The Power of Weak Incentives

Hector Lopez*

Job Market Paper

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.

JEL classification: C92; D47; D62; H21; R48.

Keywords: Weak Incentives; Behavioral Implementation; Congestion Pricing.

*University of Maryland, Department of Economics, Tydings Hall, College Park, MD 20742 (email: hlopezca@umd.edu). I am most grateful to my advisors, Peter Cramton, Erkut Ozbay, and Daniel Vincent, for their guidance and encouragement. This paper benefited from conversations with Kristian Lopez-Vargas, Naomi Utgoff, Ozlem Tonguc, Ian Chadd, Yong Paek, workshop participants at University of Maryland, and conference participants at The University of Chicago and Bilgi University. I am particularly grateful to the Department of Economics at UMD and Peter Cramton for funding the experiments presented in this paper.

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, Sönmez and Ünver (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

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

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 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

[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, Eisele and Lomax (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%.

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 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

[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%).

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 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, Harstad and Levin (1987) report an experiment in which bidders do not report their true value.[7] Attiyeh, Franciosi and Isaac (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, Sjostrom and Yamato (2007) propose double implementation, both in Nash and weakly dominant strategies. Li (2015) proposes implementation in obviously-strategy-proof strategies.[9] These notions exacerbate the

[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.

[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.

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, Che and Yasuda (2011) and Abdulkadiroğlu, Che and Yasuda (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, 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, Rockenbach and Serra-Garcia (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, Meng and Lee (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.

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 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 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 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 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. *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.

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 $\theta = (\theta_1, \theta_2)$, $\theta_j \in \mathbb{R}^S$.[12] There is a positive

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

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

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]

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

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 $x_{\theta s} = 0$.

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 $k = \lambda_\theta$,

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

$x_{\theta s} > 0$ 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 $F_s(x(\hat{x}))\theta_{1s} + \theta_{2s} + p_s(x(\hat{x}), \hat{x}) = k_\theta \text{ whenever } x(\hat{x})_{\theta s} > 0$ and $F_s(x(\hat{x}))\theta_{1s} + \theta_{2s} + p_s(x(\hat{x}), \hat{x}) \leq k_\theta \text{ whenever } x(\hat{x})_{\theta s} = 0$.

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 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

[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.

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

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:

$$\begin{array}{rl} V & \text{is Lipschitz continuous} \\ V_{\theta s}(x) \geq 0 & \text{whenever } x_{\theta s} = 0 \\ \sum\limits_{S} V_{\theta s}(x) = 0 & \text{for all } \theta \in \Theta \\ V(x) = 0 & \text{implies } x \text{ is a Nash equilibrium} \end{array}$$

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

Proposition 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

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

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 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.*

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 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 $u_\theta(x) = \theta_{2d} - \theta_{1d} t(\sum\limits_\theta x_{\theta d})$. 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.

3.1 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.

3.1.1 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 $P^* = t'(x^*)\sum_\theta \theta_{1d} x^*_{\theta d}$.

3.1.2 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: $P^D(x) = t'(x)\sum_\theta \theta_{1d} x_{\theta d}$.

3.1.3 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:$P^M(x, \hat{x}) = t'(x)\sum_\theta \theta_{1d} \hat{x}_{\theta d}$.

3.1.4 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.

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

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

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 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 1 and 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.

In figure 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

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

Figure 1: Type allocations

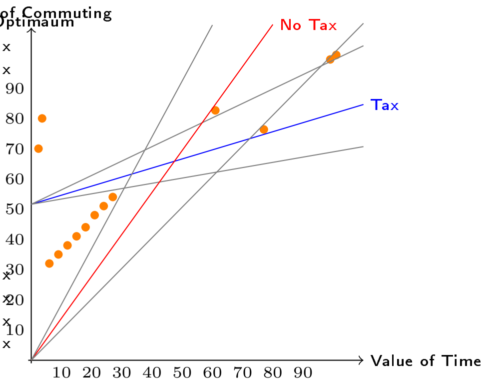

Figure 2: Experimental types

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.

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.

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.

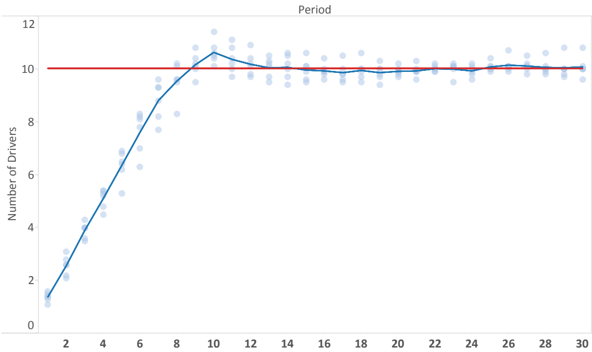

Figure 3: Number of drivers with No price

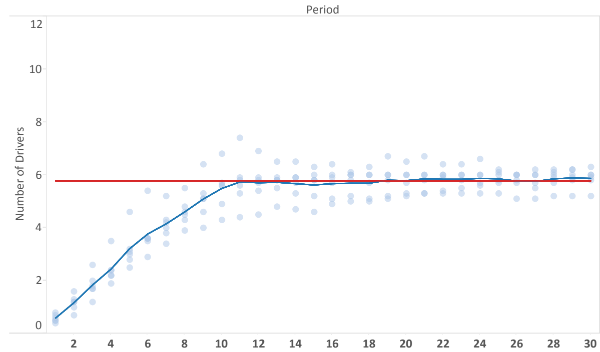

Figure 6: Number of drivers with Fixed price

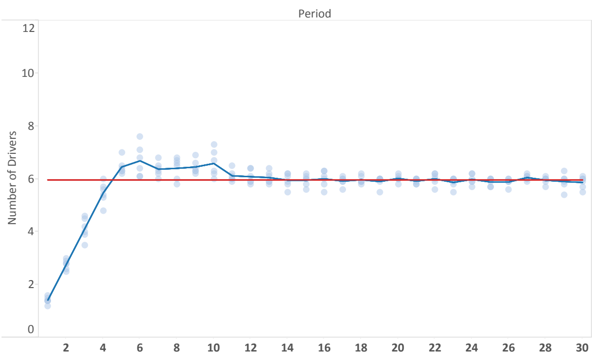

Figure 4: Number of drivers with Dynamic price

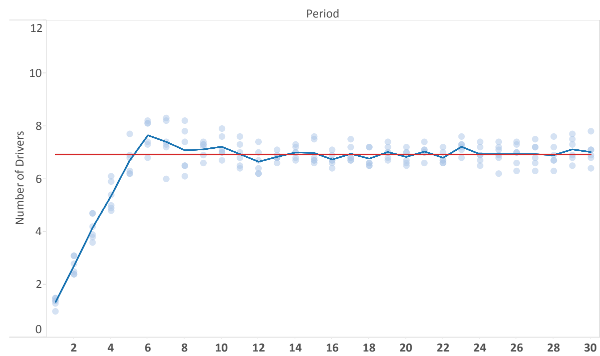

Figure 7: Number of drivers with Message price

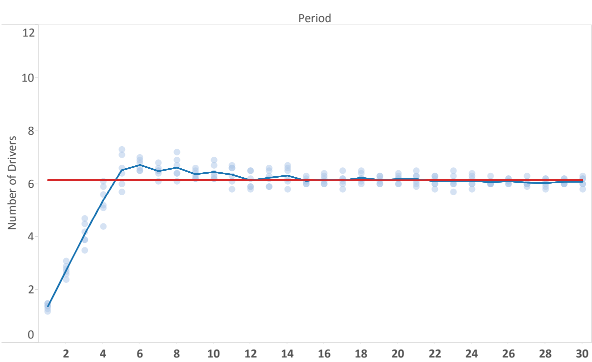

Figure 5: Number of drivers with Balanced Dynamic price

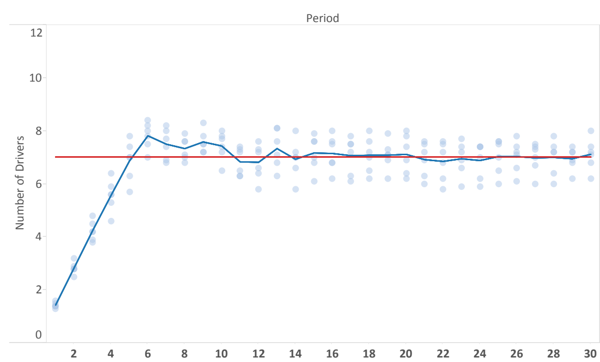

Figure 8: Number of drivers with Balanced Message price

The number of drivers of every treatment is shown in figures 3 through 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 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 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 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 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 5 and 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.

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 9 through 14 are analogous to figure 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.

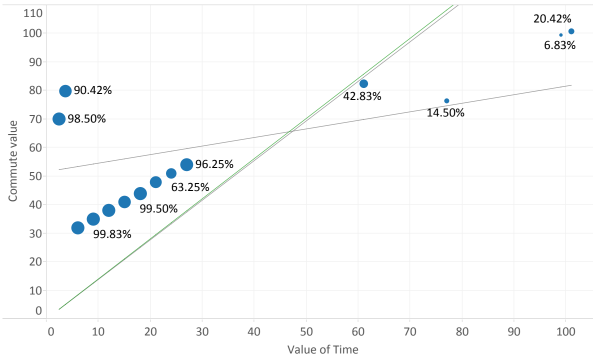

Figure 9: Identities with No price

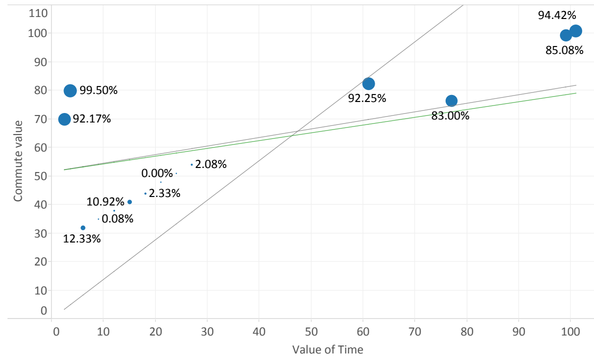

Figure 12: Identities with Fixed price

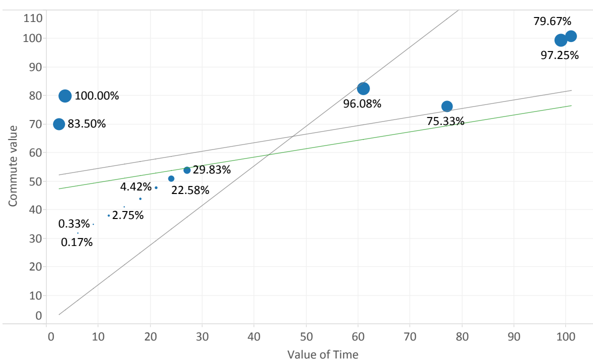

Figure 10: Identities with Dynamic price

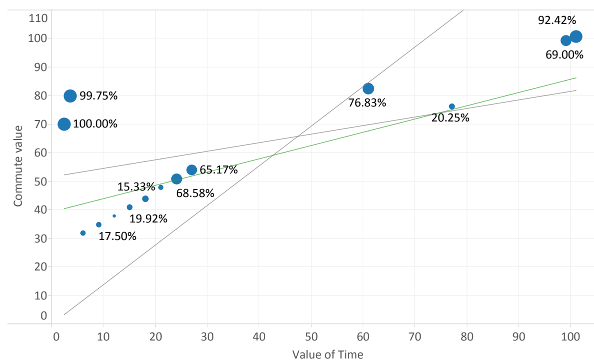

Figure 13: Identities with Message price

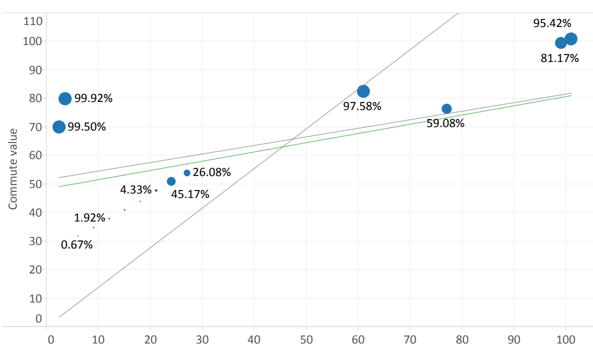

Figure 11: Identities with Balanced Dynamic price

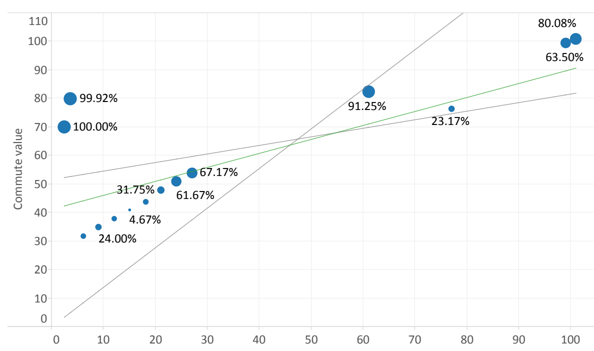

Figure 14: Identities with Balanced Message price

In figure 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 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 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 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 11 and 14.

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.

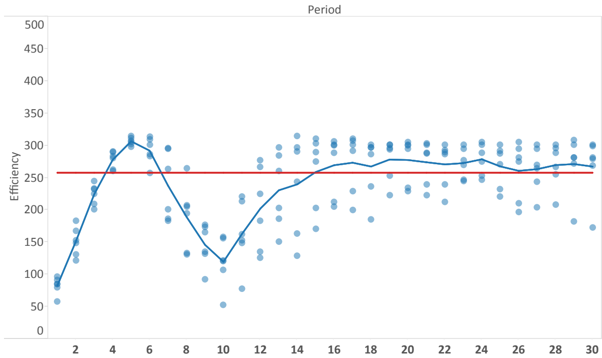

Figure 15: Efficiency with No price

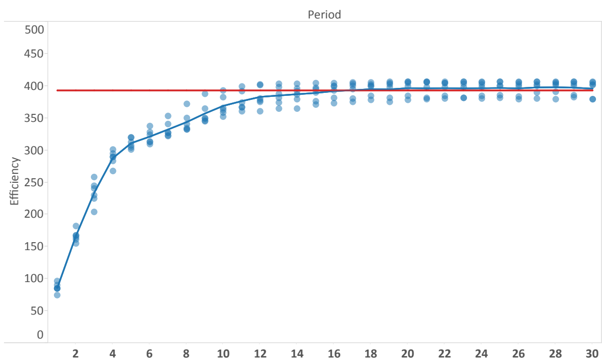

Figure 16: Efficiency with Dynamic price

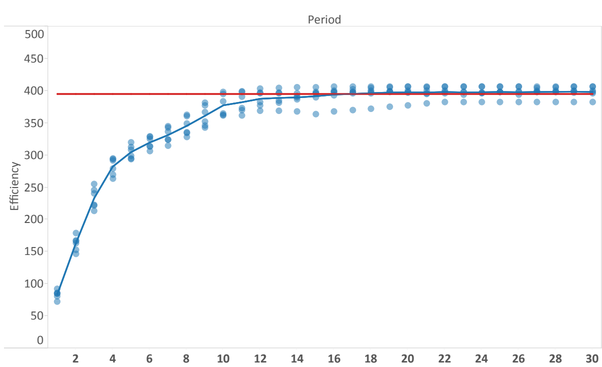

Figure 17: Efficiency with Balanced Dynamic price

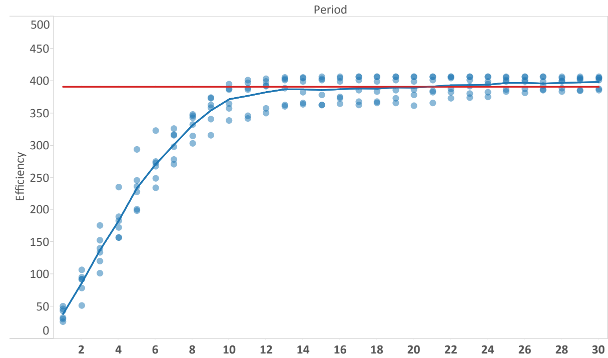

Figure 18: Efficiency with Fixed price

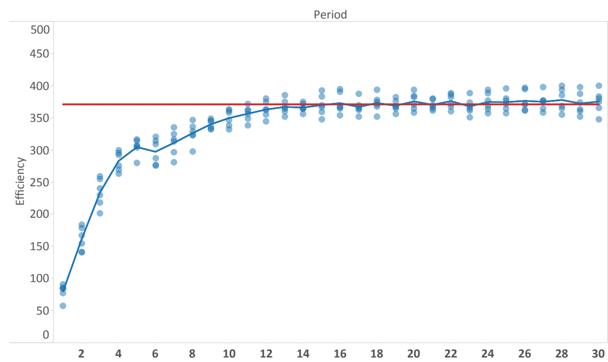

Figure 19: Efficiency with Message price

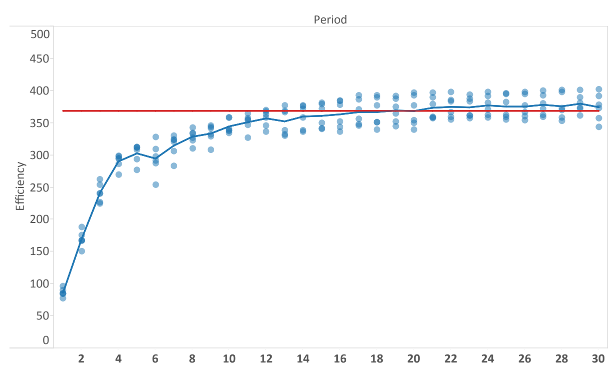

Figure 20: Efficiency with Balanced Message price

The efficiency of every treatment is shown in figures 15 through 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 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 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 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 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 17 and 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.

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 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 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 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 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 17 and 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 1 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.

					P values						
Period	Treatment	Measure	Mean	SD	No Price	Fixed	Dynamic	Message	Bdynamic	Bmessage	Epsilon
11	No Price	efficiency	65.91%	13.01%							4.88%
11	Fixed	efficiency	100.00%	4.12%	<1%			<1%		<1%	1.19%
11	Dynamic	efficiency	100.53%	3.26%	<1%			<1%		<1%	1.14%
11	Message	efficiency	95.00%	3.44%	<1%						1.14%
11	Bdynamic	efficiency	101.05%	2.85%	<1%	<1%		<1%		<1%	0.91%
11	Bmessage	efficiency	94.35%	4.86%	<1%						1.94%
11	No Price	N. of Drivers	10.030	0.323							15.44%
11	Fixed	N. of Drivers	5.770	0.493	<1%		<1%	<1%	<1%	<1%	22.74%
11	Dynamic	N. of Drivers	5.963	0.202	<1%			<1%	<1%	<1%	22.31%
11	Message	N. of Drivers	6.926	0.356	<1%						22.36%
11	Bdynamic	N. of Drivers	6.147	0.214	<1%			<1%		<1%	25.84%
11	Bmessage	N. of Drivers	7.015	0.606	<1%						17.06%

Table 1: Estimates for Period ≥ 11

Figures 21 and 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% confidence interval are shown in Figures 23 and 24.

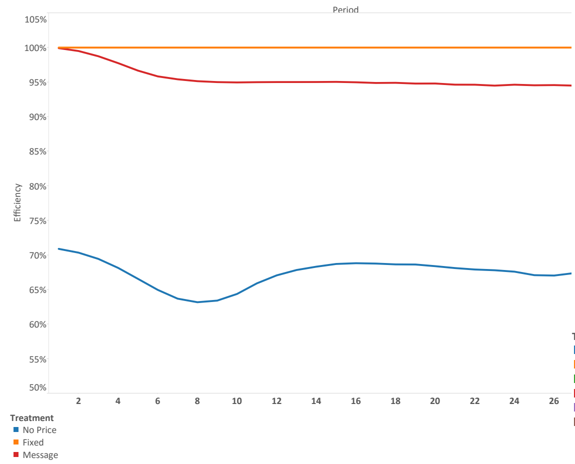

Figure 21: Efficiency estimates by period

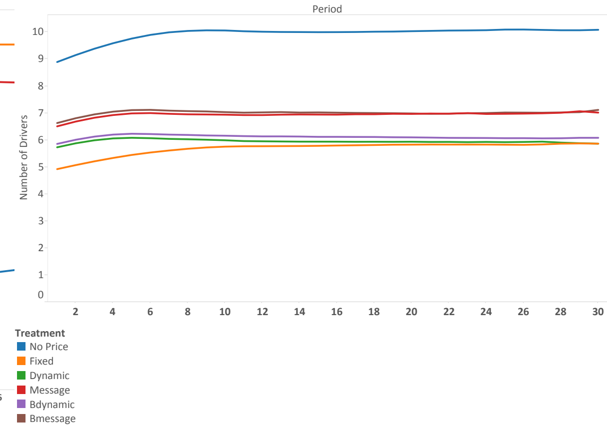

Figure 22: No. of Drivers estimates by period

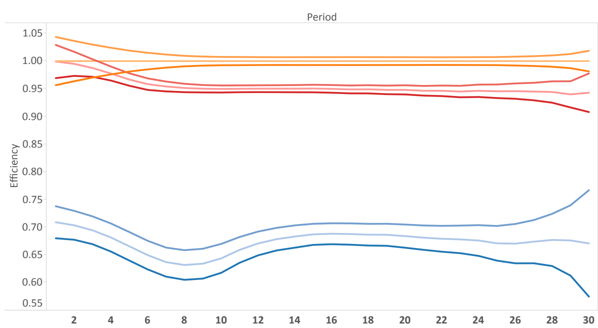

Figure 23: 95% Confidence Intervals for efficiency estimates. Blue: No price; Red: Message; Orange: Fixed

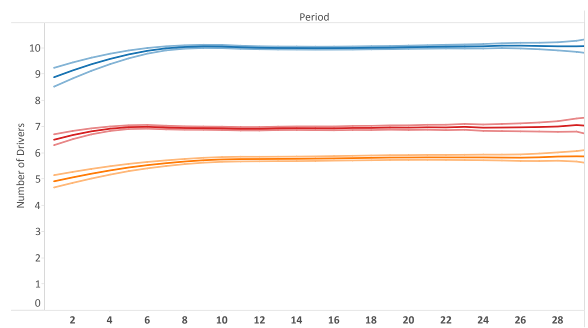

Figure 24: 95% Confidence Intervals for Number of Drivers estimates. Blue: No price; Red: Message; Orange: Fixed

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 25 shows the efficiency levels of all Nash equilibria in the game induced by the message congestion price by average message: $z(\hat{x}) = \frac{\sum_{\theta} \theta_{1i} \hat{x}_{\theta}}{x}$.

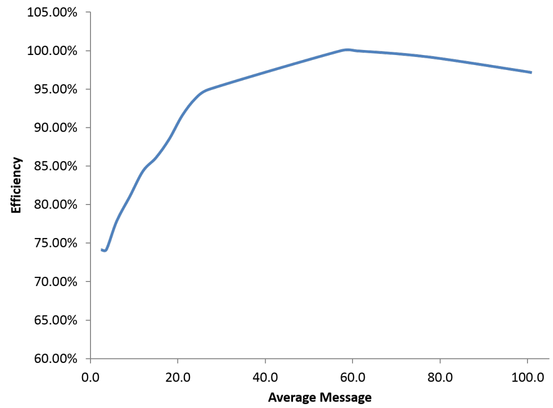

Figure 25: Nash Equilibria Efficiency by average message

In figure 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 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 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.

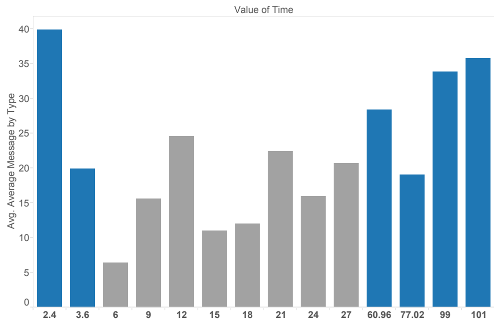

Figure 26: Average Message Sent by Type

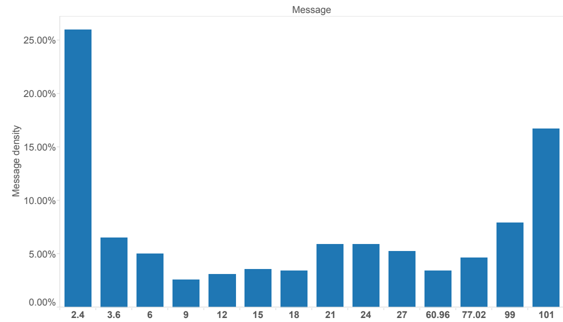

Figure 27: Messages Received

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 28 and 29 show the relationship between the average message and the real average message, as if all subjects reported their true value of time. Figure 28 shows their evolution over time (all sessions aggregated) and figure 29 shows all data points.

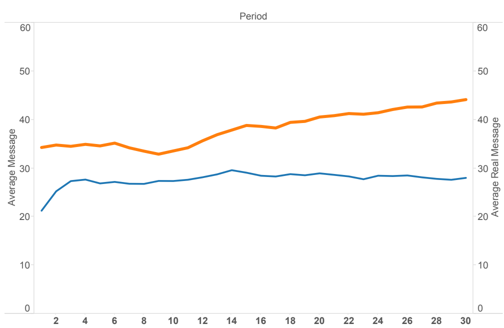

Figure 28: Average Message over time

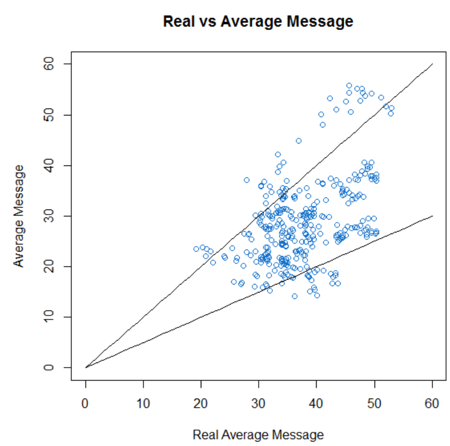

Figure 29: Real vs Sent Average Message

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 30 and 31 shows the empirical unconditional density and distribution.

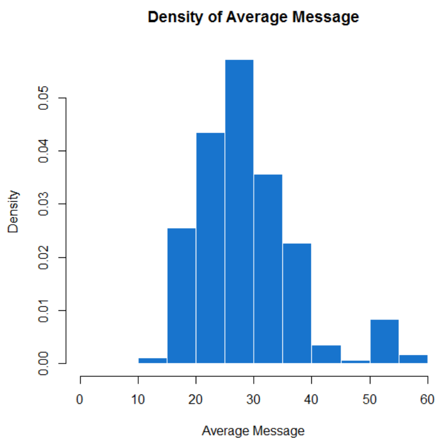

Figure 30: Density of Average Message

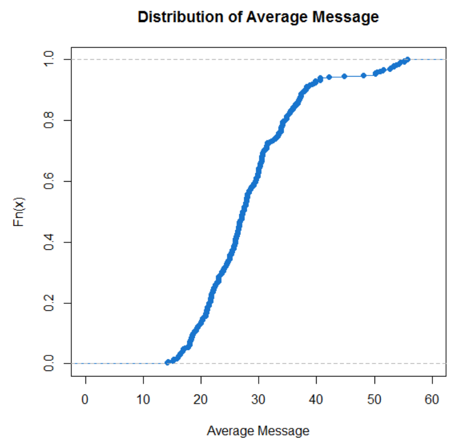

Figure 31: Distribution of Average Message

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 $\hat{V} : \hat{X} \to \mathbb{R}^{|\Theta| \times |S \times \Theta|}$ 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.

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 32 and 33 show observed covariance in the message price treatment. Every

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

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 34 and 35 show the covariance for the no price treatment. The null was not rejected (p-value >10%).

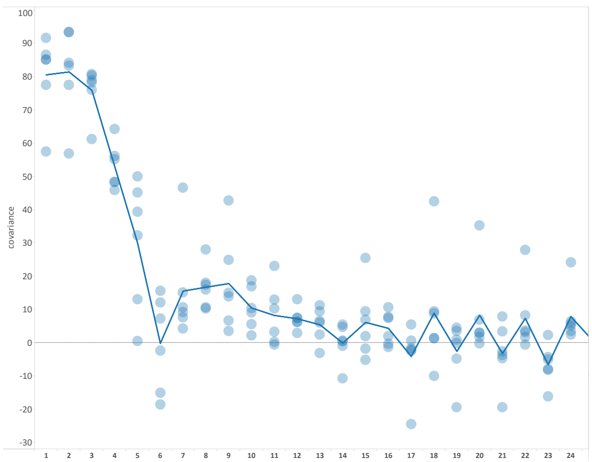

Figure 32: Average-truth-telling - message treatment

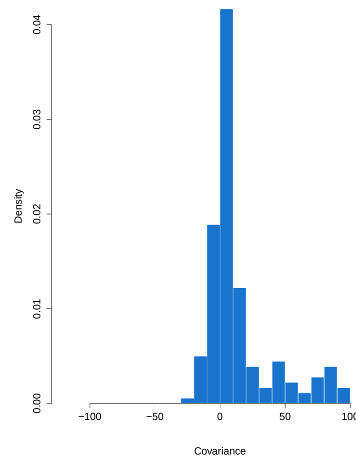

Figure 33: Average-truth-telling histogram - message treatment

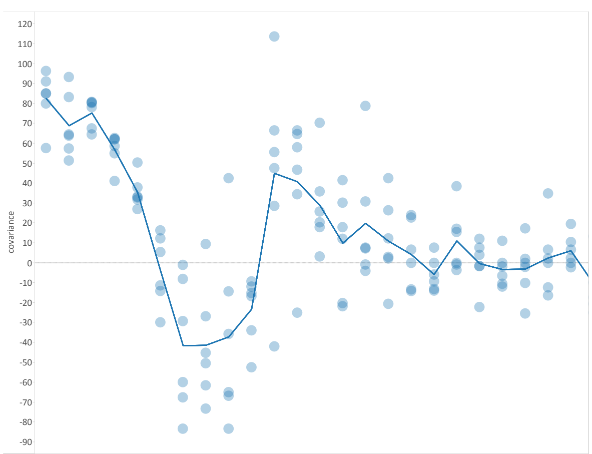

Figure 34: Average-truth-telling - no price treatment

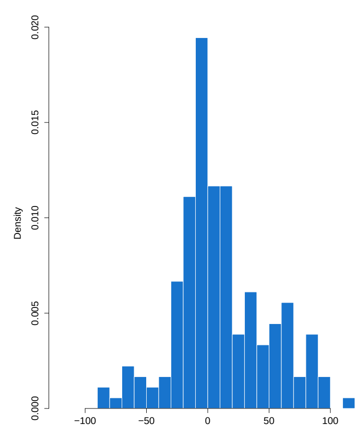

Figure 35: Average-truth-telling histogram - no price treatment

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 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.

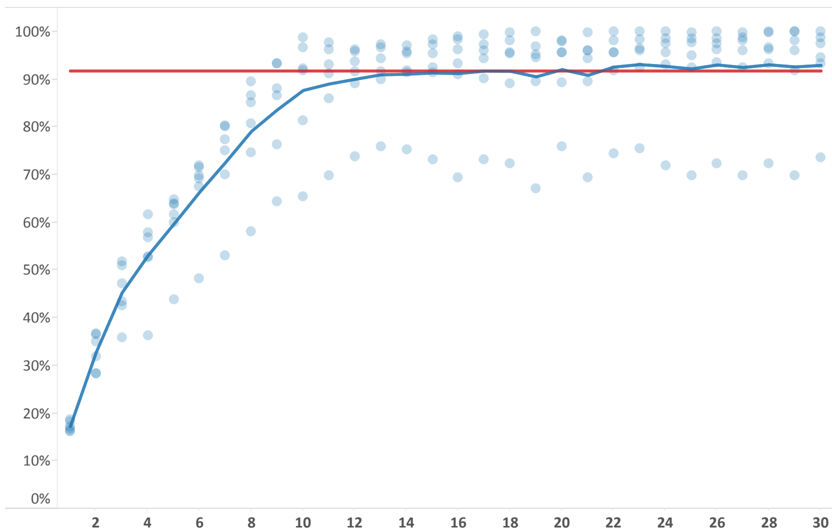

Figure 36: Efficiency of Random Treatments

Figures 37 to 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	**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 2: Random Types Efficiency for Periods 11-30

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 equilibrium 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.

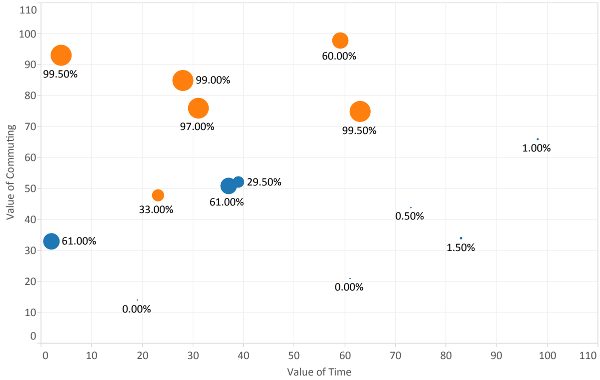

Figure 37: Types Random Treatment 1

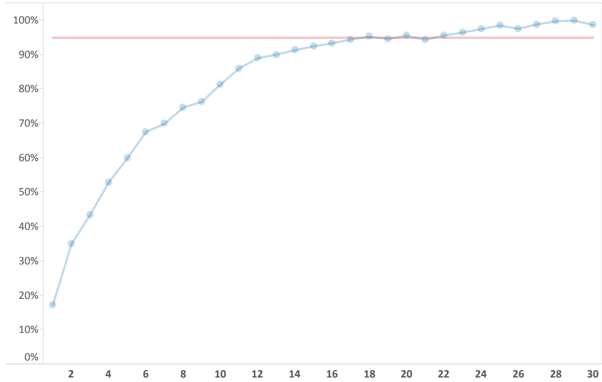

Figure 40: Efficiency Random Treatment 1

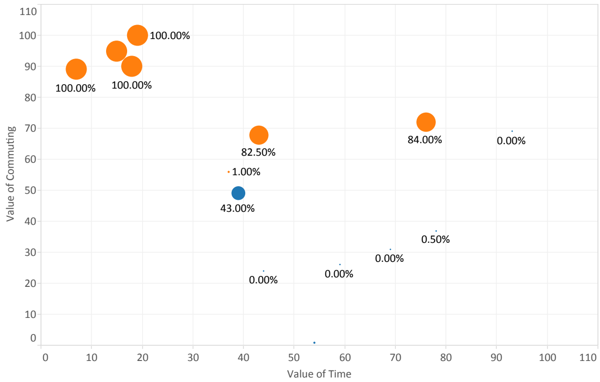

Figure 38: Types Random Treatment 2

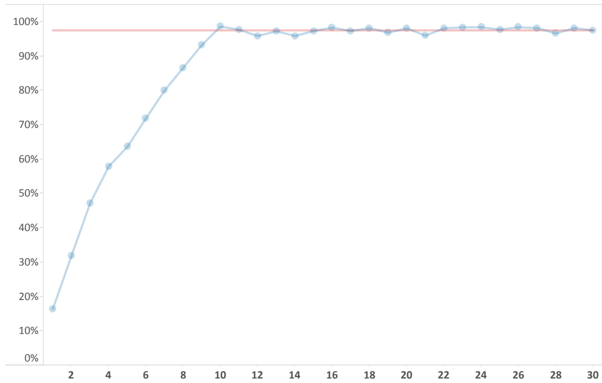

Figure 41: Efficiency Random Treatment 2

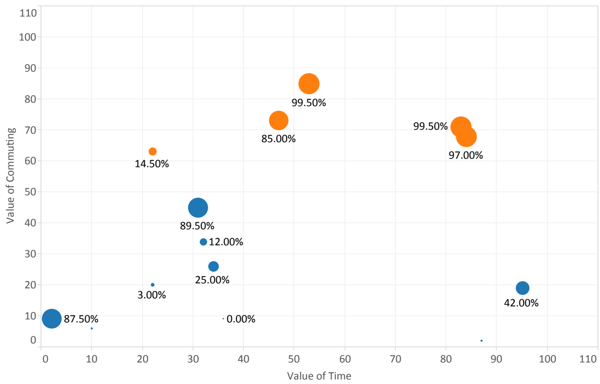

Figure 39: Types Random Treatment 3

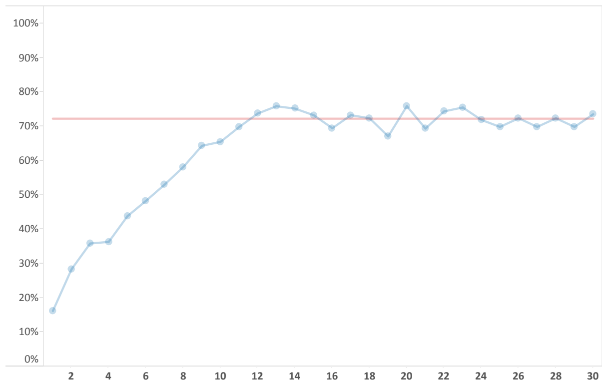

Figure 42: Efficiency Random Treatment 3

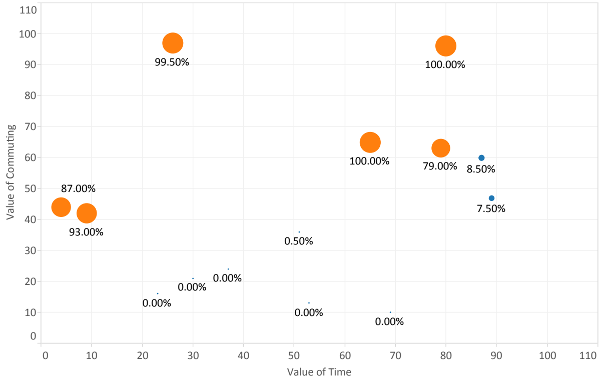

Figure 43: Types Random Treatment 4

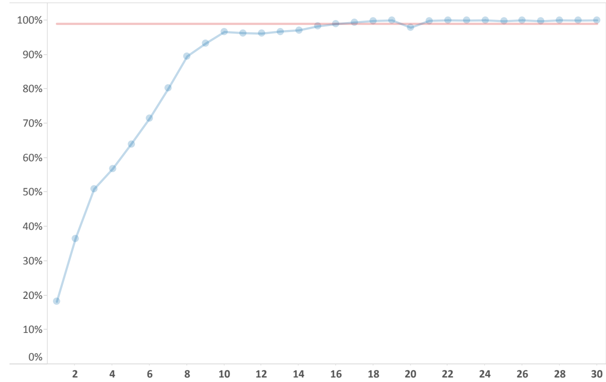

Figure 46: Efficiency Random Treatment 4

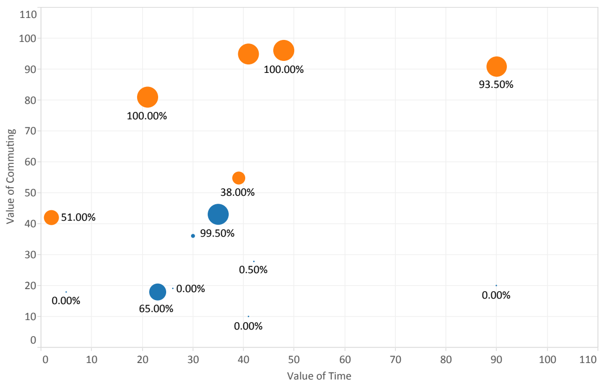

Figure 44: Types Random Treatment 5

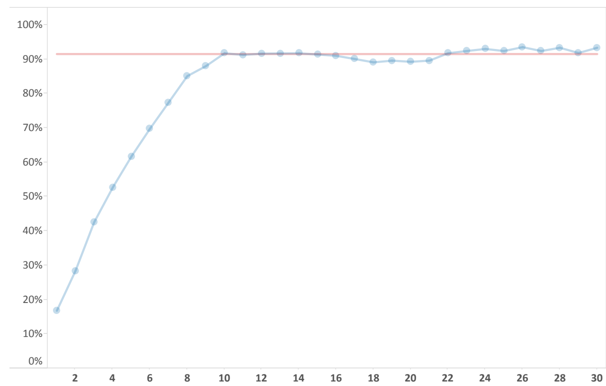

Figure 47: Efficiency Random Treatment 5

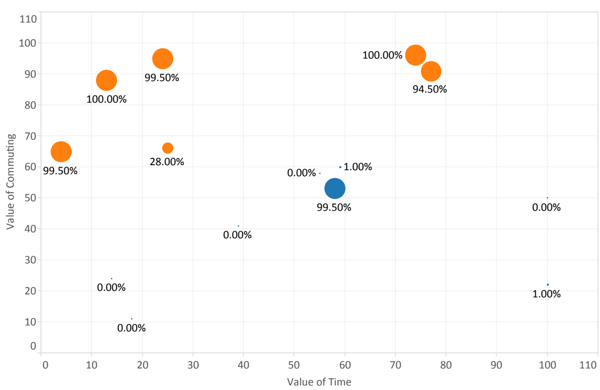

Figure 45: Types Random Treatment 6

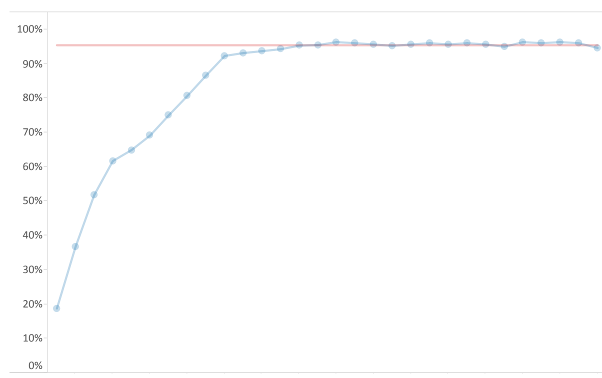

Figure 48: Efficiency Random Treatment 6

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 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

[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.

[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.

incentives and behavioral traits, opening the door to new methods for solving problems in practice.

6 Appendix

Proposition 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 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. □

Proposition 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 $u(x_i, x^*_{-i}, \theta_i) + \sum_{N\backslash i} u_j(x_i, x^*_{-i}, \theta_j) - h^w_i(x_{-i}, \theta_{-i})$, hence i maximizes his payoff by selecting x^*_ias an action. Thus (x^*, θ) is a subgame perfect Nash equilibrium. $\square$

Proposition 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^*.

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). $\square$

Proposition 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.

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 $x^*_{\theta s} = \sum_{\hat{\theta}} \hat{x}^*_{\theta s \hat{\theta}}$. $\square$

References

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

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

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

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

Aumann, Robert, and Adam Brandenburger. 1995. "Epistemic Conditions for Nash Equilibrium." *Econometrica*, *Econometrica*, 63(5): 1161–1180.

Bierbrauer, Felix, Axel Ockenfels, Andreas Pollak, and Désirée Rueckert. 2014. "Robust Mechanism Design and Social Preferences." Social Science Research Network SSRN Scholarly Paper ID 2425537, Social Science Research Network SSRN Scholarly Paper ID 2425537, Rochester, NY.

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

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

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

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

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

Erat, Sanjiv, and Uri Gneezy. 2012. "White Lies." *Management Science*, *Management Science*, 58(4): 723–733.

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

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

Fischbacher, Urs. 2007. "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics*, *Experimental Economics*, 10(2): 171–178.

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

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

Gneezy, Uri, Bettina Rockenbach, and Marta Serra-Garcia. 2013. "Measuring lying aversion." *Journal of Economic Behavior & Organization*, *Journal of Economic Behavior & Organization*, 93: 293–300.

Groves, Theodore. 1973. "Incentives in Teams." *Econometrica*, *Econometrica*, 41(4): 617–631.

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

Hartman, John L. 2012. "Special Issue on Transport Infrastructure: A Route Choice Experiment with an Efficient Toll." *Networks and Spatial Economics*, *Networks and Spatial Economics*, 12(2): 205–222.

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

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

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

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

Li, Michael Z. F. 2002. "The role of speed-flow relationship in congestion pricing implementation with an application to Singapore." *Transportation Research Part B: Methodological*, *Transportation Research Part B: Methodological*, 36(8): 731–754.

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

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

London, Transport for. 2003. *Central London congestion charging: impacts monitoring : first annual report, conditions before charging.* [London]:Transport for London, Street Management.

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

Nie, Yu (Marco). 2012. "Transaction costs and tradable mobility credits." *Transportation Research Part B: Methodological*, *Transportation Research Part B: Methodological*, 46(1): 189–203.

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

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

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

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

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

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

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

Sandholm, William H. 2010. *Population Games and Evolutionary Dynamics.* Cambridge, Mass:The MIT Press.

Schneider, Kerstin, and Joachim Weimann. 2004. "Against all Odds: Nash Equilibria in a Road Pricing Experiment." In *Human Behaviour and Traffic Networks.* , ed. Michael Schreckenberg and Reinhard Selten, 133–153. , ed. Michael Schreckenberg and Reinhard Selten, 133–153. Springer Berlin Heidelberg.

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

Selten, R., T. Chmura, T. Pitz, S. Kube, and M. Schreckenberg. 2007. "Commuters route choice behaviour." *Games and Economic Behavior*, *Games and Economic Behavior*, 58(2): 394–406.

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

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

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

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

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

Yang, Hai, and Xiaolei Wang. 2011. "Managing network mobility with tradable credits." *Transportation Research Part B: Methodological*, *Transportation Research Part B: Methodological*, 45(3): 580–594.

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

Z.F. Li, Michael. 1999. "Estimating congestion toll by using traffic count data - Singapore area licensing scheme." *Transportation Research Part E: Logistics and Transportation Review*, *Transportation Research Part E: Logistics and Transportation Review*, 35(1): 1–10.

INSTRUCTIONS

Thank you for participating in today's experiment. This is an experiment in the economics of decision-making. Various research foundations have provided funds for this research. The instructions are simple. If you follow them carefully and make good decisions, you will earn money.

The entire experiment should be complete within ninety minutes. You will be paid a $3 show-up for showing up on time. In addition, you will be paid $3 if you complete the session. You may collect the show-up fee at any moment if you decide to terminate your participation. If you complete the session, you will be privately paid both fees at the end. In addition, you will be paid a performance fee depending on your decisions. The mechanics for this payment are detailed below. All quantities in the experiment are measured in experimental units (EU). The connection between experimental units and real dollars is explained below.

This experiment is organized in 6 sections. Every section is divided into 30 rounds. In each round, you and other participants will decide whether to "drive" or "not drive" on a fixed road. The length of your commute will depend on the number of drivers on the road. At the beginning of each section, you will be assigned two numbers. These numbers are your private information.

1. Your value of commuting
2. The value of your time

In every section, you may or may not face a tax to use the road. In addition, a subsidy could be offered not to drive. You will be informed whether or not a tax will be charged or a subsidy offered before you make your decision every round. In some sections, you will be able to affect the value of the tax or subsidy by sending a message regarding the value of your time.

At the end of every round you will be informed of the following elements:

i. The time spent on the road
ii. Your payoff (in experimental units)

Your payoff will depend on your decisions, your values, the time spent on the road and the tax.

Not driving guarantees a payoff (in experimental units) of S, where S is the value (possibly 0) of the subsidy offered not to drive.

Driving has a payoff (in experimental units) equal to $U - \frac{V}{60}t - T$, where U is your value of commuting, V is your value of time, T is a tax (possibly 0) to use the road and t is the time spent on the road.

Performance fee

All six sections have the same performance fee structure. In every round you will be faced with two decisions: to drive or not to drive.

Depending on your values, the time, the tax and the subsidy, driving could be better, equal or worse than not driving. Your payment in each round will only depend on the optimality of your decision. If you select the option with the higher payoff in round r, you will earn $x_r = (0.9764)^{30-r}$ experimental units; otherwise you will earn \$0.

Your performance fee will be then calculated as the sum of your profit in every round multiplied by a factor of 0.107675 .

Example

Suppose that your value of commute U is equal to 20 and your value of time V is equal to 2. In round 1 the time spent on the road is 10, there is a tax of 3 and a subsidy not to drive of 1. Then:

Not driving payoff: 1

Driving payoff: $20 - \frac{2}{60} * 10 - 3 = 16.666$

In this case, the better option is to drive. If you decide not to drive you will earn 0 EU. If you decide to drive you will earn $x_r = (0.9764)^{30-1} = 0.5$.

Questions

If you have any questions right now, please share it with all.

If you have any question during the experiment, please quietly raise your hand and one of the experimenters will come to you to answer your question. It is important that you do not talk with any of the other participants.