was provided.

Subjects were informed that in some sections (treatments) they could be asked for their value of time and were instructed to "send one of the available messages". Subjects were informed that messages would be used to calculate the congestion price for the next period, but the exact mechanism was not explained because in the experimental setting, due to the small number of participants, every message had a measurable impact on the congestion price.

Subjects were explained in detail how earnings were calculated. In every round r, subjects received $x_r = (0.9764)^{30-r}$ ($x_{30} = 1$,$x_1 = 0.5$) "points" for a conditionally optimal action and 0 otherwise. This payment scheme fulfills two purposes. First, no Nash equilibrium is favored; remember that for the message treatment there are many equilibria for this game. Second, it provides incentives for agents to adjust their strategies over time. Dollar earnings were calculated by adding up all points and multiplying this quantity by 0.107675. This constant was calculated, and explained as such, to produce a range from \$0 to \$14 dollars. In addition, subjects were paid a \$6 show up fee. Subjects received an average payment of \$18.28. The following section present the results of the experiment and gives a general description of some stylized facts.

3.4.1 Experimental results

The results of the experiment are presented in this section. For every treatment, three different dimensions are described: the number of drivers on the road, their types, and the efficiency. The analysis of the results is included in the following section.

3.4.1.1 Number of drivers

The main objective of a congestion price is to achieve an efficient congestion level. In every round, the number of drivers is measured by $x_s = \sum x_{is}$, where x_{is} is the proportion of subject i's 10 drivers currently on the road in round s. The Nash equilibrium quantity of drivers with no congestion price is 10. The socially optimal quantity of drivers is 6.

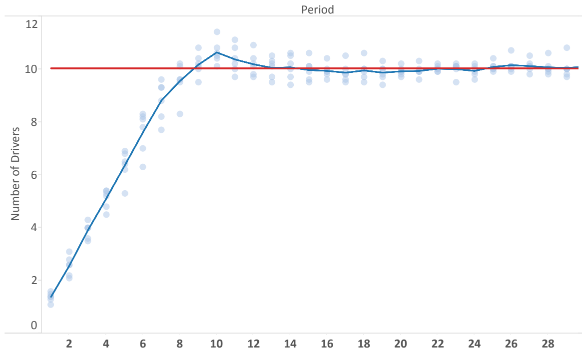


Figure 3.3: Number of drivers with No price

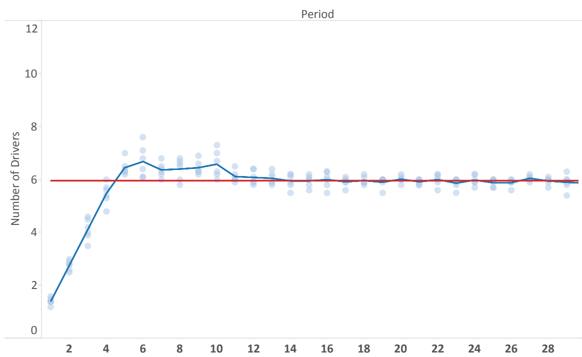


Figure 3.4: Number of drivers with Dynamic price

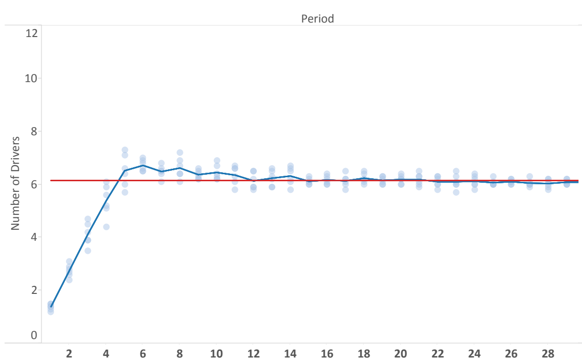


Figure 3.5: Number of drivers with Balanced Dynamic price

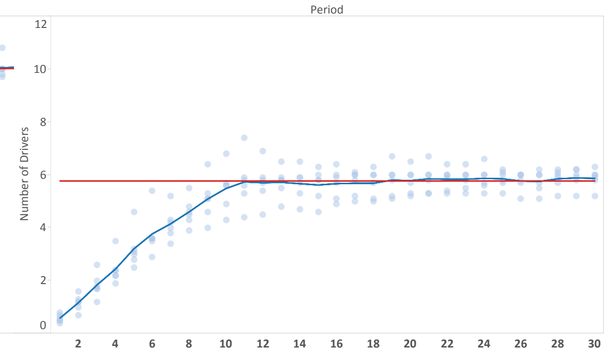


Figure 3.6: Number of drivers with Fixed price

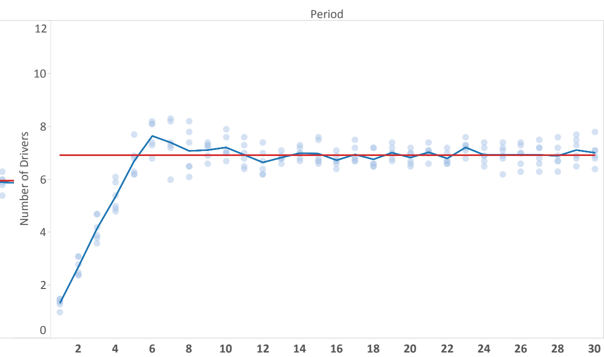


Figure 3.7: Number of drivers with Message price

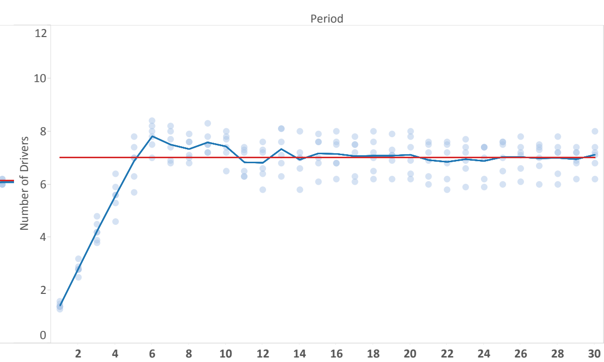


Figure 3.8: Number of drivers with Balanced Message price

The number of drivers of every treatment is shown in figures 3.3 through 3.8. In every figure, every blue dot is the observed number of drivers in each period in each session. The blue line is the average over sessions. The read line is the simple average of each blue dot's number of drivers for periods equal or greater than 11.

Figure 3.3 shows the evolution of the number of drivers for the treatment without congestion pricing. In this treatment, the Nash equilibrium quantity of drivers is 10. In the experiment, 10.03 was observed.

In figure 3.6, the results of the fixed congestion price are shown. This treatment represents the theoretical best option, as it assumes the policy maker knows all the information, in this case, θ_{1i} and θ_{2i} for every subject. The social optimum is associated with 6 drivers. In the experiment, the observed number of drivers was 5.77.

In figure 3.4 the results of the dynamic price are shown. In the experiment, the number of drivers was 5.96. It can be observed that the number of drivers fluctuates less around the average and converges faster to the average value when compared with the fixed congestion price or with the no price treatments. In this treatment it is assumed that the policy maker knows v_i for every subject and can perfectly identify each driver on the road.

Figure 3.7 shows the results for the message price. The observed number of drivers was 6.92. In this treatment, the policy maker has no information about D_i and v_i.

Figures 3.5 and 3.8 shown the balanced versions of the dynamic and message price treatments. It can be observed that the effectiveness of the systems is not decreased by charging lower congestion prices and distributing all the proceeds to subjects who decide not to drive. In the balanced dynamic price treatment, the observed number of drivers is 6.14. In the balanced message price treatment, the observed number of drivers is 7.01.

3.4.1.2 Identities

An effective system would not only produce the right level of traffic congestion, but also the right set of drivers. Figures 3.9 through 3.14 are analogous to figure 3.2. They show the types in a Cartesian plane where the "x-axis" is the value of time and the "y-axis" is the value of commuting. Every blue dot represents a type. The size and the number next to each dot represent the frequency that type was driving for periods equal to or greater than eleven. The two gray lines represent the Nash equilibrium time without congestion price and the social optimum time. The green line represents the observed average time. When all subjects play a Nash equilibrium strategy, the frequency of each blue dot is 100% for types above the green line and 0% for types below the green line.

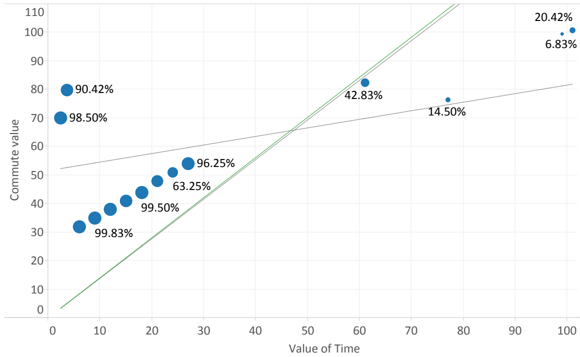


Figure 3.9: Identities with No price

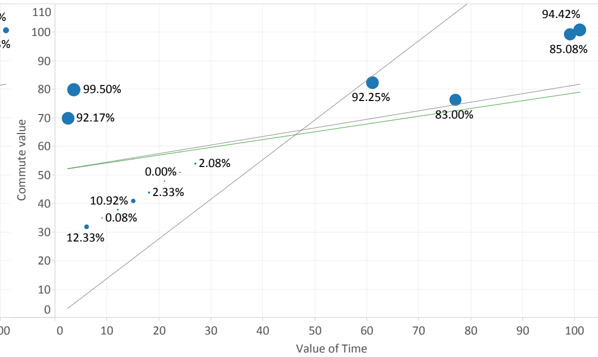


Figure 3.12: Identities with Fixed price

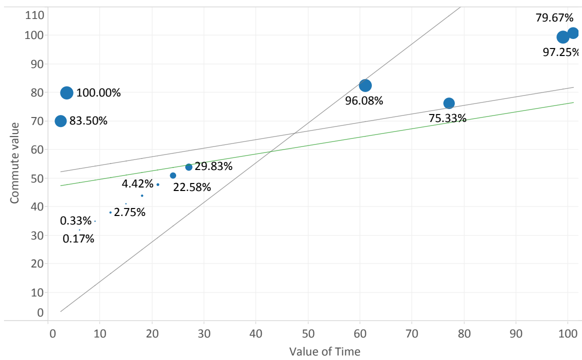


Figure 3.10: Identities with Dynamic price

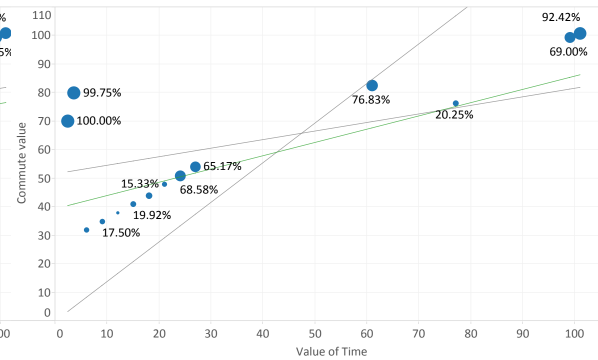


Figure 3.13: Identities with Message price

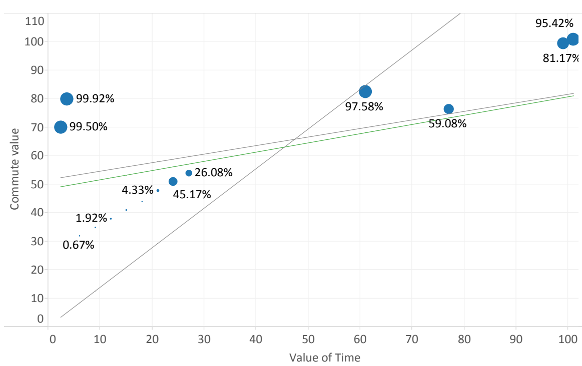


Figure 3.11: Identities with Balanced Dynamic price

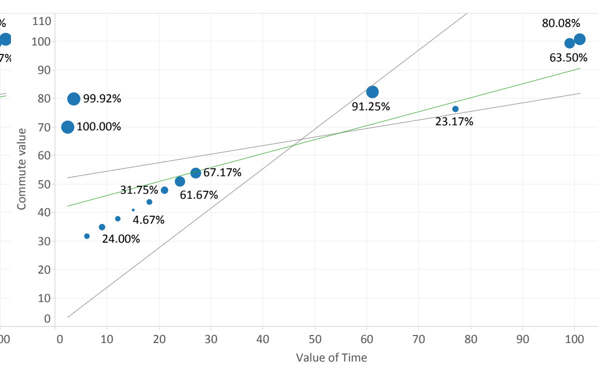


Figure 3.14: Identities with Balanced Message price

In figure 3.9 the types for the No price treatment are shown. It can be observed that all types that, in equilibrium, should drive are driving, but not in 100% of the periods. On the other hand, some types that should not drive, in equilibrium, drive some of the periods. In particular, type 11 (value of time = 60.96, value of commuting = 82.65) fails to stop driving in 42% of the periods.

In figure 3.12, the fixed congestion price has been imposed. The types who would benefit from driving do, but not in 100% of the periods. In particular, type 1 (value of time = 2.4, value of commuting = 70) drives in 92% of the periods, despite having strong incentives to keep driving. Similarly, type 12 (value of time = 77.02, value of commuting = 76.35) does not drive in 100% of the periods and forgoes positive payoffs (and payments).

Figure 3.10 shows the dynamic price treatment. In this treatment, types 1 and 12 display a behavior similar to their behavior in the treatment with the fixed congestion price: they fail to drive 100% of the time, despite being profitable. In the fixed congestion price treatment, this behavior had consequences only for the subject making the suboptimal decision. However, in this treatment, their actions had an impact on the congestion price charged to others. In particular, types 9 (value of time = 24, value of commuting = 51) and 10 (value of time = 27, value of commuting = 54) benefited from this behavior. On average, when type 12 failed to drive, despite being profitable, types 9 and 10 entered the road.

Figure 3.13 shows the message price treatment. It can be observed that, conditional on observed times and congestion prices, most types who would benefit from driving do. However, in this treatment type 12 drove even less than in the treatment with the dynamic price and this opportunity was seized by types 9 and 10. Balanced treatments are shown in figures 3.11 and 3.14.

3.4.1.3 Efficiency

Efficiency is measured as the sum of experimental payoffs in very round. Every subject received two numbers: a value of commuting θ_{2i} and a value of time θ_{1i}. Efficiency in round s is defined by $E_s = \sum_{i=1}^{14}(\theta_{2i} - \frac{\theta_{1i}}{60}t_s)x_i$, where x_i is the proportion of subject i's 10 drivers currently on the road and t_s is the observed time in round s. In every round, the time was calculated using the function $t_s(x_s) = \frac{x_s^3}{12}$, where $x_s = \sum x_i$. All treatments are initialized with $x_i = 0$ for all subjects.

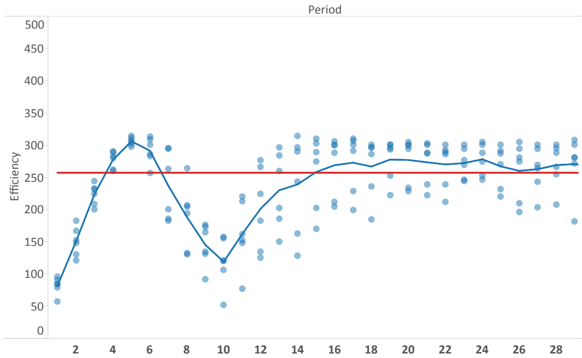


Figure 3.15: Efficiency with No price

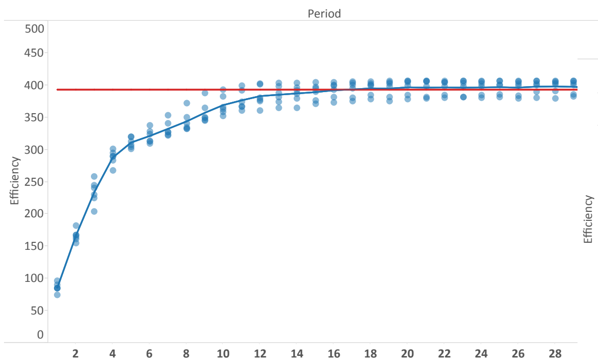


Figure 3.16: Efficiency with Dynamic price

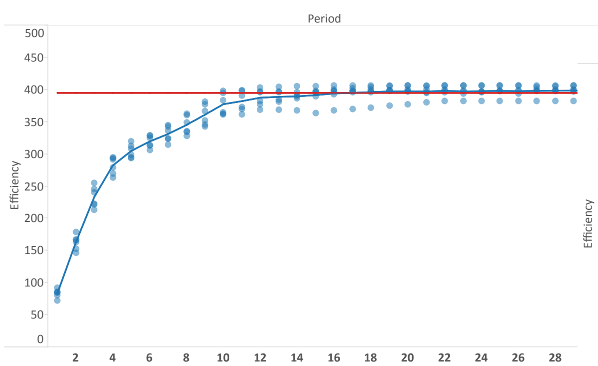


Figure 3.17: Efficiency with Balanced Dynamic price

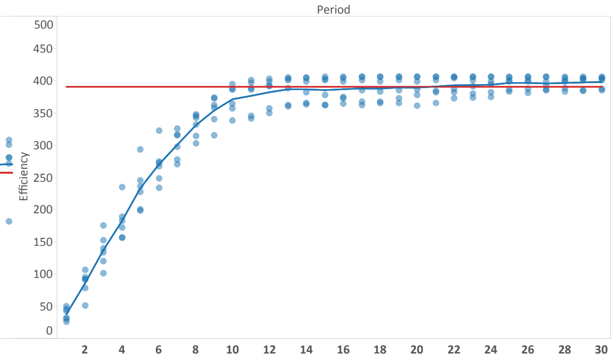


Figure 3.18: Efficiency with Fixed price

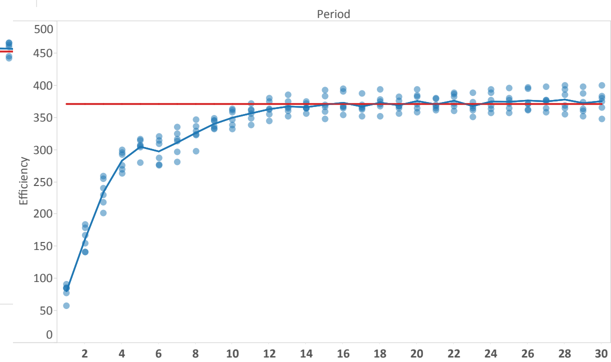


Figure 3.19: Efficiency with Message price

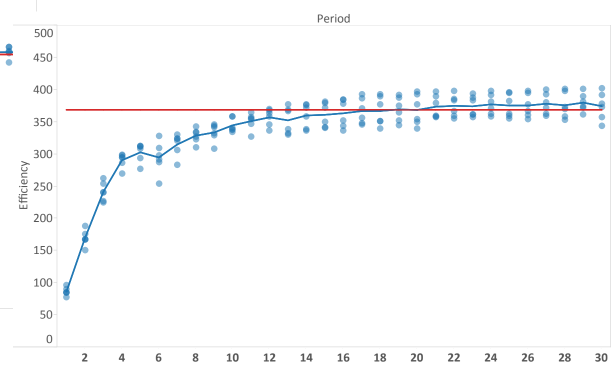


Figure 3.20: Efficiency with Balanced Message price

The efficiency of every treatment is shown in figures 3.15 through 3.20. In every figure, every blue dot is the observed efficiency in each period in each session. The blue line is the average over sessions. The read line is the simple average of each blue dot's efficiency for periods equal or greater than 11.

Figure 3.15 shows the evolution of efficiency of the treatment without congestion pricing. In this treatment, the Nash equilibrium is associated with an efficiency of 301.3 experimental dollars. In the experiment, the observed efficiency was 257.7.

In figure 3.18, the results of the fixed congestion price are shown. This treatment represents the theoretical maximum efficiency that can be achieved. It assumes the policy maker knows all the information, in this case, D_i and v_i for every subject. The social optimum achieves an efficiency of 406.6 experimental dollars. In the experiment, an efficiency of 390.9 was observed.

In figure 3.16 the results of the dynamic price are shown. In the experiment, an efficiency of 393.0 was observed. It can be observed that the efficiency fluctuates less around the average and converges faster to the average value. Both characteristics are consequences of the the stability of the game. In this treatment it is assumed that the policy maker knows v_i for every subject and can perfectly identify each driver on the road.

Figure 3.19 shows the results for the message price. The observed efficiency is 371.4 experimental dollars. This is a high level of efficiency, considering the fact that in this treatment D_i and v_i are unknown.

Figures 3.17 and 3.20 show the balanced versions of the dynamic and message price treatments. It can be observed that efficiency is not hurt by charging lower congestion prices and distributing all the proceeds to subjects who decide not to drive. In the balanced dynamic price treatment, the observed efficiency is 395.0. In the balanced message price treatment, the observed efficiency is 368.8.

3.4.2 Analysis

This section evaluates the hypothesis derived from the theory. The main objective of the experiment is to test whether the message system allows drivers to converge to the socially optimal traffic congestion level. Other treatments are design to put the results of the message price treatment in context. In this section it is considered that a treatment has converged in period p whenever the average absolute deviation from the mean efficiency is less or equal to 5% for all consecutive periods. The mean efficiency in period p is $m_p = \frac{1}{30-p+1}\sum_{i=p}^{30} E_i$, the absolute deviation in period $w \geq p$ with respect to the mean efficiency at p is $e_{w,p} = |E_w - m_p|$ and the average absolute deviation is $e_p = \frac{1}{30-p+1}\sum_{i=p}^{30} \frac{e_{i,p}}{m_p}$. A treatment converged in period p whenever $e_s \leq 5\%$ for all $s > p$. All treatments, but the no price treatment, converged on period 6. The no price treatment converged on period 11.

Hypothesis 1. *The no congestion price treatment will achieve the theoretical efficiency associated with no congestion price*

This is a standard hypothesis supported the rational model. The theoretical efficiency associated with no congestion price is 301.3 experimental dollars. Figure 3.15 shows that $m_{11} = 257.6$. Assuming that $E_s = m_{11} + \epsilon_s$, where ϵ is i.i.d $E[\epsilon_s] = 0$ for all periods $s \geq 11$, a t-test was used to evaluate the null hypothesis of $m_{11} = 301.3$ versus the alternative $m_{11} \neq 301.3$. The null was rejected with confidence of 99%. In the experiment, the no congestion price achieved a lower efficiency than the rational model. This fact is at odds with a purely rational model of human behavior. This deviation could have happened in the opposite direction, and after all, a congestion price might not be needed.

Hypothesis 2. *The fixed price treatment will achieve the theoretical optimal efficiency*

This is a standard hypothesis supported the rational model: a social planner would be able to solve the congestion problem with a pigouvian price. The theoretical efficiency associated with the optimal congestion is 406.6 experimental dollars. Figure 3.18 shows that $m_{11} = 390.92$. Assuming that $E_s = m_{11} + \epsilon_s$, where ϵ is i.i.d $E[\epsilon_s] = 0$ for all periods $s \geq 11$, a t-test was used to evaluate the null hypothesis of $m_{11} = 406.6$ versus the alternative $m_{11} < 406.6$. The null was rejected with confidence of 99%.

Hypothesis 3. *The dynamic price treatment will achieve the same efficiency as the fixed price treatment*

The message treatment differs from the fixed price treatment in two aspects: it changes over time and depends on reports. The dynamic price treatment bridges these differences by changing over time, but is independent of agents' reports. Congestion prices in this treatment behave as if all subjects told the truth all the time. Figure 3.16 shows that efficiency observed in the dynamic message treatment was $m_{11} = 393$. A paired Wilcoxon signed-rank test was used to evaluate the null hypothesis that the differences between the dynamic price and the fixed price efficiencies were symmetric around zero. This test does not require additional assumptions about error terms. The null was not rejected (p $>$ 10%).

The efficiency results of the no price and fixed price treatments show that the conclusions of the rational model are likely to fail in a real-world situation. The results from the fixed price and dynamic price treatments are evidence that theoretical efficiencies might not be achievable in real life.

Hypothesis 4.b *The message treatment will achieve the same efficiency as the fixed price treatment*

Figure 3.19 shows that the message treatment achieved an average efficiency $m_{11} = 371.36$. A paired Wilcoxon signed-rank test was used to evaluate the null hypothesis that the differences between the message price and the fixed price efficiencies were symmetric around zero. The null was not rejected (p $>$ 10%). The efficiency observed in this treatment is 95% of the efficiency achievable by a social planner with full information.

Hypothesis 5 *The balanced treatments will achieve the same efficiency as the unbalanced treatments*

Figures 3.17 and 3.20 show the results of the balanced treatments. The balanced dynamic price treatment obtained an average efficiency $m_{11} = 395.03$. The balanced message price treatment obtained an average efficiency $m_{11} = 368.84$. In both cases, the null hypothesis was that the balanced treatments would achieve an efficiency equal to their unbalanced versions. The null hypothesis was not rejected in both cases (p $>$ 10%).

Table 3.2 contains a summary of the mean efficiency achieved in every treatment as a percentage of the mean efficiency obtained by the fixed tax treatment. The standard deviation has been scaled accordingly. The table in the middle contains p-values for the null hypothesis that the row treatment and the column treatment have the same efficiency against the alternative that the row has a higher efficiency. A paired Wilcoxon signed-rank test was used. The lower portion of the table shows the results for the number of drivers on the road. Estimates of the average number of drivers have not been scaled because units represent subjects' decisions directly. P-values are also reported for the number of drivers. The alternative hypothesis is that the row treatment has a lower number of drivers than the column treatment. The last column shows e_p, and an analogous measure for the number of drivers, for every treatment. All estimates are calculated using data from periods 11 to 30. Figures 3.21 and 3.22 show estimates for efficiency and the number of drivers for different choices of initial period of analysis. All treatments are significantly (p-values $<$ 1% for all periods of analysis) more efficient than the no price treatment. Dynamic treatments and the fixed treatment achieve a significantly (p-values $<$ 1% for all periods of analysis) higher efficiency than message treatments. 95% confi-

Period	Treatment	Measure	Mean	SD	Epsilon
11	No Price	efficiency	65.91%	13.01%	4.88%
11	Fixed	efficiency	100.00%	4.12%	1.19%
11	Dynamic	efficiency	100.53%	3.26%	1.14%
11	Message	efficiency	95.00%	3.44%	1.14%
11	Bdynamic	efficiency	101.05%	2.85%	0.91%
11	Bmessage	efficiency	94.35%	4.86%	1.94%
11	No Price	N. of Drivers	10.030	0.323	15.44%
11	Fixed	N. of Drivers	5.770	0.493	22.74%
11	Dynamic	N. of Drivers	5.963	0.202	22.31%
11	Message	N. of Drivers	6.926	0.356	22.36%
11	Bdynamic	N. of Drivers	6.147	0.214	25.84%
11	Bmessage	N. of Drivers	7.015	0.606	17.06%

Table 3.1: Estimates for period ≥ 11

Period	Treatment	Measure	No Price	Fixed	Dynamic	Message	Bdynamic	Bmessage
11	No Price	efficiency						
11	Fixed	efficiency	<1%			<1%		<1%
11	Dynamic	efficiency	<1%			<1%		<1%
11	Message	efficiency	<1%					
11	Bdynamic	efficiency	<1%	<1%		<1%		<1%
11	Bmessage	efficiency	<1%					
11	No Price	N. of Drivers						
11	Fixed	N. of Drivers	<1%		<1%	<1%	<1%	<1%
11	Dynamic	N. of Drivers	<1%			<1%	<1%	<1%
11	Message	N. of Drivers	<1%					
11	Bdynamic	N. of Drivers	<1%			<1%		<1%
11	Bmessage	N. of Drivers	<1%					

Table 3.2: P values for estimates for period ≥ 11

dence interval are shown in Figures 3.23 and 3.24.

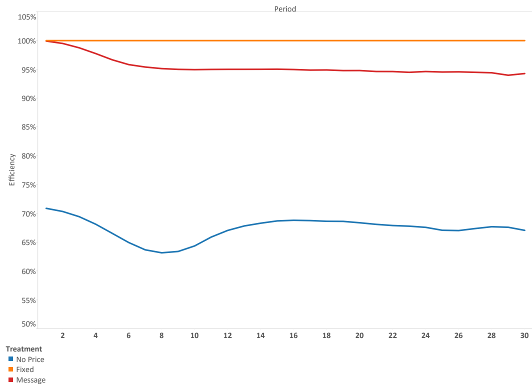


Figure 3.21: Efficiency estimates by period

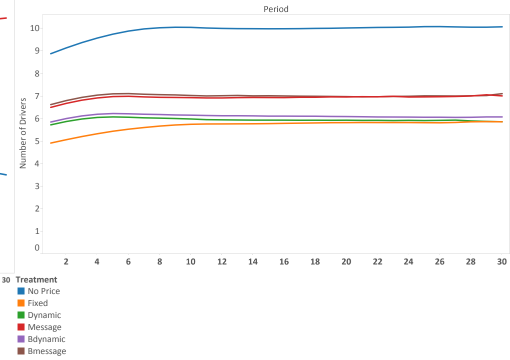


Figure 3.22: No. of Drivers estimates by period

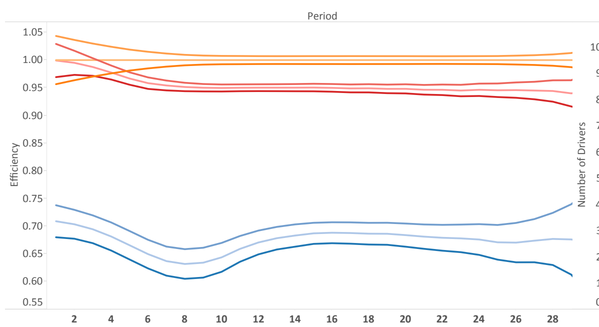


Figure 3.23: 95% Confidence Intervals for efficiency estimates. Blue: No price; Red: Message; Orange: Fixed

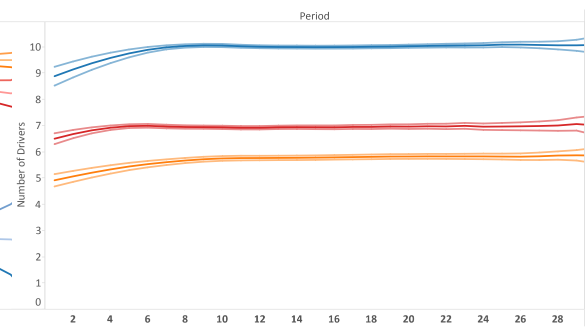


Figure 3.24: 95% Confidence Intervals for Number of Drivers estimates. Blue: No price; Red: Message; Orange: Fixed

3.4.2.1 Message Congestion Price

This section describes the observed messages and confirms that subjects followed an average truth-telling mean dynamic, hence the high levels of efficiency. In principle, even assuming that subjects would play a Nash equilibrium, efficiency gains are not guaranteed. Figure 3.25 shows the efficiency levels of all Nash equilibria in the game induced by the message congestion price by average message: $z(\hat{x}) = \frac{\sum_{\theta} \theta_{1i} \hat{x}_{\theta}}{x}$.

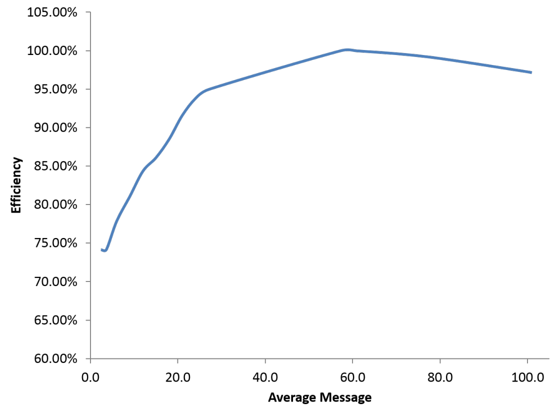


Figure 3.25: Nash Equilibria Efficiency by average message

In figure 3.25, when all subjects send the lowest possible value of time, the congestion price is sufficiently low to be completely ineffective i.e. the Nash equilibrium with no congestion price is also a Nash equilibrium of the message congestion price system. However, as argued before, the final outcome of the system does not only depend on its Nash equilibria, but also (and more importantly) on the non-equilibrium behavior. In particular, the outcome of the system is tied to the aggregate message, which is determined by individual messages.

Figure 3.26 shows the average message sent by type. Types who drive in the social optimum are shown in blue, types who do not drive in the social optimum are shown in gray. It can be observed that those types who drive in the social optimum send higher messages than those who don't.[19] In addition, it can be observed that some types send higher values than their true values while other types do the opposite. Figure 3.27 shows the number of times a particular message was received by the system as a proportion of the total number of messages received. It can be observed that the lowest and highest messages are the most often used.

[19] The mean message sent by those types who drive in the optimal allocation is 29.64 (30.17), the mean message for other types is 16.06 (24.21). The average message of the optimal group is greater with a confidence level of 99% using a Welch's t-test.

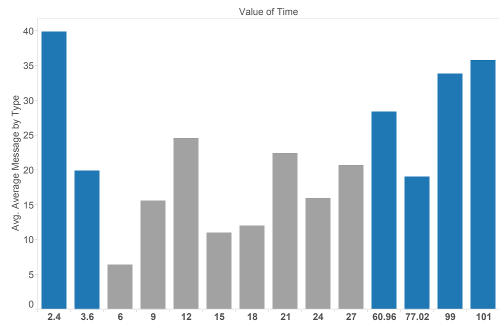


Figure 3.26: Average Message Sent by Type

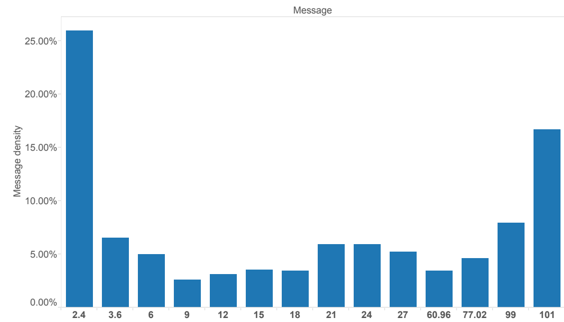


Figure 3.27: Messages Received

Individual messages are important, but they have a very limited impact on the system's outcome, as the congestion price depends on the average message. Figures 3.28 and 3.29 show the relationship between the average message and the real average message, as if all subjects reported their true value of time. Figure 3.28 shows their evolution over time (all sessions aggregated) and figure 3.29 shows all data points.

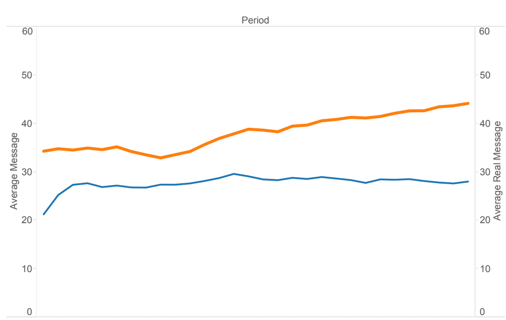


Figure 3.28: Average Message over time

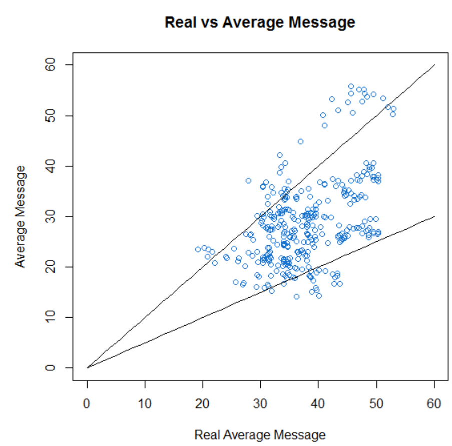


Figure 3.29: Real vs Sent Average Message

In the previous two figures, two stylized facts about the average message are readily observable: (i) the population understates its value of time, (ii) but not to the lowest possible extent. These behavioral regularities guarantee efficiency gains in the message treatment. Consider the unconditional distribution of messages sent G and let z^* be the equilibrium average message when all drivers send their true value of time. Since average sent messages are smaller than average real messages we have that $G(z^*) = 1$ i.e. the highest observed average message will always be below the real equilibrium average message. Let $f(z)$ be the achieved efficiency when z is sent to the system. Then, unless G is degenerate, $E[f(z)] > f(0)$ i.e. the implementation of the message system is guaranteed to generate efficiency gains, unlike policy guesses about the value of time.[20] Figures 3.30 and 3.31 shows the empirical unconditional density and distribution.

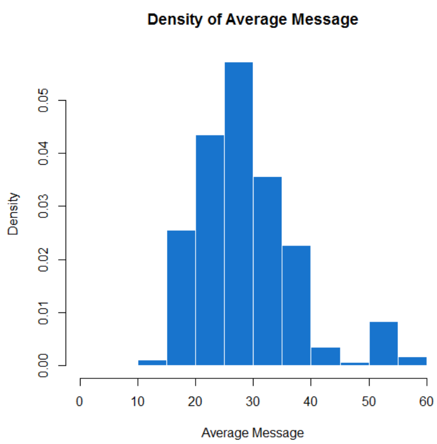


Figure 3.30: Density of Average Message

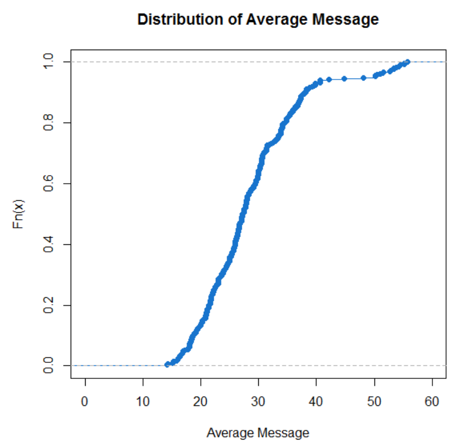


Figure 3.31: Distribution of Average Message

The average message can explain that the observed efficiency gains are positive, but not their high level. In theory, whenever agents play an average truth-telling mean dynamic in the presence of the message congestion price, the efficient outcome is expected. Recall from previous sections that a mean dynamic $\hat{V} : \hat{X} \rightarrow \mathbb{R}^{|\Theta|\times|S\times\Theta|}$ describes what actions and messages are sent and the mean dynamic $V_{\theta s}(x) = \sum_{\hat{\theta}} x_{\theta s\hat{\theta}}$ describes all actions as if all agents reported the truth.

[20] As an example, suppose G is uniform, then the minimum efficiency of the message system would be $\frac{1}{2} + \frac{f(0)}{2}$.

Hypothesis 4.a *Subjects will play an average-truth-telling mean dynamic in the message price treatment.*

An average truth-telling mean dynamic is characterized by one inequality: $0 < V(x) \cdot \nabla W(x) = \sum \dot{x}_\theta(\theta_{2d} - \theta_{1d} t(\sum_\theta x_{\theta d}) - t'(x) \sum_\theta \theta_{1d} x_{\theta d})$ whenever $V(x) \neq 0$. This is the covariance between the direction taken by agents and the direction of greatest increase on welfare. Proposition 5 shows that as long as this covariance is positive, agents are guaranteed to arrive to the social optimum. Figures 3.32 and 3.33 show observed covariance in the message price treatment. Every observation is calculated as $\sum (x_{t,\theta} - x_{t-1,\theta})(\theta_{2d} - \theta_{1d} t(\sum_\theta x_{t,\theta d}) - t'(x_t) \sum_\theta \theta_{1d} x_{t,\theta d})$ for periods $t = 1 \ldots 30$. A binomial test was used to reject the hypothesis that the covariance was zero against the alternative of being greater than zero. The null was rejected at a confidence level of 99%. Figures 3.34 and 3.35 show the covariance for the no price treatment. The null was not rejected (p-value >10%).

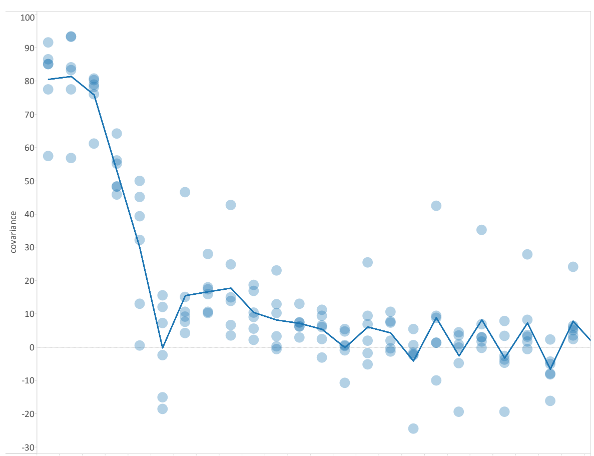


Figure 3.32: Average-truth-telling - message treatment

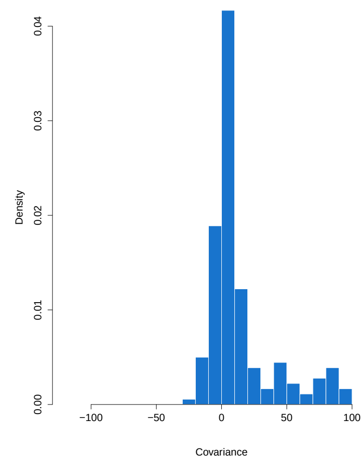


Figure 3.33: Average-truth-telling histogram - message treatment

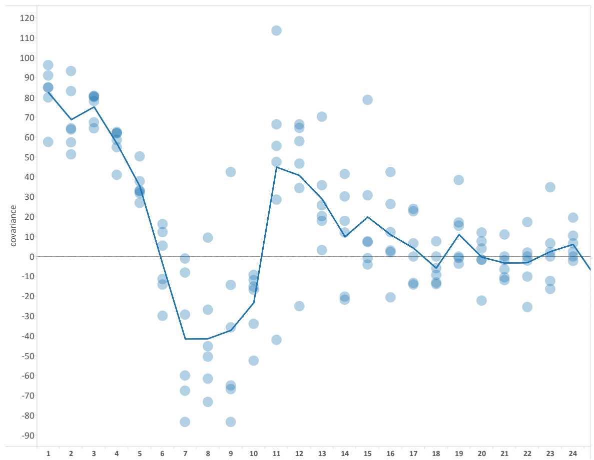


Figure 3.34: Average-truth-telling - no price treatment

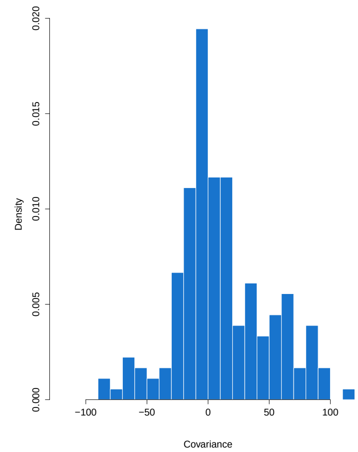


Figure 3.35: Average-truth-telling histogram - no price treatment

3.4.3 Robustness

The experimental design pursued in this paper relied on a particular selection of types. However, it is important to test the robustness of the message price mechanism to different sets of types. Figure 3.36 shows the efficiency achieved in six different random treatments in which 14 subjects received a random value of time and a random value of commuting, both sampled from a uniform distribution with support $[1, 100]$. These random treatments are otherwise identical to the message price treatment discussed above.

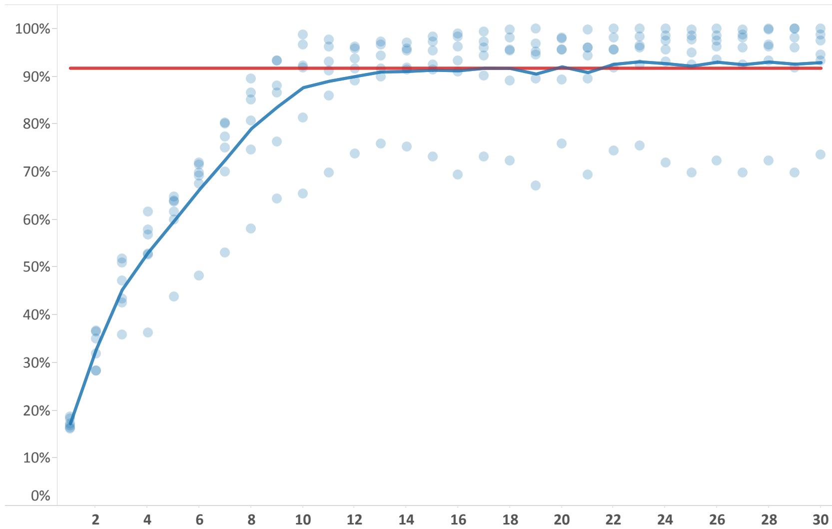


Figure 3.36: Efficiency of Random Treatments

Figures 3.37 to 3.48 show the experimental results of every random treatment. Figures on the left display driving frequency by type. Those types who drive in the social optimum are depicted in orange. Figures on the right show efficiency over time.

Hypothesis 6.a. *Subjects will play an average-truth-telling mean dynamic*

Hypothesis 6.b. *The random treatment will achieve the same level of efficiency as the message price treatment*

The following table shows the average efficiency achieved. The message price treatment achieved an efficiency of 91.46% (3.31%) (with respect to the theoretical optimum). A Welch's t-test was used to test the null hypothesis that the efficiency in each random treatment is equal to 91.46% against the alternative that the efficiency in the random treatment was smaller. In all random treatments, but the third, the null was not rejected i.e. the message congestion price performed equally on random types as in designed types. A binomial test was used to reject the hypothesis that the covariance was zero against the alternative of being greater than zero. Random treatments 2 and 3 highlight the importance of the careful selection of types in the main message treatment. In random treatment 2 the Nash equilib-

Random	Mean	SD	Equilibrium	Message price	Avg-truth-telling
1	94.92%	3.74%	77.90%	>10%	<1%
2	97.51%	0.86%	93.75%	>10%	5%
3	72.21%	2.53%	63.34%	<1%	5%
4	98.96%	1.38%	91.51%	>10%	<1%
5	91.48%	1.37%	82.89%	>10%	2.13%
6	95.38%	0.88%	67.66%	>10%	2.13%

Table 3.3: Random Types Efficiency for Periods 11-30

rium efficiency without congestion pricing is high, reducing the potential gains of the message mechanism and hence the ability to identify them. Random treatment 3, on the other hand, displays 4 types who are aligned and hence poorly represent a situation with a large number of drivers.

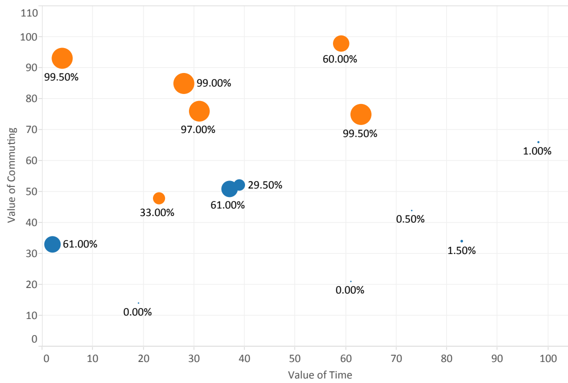


Figure 3.37: Types Random Treatment 1

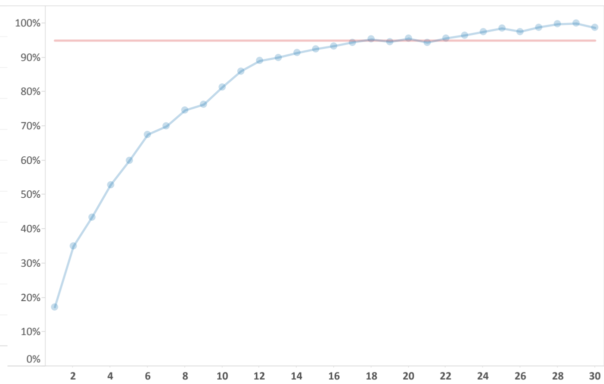


Figure 3.40: Efficiency Random Treatment 1

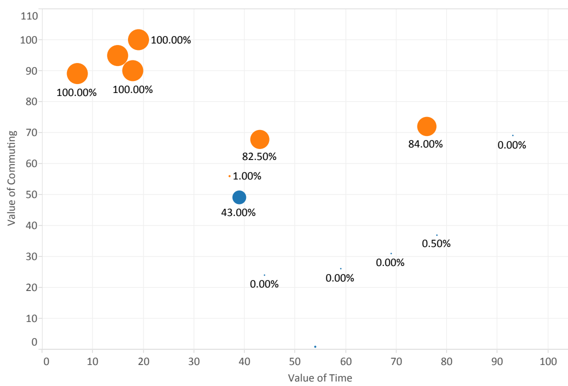


Figure 3.38: Types Random Treatment 2

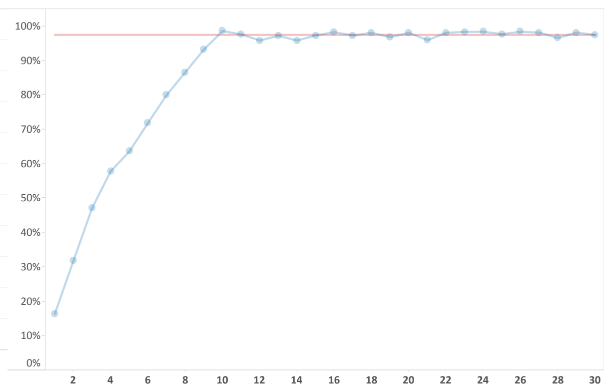


Figure 3.41: Efficiency Random Treatment 2

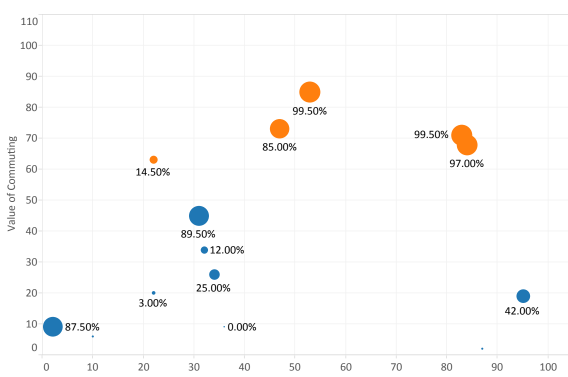


Figure 3.39: Types Random Treatment 3

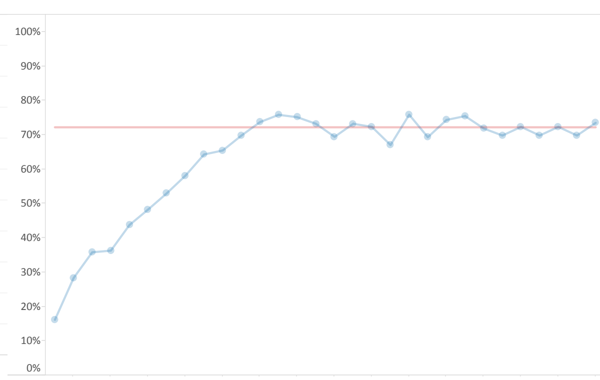


Figure 3.42: Efficiency Random Treatment 3

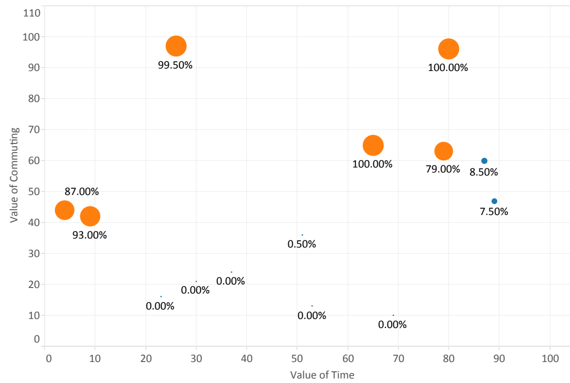


Figure 3.43: Types Random Treatment 4

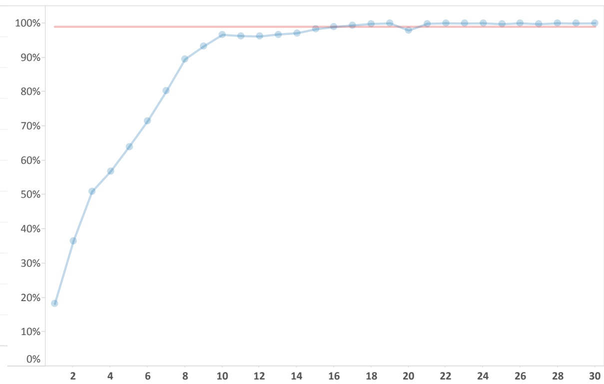


Figure 3.46: Efficiency Random Treatment 4

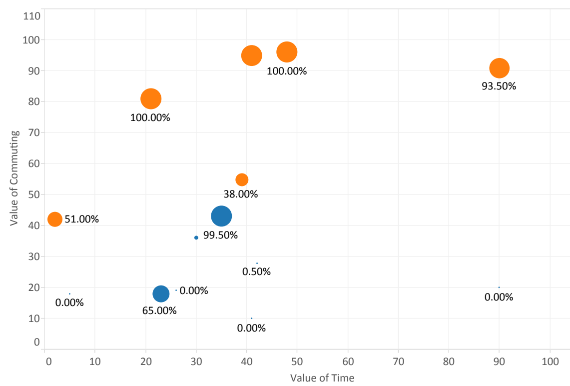


Figure 3.44: Types Random Treatment 5

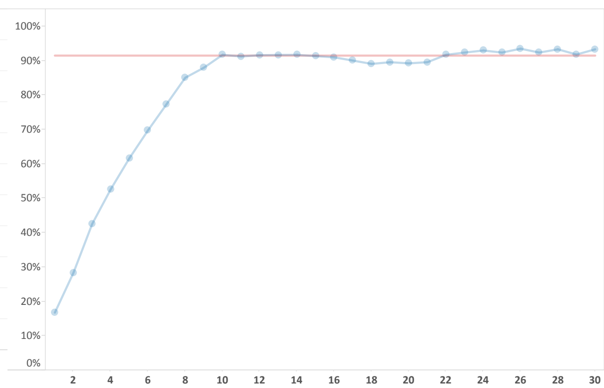


Figure 3.47: Efficiency Random Treatment 5

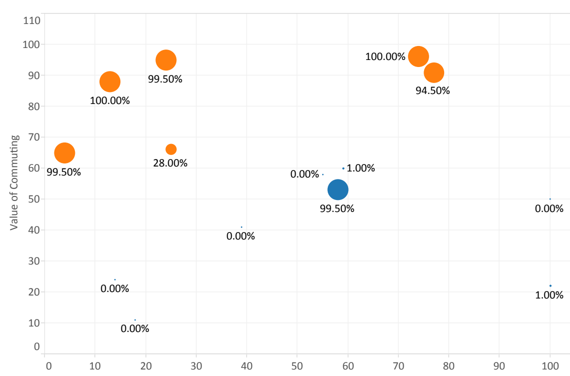


Figure 3.45: Types Random Treatment 6

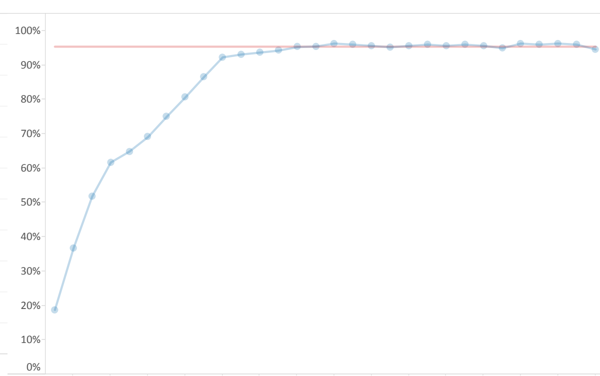


Figure 3.48: Efficiency Random Treatment 6

3.5 Discussion

A social planner would like a socially optimal outcome $x^*(\theta)$ to be chosen in every state of the world $\theta \in \Theta$. In general, this can be done in two steps: i) using a mechanism M to make $x^*(\theta)$ a rational choice (a Nash equilibrium), and ii) providing M with nice properties that facilitate coordination in $x^*(\theta)$. This has been the objective of mechanism design.[21] However, most mechanisms assume that agents are fully rational all the time and possess common knowledge of types and the structure of the game induced by the mechanism. These assumptions have proven extremely useful and powerful as they have allowed the study of very complex problems as well as the development of many successful mechanisms, but has well identified limitations.

This paper addresses one of those limitations by incorporating behavioral traits as a mechanism designer tool and showing that it can be as effective as strong incentive properties in solving social problems.

The introduction of behavioral traits to the mechanism design framework enables the study of questions typically outside the scope of the purely rational model: Are mechanisms with the same incentive properties equally effective?[22] Are incentives more effective the stronger they are?[23] Are incentives more effective the simpler they are?[24] What considerations, other than incentives, affect the effectiveness of a

[21] Maskin (2008)

[22] There might be two efficient and incentive compatible mechanisms for the same problem, of which only one is effective.

[23] A measure of incentive strength could be the difference in payoff between truth-telling and the best misrepresentation.

[24] Consider, for example, truth-telling as a dominant strategy and as a Nash equilibrium, the former being simpler.

mechanism?[25] When is it efficient to provide incentives?[26] Can non-incentive compatible mechanisms be more effective than incentive compatible ones?[27]

The answers to these questions will most likely unveil an intricate relationship between rational incentives and behavioral traits, opening the door to new methods for solving problems in practice.

3.6 Appendix

Proposition 3.1. There is an efficient mechanism with weak incentives, namely the wVCG.

Proof. Let x^*be an efficient profile of actions and θ be the true profile of types. Suppose all agents other than i select x_j^* and report their true type θ_j. For i, the payoff associated with doing x_i and reporting θ_i' is $u(x_i, x_{-i}^*, \theta_i) + \sum_{N \setminus i} u_j(x_i, x_{-i}^*, \theta_j) - h_i^w(x_{-i}, \theta_{-i})$ which is maximized by selection x_i^* as an action and θ_i as a report. □

Proposition 3.2. There is a budget balanced mechanism with weak incentives for any profile of actions. In particular, any efficient profile of actions can be supported as a budget balanced Nash equilibrium.

Proof. Let x^0 be any profile of actions and let prices be defined as $p_i(x, \theta) = \sum_{N \setminus i} u_j(x, \theta_j) - \sum_{N \setminus i} u_j((x_i^0, x_{-i}), \theta_j)$, thus $p_i(x^0, \theta) = 0$ for all $i \in N$ and $\theta \in \Theta$. In particular, let $x^0 = x \in x^*(\theta)$, then the efficient profile of actions can be supported as a budget balanced Nash equilibrium. □

---

[25] For example, a mechanism that converges to the efficient Nash equilibrium under a wide class of behavioral procedures have a better change of being effective than a mechanism that cannot guarantee such convergence.

[26] Usually, the efficiency of a mechanism is measured by the efficiency attained within the mechanism i.e. by the outcome it produces, however, this measure leaves other considerations out of the analysis. For example, how expensive is to implement and run the mechanism.

[27] It is possible that some effective mechanisms support $x^*(\theta)$ as a non-equilibrium but sensible profile of actions.

Proposition 3.3. Any efficient profile of actions can be supported as a subgame perfect Nash equilibrium of a sequential mechanism with weak incentives.

Proof. The timing is as follows: i) agents select an action, ii) the profile of actions is revealed, and iii) agents send a report. Suppose a profile of actions x was chosen in the first stage of the game. Suppose other agents have sent θ_{-i}, sending report θ_i' is associated with a payoff equal to $u(x_i, x_{-i}^*, \theta_i) + \sum_{N\setminus i} u_j(x_i, x_{-i}^*, \theta_j) - h_i^w(x_{-i}, \theta_{-i})$, hence sending θ_i is a best response. Thus θ constitutes a Nash equilibrium in the second stage. Suppose agents have chosen x_{-i}^* in the first stage, the payoff associated with x_i subject to selecting the Nash equilibrium θ in the second stage is $u(x_i, x_{-i}^*, \theta_i) + \sum_{N\setminus i} u_j(x_i, x_{-i}^*, \theta_j) - h_i^w(x_{-i}, \theta_{-i})$, hence i maximizes his payoff by selecting x_i^*as an action. Thus (x^*, θ) is a subgame perfect Nash equilibrium. □

Proposition 3.4. Let $v_{\theta s}(x) = F_s(x)\theta_{1s} + \theta_{2s} + \sum_j \frac{\partial F_j}{\partial x_s} \sum_\theta x_{\theta j}\theta_{1j}$ for all $\theta \in \Theta$ and $s \in S$ and V an admissible mean dynamic such that $V(x) \cdot \nabla W(x) > 0$ whenever $V(x) \neq 0$, then every solution trajectory of V converges to the efficient distribution of actions x^*.

Proof. Let $x : \mathbb{R}_+ \to X$ be a solution trajectory of V, then all of its accumulation points are critical points of $W \circ x$. Since W is concave it has a unique maximizer x^* and $\nabla W(x) = 0$ only when $x = x^*$. x^* is also the unique Nash equilibrium. Since $V(x) \cdot \nabla W(x) > 0$ whenever $V(x) \neq 0$, then x^* becomes the only accumulation point of $W \circ x$ (since it is a monotone function). □

Proposition 3.5. Let $\hat{V}$ be an average truth-telling mean dynamic, then the mechanism with weak incentives defined by $p_s(x, \hat{x}) = \sum_j \frac{\partial F_j}{\partial x_s} \sum_\theta \hat{x}_{\theta j}\theta_{1j}$ converges to the efficient x^*distribution of actions.

Proof. The induced mean dynamic V satisfies all the assumptions of the previous theorem, hence x will converge to x^*. Thus actions will converge to the efficient outcome and strategies will converge to any $\hat{x}^*$such that $x_{\theta s}^* = \sum_{\hat{\theta}} \hat{x}_{\theta s \hat{\theta}}^*$. □

Bibliography

Atila Abdulkadiroğlu. College admissions with affirmative action. *International Journal of Game Theory*, 33(4):535–549, November 2005. doi: 10.1007/s00182-005-0215-7.

Atila Abdulkadiroğlu, Parag A. Pathak, Alvin E. Roth, and Tayfun Sönmez. The Boston Public School Match. *The American Economic Review*, 95(2):368–371, May 2005.

Atila Abdulkadiroglu, Parag Pathak, Alvin E. Roth, and Tayfun Sonmez. Changing the Boston school choice mechanism. Technical report, National Bureau of Economic Research, 2006.

Atila Abdulkadiroglu, Parag A. Pathak, and Alvin E. Roth. Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match. *American Economic Review*, 99(5):1954–1978, December 2009. doi: 10.1257/aer.99.5.1954.

Atila Abdulkadiroğlu, Yeon-Koo Che, and Yosuke Yasuda. Resolving Conflicting Preferences in School Choice: The "Boston Mechanism" Reconsidered. *The American Economic Review*, 101(1):399–410, February 2011.

Atila Abdulkadiroğlu, Yeon-Koo Che, and Yosuke Yasuda. Expanding "Choice" in School Choice. *American Economic Journal: Microeconomics*, 7(1):1–42, 2015. doi: 10.1257/mic.20120027.

Atila Abdulkadirouglu and Tayfun Sönmez. School choice: A mechanism design approach. *American economic review*, pages 729–747, 2003.

Atila Abdulkadirouglu, Parag A. Pathak, and Alvin E. Roth. The New York City High School Match. *The American Economic Review*, 95(2):364–367, May 2005. doi: 10.2307/4132848.

José Alcalde and Salvador Barberà. Top dominance and the possibility of strategy-proof stable solutions to matching problems. *Economic Theory*, 4(3):417–435, May 1994. doi: 10.1007/BF01215380.

Greg Attiyeh, Robert Franciosi, and R. Mark Isaac. Experiments With the Pivot Process for Providing Public Goods. *Public Choice*, 102(1-2):93–112, January 2000. doi: 10.1023/A:1005025416722.

Robert Aumann and Adam Brandenburger. Epistemic Conditions for Nash Equilibrium. *Econometrica*, 63(5):1161–1180, September 1995. doi: 10.2307/2171725.

Lawrence M. Ausubel. An Efficient Dynamic Auction for Heterogeneous Commodities. *The American Economic Review*, 96(3):602–629, June 2006.

Lawrence M. Ausubel and Paul R. Milgrom. Ascending auctions with package bidding. *Advances in Theoretical Economics*, 1(1), 2002.

Orhan Aygün and Tayfun Sönmez. Matching with Contracts: Comment. *American Economic Review,(forthcoming)*, 2013.

Michel Balinski and Tayfun Sönmez. A Tale of Two Mechanisms: Student Placement. *Journal of Economic Theory*, 84(1):73–94, January 1999. doi: 10.1006/jeth.1998.2469.

Felix Bierbrauer, Axel Ockenfels, Andreas Pollak, and Désirée Rueckert. Robust

Mechanism Design and Social Preferences. SSRN Scholarly Paper ID 2425537, Social Science Research Network, Rochester, NY, March 2014.

Sushil Bikhchandani and Joseph M. Ostroy. The Package Assignment Model. *Journal of Economic Theory*, 107(2):377–406, December 2002. doi: 10.1006/jeth.2001.2957.

Eric Budish. The Combinatorial Assignment Problem: Approximate Competitive Equilibrium from Equal Incomes. *Journal of Political Economy*, 119(6):1061–1103, December 2011. doi: 10.1086/664613.

Eric B. Budish and Judd B. Kessler. Changing the Course Allocation Mechanism at Wharton. SSRN Scholarly Paper ID 2579107, Social Science Research Network, Rochester, NY, October 2014.

Yeon-Koo Che and Olivier Tercieux. Efficiency and Stability in Large Matching Markets. SSRN Scholarly Paper ID 2638009, Social Science Research Network, Rochester, NY, July 2015.

Yan Chen and Tayfun Sönmez. School choice: an experimental study. *Journal of Economic Theory*, 127(1):202–231, March 2006. doi: 10.1016/j.jet.2004.10.006.

Edward H. Clarke. Multipart Pricing of Public Goods. *Public Choice*, 11:17–33, October 1971.

Peter Cramton and Jesse A. Schwartz. Collusive bidding: Lessons from the FCC spectrum auctions. *Journal of regulatory Economics*, 17(3):229–252, 2000.

Vincent P. Crawford and Elsie Marie Knoer. Job Matching with Heterogeneous Firms and Workers. *Econometrica*, 49(2):437–450, March 1981. doi: 10.2307/1913320.

Geoffroy de Clippel. Behavioral implementation. *The American Economic Review*, 104(10):2975–3002, 2014.

Sven de Vries, James Schummer, and Rakesh V. Vohra. On ascending Vickrey auctions for heterogeneous objects. *Journal of Economic Theory*, 132(1):95–118, January 2007. doi: 10.1016/j.jet.2005.07.010.

Gabrielle Demange and David Gale. The Strategy Structure of Two-Sided Matching Markets. *Econometrica*, 53(4):873–888, July 1985. doi: 10.2307/1912658.

Umut M. Dur, Scott Duke Kominers, Parag A. Pathak, and Tayfun Sönmez. The demise of walk zones in Boston: Priorities vs. precedence in school choice. Technical report, National Bureau of Economic Research, 2013.

Sanjiv Erat and Uri Gneezy. White Lies. *Management Science*, 58(4):723–733, April 2012. doi: 10.1287/mnsc.1110.1449.

Aytek Erdil and Haluk Ergin. What's the Matter with Tie-Breaking? Improving Efficiency in School Choice. *The American Economic Review*, 98(3):669–689, June 2008.

Haluk Ergin and Tayfun Sönmez. Games of school choice under the Boston mechanism. *Journal of Public Economics*, 90(1–2):215–237, January 2006. doi: 10.1016/j.jpubeco.2005.02.002.

Haluk I. Ergin. Efficient Resource Allocation on the Basis of Priorities. *Econometrica*, 70(6):2489–2497, November 2002. doi: 10.2307/3081994.

Emmanuel Farhi and Xavier Gabaix. Optimal Taxation with Behavioral Agents. Working Paper 21524, National Bureau of Economic Research, September 2015.

Clayton R. Featherstone. Rank efficiency: Investigating a widespread ordinal welfare criterion. *Job Market Paper*, 2011.

Clayton R Featherstone and Muriel Niederle. Improving on Strategy-proof School Choice Mechanisms: An Experimental Investigation. *Working Paper, Stanford University.*, 2015.

Urs Fischbacher. z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178, February 2007. doi: 10.1007/s10683-006-9159-4.

Daniel Fragiadakis and Peter Troyan. Designing Mechanisms to Make Welfare-Improving Strategies Focal. SSRN Scholarly Paper ID 2599701, Social Science Research Network, Rochester, NY, November 2015.

D. Gale and L. S. Shapley. College Admissions and the Stability of Marriage. *The American Mathematical Monthly*, 120(5):386–391, May 2013. doi: 10.4169/amer.math.monthly.120.05.386.

Uri Gneezy. Deception: The role of consequences. *American Economic Review*, pages 384–394, 2005.

Uri Gneezy, Bettina Rockenbach, and Marta Serra-Garcia. Measuring lying aversion. *Journal of Economic Behavior & Organization*, 93:293–300, September 2013. doi: 10.1016/j.jebo.2013.03.025.

Theodore Groves. Incentives in Teams. *Econometrica*, 41(4):617–631, July 1973. doi: 10.2307/1914085.

Faruk Gul and Ennio Stacchetti. Walrasian Equilibrium with Gross Substitutes. *Journal of Economic Theory*, 87(1):95–124, July 1999. doi: 10.1006/jeth.1999.2531.

Xiaolei Guo and Hai Yang. Pareto-improving congestion pricing and revenue refunding with multiple user classes. *Transportation Research Part B: Methodological*, 44(8-9):972–982, September 2010. doi: 10.1016/j.trb.2009.12.009.

John L. Hartman. Special Issue on Transport Infrastructure: A Route Choice Experiment with an Efficient Toll. *Networks and Spatial Economics*, 12(2):205–222, June 2012. doi: 10.1007/s11067-009-9111-1.

John William Hatfield and Fuhito Kojima. Group incentive compatibility for matching with contracts. *Games and Economic Behavior*, 67(2):745 – 749, 2009. doi: http://dx.doi.org/10.1016/j.geb.2009.01.007.

John William Hatfield and Fuhito Kojima. Substitutes and stability for matching with contracts. *Journal of Economic Theory*, 145(5):1704–1723, 2010.

John William Hatfield and Paul R. Milgrom. Matching with Contracts. *The American Economic Review*, 95(4):913–935, September 2005. doi: 10.2307/4132699.

Leonid Hurwicz. On Informationally Decentralized Systems. In *Decision and Organization, edited by C.B. McGuire and R. Radner*. North-Holland, 1972.

Nicole Immorlica and Mohammad Mahdian. Marriage, honesty, and stability. In *In Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA*, pages 53–62, 2005.

John H. Kagel and Dan Levin. Independent Private Value Auctions: Bidder Behaviour in First-, Second- and Third-Price Auctions with Varying Numbers of Bidders. *The Economic Journal*, 103(419):868–879, July 1993. doi: 10.2307/2234706.

John H. Kagel, Ronald M. Harstad, and Dan Levin. Information Impact and Allocation Rules in Auctions with Affiliated Private Values: A Laboratory Study. *Econometrica*, 55(6):1275–1304, November 1987. doi: 10.2307/1913557.

Yuichiro Kamada and Fuhito Kojima. Efficient Matching under Distributional Constraints: Theory and Applications. 2013.

Toshiji Kawagoe and Toru Mori. Can the Pivotal mechanism induce truth-telling? An experimental study. *Public Choice*, 108(3-4):331–354, August 2001. doi: 10.1023/A:1017542406848.

Alexander S. Kelso and Vincent P. Crawford. Job Matching, Coalition Formation, and Gross Substitutes. *Econometrica*, 50(6):1483–1504, November 1982. doi: 10.2307/1913392.

Onur Kesten. On two kinds of manipulation for school choice problems. *Economic Theory*, 51(3):677–693, November 2012. doi: 10.1007/s00199-011-0618-6.

Fuhito Kojima. When Can Manipulations be Avoided in Two-Sided Matching Markets?–Maximal Domain Results. *The BE Journal of Theoretical Economics*, 7(1), 2007.

Fuhito Kojima and Parag A. Pathak. Incentives and Stability in Large Two-Sided Matching Markets. *The American Economic Review*, 99(3):608–627, June 2009.

Scott Duke Kominers and Tayfun Sönmez. Designing for diversity in matching. In *ACM Conference on Electronic Commerce*, pages 603–604, 2013.

Herman B. Leonard. Elicitation of Honest Preferences for the Assignment of Individuals to Positions. *Journal of Political Economy*, 91(3):461–479, June 1983.

Michael Z. F. Li. The role of speed-flow relationship in congestion pricing implementation with an application to Singapore. *Transportation Research Part B: Methodological*, 36(8):731–754, September 2002. doi: 10.1016/S0191-2615(01)00026-1.

Shengwu Li. Obviously Strategy-Proof Mechanisms. SSRN Scholarly Paper ID 2560028, Social Science Research Network, Rochester, NY, November 2015.

Todd Litman. *Smart Congestion Relief.* Victoria Transport Policy Institute, September 2014.

Transport for London. *Central London congestion charging: impacts monitoring : first annual report, conditions before charging.* Transport for London, Street Management, [London], 2003. ISBN 1-871829-18-6 978-1-871829-18-1.

Eric Maskin. *Mechanism Design: How to Implement Social Goals.* Nobel Foundation, 2008.

Paul Milgrom and Ilya Segal. Envelope Theorems for Arbitrary Choice Sets. *Econometrica*, 70(2):583–601, March 2002.

Paul R. Milgrom, Jeremy Bulow, and Jonathan Levin. Winning Play in Spectrum Auctions? 2009.

Debasis Mishra and David C. Parkes. Ascending price Vickrey auctions for general valuations. *Journal of Economic Theory*, 132(1):335–366, January 2007. doi: 10.1016/j.jet.2005.09.004.

Yu (Marco) Nie. Transaction costs and tradable mobility credits. *Transportation Research Part B: Methodological*, 46(1):189–203, January 2012. doi: 10.1016/j.trb.2011.10.002.

Parag A. Pathak and Tayfun Sönmez. Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism. *The American Economic Review*, 98(4):1636–1652, September 2008. doi: 10.2307/29730139.

Parag A Pathak and Tayfun Sönmez. School Admissions Reform in Chicago and

England: Comparing Mechanisms by their Vulnerability to Manipulation. *American Economic Review*, 103(1):80–106, February 2013. doi: 10.1257/aer.103.1.80.

Arthur Pigou. *The Economics of Welfare*. London: Macmillan, 1920.

Amnon Rapoport, William E. Stein, James E. Parco, and Darryl A. Seale. Equilibrium play in single-server queues with endogenously determined arrival times. *Journal of Economic Behavior & Organization*, 55(1):67–91, September 2004. doi: 10.1016/j.jebo.2003.07.003.

Amnon Rapoport, Eyran J. Gisches, Terry Daniel, and Robin Lindsey. Pre-trip information and route-choice decisions with stochastic travel conditions: Experiment. *Transportation Research Part B: Methodological*, 68:154–172, October 2014. doi: 10.1016/j.trb.2014.06.007.

Alvin E. Roth. The Economics of Matching: Stability and Incentives. *Mathematics of Operations Research*, 7(4):617–628, November 1982. doi: 10.2307/3689483.

Alvin E. Roth. The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory. *Journal of Political Economy*, 92(6):991–1016, December 1984a. doi: 10.2307/1831989.

Alvin E Roth. Misrepresentation and stability in the marriage problem. *Journal of Economic Theory*, 34(2):383–387, December 1984b. doi: 10.1016/0022-0531(84)90152-2.

Alvin E Roth. The college admissions problem is not equivalent to the marriage problem. *Journal of Economic Theory*, 36(2):277–288, August 1985. doi: 10.1016/0022-0531(85)90106-1.

Alvin E. Roth and Elliott Peranson. The Redesign of the Matching Market for

American Physicians: Some Engineering Aspects of Economic Design. *The American Economic Review*, 89(4):748–780, September 1999. doi: 10.2307/117158.

Alvin E. Roth and Marilda A. Oliveira Sotomayor. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Cambridge University Press, June 1992. ISBN 978-0-521-43788-2.

Alvin E. Roth and Xiaolin Xing. Jumping the Gun: Imperfections and Institutions Related to the Timing of Market Transactions. *The American Economic Review*, 84(4):992–1044, September 1994. doi: 10.2307/2118042.

Alvin E. Roth, Tayfun Sönmez, and M. Utku Ünver. Kidney Exchange. *The Quarterly Journal of Economics*, 119(2):457–488, May 2004.

Tatsuyoshi Saijo, Tomas Sjostrom, and Takehiko Yamato. Secure implementation. *Theoretical Economics*, 2(3):203–229, 2007.

William H. Sandholm. Evolutionary Implementation and Congestion Pricing. *The Review of Economic Studies*, 69(3):667–689, July 2002.

William H. Sandholm. Negative Externalities and Evolutionary Implementation. *The Review of Economic Studies*, 72(3):885–915, July 2005.

William H. Sandholm. *Population Games and Evolutionary Dynamics*. The MIT Press, Cambridge, Mass, December 2010. ISBN 978-0-262-19587-4.

Kerstin Schneider and Joachim Weimann. Against all Odds: Nash Equilibria in a Road Pricing Experiment. In Michael Schreckenberg and Reinhard Selten, editors, *Human Behaviour and Traffic Networks*, pages 133–153. Springer Berlin Heidelberg, January 2004. ISBN 978-3-642-05950-6 978-3-662-07809-9.

David Schrank, Bill Eisele, and Tim Lomax. TTI 2012 urban mobility report. *Proceedings of the 2012 annual urban mobility report. Texas A&M Transportation Institute, Texas, USA*, 2012.

R. Selten, T. Chmura, T. Pitz, S. Kube, and M. Schreckenberg. Commuters route choice behaviour. *Games and Economic Behavior*, 58(2):394–406, February 2007. doi: 10.1016/j.geb.2006.03.012.

Lloyd S. Shapley and Martin Shubik. The assignment game I: The core. *International Journal of Game Theory*, 1(1):111–130, 1971.

Vernon L. Smith. An Experimental Study of Competitive Market Behavior. *Journal of Political Economy*, 70(2):111–137, April 1962.

Vernon L. Smith. Experiments with a Decentralized Mechanism for Public Good Decisions. *The American Economic Review*, 70(4):584–599, September 1980.

Tayfun Sönmez. Manipulation via Capacities in Two-Sided Matching Markets. *Journal of Economic Theory*, 77(1):197–204, November 1997. doi: 10.1006/jeth.1997.2316.

Tayfun Sönmez. Can Pre-arranged Matches Be Avoided in Two-Sided Matching Markets? *Journal of Economic Theory*, 86(1):148–156, May 1999. doi: 10.1006/jeth.1999.2507.

Tayfun Sönmez. Bidding for Army Career Specialties: Improving the ROTC Branching Mechanism. *Journal of Political Economy*, 121(1):186–219, February 2013. doi: 10.1086/669915.

Tayfun Sönmez and Tobias B. Switzer. MATCHING WITH (BRANCH-OF-CHOICE) CONTRACTS AT THE UNITED STATES MILITARY ACADEMY. *Econometrica*, 81(2):451–488, March 2013.

William Vickrey. Counterspeculation, Auctions, and Competitive Sealed Tenders. *The Journal of Finance*, 16(1):8–37, March 1961. doi: 10.2307/2977633.

Xiaolei Wang, Hai Yang, Deren Han, and Wei Liu. Trial and error method for optimal tradable credit schemes: The network case. *Journal of Advanced Transportation*, 48(6):685–700, 2014. doi: 10.1002/atr.1245.

Hai Yang and Xiaolei Wang. Managing network mobility with tradable credits. *Transportation Research Part B: Methodological*, 45(3):580–594, March 2011. doi: 10.1016/j.trb.2010.10.002.

Hai Yang, Qiang Meng, and Der-Horng Lee. Trial-and-error implementation of marginal-cost pricing on networks in the absence of demand functions. *Transportation Research Part B: Methodological*, 38(6):477–493, July 2004. doi: 10.1016/S0191-2615(03)00077-8.

Michael Z.F. Li. Estimating congestion toll by using traffic count data - Singapore area licensing scheme. *Transportation Research Part E: Logistics and Transportation Review*, 35(1):1–10, March 1999. doi: 10.1016/S1366-5545(98)00019-2.