Journal of Mechanism and Institution Design

CONTENTS

Volume 5, Issue 1, December 2020
ISSN: 2399-844X (Print), 2399-8458 (Online)

Editorial board

Published by
The Society for the Promotion of Mechanism and Institution Design
Editorial office, Centre for Mechanism and Institution Design
University of York, Heslington, York YO10 5DD
United Kingdom
http://www.mechanism-design.org

ISSN: 2399-844X (Print), 2399-8458 (Online), DOI: 10.22574/jmid

The founding institutional members are

University of York, UK
Alpen-Adria-Universität Klagenfurt, Austria
Southwestern University of Economics and Finance, China.

Cover & Logo Artwork @ Jasmine Yang
LaTeX Editor & Journal Design @ Paul Schweinzer (using ‘confproc’)
LaTeX Editorial Assistants @ Theresa Marchetti & Daniel Rehsmann

Printed in York and Klagenfurt by the Centre for Mechanism and Institution Design

A Letter from the Editor

The year of 2020 will be remembered as having been a weird, eerie, and difficult one. As of December 8th, the number of deaths caused by Covid-19 has reached 1,554,491 globally, according to Worldometer. Many people lost their loved ones and have been affected terribly and hurt. We hope and expect that the situation will get better, gradually returning to normality next year.

In briefly looking back at this year's more positive events, I am happy to report on the Conference on Mechanism and Institution Design, held June 11th–13th at the University of Klagenfurt, Austria, organized by Paul Schweinzer. He left a little mark on history in organizing—to the best of our knowledge—the first ever fully online economic theory conference. There were no registration, participation, or presentation fees of any kind but many people kindly made voluntary contributions towards the Journal (thanks to Bettina Klaus for suggesting this terrific use of unused travel budgets!). Three distinguished economists, Pierpaolo Battigalli of Bocconi University, Johannes Hörner of Yale University, and Benny Moldovanu of the University of Bonn, were the keynote speakers. In total, there were 122 talks scheduled around the clock, with participants from leading universities and research institutes worldwide. A repository of most talks is available and can be freely used as a source of inspiration, teaching, or as a perfectly geeky way of spending a weekend. We are looking forward to meeting again, hopefully in social proximity, at our next conference at the National University of Singapore in the beautiful city country of Singapore, 2022.

From January 2021 Xiaotie Deng will retire from our editorial board. We wish to express our heartfelt gratitude to him for his five years of excellent service, advice, and support. Meanwhile, we are very pleased to announce that an outstanding colleague, Marek Pycia, has kindly agreed to join our editorial board from January 2021. We thank him and look forward to working with him to further advance the Journal.

Our Journal will continue to publish high quality papers in the field of mechanism and institution design and provide the best possible services to our authors and readers for the public interest without any charge of any kind. As always, we appreciate and rely on your support.

Zaifu Yang, York, December 8th, 2020.

Journal of Mechanism and Institution Design
ISSN: 2399-844X (Print), 2399-8458 (Online)
DOI:10.22574/jmid.2020.12.001

IMPLEMENTATION WITH EX POST HIDDEN ACTIONS

Caleb M. Koch

ETH Zurich, Switzerland
caleb.m.koch@gmail.com

ABSTRACT

We study implementation in settings where agents take strategic actions that influence preferences over mechanism outcomes and yet are hidden from the mechanism designer. We show that such settings can arise in entry auctions for markets, and that the Vickery-Clarke-Groves mechanism is not necessarily truthful. In this paper we first formalize so-called *ex post* hidden actions, we then characterize social choice functions that can be implemented in a way that is robust with respect to *ex post* hidden actions, and finally we propose a mechanism to do so. The model allows agents to have multi-dimensional types and to have quasi-linear utilities in money. We showcase these results by identifying social choice functions that can and cannot be implemented in entry auctions for Cournot competition models.

Keywords: *Ex post* hidden actions, dominant-strategy incentive compatibility, auctions.

JEL Classification Numbers: C72, D70, D82.

I am indebted to Heinrich Nax for his advice and encouragement. This paper is based on my PhD thesis. I am grateful to Rabah Amir, Yen-Lin Chiu, Stefano Duca, Dirk Helbing, Osama Khan, Bary Pradelski, John Quah, Alain Rossier, Niccolò Urbinati, participants of the Conference on Economic Design '19 at the University of Corvinus, as well as an excellent editor and three anonymous referees for helpful comments. Financial support from the European Commission through the ERC Advanced Investigator Grant 'Momentum' (Grant 324247) is gratefully acknowledged. The paper was previously distributed under the title, "*Implementation with secrets*" and "*Implementation with strategic private information*". Any remaining errors are my own.

1. INTRODUCTION

MECHANISM design, sometimes called "reverse game theory", is the theory of designing rules and regulations so that self-interested agents can achieve social goals. A typical mechanism design problem is formulated as follows. First, each agent has private information. Second, each agent may submit a report to the mechanism designer about his/her private information (this report may or may not be truthful). Finally, the mechanism designer makes a social decision based on agents' reports. This typical setup, however, assumes that *the only strategic action that agents take is sending the mechanism designer a report.* As such, mechanism design assumes that agents cannot influence other agents' payoffs via strategic interactions outside the mechanism, and any strategic actions other than reporting are thus irrelevant for the design of a mechanism itself. In other words, mechanisms are typically assumed to be a closed system.

However, there are economic settings that might challenge the closed system assumption. For example, consider a spectrum auction where firms compete for (scarce) spectrum licenses to enter into the telecommunications market (McAfee & McMillan, 1996; Milgrom, 2000; Klemperer, 2002a,b). What is the timing of such an auction? First, firms make bids for spectrum licenses. Second, firms receive licenses based on all firms' bids (and pay the government according to the auction payment scheme). Third – and perhaps most importantly – *firms with spectrum licenses compete for market share.* This third point is relevant because it means that firms can influence each others' payoffs via strategic market actions that are beyond the spectrum auction. However, the telecommunications market has, to the best of the author's knowledge, never been modeled explicitly when designing spectrum auctions.

Do strategic market actions after auctions make a difference for mechanism design? We show that the answer is yes: in the main text (see, e.g. Example 4 and Remark 4), we give an example to illustrate where the Vickery-Clarke-Groves Mechanism is not strategy-proof or truthful when strategic interactions after a mechanism are modeled explicitly, even if there exists a unique Nash equilibrium. More generally, we attempt to clarify how and why *ex post* hidden actions after a mechanism can play an important role in influencing mechanisms that are designed to implement social goals.

In this paper, we address these questions by studying environments in which agents can take hidden actions from the mechanism designer after mechanism outcomes. Such hidden strategic actions can play a crucial role in determining personal and/or others' preferences over social decisions. To reiterate, strategic actions differ from exogenous private information as they can (i) be selected strategically and (ii) may be subject to change after agents give information to the mechanism designer. For example, hidden actions could represent firms' market decisions after spectrum licenses are distributed, where such market decisions are clearly unknown to the mechanism designer before licenses are distributed.

This paper makes two major contributions. The first is formalizing so-called '*ex post* hidden actions'. As noted above, an *ex post* hidden action is taken by an agent *after* a mechanism outcome, is unknown to a mechanism designer, and yet can be crucial for determining personal and/or others' preferences over mechanism outcomes. An *ex post* hidden action may include pricing decisions, merger and acquisition decisions, contract negotiations, etc.

The second contribution is characterizing social choice functions that can be implemented in a way that is *robust* to *ex post* hidden actions. Concretely, the class of mechanisms considered here proceeds as follows. First, agents report preferences over mechanism outcomes (this can be thought of reporting a utility function); to reiterate, agents do not report *ex post* hidden actions. Second, the mechanism designer makes a social decision based on agents' reports, and agents pay the mechanism designer based on a pre-specified payment scheme. Finally, after the social decision, agents take their (*ex post*) hidden actions and receive their payoffs.

Such mechanisms can be viewed as a strategic counterpart to *robust mechanism design* (Bergemann & Morris, 2005). A robust mechanism asks each agent to report private information but not his/her belief, which is also private information. Such mechanisms are thus robust to unreported exogenous private information. Here, mechanisms can be view as being robust with respect to unreported *strategic* private information. We comment on this further in the related literature subsection below.

The main theorems in this paper characterize social choice functions that can be implemented in dominant-strategies when agents take *ex post* hidden actions. The characterization boils down to a single 'monotonicity' condition that must be satisfied by the social choice function. This condition requires two properties: *(i)* the social choice function gives agents with 'higher types' priority to 'better outcomes', and *(ii)* matching higher types with better outcomes exhibits increasing returns. Implementation is generally impossible without both conditions.

We show that implementation with *ex post* hidden actions is indeed possible and practical by studying entry mechanisms for a Cournot market (which builds on an example in Dasgupta & Maskin, 2000). In particular, firms bid to enter into a Cournot market, and firms' valuations of mechanism outcomes depend on *ex post* hidden actions because of strategic interactions in the Cournot market. It turns out that our characterization can delineate between social choice functions that can and cannot be implemented in a way that is robust to the Cournot market. We show that: *(i)* there exists a truthful mechanism that can maximize *consumer* surplus and that, *(ii)* even if there is a unique Nash equilibrium in the Cournot market, there are no truthful mechanisms that can maximize *producer* surplus (at Nash equilibrium), including the Vickery-Clarke-Groves Mechanism.

This paper draws primarily from two research streams in mechanism design.

The first research stream from which this paper draws is *robust mechanism design* (Bergemann & Morris, 2005). The fundamental goal of robust mechanism design is to identify conditions under which a mechanism can elicit information truthfully without asking for beliefs. A belief is modeled as exogenous private information, which means that a robust mechanism can be viewed as being robust with to exogenous private information. This literature has identified various necessary and sufficient conditions under which agents report truthfully regardless of privately held beliefs (see, e.g., Bergemann & Morris, 2011, and Bergemann et al., 2011).

This paper tries to offer a *strategic* counterpart to the current robust mechanism design literature: rather than beliefs, the main results relate to when one can design mechanisms where agents report truthfully regardless of *ex post* hidden actions. This opens up several possibilities for robust mechanism design, such as allowing beliefs to be influenced by cheap talk.

The second research stream focuses on incentive compatibility in quasi-linear environments with exogenous private information. *Ex post* hidden actions in a mechanism design environment are closely related to studying environments with interdependent preferences. Jehiel & Moldovanu (2001) show that the Bayesian incentive compatibility of the efficient allocation rule is generally impossible. We use an example that is closely related to the *ex post* hidden action context (when the Nash equilibirum is unique, preferences with *ex post* hidden action reduce to interdependent preferences).[1] This paper is closer to the strand of literature that focuses on *dominant-strategy incentive compatibility* – which effectively requires Bayesian implementation within every information structure (see, e.g. Dasgupta et al., 1979) – as dominant-strategies are the focus here. The advantage of the dominant-strategy vs. Bayesian approach is its 'detail free' nature (Wilson, 1987), in the sense that a designer does not need information about agents' priors in order to formulate expected profits.[2]

The most closely related papers from this literature focus on *deterministic* mechanisms and explore necessary and sufficient conditions for implementation. Roberts (1979) was the first to do so in settings where agents' types come from an unrestricted type domain. The main finding of Roberts is a necessary and sufficient condition, called "Positive Association Differences" (PAD), that characterizes dominant-strategy incentive compatibility on such domains. The drawback of PAD is its inapplicability in many relevant economic settings, such as the allocation of private goods, since it precludes assumptions such as free disposal and externality in consumption.[3] Bikhchandani et al. (2006) resolved these issues and characterized dominant-strategy implementation on a *restricted* type domain that admits assumptions such as free disposal. The advantage of the approach taken in Bikhchandani et al. (2006) is its applicability to a large class of economically relevant settings, including auctions.

The rest of the paper proceeds as follows. In Section 2, we introduce the model. In Section 3, we present the 'implementation with *ex post* hidden actions' problem statement. In Section 4, we present our main results. We conclude in Section 5. All proofs, unless otherwise provided, are relegated to the appendices.

[1] For other references from the Bayesian incentive compatibility literature that are relevant to this paper, see Myerson (1981), Rochet (1987), McAfee & McMillan (1988), Jehiel et al. (1999), Williams (1999), Krishna & Maenner (2001), and Milgrom & Segal (2002).

[2] Relatedly, Chung & Ely (2007) provide a theoretical foundation for dominant-strategy implementation by proving that such mechanisms are maximin optimal.

[3] See Bikhchandani et al. (2006) for an extended discussion on the limitations of using Positive Association Differences in applied economic models.

2. THE MODEL

This section presents the model underlying implementation with *ex post* hidden actions. Definitions are in **bold**.

2.1. The Environment

Consider a finite set of ***agents*** $\mathcal{N} = \{1, 2, \ldots, n\}$ and ***mechanism outcomes*** $\mathcal{Q} = \{q_1, q_2, \ldots, q_K\}$. Each agent $i \in \mathcal{N}$ has ***private information*** $V_i \in \mathcal{D}_i$ and takes an ***ex post hidden action*** $s_i \in \mathcal{S}_i$ (where $\mathcal{S}_i$ may be finite or infinite). A collection of *ex post* hidden actions is denoted as $s = (s_1, \ldots, s_n) \in \times_{j \in \mathcal{N}} \mathcal{S}_j =: \mathcal{S}$. Private information and hidden actions determine an agent's ***utility*** of an outcome $q \in \mathcal{Q}$ and money $m \in \mathbb{R}$, which is represented as $\Pi_i(\cdot, s; V_i) : \mathcal{Q} \times \mathbb{R} \to \mathbb{R}$:

$$\Pi_i(q, m, s; V_i) = U_i(q, s; V_i) + m. \tag{1}$$

It is useful to let $U_i(q, s; V_i) \equiv V_i(q, s)$ and refer to V_i as agent i's ***type***. We refer to $\mathcal{D}_i \subseteq \{V_i : \mathcal{Q} \times \mathcal{S} \to \mathbb{R}\} =: \mathscr{D}_i$ as the ***domain*** of possible types and let $\mathcal{D} := \times_j \mathcal{D}_j$.

Definition 1 (Social Choice Environment). An ***environment with ex post hidden actions*** is defined by a tuple $\{\mathcal{N}, (\mathcal{S}_j, \mathcal{D}_j, V_j)_{j \in \mathcal{N}}, \mathcal{Q}\}$.

By fixing an outcome $q \in \mathcal{Q}$, an environment with *ex post* hidden actions can be reduced to a non-cooperative game, $\{\mathcal{N}, (\mathcal{S}_j, V_j(q, \cdot))_{j \in \mathcal{N}}\}$, where $\mathcal{S}_j$ is agent j's action space.

From a standard mechanism design perspective, one could view agents as having a mix of independent and (strategically) interdependent types. On the one hand, agents have independent types insofar as $V_i(\cdot)$ does not depend on V_{-i}. On the other hand, agents have (strategically) interdependent types insofar as $V_i(\cdot)$ depends on $s_{-i} \in \mathcal{S}_{-i}$.

We give an example of an environment with *ex post* hidden actions below.

Example 1 (Oil Drilling; Dasgupta & Maskin, 2000). A mechanism designer is distributing two licenses to drill oil. Each (prospective) firm's valuation of a license equals its expected profit in the oil market. As such, the challenge for the mechanism designer is that each firm decides its strategic actions in the market *after* licenses are distributed. This means that the mechanism designer cannot distribute licenses based on firms' oil drilling plans.

To model this more formally, let $\mathcal{N} = \{1, 2, 3\}$ denote the set of prospective firms. $\mathcal{Q} = \{(1, 1, 0), (1, 0, 1), (0, 1, 1)\}$ denotes the set of possible license distributions (1 represents a firm receiving a license, 0 represents a firm not receiving a license). Let $s_i \in \mathcal{S}_i = [0, S_i]$ denote firm i's oil production, which is an *ex post* hidden action. A firm's marginal cost of production, $V_i \in \mathcal{D}_i = [0, D_i]$, is exogenous private information. We model a firm's profit in the oil market as a Cournot competition. This means that firm i's utility is given as

$$\mathcal{U}_i(q, m, s; V_i) := m + \begin{cases} \overbrace{s_i \cdot P(q \cdot s) - s_i V_i}^{\text{Profit from oil market}} & \text{if } q_i = 1 \\ \underbrace{0}_{\text{Profit = 0 without license}} & \text{otherwise} \end{cases}$$

where $P : \mathbb{R}_+ \to \mathbb{R}_+$ is the inverse demand function.

In this example, a firm's oil pumping decision in the market is an *ex post hidden action* because pumping oil: *(i)* impact a firm's preferences over mechanism outcomes and *(ii)* takes shape *after* the mechanism outcome. The setup above forms the basis of examples throughout the paper. ■

2.2. Social Choice Mechanism

An agent's *ex post* hidden action takes on a different role than the standard concept of private information. In general, an agent may choose $s_i \in \mathcal{S}_i$ based on others' types, others' strategic actions, and is determined after the mechanism outcome. This means that the mechanism designer cannot elicit *ex post* hidden actions from agents, as such information is subject to change. Instead, the designer must solely rely on exogenous private information to make social decisions.

As such, we consider social choice mechanisms that require each agent to report exogenous private information, but not *ex post* hidden actions. We thus define a ***social choice function*** $f : \mathcal{D} \to \mathcal{Q}$ as a mapping from agents' reported types to an outcome. A ***payment function*** $p = (p_1, p_2, \ldots, p_n) : \mathcal{D} \to \mathbb{R}^n$ maps agents' reported types to a money payment for each agent. Finally, a ***social choice mechanism*** is denoted by a pair (f, p).

2.3. Timing

Having spelled out the model ingredients, it is worth noting the exact timing of the mechanisms that will be considered. (Other timings are perhaps possible, but we restrict attention here for the sake of discussion.)

1. Each agent reports a type $\tilde{V}_i \in \mathcal{D}_i$ simultaneously.
2. The mechanism designer makes a social decision based on everybody's report, $f(\tilde{V}_1, \ldots, \tilde{V}_n)$.
3. Each agent pays the mechanism designer money based on everybody's reported types, $p(\tilde{V}_1, \ldots, \tilde{V}_n) = (p_1, \ldots, p_n) \in \mathbb{R}^n$.
4. With everybody knowing everybody's true type, each player takes an action, $s = (s_1, \ldots, s_n) \in \mathcal{S}$.
5. Finally, each player receives his/her payoff,

$$V_i(f(\tilde{V}_1, \ldots, \tilde{V}_n), s) - p_i(\tilde{V}_1, \ldots, \tilde{V}_n).$$

3. PROBLEM STATEMENT

In this section, we formalize the main research question of this paper.

3.1. Gathering Information Truthfully

Consider an environment with *ex post* hidden actions. The mechanism designer's goal is, first, to gather information from agents and, second, to make a social decision based on this information. The challenge is that the mechanism designer cannot make payments based on agents' *ex post* hidden actions, $s_i \in \mathcal{S}_i$, but only on agents' exogenous private information, $V_i \in \mathcal{D}_i$. Informally, our criterion for truthfulness is that each agent has a (weakly) dominant-strategy to report his/her true V_i while all agents' strategic actions follow a Nash equilibrium.

We introduce some notation to define our notion of truthfulness formally. As noted above, any outcome $q \in \mathcal{Q}$ reduces an environment with *ex post* hidden actions to a non-cooperative game, $\{\mathcal{N}, (\mathcal{S}_j, V_j(q, \cdot))_{j \in \mathcal{N}}\}$. Let $\mathcal{NE}^{(V_i, V_{-i})}(q) \subseteq \mathcal{S}$ denote the set of Nash equilibria of the non-cooperative game under outcome

$q \in \mathcal{Q}$ and with types (V_i, V_{-i}). When working with a social choice function f, it is often useful to let $\mathcal{NE}^{(V_i,V_{-i})}(f(V_i, V_{-i})) \equiv \mathcal{NE}^{(V_i,V_{-i})}(V_i, V_{-i})$.

All results in the following paper build on the following assumption, which states that there exists at least one Nash equilibrium at every combination of types $V \in \mathcal{D}$ and possible mechanism outcomes $q \in \mathcal{Q}$. Why is this assumption important? Among other reasons, it gives players credible deviations from telling the truth – otherwise the mechanism designer is not worried about non-truthful strategies because it leads to non-Nash outcomes. Relatedly, the absence of Nash equilibria renders it impossible to define a payment function, which necessarily relies on Nash outcomes to provide a foundation on which a payment function can be built.

Assumption 1 *An environment with ex post hidden actions satisfies **regularity** if $\mathcal{NE}^V(q)$ is non-empty for all $V \in \mathcal{D}$ and $q \in \mathcal{Q}$.*

The following definition formalizes what it means for an agent to always have a truthful dominant-strategy.

Definition 2 (Truthful Dominant-Strategy). For any (f, p) and an agent i, truthful reporting is a ***(weakly) dominant-strategy*** if, for every $V_{-i} \in \mathcal{D}_{-i}$, it is weakly optimal for any type $V_i \in \mathcal{D}_i$ to report truthfully, that is,

$$V_i(f(V_i, V_{-i}), s^*) - p_i(V_i, V_{-i}) \geq V_i(f(\tilde{V}_i, V_{-i}), \tilde{s}^*) - p_i(\tilde{V}_i, V_{-i}) \qquad (2)$$

for all untruthful reports $\tilde{V}_i \in \mathcal{D}_i \setminus \{V_i\}$ that yield different outcomes ($f(V_i, V_{-i}) \neq f(\tilde{V}_i, V_{-i})$), Nash equilibria that result from reporting truthfully, i.e. $s^* \in \mathcal{NE}^{(V_i,V_{-i})}(V_i, V_{-i})$, and Nash equilibria that result from mis-reporting, i.e. $\tilde{s}^* \in \mathcal{NE}^{(V_i,V_{-i})}(\tilde{V}_i, V_{-i})$.

Notice that a truthful dominant-strategy is defined with respect to the *true* Nash equilibria of the underlying game (that is, $\mathcal{NE}^{(V_i,V_{-i})}(\cdot)$) and not Nash equilibria based on reports (that is, $\mathcal{NE}^{(\tilde{V}_i,V_{-i})}(\cdot)$ on the right-hand side of (2)).

Dominant-strategy incentive compatibility extends the concept of a truthful dominant-strategy to all agents.

Definition 3 (Nash Dominant-Strategy Incentive Compatibility). A social choice mechanism (f, p) satisfies ***Nash dominant-strategy incentive compatibility (DSIC)*** if truthful reporting is a dominant strategy for all agents.

In other words, if a social choice mechanism satisfies Nash DSIC, then no agent can strictly gain by misreporting ($\tilde{V}_i$) to induce a new Nash equilibrium outcome ($\tilde{s}^*$).

Nash DSIC underlies a so-called "truthful" social choice function.

Definition 4 (Truthful). A social choice function f is ***truthful*** if there exists a payment function p such that (f, p) satisfies Nash DSIC; p is said to ***implement*** f.

The main goal of this paper is to analyze social choice functions that can be implemented in environments with *ex post* hidden actions.

Before proceeding, we make three remarks about the definition of 'truthful'.

Remark 1. All results in this paper are formulated in terms of (i) reporting V_i truthfully as a dominant-strategy and (ii) $s^* \in \mathcal{S}$ following a Nash equilibrium of the *true* underlying game. Concerning the latter, everything can be easily adapted to represent *rationalizable strategies* rather than Nash equilibria.[4] All of the proofs would proceed as stated (with a few modifications).

Remark 2. What doss the economic interpretation of 'truthful' mean in light of *ex post* hidden actions? We have two interpretations in mind. First, it may be that private information is common knowledge (in the sense of Aumann, 1976), and the mechanism designer knows that agents have common knowledge but does not know agents' private information – such a setting might be appropriate for small-n environments, or when agents have interacted in mechanism settings together repeatedly. Second, agents may report while not knowing others' private information, and then types of all agents are revealed *after* the mechanism outcome – such a setting may be appropriate for, e.g. the oil drilling license auction or some spectrum auction settings. It is our hope that main results and proofs can be adapted to different definitions of truthfulness that may be more appropriate in specific economic settings.

4. A CHARACTERIZATION OF IMPLEMENTATION WITH *EX POST* HIDDEN ACTIONS

This section contains the main results of the paper, which are presented as follows. First, we derive a straightforward condition that is necessary for

4 See Bernheim (1984) and Pearce (1984) for the definition of rationalizable strategies and Bergemann et al. (2011) for a characterization of implementation using rationalizable strategies in robust mechanism design.

implementation. Second, we show that this condition is also sufficient when we focus on unit-demand auctions with identical goods, and third we use this to study different social choice functions in a Cournot entry market. Finally, we present a generalized characterization of implementation with *ex post* hidden actions.

4.1. A Necessary Condition for Implementation

In implementation theory, truthfulness is often characterized by a 'monotonicity condition' (e.g., 'Maskin monotonicity', Maskin, 1999). Here, we take the same approach.

The next definition introduces a monotonicity condition for social choice functions.

Definition 5 (Nash Monotonicity, Nash-MON). In an environment with *ex post* hidden actions, a social choice function f satisfies ***Nash monotonicity (Nash-MON)*** if, for all $i \in \mathcal{N}$, $V_{-i} \in \mathcal{D}_{-i}$, and $V_i, \tilde{V}_i \in \mathcal{D}_i$ such that $f(V_i, V_{-i}) \neq f(\tilde{V}_i, V_{-i})$,

$$\overbrace{V_i(f(V_i, V_{-i}), s^*) - V_i(f(\tilde{V}_i, V_{-i}), s^{**})}^{\text{marginal gain for } V_i \text{ to report truthfully vs. report } \tilde{V}_i} \geq \underbrace{\tilde{V}_i(f(V_i, V_{-i}), \tilde{s}^{**}) - \tilde{V}_i(f(\tilde{V}_i, V_{-i}), \tilde{s}^*)}_{\text{marginal gain for } \tilde{V}_i \text{ to misreport with } V_i \text{ vs. report truthfully}} \tag{3}$$

for all Nash equilibria that result from:

- player V_i reporting truthfully, i.e. $s^* \in \mathcal{NE}^{(V_i, V_{-i})}(V_i, V_{-i})$,
- player V_i misreporting, i.e. $s^{**} \in \mathcal{NE}^{(V_i, V_{-i})}(\tilde{V}_i, V_{-i})$,
- player $\tilde{V}_i$ reporting truthfully, i.e. $\tilde{s}^* \in \mathcal{NE}^{(\tilde{V}_i, V_{-i})}(\tilde{V}_i, V_{-i})$,
- and player $\tilde{V}_i$ misreporting, i.e. $\tilde{s}^{**} \in \mathcal{NE}^{(\tilde{V}_i, V_{-i})}(V_i, V_{-i})$.

How does Nash-MON work? The key idea is that the marginal gain from reporting truthfully must always be greater than the marginal gain from misreporting, and this must be true for all (relevant) Nash equilibria. Why is Nash-MON a monotonicity condition? The reason is made clearer below (that is, after introducing additional assumptions). But for now, suppose that V_i is a 'higher' type that $\tilde{V}_i$. Then Nash-MON can be viewed as requiring that the marginal gain of matching V_i with a 'better' outcome—that is $f(V_i, V_{-i})$—is greater than matching $\tilde{V}_i$ with a 'worse' outcome—that is $f(\tilde{V}_i, V_{-i})$. We refer the interested reader to Examples 3 and 4, where the role of matching higher types with better outcomes is clearly demonstrated.

It is important to note that Nash-MON only checks those Nash equilibria that are relevant for players' *true* types. This is closely related to the assumption that players' true types are revealed after the mechanism designer makes a social decision. What would happen if players' types are not revealed (which would be akin to firms entering the oil market without knowing other firms' types)? In general, an entirely different condition than Nash-MON would be required. One would not only need to check conditions for the true Nash equilibria but also all possible Nash equilibria that could emerge for any combination of players. This would be a much stronger condition than Nash-MON, and it is unlikely that the positive results that are shown below would be possible.

It is straightforward to show that Nash-MON is a necessary condition for implementation in any environment with *ex post* hidden actions.

Lemma 1 (Necessity of Nash-MON) *In an environment with ex post hidden actions, a social choice function f is truthful only if f satisfies Nash-MON.*

Proof of Lemma 1. Suppose that $f : \mathcal{D} \to \mathcal{Q}$ is truthful, and let $p : \mathcal{D} \to \mathbb{R}^n$ implement f. Consider agent $i \in \mathcal{N}$. Fix $V_{-i} \in \mathcal{D}$. Consider types $V_i, \tilde{V}_i \in \mathcal{D}_i$ that yield different outcomes ($f(V_i, V_{-i}) \neq f(\tilde{V}_i, V_{-i})$). If type V_i participates in the social choice mechanism (f, p), then Nash DSIC implies that

$$V_i(f(V_i, V_{-i}), s^*) - p_i(V_i, V_{-i}) \geq V_i(f(\tilde{V}_i, V_{-i}), s^{**}) - p_i(\tilde{V}_i, V_{-i}) \qquad (4)$$

for all $s^* \in \mathcal{NE}^{(V_i, V_{-i})}(V_i, V_{-i})$ and $s^{**} \in \mathcal{NE}^{(V_i, V_{-i})}(\tilde{V}_i, V_{-i})$ (supposing that such equilibria exist). If type $\tilde{V}_i$ participates in (f, p), then Nash DSIC implies that

$$\tilde{V}_i(f(\tilde{V}_i, V_{-i}), \tilde{s}^*) - p_i(\tilde{V}_i, V_{-i}) \geq \tilde{V}_i(f(V_i, V_{-i}), \tilde{s}^{**}) - p_i(V_i, V_{-i}) \qquad (5)$$

for all $\tilde{s}^* \in \mathcal{NE}^{(\tilde{V}_i, V_{-i})}(\tilde{V}_i, V_{-i})$ and $\tilde{s}^{**} \in \mathcal{NE}^{(\tilde{V}_i, V_{-i})}(V_i, V_{-i})$ (again supposing that such equilibria exist). Combining (4) and 5 yields

$$\begin{aligned} V_i(f(V_i, V_{-i}), s^*) - V_i(f(\tilde{V}_i, V_{-i}), s^{**}) & \qquad (6) \\ \geq p_i(V_i, V_{-i}) - p_i(\tilde{V}_i, V_{-i}) & \qquad (7) \\ \geq \tilde{V}_i(f(V_i, V_{-i}), \tilde{s}^{**}) - \tilde{V}_i(f(\tilde{V}_i, V_{-i}), \tilde{s}^*). & \qquad (8) \end{aligned}$$

Finally, (6) ≥ (8) implies that Nash-MON is necessary. □

4.1.1. Nash-MON Alone Is not Sufficient

Given the many impossibility theorems in implementation theory, it is perhaps unsurprising to find that Nash-MON alone is not sufficient for truthfulness. We show in the following example that, even in a single-agent environment, *ex post* hidden actions make it impossible to find a payment function that implements a social choice function.[5]

Example 2 (By itself, Nash-MON is not sufficient for truthfulness). Suppose there are three outcomes, $\mathcal{Q} = \{A, B, C\}$, and one agent, $\mathcal{N} = \{1\}$. Let the domain of types be $\mathcal{D}_1 = \{V^A, V^B, V^C\}$. Let the set of possible *ex post* hidden actions be $\mathcal{S}_1 = \{s_1, s_2\}$. Consider a social choice function such that $f(V^A) = A$, $f(V^B) = B$ and $f(V^C) = C$. The agent's preferences over outcomes, as a function of private information, is given as

$$V^A = \begin{array}{cc} & \begin{array}{cc} s_1 & s_2 \end{array} \\ \begin{array}{c} A \\ B \\ C \end{array} & \begin{bmatrix} 0 & -5 \\ 55 & 40 \\ 60 & 70 \end{bmatrix} \end{array} \quad V^B = \begin{array}{cc} & \begin{array}{cc} s_1 & s_2 \end{array} \\ \begin{array}{c} A \\ B \\ C \end{array} & \begin{bmatrix} -10 & 0 \\ 25 & 60 \\ 70 & 85 \end{bmatrix} \end{array} \quad V^C = \begin{array}{cc} & \begin{array}{cc} s_1 & s_2 \end{array} \\ \begin{array}{c} A \\ B \\ C \end{array} & \begin{bmatrix} 0 & -5 \\ 40 & 35 \\ 75 & 20 \end{bmatrix} \end{array}.$$

As such, if V^A reports truthfully then his/her optimal action is s_1 and $V^A(A, s_1) = 0$. If V^B reports truthfully then his/her optimal action is s_2 and $V^B(B, s_2) = 60$. If V^C reports truthfully then his/her optimal action is s_1 and $V^C(C, s_1) = 75$.

In a single-agent environment, Nash-MON checks over the agent's optimal action at each outcome, e.g. $s_1 = \arg\max_{s \in \{s_1, s_2\}} V^A(A, s)$ and $s_2 = \arg\max_{s \in \{s_1, s_2\}} V^A(C, s)$. This means that Nash-MON imposes three conditions:

$$\begin{aligned} V^A(A, s_1) - V^A(B, s_1) = 0 - 55 \;&\geq\; 0 - 60 = V^B(A, s_2) - V^B(B, s_2) \\ V^A(A, s_1) - V^A(C, s_2) = 0 - 70 \;&\geq\; 0 - 75 = V^C(A, s_1) - V^C(C, s_1) \\ V^B(B, s_2) - V^B(C, s_2) = 60 - 85 \;&\geq\; 40 - 75 = V^C(B, s_1) - V^C(C, s_1). \end{aligned}$$

As such, Nash-MON is satisfied in this setting.

[5] The following example builds on Example 1 from Bikhchandani et al. (2006) directly.

However, there is no payment function that can implement f. Suppose that the agent pays p^A for report V^A, p^B for report V^B, and p^C for report V^C. Without loss of generality let $p^A = 0$. If it is optimal for type V^A to report truthfully vs. report V^B then

$$V^A(A, s_1) - p^A \geq V^A(B, s_1) - p^B \implies p^B \geq 55. \tag{9}$$

If it is optimal for type V^B to report truthfully vs. report V^C then

$$V^B(B, s_2) - p^B \geq V^B(C, s_2) - p^C \implies p^C - p^B \geq 25. \tag{10}$$

Taken together, (9) and (10) imply that $p^C \geq 80$. But if type V^C prefers reporting truthfully vs. reporting V^A, then

$$V^C(C, s_1) - p^C \geq V^C(A, s_1) - p^A \iff 75 - p^C \geq 0 - 0 \iff 75 \geq p^C$$

which contradicts $p^C \geq 80$. Consequently, even though f satisfies Nash-MON, there exists no payment function that can implement f. ■

Because Nash-MON alone is not sufficient, we pursue two strategies that circumvent the kind of scenarios presented above. The first strategy focuses on unit-demand auctions with identical goods. The second strategy uses restrictions on the domain of types in a similar manner to Bikhchandani et al. (2006).

4.2. Unit-Demand Auctions with Identical Goods

Our first result focuses on unit-demand auctions with identical goods, which includes the oil drilling license auction in Example 1 (we show this in Remark 3 after we present the definition of a unit-demand auction).

In a unit-demand auction with identical goods, each agent only cares about acquiring an item and does not necessarily care about *which* item is acquired.[6] This means that the space of mechanism outcomes can be represented as $Q \subseteq \{0, 1\}^n$, where 1 represents an agent receiving an item and 0 represents an agent not receiving an item. Consequently, if other agents report V_{-i}, then i's report can only render two possible mechanism outcomes: report a 'high type' and win an item or report a 'low type' and not win an item.

[6] Note that this does not mean that agents are symmetric, or that agents equally value winning an auction. In contrast, it only means that all goods being auctioned can be treated as equivalent.

To formalize these ideas in a definition, we endow the set of outcomes $\mathcal{Q} \subseteq \{0,1\}^n$ with an order $\succeq_i$ that represents whether or not agent i wins an item. That is, for any $q', q'' \in \mathcal{Q}$, we let $q' \succeq_i q'' \iff q'_i > q''_i \iff q'_i = 1 > 0 = q''_i$.

Definition 6 (Unit-Demand Auction with Identical Goods). An environment with *ex post* hidden actions is a ***unit-demand auction with identical goods*** if:

(i) The mechanism outcome space is $\mathcal{Q} \subseteq \{0,1\}^n$,

(ii) Each agent i can either win or not win an item, i.e. $|\{f(V_i, V_{-i}) : V_i \in \mathcal{D}_i\}| = 2$ for all $V_{-i} \in \mathcal{D}_{-i}$,

(iii) Each agent prefers winning an item to not winning an item, that is, $\overline{q} \succeq_i \underline{q}$ implies that $V_i(\overline{q}, \overline{s}) \geq V_i(\underline{q}, \underline{s})$ for every $\overline{s} \in \mathcal{NE}^V(\overline{q})$, $\underline{s} \in \mathcal{NE}^V(\underline{q})$ and $V \in \mathcal{D}$.

Remark 3. Under certain restrictions, it is straightforward to show that the oil drilling license auction in Example 1 is a unit-demand auction with identical goods. *(i)* By definition, the distribution of licenses can be represented as $\mathcal{Q} \subseteq \{0,1\}^n$. *(ii)* By setting $\mathcal{D}_i = \mathcal{D}_j$ for all $i, j \in \mathcal{N}$, it means that a player i may not win a good because he/she may be outbid by player j—different type spaces could otherwise make j win *de facto*. *(iii)* Finally, by choosing an appropriate pricing function, it can be guaranteed that the Nash equilibrium profits are always greater than zero for winners. This in turn satisfies condition *(iii)* of Definition 6. It should also be noted that, by choosing the inverse demand function appropriately, it can be guaranteed that there exists a Nash equilibrium at every (V_i, V_{-i}) combination (e.g. see Examples 3 and 4 below).

Our first main result shows that Nash-MON is necessary and sufficient for implementing a social choice function in a unit-demand auction with identical goods (see Appendix A for the proof).

Theorem 1 *In a unit-demand auction environment with identical goods and ex post hidden actions, a social choice function is truthful if and only if it satisfies Nash-MON.*

There are perhaps two surprising features of Theorem 1 that are worth noting. First, besides *(iii)* in Definition 6, no other assumptions are made regarding the space of exogenous private information, $\mathcal{D}$, and *ex post* hidden actions, $\mathcal{S}$. This generality is in contrast to most implementation results that impose specific assumptions on the type space (as in Roberts, 1979, and Bikhchandani et al., 2006). Second, the theorem is a 'positive' result for a problem that allows types to be (strategically) interdependent, which contrasts with many negative results in implementation theory, especially when types are interdependent (see, e.g., Jehiel et al., 2006).

Below, we apply Theorem 1 to the oil drilling license auction from above. In Example 3, we show that there exists a truthful mechanism that can allocate licenses to the most production-efficient firms. In Example 4, we show that the welfare maximizing social choice function cannot be implemented, even by the Vickery-Clarke-Groves mechanism.

Example 3 (*continued from Example 1*: A social choice function that satisfies Nash-MON). This example builds on the oil drilling license auction from Example 1. To make it more concrete, we assume that the inverse demand function is linear, $P(X) = a - bX$. Then for appropriately selected (a, b), there exists a unique Nash equilibrium.

In this example, we study the *efficiency social choice function*, f^{EFF} : $\mathcal{D} \to \mathcal{Q}$, which gives licenses to the most production-efficient firms. This means that, if firms report $(V_1, V_2, V_3) = (4, 5, 6)$ then $f^{EFF}(4, 5, 6) = (1, 1, 0)$ (firm 3 does not receive a license because it reported the highest marginal cost of production). Because this example is a unit-demand auction with identical goods (see Remark 3), it follows from Theorem 1 that f^{EFF} can be implemented if and only if it satisfies Nash-MON.

To see whether f^{EFF} satisfies Nash-MON, fix $V_{-i} \in \mathcal{D}_{-i}$. Consider $V_i, \tilde{V}_i \in \mathcal{D}_i$ such that $f^{EFF}(V_i, V_{-i})$ gives i an oil drilling license and $f^{EFF}(\tilde{V}_i, V_{-i})$ does not give i an oil drilling license (which means that $V_i < \tilde{V}_i$). Then Nash-MON is satisfied if and only if

$$\begin{aligned} V_i(f^{EFF}(V_i, V_{-i}), s^*) - V_i(f^{EFF}(\tilde{V}_i, V_{-i}), s^{**}) \\ \geq \tilde{V}_i(f^{EFF}(V_i, V_{-i}), \tilde{s}^{**}) - \tilde{V}_i(f^{EFF}(\tilde{V}_i, V_{-i}), \tilde{s}^*) \end{aligned} \tag{11}$$

where $(s^*, s^{**}, \tilde{s}^{**}, \tilde{s}^*)$ are the unique Nash equilibria in the oil drilling market in the respective setting (see Definition 5 for the precise definition). Because firm i's payoff is zero if it does not win a license, (11) reduces to

$$V_i(f^{EFF}(V_i, V_{-i}), s^*) \geq \tilde{V}_i(f^{EFF}(V_i, V_{-i}), \tilde{s}^{**}) \tag{12}$$

where $s^* \in \mathcal{NE}^{(V_i,V_{-i})}(V_i, V_{-i})$ and $\tilde{s}^* \in \mathcal{NE}^{(\tilde{V}_i,V_{-i})}(V_i, V_{-i})$. If V_j also wins a license under $f^{EFF}(V_i, V_{-i})$, then we can plug in i's Nash equilibrium profits to re-write (12) as

$$\begin{aligned} V_i(f^{EFF}(V_i, V_{-i}), s^*) &= \frac{(a - 2V_i + V_j)^2}{9b} \\ &\geq \frac{(a - 2\tilde{V}_i + V_j)^2}{9b} = \tilde{V}_i(f^{EFF}(V_i, V_{-i}), \tilde{s}^{**})) \end{aligned} \tag{13}$$

(this is the standard Nash equilibrium result from a Cournot model competition with a linear inverse demand). Because $V_i < \tilde{V}_i$ it follows that (13) holds, which means that Nash-MON also holds. It thus follows from Theorem 1 that f^{EFF} can be implemented. We refer the interested reader to the proof of Theorem 1 (Appendix A) where we propose a payment function that implements f^{EFF}. ■

Example 4 (*continued from Example 1*: Maximizing welfare is not implementable). This example builds on the previous one. But here, we present a social choice function that cannot be implemented.

Consider the same Cournot entry environment as above, except now, we study whether the welfare-maximizing social choice function $f^{WM} : \mathcal{D} \to \mathcal{Q}$ can be implemented: that is, f^{WM} gives licenses to the two firms that, at the unique Nash equilibrium, maximize total industry profits. Because this setting is a unit-demand auction with identical goods, it follows from Theorem 1 that f^{WM} can be implemented if and only if f^{WM} satisfies Nash-MON.

To see if f^{WM} satisfies Nash-MON, fix $V_{-i} \in \mathcal{D}_{-i}$. Consider $V_i, \tilde{V}_i \in \mathcal{D}_i$ such that $f^{WM}(V_i, V_{-i})$ gives i an oil drilling license and $f^{WM}(\tilde{V}_i, V_{-i})$ does not give i an oil drilling license. Following the same steps as in Example 3, Nash-MON holds if and only if

$$\begin{aligned} &\overbrace{V_i(f^{WM}(V_i, V_{-i}), s^*)}^{\text{Profit when } V_i \text{ reports truthfully and gets a license}} \\ &\qquad \geq \underbrace{\tilde{V}_i(f^{WM}(V_i, V_{-i}), \tilde{s}^{**})}_{\text{Profit when } \tilde{V}_i \text{ misreports to get a license}} \end{aligned} \tag{14}$$

where $(s^*, \tilde{s}^{**},)$ are the unique Nash equilibria in the oil drilling market with agents (V_i, V_j) and $(\tilde{V}_i, V_j)$, respectively. As such, Nash-MON boils down to requiring that, at the unique Nash equilibrium, reporting truthfully always yields a higher payoff than misreporting in order to get a license.

We claim that (14) does not hold in general. To see this, suppose that $(a,b) = (10,1)$ and $(V_1,V_2,V_3) = (1,2,4)$. Let Π_i^{ij} represent the unique Nash equilibrium payoff of agent i if i and j enter the Cournot competition. Suppose that all agents report truthfully. Because

$$\Pi_1^{12} + \Pi_2^{12} = 16.6 \quad < \quad \Pi_1^{13} + \Pi_3^{13} = 17.0 \quad > \quad \Pi_2^{23} + \Pi_3^{23} = 13.0$$

it follows that $f^{WM}(V_1,V_2,V_3) = (1,0,1)$ (that is, firms 1 and 3 win licenses).

Is there any $\tilde{V}_3 \in \mathcal{D}_3 \subseteq \mathbb{R}_+$ such that (14) does not hold? Consider $\tilde{V}_3 = 3$. Since the unique Nash equilibrium payoffs with types $(V_1,V_2,\tilde{V}_3)$ are

$$\Pi_1^{12} + \Pi_2^{12} = 16.6 \quad > \quad \Pi_1^{1\tilde{3}} + \Pi_{\tilde{3}}^{1\tilde{3}} = 16.2 \quad > \quad \Pi_2^{2\tilde{3}} + \Pi_{\tilde{3}}^{2\tilde{3}} = 13.0$$

it follows that $f^{WM}(V_1,V_2,\tilde{V}_3) = (1,1,0)$. However, it is easy to show that (14) is not satisfied with $V_3 = 4$ and $\tilde{V}_3 = 3$:

$$V_3(f^{WM}(V_1,V_2,V_3),s^*) = 1.0 \;\ngeq\; 2.8 = \tilde{V}_3(f(V_1,V_2,V_3),\tilde{s}^{**})$$

(1.0 and 2.8 are calculated using (13)). Consequently, follows from Theorem 1 that there exists no payment function $p : \mathcal{D} \to \mathbb{R}^3$ that can implement the welfare-maximizing social choice function, f^{WM}. ■

Remark 4. Because there exists no payment function $p : \mathcal{D} \to \mathbb{R}^3$ that can implement the welfare-maximizing social choice function, it immediately follows that the Vickery-Clarke-Groves Mechanism cannot implement the efficient allocation rule shown above.

4.3. Generalized Implementation with *ex post* Hidden Actions

Here, we explore one way that Theorem 1 can be generalized beyond unit-demand auctions with identical goods. Our strategy for doing so is extending the concept of a *"rich"* type space introduced by Bikhchandani et al. (2006).

4.3.1. Restricting the Type Domain, $\mathcal{D}$

For each agent, we endow the set of mechanism outcomes $\mathcal{Q}$ with an order, $\succeq_i$. We assume that $(\mathcal{Q},\succeq_i)$ is a *quasi-ordered space*, which means that it is:

(i) *reflexive*, i.e. $q \succeq_i q$ for every $q \in \mathcal{Q}$, and (ii) *transitive*, i.e. $q \succeq_i q'$ and $q' \succeq_i q'' \implies q \succeq_i q''$ for every $q, q', q'' \in \mathcal{Q}$. In practice, $\succeq_i$ represents agent i's ranking of mechanism outcomes (this is made clear below).

Our notion of a rich environment builds on the idea that all agents prefer 'better' outcomes. To make this more concrete, we introduce an order by which players rank the outcome space.

Definition 7 (Consistency). In an environment with *ex post* hidden actions, we say that a type $V_i \in \mathscr{D}_i$ is ***consistent with*** $(\mathcal{Q}, \succeq_i)$ if, for all $V_{-i} \in \mathcal{D}_{-i}$:

$$q_k \succeq_i q_l \implies V_i(q_k, s^k) \geq V_i(q_l, s^l) \text{ for all Nash equilibria } s^k \in \mathcal{NE}^{(V_i, V_{-i})}(q_k) \text{ and } s^l \in \mathcal{NE}^{(V_i, V_{-i})}(q_l)\ .$$

In other words, a consistent type will always prefer 'better' mechanism outcomes ($q_k \succeq_i q_l$) with respect to all relevant Nash equilibria. One can also think about a consistent type as having preferences that follow his/her order over types, $\succeq_i$.

A *rich* type space is simply the collection of all consistent types.

Definition 8 (Rich Type Space, $\mathcal{D}_i$). In an environment with *ex post* hidden actions, $\mathcal{D}_i$ is ***rich type space*** if there is a quasi-order $\succeq_i$ on $\mathcal{Q}$ such that every $V_i \in \mathscr{D}_i$ that is consistent with $(\mathcal{Q}, \succeq_i)$ is in $\mathcal{D}_i$.

A *rich environment* extends the definition of a rich type space to each agent.

Definition 9 (A Rich Environment). An environment with *ex post* hidden actions is ***rich*** if there are quasi-orders $(\succeq_j)_{j \in \mathcal{N}}$ on $\mathcal{Q}$ such that each $\mathcal{D}_j$ is a rich type space.

Remark 5. It is worth noting the differences between a rich environment and a unit-demand auction (Definition 6). Because of Definition 6(iii), a unit-demand requires that every type $V_i \in \mathcal{D}_i$ is *consistent* (this follows from defining a quasi-order that orders an outcome based on whether i wins or does not win an item). However, a unit-demand auction environment does not require the type space to be rich.

While a unit-demand auction environment permits a more general type space, the trade-off comes in the definition of the mechanism outcome space, $\mathcal{Q}$. Unit-demand auctions require that $\mathcal{Q} \subseteq \{0, 1\}^n$. Rich environments, however, impose no constraints on $\mathcal{Q}$.

It turns out that 'richness' provides the key to proving the main result of this paper. In the following theorem, we show that Nash-MON is necessary *and* sufficient for implementation when restricting attention to rich environments (see Appendix B for the proof).

Theorem 2 *In a rich environment with ex post hidden actions, a social choice function is truthful if and only if it satisfies Nash-MON.*

There are two points worth raising that help unpack Theorem 2. First, Theorem 2 identifies the necessary and sufficient features a social choice function must have if we hope to find a mechanism that implements it. These features—summarized by Nash-MON—boil down to requiring that matching 'higher types' with 'better outcomes' exhibits increasing returns. Without this feature, it is generally impossible to implement a social choice function in a way that is robust with respect to *ex post* hidden actions.

Second, it is worth asking what kind of payment function can implement social choice functions that satisfy Nash-MON. We propose such a function in the proof of Theorem 2 (see Appendix B). Broadly speaking, the payment function builds on the 'increasing returns' property from Nash-MON and links it to price reports in a way that renders truth-telling as optimal. Also, similar to Bikhchandani et al. (2006), each agent i's payment function only depends on V_{-i} and the mechanism outcome $f(V_i, V_{-i})$. This allows us to ensure that truth-telling is not only optimal but is also a *dominant-strategy*.

5. CONCLUDING REMARKS

The key message of this paper is that strategic actions are important for designing institutions that implement societal goals. With a few exceptions, previous mechanism design models do not incorporate strategic actions that influence personal and/or others' preferences. The paper argues that strategic actions matter, and that current mechanisms—such as the Vickery-Clarke-Groves Mechanism—are not guaranteed to be truthful when strategic actions are present.

This paper makes two contributions. The first is introducing *ex post hidden actions*, which is one way of embedding mechanisms in a large space of real-world environments that the mechanism cannot control. The example used throughout the paper is an auction that is distributing oil drilling licenses, and the '*ex post* hidden actions' represent the strategic market decisions that firms take after winning oil drilling licenses. The second is characterizing social choice functions that can be implemented in a way that is robust with respect to *ex post* hidden actions. We complement this characterization by proposing a mechanism that does so. By means of examples, we show that the tools in this paper can be used in applied settings.

We present two characterizations in the main text. The first characterization can be viewed as an 'off-the-shelf' result that can be easily amended to applied economic settings. It applies directly to unit-demand auctions with identical goods and admits almost any combination of agent types and strategic actions. The second can be viewed as one way to generalize the first. This result falls somewhere between adding *ex post* hidden actions to the model in Bikhchandani et al. (2006) and offering a strategic counterpart to the work by Bergemann & Morris (2005) on robust mechanism design.

Looking forward, there are several avenues for future work. One possibility is exploring other concepts of truthfulness. This paper focused on designing mechanisms that are robust with respect to Nash equilibrium, but this requires that agents have sufficient information in order to correctly deduce a Nash equilibrium. Different assumptions with weaker information criteria are possible, such as a maximin approach (Wolitzky, 2016). Another possibility is studying the tension between implementation with *ex post* hidden actions vs. individual rationality (Ma et al., 1988; Jackson & Palfrey, 2001), which was not touched on in this paper. Finally, there is the possibility of developing more applied-oriented results. This paper showed that, by focusing on unit-demand auctions with identical goods, a sharp and useful characterization was possible. Future work could explore other applied-oriented environments to uncover more economic insights into implementation with *ex post* hidden actions.

A. PROOF OF THEOREM 1

The necessity of Nash-MON follows from Lemma 1. Therefore, all that remains is to show that Nash-MON is also sufficient when we restrict attention to unit-demand auction environments with identical goods and *ex post* hidden actions.

The proof of sufficiency proceeds as follows. First, we define a pricing function. Second, we identify a useful inequality relationship of the pricing function (Lemma A.2). Third, we show that the pricing function is truthful at 'end points' (Lemma A.3). Finally, we build on these two results to show that the pricing function implements f (Lemma A.1).

Before we define the pricing function, some definitions are in order.

In what follows, $V_{-i} \in \mathcal{D}_{-i}$ is fixed.

Inverse social choice function

Let $\mathcal{Y}_i(\cdot \mid V_{-i}) : \mathcal{Q} \rightrightarrows \mathcal{D}_i$ be the inverse social choice function,

$$\mathcal{Y}_i(q_k \mid V_{-i}) \equiv \mathcal{Y}_i(q_k) = \{V_i \in \mathcal{D}_i : f(V_i, V_{-i}) = q_k\}.$$

In other words, $\mathcal{Y}_i(q_k \mid V_{-i}) \subseteq \mathcal{D}_i$ is the set of all types from $\mathcal{D}_i$ such that $f(V_i, V_{-i}) = q_k$.

Differencing function

Consider the following differencing function between any two outcomes $q_k, q_l \in \mathcal{Q}$ that depends on V_{-i}:

$$\delta_{q_k q_l}(V_{-i}) = \begin{cases} \inf \big\{ V_i(q_k, s^k) - V_i(q_l, s^l) : V_i \in \mathcal{Y}_i(q_k \mid V_{-i}), \\ \qquad s^k \in \mathcal{NE}^{(V_i, V_{-i})}(q_k), s^l \in \mathcal{NE}^{(V_i, V_{-i})}(q_l) \big\} & \text{if } q_k \neq q_l \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

We often let $\delta_{q_k q_l}(V_{-i}) \equiv \delta_{kl}(V_{-i})$ unless otherwise ambiguous.

Identifying the ‘best’ and ‘worst’ outcome

By assumption, $|\{f(V_i, V_{-i}) : V_i \in \mathcal{D}_i\}| = 2$ (see *(ii)* in Definition 6). Let

$$\{\overline{q}(V_{-i}), \underline{q}(V_{-i})\} = \{f(V_i, V_{-i}) : V_i \in \mathcal{D}_i\}$$

such that, for any $V_{-i} \in \mathcal{D}_{-i}$,

$$V_i(\overline{q}(V_{-i}), \overline{s}) \geq V_i(\underline{q}(V_{-i}), \underline{s})$$

for all $\overline{s} \in \mathcal{NE}^{(V_i, V_{-i})}(\overline{q}(V_{-i}))$ and $\underline{s} \in \mathcal{NE}^{(V_i, V_{-i})}(\underline{q}(V_{-i}))$ ($\overline{q}$ and $\underline{q}$ have this property by *(iv)* of Definition 6).

In what follows below, we often let $\overline{q}(V_{-i}) \equiv \overline{q}$ and $\underline{q}(V_{-i}) \equiv \underline{q}$ unless otherwise ambiguous.

Pricing function

Consider the following payment function (where $f(V_i, V_{-i}) = q$):[7]

$$p_i(V_i, V_{-i}) \equiv p_i(q) = \begin{cases} -\delta_{\overline{q}(V_{-i})f(V_i,V_{-i})} & \text{if } f(V_i, V_{-i}) = \underline{q}(V_{-i}) \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

What is the intuition underlying $p_i(V_i, V_{-i})$? If V_i does not win an item (i.e. $f(V_i, V_{-i}) = \underline{q}(V_{-i})$) then V_i is compensated according to $-\delta(\geq 0)$. If V_i wins an item then V_i is not compensated ($p_i = 0$).

In the following lemma, we establish that the payment function given by (16) implements f.

Lemma A.1 *In a unit-demand auction environment with identical goods and ex post hidden actions, $(p_1, p_2, \ldots, p_n) : \mathcal{D} \to \mathbb{R}^n$ as defined in* (16) *implements f if f satisfies Nash-MON.*

It follows directly from Lemma A.1 that f is truthful. Before we prove this lemma, we require two supporting lemmas.

First preliminary result: Inequality relationship for δ

In the following lemma, we show that Nash-MON gives $\delta(\cdot)$ an inequality relationship that turns out to be useful for proving Lemma A.1.

Lemma A.2 *Let $V_{-i} \in D_{-i}$ and $q \in \{f(V_i, V_{-i}) : V_i \in \mathcal{D}_i\}$. Then $\delta_{\overline{q}q}(V_{-i}) \geq -\delta_{q\overline{q}}(V_{-i})$ if f satisfies Nash-MON.*

Proof of Lemma A.2. If $q = \overline{q}(V_{-i})$ then $\delta_{\overline{q}\,\overline{q}}(V_{-i}) = 0 \geq 0 = -\delta_{\overline{q}\,\overline{q}}(V_{-i})$.

Therefore, suppose that $q = \underline{q}(V_{-i})$. By Nash-MON,

$$V_i(\underline{q}, s^*) - V_i(\overline{q}, s^{**}) \geq \tilde{V}_i(\underline{q}, \tilde{s}^{**}) - \tilde{V}_i(\overline{q}, \tilde{s}^*)$$

[7] It is straightforward to show that this payment function is finite for all possible values. Fix $V_{-i} \in \mathcal{D}_{-i}$. First, $p_i^K = 0$. Second, for $k \neq K$, select $V_i \in \mathcal{Y}(q_K \mid V_{-i})$ and $\tilde{V}_i \in \mathcal{Y}_i(q_k \mid V_{-i})$, whereby Nash-MON implies that

$$\infty > V_i(q_K, s^*) - V_i(q_k, s^{**}) \geq \tilde{V}_i(q_K, \tilde{s}^{**}) - \tilde{V}_i(q_k, \tilde{s}^*) > -\infty$$

for all $s^* \in \mathcal{NE}^{(V_i,V_{-i})}(q_K)$, $s^{**} \in \mathcal{NE}^{(V_i,V_{-i})}(q_k)$, $\tilde{s}^{**} \in \mathcal{NE}^{(\tilde{V}_i,V_{-i})}(q_K)$, and $\tilde{s}^* \in \mathcal{NE}^{(\tilde{V}_i,V_{-i})}(q_k)$. Therefore, $-\delta_{Kk} = p_i^K$ is finite.

for all types $V_i \in \mathcal{Y}_i(\underline{q} \mid V_{-i})$ and $\tilde{V}_i \in \mathcal{Y}_i(\overline{q} \mid V_{-i})$, and Nash equilibria $s^* \in \mathcal{NE}^{(V_i,V_{-i})}(q)$, $s^{**} \in \mathcal{NE}^{(V_i,V_{-i})}(\overline{q})$, $\tilde{s}^{**} \in \mathcal{NE}^{(\tilde{V}_i,V_{-i})} \in (q)$, and $\tilde{s}^* \in \mathcal{NE}^{(\tilde{V}_i,V_{-i})}(\overline{q})$. Consequently,

$$\begin{aligned}
\delta_{\underline{q}\overline{q}}(V_{-i}) &= \inf\big\{V_i(\underline{q}, s^*) - V_i(\overline{q}, s^{**}) : V_i \in \mathcal{Y}_i(\underline{q} \mid V_{-i}), \\
&\qquad\qquad s^* \in \mathcal{NE}^{(V_i,V_{-i})}(\underline{q}), s^{**} \in \mathcal{NE}^{(V_i,V_{-i})}(\overline{q})\big\} \\
&\geq \sup\big\{\tilde{V}_i(\underline{q}, \tilde{s}^{**}) - \tilde{V}_i(\overline{q}, \tilde{s}^*) : \tilde{V}_i \in \mathcal{Y}_i(\overline{q} \mid V_{-i}), \\
&\qquad\qquad \tilde{s}^{**} \in \mathcal{NE}^{(\tilde{V}_i,V_{-i})}(\underline{q}), \tilde{s}^* \in \mathcal{NE}^{(\tilde{V}_i,V_{-i})}(\overline{q})\big\} \\
&= -\inf\big\{\tilde{V}_i(\overline{q}, \tilde{s}^*) - \tilde{V}_i(\underline{q}, \tilde{s}^{**}) : \tilde{V}_i \in \mathcal{Y}_i(\overline{q} \mid V_{-i}), \\
&\qquad\qquad \tilde{s}^* \in \mathcal{NE}^{(\tilde{V}_i,V_{-i})}(\overline{q}), \tilde{s}^{**} \in \mathcal{NE}^{(\tilde{V}_i,V_{-i})}(\underline{q})\big\} \\
&= -\delta_{\overline{q}\underline{q}}(V_{-i}).
\end{aligned}$$

□

Second preliminary result: Implementing the 'end points' of Q

In the following lemma, we show that the pricing function implements $\overline{q}$ and $\underline{q}$ (the proof of Lemma A.1 is a small step after this result).

Lemma A.3 *If f satisfies Nash-MON then for any types $(V_i, V_{-i}) \in \mathcal{D}$, and corresponding Nash equilibria $\underline{s} \in \mathcal{NE}^{(V_i,V_{-i})}(\underline{q}(V_{-i}))$ and $\overline{s} \in \mathcal{NE}^{(V_i,V_{-i})}(\overline{q}(V_{-i}))$:*

(i) if $V_i(\underline{q}, \underline{s}) - p_i(\underline{q}) < V_i(\overline{q}, \overline{s}) - p_i(\overline{q})$ then $f(V_i, V_{-i}) \neq \underline{q}(V_{-i})$;

(ii) if $V_i(\underline{q}, \underline{s}) - p_i(\underline{q}) > V_i(\overline{q}, \overline{s}) - p_i(\overline{q})$ then $f(V_i, V_{-i}) \neq \overline{q}(V_{-i})$.

Proof of Lemma A.3. We prove each claim separately.

Proof of claim (i)

By way of contradiction, suppose that $f(V_i, V_{-i}) = \underline{q}$. Because $p_i(\underline{q}) = -\delta_{\overline{q}\underline{q}}$ and $p_i(\overline{q}) = 0$,

$$V_i(\underline{q}, \underline{s}) - V_i(\overline{q}, \overline{s}) < p_i(\underline{q}) - p_i(\overline{q}) = \underbrace{-\delta_{\overline{q}\underline{q}}(V_{-i}) \leq \delta_{\underline{q}\overline{q}}(V_{-i})}_{\text{Lemma A.2}}$$

where the last inequality follows from Lemma A.2. Consequently,

$$\begin{aligned} V_i(\underline{q}, \underline{s}) - V_i(\overline{q}, \overline{s}) &< \delta_{\underline{q}\overline{q}}(V_{-i}) \\ &= \inf \big\{V_i'(\underline{q}, \underline{s}') - V_i'(\overline{q}, \overline{s}') : V_i' \in \mathcal{Y}_i(\underline{q} \mid V_{-i}), \\ &\qquad \underline{s}' \in \mathcal{NE}^{(V_i', V_{-i})}(\underline{q}), \overline{s}' \in \mathcal{NE}^{(V_i', V_{-i})}(\overline{q})\big\} \\ &\leq V_i(\underline{q}, \underline{s}) - V_i(\overline{q}, \overline{s}) \end{aligned}$$

where the last inequality follows because $f(V_i, V_{-i}) = \underline{q} \implies V_i \in \mathcal{Y}_i(\underline{q} \mid V_{-i})$. However, $V_i(\underline{q}, \underline{s}) - V_i(\overline{q}, \overline{s}) < V_i(\underline{q}, \underline{s}) - V_i(\overline{q}, \overline{s})$ is an impossibility. Therefore, $f(V_i, V_{-i}) \neq \underline{q}$.

Proof of claim (ii)

The proof proceeds in a similar manner. By way of contradiction, suppose that $f(V_i, V_{-i}) = \overline{q}$. By assumption,

$$V_i(\overline{q}, \overline{s}) - V_i(\underline{q}, \underline{s}) < p_i(\overline{q}) - p_i(\underline{q}) = \delta_{\overline{q}\underline{q}}(V_{-i}).$$

Consequently,

$$\begin{aligned} V_i(\overline{q}, \overline{s}) - V_i(\underline{q}, \underline{s}) &< \delta_{\overline{q}\underline{q}}(V_{-i}) \\ &= \inf \big\{V_i'(\overline{q}, \overline{s}') - V_i'(\underline{q}, \underline{s}') : V_i' \in \mathcal{Y}_i(\overline{q} \mid V_{-i}), \\ &\qquad \overline{s}' \in \mathcal{NE}^{(V_i', V_{-i})}(\overline{q}), \underline{s}' \in \mathcal{NE}^{(V_i', V_{-i})}(\underline{q})\big\} \\ &\leq V_i(\overline{q}, \overline{s}) - V_i(\underline{q}, \underline{s}) \end{aligned}$$

where the last inequality follows because $f(V_i, V_{-i}) = \overline{q} \implies V_i \in \mathcal{Y}_i(\overline{q} \mid V_{-i})$. However, $V_i(\overline{q}, \overline{s}) - V_i(\underline{q}, \underline{s}) < V_i(\overline{q}, \overline{s}) - V_i(\underline{q}, \underline{s})$ is an impossibility. Therefore, $f(V_i, V_{-i}) \neq \overline{q}$. □

Finishing the proof of Theorem 1

We conclude the proof of Theorem 1 by proving Lemma A.1.

Proof of Lemma A.1. We show that p implements f. By way of contradiction, suppose that there exists $V_{-i} \in \mathcal{D}_{-i}$, outcomes $q', q'' \in \{f(V_i', V_{-i}) : V_i \in \mathcal{D}_i\}$ such that $q' \neq q''$, a type $V_i \in \mathcal{Y}_i(q' \mid V_{-i})$ that implements q', a type $\tilde{V}_i \in \mathcal{Y}_i(q'' \mid V_{-i})$ that implements q'', and Nash equilibria $s' \in \mathcal{NE}^{(V_i, V_{-i})}(q')$ and $s'' \in \mathcal{NE}^{(V_i, V_{-i})}(q'')$ such that

$$\begin{aligned} & V_i(q', s') - p_i(V_i, V_{-i}) < V_i(q'', s'') - p_i(\tilde{V}_i, V_{-i}) \\ \Longleftrightarrow \quad & V_i(q', s') - p_i(q') < V_i(q'', s'') - p_i(q'') \end{aligned} \tag{17}$$

In other words, (17) means that V_i can strictly gain by reporting $\tilde{V}_i (\neq V_i)$ in order to induce outcome q'' and Nash equilibrium s''.

We must consider two possibilities:

(i) Suppose that $q'' = \overline{q}(V_{-i})$. Then $q' = \underline{q}(V_{-i})$. By Lemma A.3(i), $V_i(\underline{q}, s') - p_i(\underline{q}) < V_i(\overline{q}, s'') - p_i(\overline{q})$ implies that $f(V_i, V_{-i}) \neq \underline{q}(V_{-i}) = q'$. But then $V_i \notin \mathcal{Y}_i(q' \mid V_{-i})$, which is a contradiction.

(ii) Suppose that $q' = \overline{q}(V_{-i})$. Then $q'' = \underline{q}(V_{-i})$. By Lemma A.3(ii), $V_i(\overline{q}, s') - p_i(\overline{q}) < V_i(\underline{q}, s'') - p_i(\underline{q})$ implies that $f(V_i, V_{-i}) \neq \overline{q}(V_{-i}) = q'$. But then $V_i \notin \mathcal{Y}_i(q' \mid V_{-i})$, which is a contradiction.

Taken together, *(i)* and *(ii)* imply that $q', q'' \neq \overline{q}(V_{-i})$. However, having assumed that $f(V_i, V_{-i}) \neq f(\tilde{V}_i, V_{-i})$, it must be that $q' = \overline{q}(V_{-i})$ or $q'' = \overline{q}(V_{-i})$. This impossibility means that p implements f. □

B. PROOF OF THEOREM 2

The necessity of Nash-MON follows from Lemma 1. Therefore, all that remains is to prove that Nash-MON is also sufficient when each agent's type space is rich.

The proof of sufficiency proceeds in a similar manner to that shown above. First, we propose a pricing function. Second, we identify two useful inequality relationships of the pricing function (Lemma B.2). Third, we show that the pricing function implements the 'end points' of $\mathcal{Q}$ (Lemma B.3). Finally, we build on these results to show that the pricing function implements f (Lemma B.1).

Before we define the pricing function, some definitions are in order.

In what follows, $V_{-i} \in \mathcal{D}_{-i}$ is fixed.

Inverse social choice function

Let $\mathcal{Y}_i(\cdot \mid V_{-i}) : \mathcal{Q} \rightrightarrows \mathcal{D}_i$ be the inverse social choice function:

$$\mathcal{Y}_i(q_k \mid V_{-i}) = \{V_i \in \mathcal{D}_i : f(V_i, V_{-i}) = q_k\}.$$

In other words, $\mathcal{Y}_i(q_k \mid V_{-i})$ are all the types in $\mathcal{D}_i$ such that $f(V_i, V_{-i}) = q_k$.

Differencing function between outcomes

Consider the following differencing function between any two outcomes $q_k, q_l \in \mathcal{Q}$ (which is similar to that defined in (15)):

$$\delta_{q_k q_l}(V_{-i}) = \begin{cases} \inf \big\{V_i(q_k, s^k) - V_i(q_l, s^l) : V_i \in \mathcal{Y}_i(q_k \mid V_{-i}), & \\ \qquad s^k \in \mathcal{NE}^{(V_i, V_{-i})}(q_k), s^l \in \mathcal{NE}^{(V_i, V_{-i})}(q_l)\big\} & \text{if } q_k \neq q_l \\ 0 & \text{otherwise.} \end{cases}$$

Payment function

Let $(\succeq_j)_{j \in \mathcal{N}}$ be the quasi-orders on $\mathcal{Q}$ such that $\mathcal{D}_j$ is a rich type space (such $(\succeq_j)_{j \in \mathcal{N}}$ exist by assumption). For every $V_{-i} \in \mathcal{D}_{-i}$, and with relabeling if necessary, let $q_K \in \mathcal{Q}$ denote an outcome in $\mathcal{Q}$ such that:

(i) $q_K \in \{f(V_i, V_{-i}) : V_i \in \mathcal{D}_i\}$ and

(ii) $q \not\succeq_i q_K$ for each $q \in \{f(V_i, V_{-i}) : V_i \in \mathcal{D}_i\} \backslash \{q_K\}$.

The interpretation is that q_K is the 'best' possible outcome that V_i could obtain given that others reported V_{-i}.

Consider the following payment function (supposing that $f(V_i, V_{-i}) = q_k$):[8]

$$p_i(V_i, V_{-i}) \equiv p_i(q_k, V_{-i}) \equiv p_i^k = \begin{cases} -\delta_{Kk}(V_{-i}) & \text{if } k \neq K \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

In the following lemma, we claim that this payment function implements f.

Lemma B.1 *In a rich environment with ex post hidden actions,* $(p_1, p_2, \ldots, p_n) : \mathcal{D} \to \mathbb{R}^n$ *as defined in* (18) *implements* f *if* f *satisfies Nash-MON.*

It follows directly from Lemma B.1 that f is truthful. But to prove Lemma B.1, two additional lemmas are needed.

[8] $p_i(\cdot)$ is finite for reasons similar to those given in fn. 7.

First supporting lemma: Inequality relationships of δ

It turns out that the following inequality relationships are essential for proving that $(p_1, \ldots, p_n) : \mathcal{D} \to \mathbb{R}^n$ from (18) implements f.

Lemma B.2 *For any social choice function f, $V_{-i} \in \mathcal{D}_{-i}$, and $q_k, q_l, q_r \in \mathcal{Q}$:*

(i) if $q_k \succeq_i q_l$ then $\delta_{rk}(V_{-i}) \leq \delta_{rl}(V_{-i})$;

(ii) if f satisfies Nash-MON then $\delta_{kl}(V_{-i}) \geq -\delta_{lk}(V_{-i})$.

Proof of Lemma B.2. We prove each claim separately.

Proof of claim (i)

By the rich domain assumption, $V_i(q_k, s^k) \geq V_i(q_l, s^l)$ for all types $V_i \in \mathcal{D}_i$ and Nash equilibria $s^k \in \mathcal{NE}^{(V_i,V_{-i})}(q_k)$ and $s^l \in \mathcal{NE}^{(V_i,V_{-i})}(q_l)$. Consequently,

$$V_i(q_r, s^r) - V_i(q_l, s^l) \geq V_i(q_r, s^r) - V_i(q_k, s^k)$$

for all $V_i \in \mathcal{Y}_i(q_r \mid V_{-i})$ and $s^r \in \mathcal{NE}^{(V_i,V_{-i})}(q_r)$. Therefore, $\delta_{rl}(V_{-i}) \geq \delta_{rk}(V_{-i})$.

Proof of claim (ii)

If q_k and q_l represent the same outcome then the result is trivial.

Therefore, suppose that q_k and q_l are different outcomes. By Nash-MON,

$$V_i(q_k, s^*) - V_i(q_l, s^{**}) \geq \tilde{V}_i(q_k, \tilde{s}^{**}) - \tilde{V}_i(q_l, \tilde{s}^*)$$

for all types $V_i \in \mathcal{Y}_i(q_k \mid V_{-i})$ and $\tilde{V}_i \in \mathcal{Y}_i(q_l \mid V_{-i})$, and all Nash equilibria $s^* \in \mathcal{NE}^{(V_i,V_{-i})}(q_k)$, $s^{**} \in \mathcal{NE}^{(V_i,V_{-i})}(q_l)$, $\tilde{s}^{**} \in \mathcal{NE}^{(\tilde{V}_i,V_{-i})}(q_k)$ and $\tilde{s}^* \in$

$\mathcal{NE}^{(\tilde{V}_i, V_{-i})}(q_l)$. Consequently,

$$\begin{aligned}
\delta_{kl}(V_{-i}) &= \inf \Big\{ V_i(q_k, s^*) - V_i(q_l, s^{**}) : V_i \in \mathcal{Y}_i(q_k \mid V_{-i}), \\
&\qquad\qquad s^* \in \mathcal{NE}^{(V_i, V_{-i})}(q_k), s^{**} \in \mathcal{NE}^{(V_i, V_{-i})}(q_l) \Big\} \\
&\geq \sup \Big\{ \tilde{V}_i(q_k, \tilde{s}^{**}) - \tilde{V}_i(q_l, \tilde{s}^*) : \tilde{V}_i \in \mathcal{Y}_i(q_l \mid V_{-i}), \\
&\qquad\qquad \tilde{s}^{**} \in \mathcal{NE}^{(\tilde{V}_i, V_{-i})}(q_k), \tilde{s}^* \in \mathcal{NE}^{(\tilde{V}_i, V_{-i})}(q_l) \Big\} \\
&= -\inf \Big\{ \tilde{V}_i(q_l, \tilde{s}^*) - \tilde{V}_i(q_k, \tilde{s}^{**}) : \tilde{V}_i \in \mathcal{Y}_i(q_l \mid V_{-i}), \\
&\qquad\qquad \tilde{s}^* \in \mathcal{NE}^{(\tilde{V}_i, V_{-i})}(q_l), \tilde{s}^{**} \in \mathcal{NE}^{(\tilde{V}_i, V_{-i})}(q_k) \Big\} \\
&= -\delta_{lk}(V_{-i}).
\end{aligned}$$

□

Second supporting lemma: Implementing q_K

In the following lemma, we show that the pricing function implements q_K.

Lemma B.3 *Let f be a social choice function that satisfies Nash-MON. Then for any outcome $q_l \in \mathcal{Q}$ such that $q_l \neq q_K$, types $(V_i, V_{-i}) \in \mathcal{D}$, and Nash equilibria $s^l \in \mathcal{NE}^{(V_i, V_{-i})}(q_l)$ and $s^K \in \mathcal{NE}^{(V_i, V_{-i})}(q_K)$:*

(i) if $V_i(q_l, s^l) - p_i^l < V_i(q_K, s^K) - p_i^K$ then $f(V_i, V_{-i}) \neq q_l$;

(ii) if $V_i(q_l, s^l) - p_i^l > V_i(q_K, s^K) - p_i^K$ then $f(V_i, V_{-i}) \neq q_K$.

Proof of Lemma B.3. We prove each claim separately.

Proof of claim (i)

By way of contradiction, suppose that $f(V_i, V_{-i}) = q_l$. By definition, $p_i^K = 0$ and $p_i^l = -\delta_{Kl}(V_{-i})$. Therefore,

$$V_i(q_l, s^l) - V_i(q_K, s^K) < p_i^l - p_i^K = \underbrace{-\delta_{Kl}(V_{-i}) \leq \delta_{lK}(V_{-i})}_{\text{Lemma B.2(ii)}}$$

where the last inequality follows from Lemma B.2(ii). Because $f(V_i, V_{-i}) = q_l \implies V_i \in \mathcal{Y}_i(q_l \mid V_{-i})$, it follows that

$$V_i(q_l,s^l) - V_i(q_K, s^K) \tag{19}$$

$$< \delta_{lK}(V_{-i}) \tag{20}$$

$$= \inf\{V_i'(q_l, \hat{s}^l) - V_i'(q_K, \hat{s}^K) : V_i' \in \mathcal{Y}_i(q_l \mid V_{-i}), \tag{21}$$

$$\hat{s}^l \in \mathcal{NE}^{(V_i',V_{-i})}(q_l), \hat{s}^K \in \mathcal{NE}^{(V_i',V_{-i})}(q_K)\} \tag{22}$$

$$\leq V_i(q_l, s^l) - V_i(q_K, s^K), \tag{23}$$

where (19) < (23) = (19) is a contradiction. Therefore, $f(V_i, V_{-i}) \neq q_l$.

Proof of claim (ii)

By way of contradiction, suppose that $f(V_i, V_{-i}) = q_K$. By assumption,

$$V_i(q_K, s^K) - V_i(q_l, s^l) < p_i^K - p_i^l = \delta_{Kl}(V_{-i}).$$

Because $f(V_i, V_{-i}) = q_K \implies V_i \in \mathcal{Y}_i(q_K \mid V_{-i})$, it follows that

$$V_i(q_K,s^K) - V_i(q_l, s^l) \tag{24}$$

$$< \delta_{Kl}(V_{-i}) \tag{25}$$

$$= \inf\left\{V_i'(q_K, \hat{s}^K) - V_i'(q_l, \hat{s}^l) : V_i' \in \mathcal{Y}_i(q_K \mid V_{-i}), \right. \tag{26}$$

$$\left. \hat{s}^K \in \mathcal{NE}^{(V_i',V_{-i})}(q_K), \hat{s}^l \in \mathcal{NE}^{(V_i',V_{-i})}(q_l)\right\} \tag{27}$$

$$\leq V_i(q_K, s^K) - V_i(q_l, s^l), \tag{28}$$

where (24) < (28) = (24) is a contradiction. Therefore, $f(V_i, V_{-i}) \neq q_K$. □

Finishing the proof of Theorem 2

We conclude the proof of Theorem 2 by proving Lemma B.1.

Proof of Lemma B.1. Fix $V_{-i} \in \mathcal{D}_{-i}$. By way of contradiction, suppose there exist outcomes $q_j, q_k \in \mathcal{Q}$ ($q_j \neq q_k$), a type $V_i \in \mathcal{Y}_i(q_j \mid V_{-i})$ that implements q_j, a type $\tilde{V}_i \in \mathcal{Y}_i(q_k \mid V_{-i})$ that implements q_k, and Nash equilibria $s^j \in \mathcal{NE}^{(V_i,V_{-i})}(q_j)$ and $s^k \in \mathcal{NE}^{(V_i,V_{-i})}(q_k)$ such that

$$\begin{aligned} & V_i(q_j, s^j) - p_i(V_i, V_{-i}) < V_i(q_k, s^k) - p_i(\tilde{V}_i, V_{-i}) \\ \iff \quad & V_i(q_j, s^j) - p_i^j < V_i(q_k, s^k) - p_i^k. \end{aligned} \tag{29}$$

In other words, (29) says that V_i can strictly gain by reporting $\tilde{V}_i$ ($\neq V_i$) in order to induce outcome q_k and Nash equilibrium s^k.

It follows that $k \neq K$; otherwise by Lemma B.3(i), $V_i(q_j, s^j) - p_i^j < V_i(q_K, s^K) - p_i^K$ implies $f(V_i, V_{-i}) \neq q_j$, which contradicts the initial assumption. Similarly, it also follows that $j \neq K$; otherwise by Lemma B.3(ii), $V_i(q_K, s^K) - p_i^K < V_i(q_k, s^k) - p_i^k$ implies $f(V_i, V_{-i}) \neq q_K$, which contradicts the initial assumption.

Since $f(V_i, V_{-i}) = q_j$, it follows from the contrapositive of Lemma B.3(i) that $V_i(q_j, s^j) - p_i^j \geq V_i(q_K, s^K) - p_i^K$ for any $s^K \in \mathcal{NE}^{(V_i,V_{-i})}(q_K)$. Consider $\gamma > \varepsilon > 0$ such that

$$V_i(q_j, s^j) + \varepsilon - p_i^j < V_i(q_K, s^K) + \gamma - p_i^K < V_i(q_k, s^k) - p_i^k. \tag{30}$$

Define

$$\widehat{Q} = \{q_j\} \cup \{q_l \in Q : q_l \succeq_i q_j \text{ and } \max_{s^l \in \mathcal{NE}^{(V_i,V_{-i})}(q_l)} V_i(q_l, s^l) = V_i(q_j, s^j)\}.$$

Define the type $\widehat{V}_i$ in the following way:

$$\widehat{V}_i(q_r, s) = \begin{cases} V_i(q_j, s^j) + \varepsilon & \text{if } q_r \in \widehat{Q} \backslash \{q_K\} \\ \max_{s^r \in \mathcal{NE}^{(V_i,V_{-i})}(q_r)} V_i(q_r, s^r) + \gamma & \text{if } q_r = q_K \\ \max_{s^r \in \mathcal{NE}^{(V_i,V_{-i})}(q_r)} V_i(q_r, s^r) & \text{otherwise.} \end{cases} \tag{31}$$

Note that, for any $q_l \in \widehat{Q}$, any $\mathcal{NE}^{(V_i,V_{-i})}(q_l) = \mathcal{NE}^{(\widehat{V}_i,V_{-i})}(q_l)$.

We make the following claim about $\widehat{V}_i$.

Claim 1 *$\widehat{V}_i$ as defined in (31) is consistent with $(Q, \succeq_i)$.*

Supposing Claim 1 is true (we prove this below), it follows from the rich environment assumption that $\widehat{V}_i \in \mathcal{D}_i$. Since $q_l \succeq_i q_j$ for all $q_l \in \widehat{Q}$, it follows from Lemma B.2(i) that $\delta_{Kl}(V_{-i}) \leq \delta_{Kj}(V_{-i})$ and $p_i^l \geq p_i^j$ for all

$q_l \in \widehat{Q}$. Because $\max_{s^l \in \mathcal{NE}^{(V_i,V_{-i})}(q_l)} V_i(q_l, s^l) = \widehat{V}_i(q_l, \overline{s}^l) = \widehat{V}_i(q_j, s^j)$ for all $q_l \in \widehat{Q} \backslash \{q_K\}$,

$$\begin{aligned} \widehat{V}_i(q_l, \overline{s}^l) - p_i^l &\leq \widehat{V}_i(q_j, s^j) - p_i^j \\ &< \widehat{V}_i(q_K, s^K) - p_i^K < V_i(q_k, s^k) - p_i^k \end{aligned} \tag{32}$$

where the strict inequalities follow from (30).

We make three observations.

(i) If $q_k \in \widehat{Q}$ (or $= q_K$), then the first and last terms from (32) would yield

$$\begin{aligned} \max_{s \in \mathcal{NE}^{(V_i,V_{-i})}(q_k)} V_i(q_k, s) + \varepsilon(\text{or } +\gamma) - p_i^k &< V_i(q_k, s^k) - p_i^k \\ \implies V_i(q_k, s^k) + \varepsilon(\text{or } +\gamma) - p_i^k &< V_i(q_k, s^k) - p_i^k, \end{aligned}$$

which is a contradiction. Thus, $q_k \notin \widehat{Q} \cup \{q_K\}$.

(ii) Because $q_k \notin \widehat{Q} \cup \{q_K\}$, it follows from Lemma B.3(ii) and (32) that $\widehat{V}_i(q_K, s^K) - p_i^K < \widehat{V}_i(q_k, s^k) - p_i^K$ implies that $f(\widehat{V}_i, V_{-i}) \neq q_K$.

(iii) Finally, since $\widehat{V}_i(q_l, \overline{s}^l) - p_i^l < \widehat{V}_i(q_K, s^K) - p_i^K$, it follows from Lemma B.3(i) that $f(\widehat{V}_i, V_{-i}) \neq q_l$ for any $q_l \in \widehat{Q} \backslash \{q_K\}$.

Based on these three observations, it follows that $f(\widehat{V}_i, V_{-i}) = q_{\widehat{k}} \notin \widehat{Q} \cup \{q_K\}$. However,

$$0 = \widehat{V}_i(q_{\widehat{k}}, s^{\widehat{k}*}) - V_i(q_{\widehat{k}}, s^{\widehat{k}*}) < \widehat{V}_i(q_j, s^j) - V_i(q_j, s^j) = \varepsilon,$$

which violates Nash-MON. This contradiction means that Nash-MON is sufficient for f to be truthful in a rich environment with *ex post* hidden actions. □

Proof of Claim 1. Suppose that $q_r \succeq_i q_t$. Then $q_t \neq q_K$. Let $s^r \in \mathcal{NE}^{(V_i,V_{-i})}(q_r)$ and $s^t \in \mathcal{NE}^{(V_i,V_{-i})}(q_t)$. If $q_t \notin \widehat{Q}$ then

$$\widehat{V}_i(q_t, s^t) = \underbrace{V_i(q_t, s^t) \leq V_i(q_r, s^r)}_{\text{Consistency of } V_i} \leq V_i(q_r, s^r) + 0(\text{ or } +\gamma \text{ or } +\varepsilon)$$

where the first inequality follows from the consistency of V_i, and $(+\gamma)$ if $q_r = q_K$ or $(+\varepsilon)$ if $q_r \in \widehat{Q} \backslash \{q_K\}$. Consistency of $\widehat{V}_i$ thus follows.

Therefore, suppose that $q_t \in \widehat{Q}$. Three cases must be considered.

Case 1

If $q_r \in \widehat{Q} \backslash \{q_K\}$ then

$$\widehat{V}_i(q_t, s^t) = \underbrace{V_i(q_t, s^t) + \varepsilon \leq V_i(q_r, s^r) + \varepsilon}_{\text{Consistency of } V_i} = \widehat{V}_i(q_r, s^r)$$

where the inequality follows from the consistency of V_i. Consistency of $\widehat{V}_i$ thus follows.

Case 2

If $q_r = q_K \succeq_i q_t$ then, by the consistency of V_i and $\gamma > \varepsilon$,

$$\begin{aligned}
\widehat{V}_i(q_r, s^r) &= \max_{s^K \in \mathcal{NE}^{(V_i, V_{-i})}(q_K)} V_i(q_K, s^K) + \gamma \\
&> \max_{s^K \in \mathcal{NE}^{(V_i, V_{-i})}(q_K)} V_i(q_K, s^K) + \varepsilon \\
&\geq \max_{s^t \in \mathcal{NE}^{(V_i, V_{-i})}(q_t)} V_i(q_t, s^t) + \varepsilon = \widehat{V}(q_t, s^t).
\end{aligned}$$

where the weak inequality follows from the consistency of V_i. Consistency of $\widehat{V}_i$ thus follows in this case.

Case 3

If $q_r \notin \widehat{Q} \cup \{q_K\}$ then

$$\max_{s^r \in \mathcal{NE}^{(V_i, V_{-i})}(q_r)} V_i(q_r, s^r) \;\neq\; V_i(q_j, s^j) = \max_{s^t \in \mathcal{NE}^{(V_i, V_{-i})}(q_t)} V_i(q_t, s^t).$$

This means that

$$V_i(q_j, s^j) = \underbrace{\max_{s^t \in \mathcal{NE}^{(V_i, V_{-i})}(q_t)} V_i(q_t, s^t) < \max_{s^r \in \mathcal{NE}^{(V_i, V_{-i})}(q_r)} V_i(q_r, s^r)}_{\text{Consistency of } V_i} \neq V_i(q_j, s^j),$$

where strictness is necessary for ($\neq$) to be true. As such, a small enough $\varepsilon > 0$ can be selected to ensure that

$$\widehat{V}_i(q_t, s^t) = \max_{s^t \in \mathcal{NE}^{(V_i, V_{-i})}(q_l)} V_i(q_t, s^t) + \varepsilon < \max_{s^r \in \mathcal{NE}^{(V_i, V_{-i})}(q_r)} V_i(q_r, s^r) = \widehat{V}_i(q_r, s^r).$$

Consistency of $\widehat{V}_i$ thus follows.

□

References

Aumann, R. J. (1976). Agreeing to disagree. *The Annals of Statistics*, *4*(6), 1236–1239.

Bergemann, D., & Morris, S. (2005). Robust mechanism design. *Econometrica*, *73*(6), 1771–1813.

Bergemann, D., & Morris, S. (2011). Robust implementation in general mechanisms. *Games and Economic Behavior*, *71*(2), 261–281.

Bergemann, D., Morris, S., & Tercieux, O. (2011). Rationalizable implementation. *Journal of Economic Theory*, *146*(3), 1253–1274.

Bernheim, D. (1984). Rationalizable strategic behavior. *Econometrica*, *52*(4), 1007–1028.

Bikhchandani, S., Chatterji, S., Lavi, R., Mu'alem, A., Nisan, N., & Sen, A. (2006). Weak monotonicity characterizes deterministic dominant-strategy implementation. *Econometrica*, *74*(4), 1109–1132.

Chung, K.-S., & Ely, J. C. (2007). Foundations of dominant-strategy mechanisms. *Review of Economic Studies*, *74*(2), 447–476.

Dasgupta, P., Hammond, P., & Maskin, E. (1979). The implementation of social choice rules: Some general results on incentive compatibility. *Review of Economic Studies*, *46*(2), 185–216.

Dasgupta, P., & Maskin, E. (2000). Efficient auctions. *Quarterly Journal of Economics*, *115*(2), 341–388.

Jackson, M. O., & Palfrey, T. R. (2001). Voluntary implementation. *Journal of Economic Theory*, *98*(1), 1–25.

Jehiel, P., Meyer-ter-Vehn, M., Moldovanu, B., & Zame, W. R. (2006). The limits of ex post implementation. *Econometrica*, *74*(3), 585–610.

Jehiel, P., & Moldovanu, B. (2001). Efficient design with interdependent valuations. *Econometrica*, *69*(5), 1237–1259.

Jehiel, P., Moldovanu, B., & Stacchetti, E. (1999). Multidimensional mechanism design for auctions with externalities. *Journal of Economic Theory*, *85*(2), 258–293.

Klemperer, P. (2002a). How (not) to run auctions: The European 3G telecom auctions. *European Economic Review*, *46*(4-5), 829–845.

Klemperer, P. (2002b). What really matters in auction design. *Journal of Economic Perspectives*, *16*(1), 169–189.

Krishna, V., & Maenner, E. (2001). Convex potentials with an application to mechanism design. *Econometrica*, *69*(4), 1113–1119.

Ma, C.-T., Moore, J., & Turnbull, S. (1988). Stopping agents from "cheating". *Journal of Economic Theory*, *46*(2), 355–372.

Maskin, E. (1999). Nash equilibrium and welfare optimality. *Review of Economic Studies*, *66*(1), 23–38.

McAfee, R. P., & McMillan, J. (1988). Multidimensional incentive compatibility and mechanism design. *Journal of Economic Theory*, *46*(2), 335–354.

McAfee, R. P., & McMillan, J. (1996). Analyzing the airwaves auction. *Journal of Economic Perspectives*, *10*(1), 159–175.

Milgrom, P. (2000). Putting auction theory to work: The simultaneous ascending auction. *Journal of Political Economy*, *108*(2), 245–272.

Milgrom, P., & Segal, I. (2002). Envelope theorems for arbitrary choice sets. *Econometrica*, *70*(2), 583–601.

Myerson, R. B. (1981). Optimal auction design. *Mathematics of Operations Research*, *6*(1), 58–73.

Pearce, D. G. (1984). Rationalizable strategic behavior and the problem of perfection. *Econometrica*, *52*(4), 1029–1050.

Roberts, K. (1979). The characterization of implementable choice rules. In J.-J. Laffont (Ed.), *Aggregation and revelation of preferences* (pp. 321–345). Amsterdam: North-Holland.

Rochet, J.-C. (1987). A necessary and sufficient condition for rationalizability in a quasi-linear context. *Journal of Mathematical Economics*, *16*(2), 191–200.

Williams, S. R. (1999). A characterization of efficient, Bayesian incentive compatible mechanisms. *Economic Theory*, *14*(1), 155–180.

Wilson, R. (1987). Game-theoretic approaches to trading processes. In T. Bewley (Ed.), *Advances in economic theory: Fifth world congress* (pp. 33–77). Cambridge: Cambridge University Press.

Wolitzky, A. (2016). Mechanism design with maxmin agents: Theory and an application to bilateral trade. *Theoretical Economics*, *11*(3), 971–1004.

Journal of Mechanism and Institution Design
ISSN: 2399-844X (Print), 2399-8458 (Online)
DOI:10.22574/jmid.2020.12.002

ON OPTIMAL TAXES AND SUBSIDIES: A DISCRETE SADDLE-POINT THEOREM WITH APPLICATION TO JOB MATCHING UNDER CONSTRAINTS

Koji Yokote

Waseda University, Japan
sidehand@toki.waseda.jp

ABSTRACT

When a government intervenes in markets by setting a target amount of goods/services traded, its tax/subsidy policy is optimal if it entices the market participants to obey the policy target while achieving the highest possible social welfare. For the model of job market interventions by Kojima et al. (2019), we establish the existence of optimal taxes/subsidies as well as their characterization. Our methodological contribution is to introduce a discrete version of Karush-Kuhn-Tucker's saddle-point theorem based on the techniques in discrete convex analysis. We have two main results: we (i) characterize the optimal taxes/subsidies and the corresponding equilibrium salaries as the minimizers of a Lagrange function, and (ii) prove that the function satisfies a notion of discrete convexity (called L$^\natural$-convexity). These results together with others imply that an optimal tax/subsidy level exists and can be calculated via a computationally efficient algorithm.

Keywords: Job matching, taxation, discrete convex analysis.

JEL Classification Numbers: C61, C78, D47, H21.

The author is grateful to the editor, two anonymous referees, and seminar participants at Osaka University of Economics and Tokyo Metropolitan University, for their valuable comments. This work was supported by JSPS KAKENHI Grant Numbers 17H07179, 20K13458, and Waseda University Grants for Special Research Projects Grant Number 2019C-015.

1. INTRODUCTION

Governments often intervene in markets in order to address socio-economic problems such as environmental pollution or inequality. A representative form of government interventions is that the government (or international organizations) first determines a target amount of goods/services traded and then introduces taxes/subsidies accordingly.[1] Three examples are as follows:

- Some countries/cities set a policy goal regarding the usage of fossil fuel-powered cars and electric cars in an attempt to decrease the number of the former and increase the number of the latter (see Burch and Gilchrist 2018 for a survey). Taxes/subsidies are introduced to achieve the goal.

- In 2019, the G20 countries reached an agreement to reduce additional marine plastic waste to zero by 2050.[2] Imposing consumption taxes on plastic products is a hot environmental issue lately.[3]

- In the job market, the government often introduces taxes/subsidies to firms in an attempt to achieve a policy goal concerning diversity and/or equality.[4]

In these situations, an ideal tax/subsidy policy would entice the market participants to obey the policy target while achieving the highest possible social welfare; such a policy is called *optimal* hereafter. The purpose of this paper is to take a discrete mathematics approach to optimal taxes/subsidies and address theoretical issues of existence, characterization, and efficient computation. We consider the model of job market interventions by Kojima et al. (2019), which is suitable for analyzing the third example mentioned above.

---

1 The basic idea of this tax scheme can be traced back to Baumol and Oates (1971). Note the difference from the well-known *Pigovian taxes* determined by the level of externalities. For an application of the latter tax scheme, see Parry et al. (2007).

2 The Japan Times, 2020-7-16 (accessed on 2020-9-7).

3 In Chicago, a 7-cent tax on all grocery bags led to a 42 percent drop in usage; see Newsday, 2019-8-11 (accessed on 2020-9-7).

4 The Japanese government provides a subsidy to a firm if 2.2% of its workforce consists of people with disabilities; see Japanese Ministry of Health, Labour and Welfare (in Japanese, accessed on 2020-9-7). Other related examples in job markets can be found in Kojima et al. (2019).

Our analysis depends on admittedly restrictive assumptions; we assume that (i) all the participants have a quasi-linear utility function (i.e., no income effect is allowed),[5] (ii) all the firms view workers as substitutes (i.e., no complementarity is allowed), and (iii) the participants' valuations/revenues are known to the social planner. However, we believe that our analysis serves as a preliminary step for solving more complicated problems. To facilitate applications in future research, we provide a detailed discussion of the underlying mathematical tool.

In the first part of the paper, we introduce a preliminary mathematical tool. We borrow techniques from discrete convex analysis (Murota 2003a) and introduce a discrete version of Karush-Kuhn-Tucker's theorem, termed the *discrete saddle-point theorem*.[6] This theorem concerns a constrained maximization problem of a function defined over a discrete domain. Its major advantage is that it enables a unified treatment of constraints and discrete variables: the former is used to capture the government's policy goal and the latter is used to deal with indivisible commodities and discuss computational problems.

In the second part, we apply the discrete saddle-point theorem to the model of job market interventions developed by Kojima et al. (2019).[7] In this model, workers and firms participate in bilateral contracts associated with wages and the government sets the maximum/minimum number of workers allowed to be hired by each firm.[8] Applying the discrete saddle-point theorem to a suitably defined maximization problem, we (i) characterize the optimal taxes/subsidies and the corresponding equilibrium salaries as the minimizers of a Lagrange function, and (ii) prove that the function satisfies a notion of discrete convexity (called $L^\natural$-convexity). These results together with Kojima et al.'s (2019) result imply that an optimal tax/subsidy level exists and can be calculated via a computationally efficient algorithm.

This paper is part of the literature on the application of discrete convex

[5] Quasi-linearity means that utilities are linear with respect to money. We note that Kojima et al. (2019) do not assume quasi-linearity for doctors.

[6] Mathematically, this theorem is an immediate corollary of the well-known *intersection theorem* in discrete convex analysis but is tailored to the use of economists by emphasizing its parallelism with the original theorem.

[7] This model is a generalization of that of Kelso and Crawford (1982).

[8] These constraints are called "interval constraints." Precisely, Kojima et al. (2019) (see also Kojima et al. (2020a)) consider a general form of constraints and prove that the substitutes condition is preserved if and only if the constraints are "generalized interval constraints," a slight generalization of interval constraints.

analysis in economics; see Murota (2016) for a survey. Examples or prior applications include matching models (Kojima et al. 2018) and trading networks (Candogan et al. 2016). To the best of our knowledge, this paper is the first to apply discrete convex analysis to the analysis of taxes/subsidies.

This paper also contributes to the literature on matching with constraints. One can find different theoretical approaches from ours, as well as various forms of constraints in real-world problems, in Kamada and Kojima (2017) or Kamada and Kojima (2020).

Dupuy et al. (2020) develop a model of matching with taxation that accommodates a general form of taxes beyond linear ones and conduct a welfare analysis from both theoretical and econometric perspectives. Our approach is distinguished from theirs in that constraints on hiring and the computation of taxes are discussed. The extensive literature on taxation can be found in the reference therein.

The remainder of this paper is organized as follows. Section 2 introduces basic concepts in discrete convex analysis and presents the discrete saddle-point theorem. Section 3 analyzes optimal taxes/subsidies in a job-matching model. Section 4 presents concluding remarks. All proofs are provided in Section 5.

2. MATHEMATICAL PRELIMINARY: A DISCRETE SADDLE-POINT THEOREM

In the handling of mathematical problems with constraints, Karush-Kuhn-Tucker's saddle-point theorem plays a central role. In this section, we introduce a discrete version of the theorem by utilizing the concepts in discrete convex analysis (Murota 2003a).

Let K be a finite set. Let $\mathbb{R}^K$ denote the real vector space indexed by the elements in K. Let $\mathbb{Z}^K \subseteq \mathbb{R}^K$ be the set of vectors with integer coordinates. For a function $f : \mathbb{Z}^K \to \mathbb{Z} \cup \{-\infty\}$, we define the **effective domain of $f(\cdot)$** by

$$\mathrm{dom} f = \{x \in \mathbb{Z}^K : f(x) > -\infty\}.$$

For $A \subseteq K$, let $\mathbb{1}^A \in \{0,1\}^K$ denote the **characteristic vector of A**, i.e.,

$$\mathbb{1}^A_k = \begin{cases} 1 & \text{if } k \in A, \\ 0 & \text{otherwise.} \end{cases}$$

For a singleton set $\{k\} \subseteq K$, we write $\mathbb{1}^k$ for $\mathbb{1}^{\{k\}}$.

For $x \in \mathbb{Z}^K$, we define

$$\mathrm{supp}^+ x = \{k \in K : x_k > 0\},\ \mathrm{supp}^- x = \{k \in K : x_k < 0\}.$$

A function $f : \mathbb{Z}^K \to \mathbb{Z} \cup \{-\infty\}$ with $\mathrm{dom} f \neq \emptyset$ is said to be **M$^\natural$-concave** (Murota 2003a) if for any $x, y \in \mathbb{Z}^K$ and $k \in \mathrm{supp}^+(x - y)$, we have

(i) $f(x) + f(y) \leq f(x - \mathbb{1}^k) + f(y + \mathbb{1}^k)$, or

(ii) there exists $\ell \in \mathrm{supp}^-(x - y)$ such that $f(x) + f(y) \leq f(x - \mathbb{1}^k + \mathbb{1}^\ell) + f(y + \mathbb{1}^k - \mathbb{1}^\ell)$.

Intuitively, this condition states that if two points x and y approach each other, then the sum of the function values weakly increases. For a discussion of this property, see Chapter 6 of Murota (2003a) or Section 3 of Kojima et al. (2018). The class of M$^\natural$-concave functions includes various standard functions such as linear functions and quadratic functions; see Section 6.3 of Murota (2003a) for other examples.

For $x, y \in \mathbb{Z}^K$, we define $x \wedge y \in \mathbb{Z}^K$ and $x \vee y \in \mathbb{Z}^K$ by

$$(x \wedge y)_k = \min\{x_k, y_k\} \text{ for all } k \in K,\ (x \vee y)_k = \max\{x_k, y_k\} \text{ for all } k \in K.$$

A function $f : \mathbb{Z}^K \to \mathbb{Z} \cup \{-\infty\}$ with $\mathrm{dom} f \neq \emptyset$ is said to be **L$^\natural$-concave** (Murota 2003a) if for any $x, y \in \mathrm{dom} f$ and[9] $\lambda \in \mathbb{Z}_+$, we have

$$f(x) + f(y) \leq f((x + \lambda \cdot \mathbb{1}^K) \wedge y) + f(x \vee (y - \lambda \cdot \mathbb{1}^K)).$$

M$^\natural$-concavity and L$^\natural$-concavity are connected via the conjugacy relationship; see Theorem 8.12 of Murota (2003a). A function $f(\cdot)$ is said to be **L$^\natural$-convex** if $-f(\cdot)$ is L$^\natural$-concave. M$^\natural$-convexity is defined analogously.

We now turn our attention to constrained maximization problems with discrete variables. Recall that Karush-Kuhn-Tucker's saddle-point theorem holds if the objective and constraint functions are concave. With this in mind, we can consider replacing "concavity" with "M$^\natural$-concavity." This idea is stated below, where all the functions are assumed to be defined on $\mathbb{Z}^K$.

Naive idea. *Consider the problem of maximizing an objective function subject to constraint functions. If the objective and constraint functions satisfy M$^\natural$-concavity, then the solutions to the problem are translated into the saddle points of a Lagrange function.*

[9] $\mathbb{Z}_+$ denotes the set of non-negative integers.

Unfortunately, this idea does not work, even if the constraint functions are linear; we provide a counterexample in Section S.1 of the Supplemental Material (Yokote 2020). To circumvent this difficulty, we restrict the class of constraint functions to linear functions with a certain discrete structure. We say that $\mathcal{K} \subseteq 2^K \setminus \{\emptyset\}$ is a **hierarchy**[10] if, for every pair of elements A and A' in $\mathcal{K}$, we have $A \subseteq A'$ or $A' \subseteq A$ or $A \cap A' = \emptyset$. We say that a set of functions $g_1, \ldots, g_q : \mathbb{Z}^K \to \mathbb{Z}$ is a **hierarchical set of affine functions** if

(i) for each $j = 1, \ldots, q$, there exist $a_j \in \mathbb{Z}$ and $A_j \subseteq K$ with $A_j \neq \emptyset$ such that $g_j(x) = a_j - x \cdot \mathbb{1}^{A_j}$ for all $x \in \mathbb{Z}^K$, and

(ii) the set $\{A_j : j = 1, \ldots, q\}$ is a hierarchy.

We are in a position to state our first theorem.

Theorem 1 (Discrete Saddle-Point Theorem)**.** *Let $f : \mathbb{Z}^K \to \mathbb{Z} \cup \{-\infty\}$ be an $M^\natural$-concave function and $g_1, \ldots, g_q : \mathbb{Z}^K \to \mathbb{Z}$ be a hierarchical set of affine functions. Then, for $x^* \in \mathbb{Z}^K_+$, the following are equivalent:*

(i) *x^* is a solution to* $\max f(x)$ *subject to $g_j(x) \geq 0$ for all $j = 1, \ldots, q$.*

(ii) *There exists $(\lambda_1^*, \ldots, \lambda_q^*) \in \mathbb{Z}^q_+$ such that*

$$L(x, \lambda^*) \leq L(x^*, \lambda^*) \leq L(x^*, \lambda) \text{ for all } x \in \mathbb{Z}^K, \lambda \in \mathbb{Z}^q_+, \tag{1}$$

where $L(\cdot, \cdot) : \mathbb{Z}^K \times \mathbb{Z}^q_+ \to \mathbb{Z} \cup \{-\infty\}$ is given by

$$L(x, \lambda) = f(x) + \sum_{j=1}^{q} \lambda_j g_j(x) \text{ for all } x \in \mathbb{Z}^K, \lambda \in \mathbb{Z}^q_+. \tag{2}$$

Moreover, $(\lambda_1^*, \ldots, \lambda_q^*)$ satisfies

$$\lambda_j^* \cdot g_j(x^*) = 0 \text{ for all } j = 1, \ldots, q. \tag{3}$$

Proof. See Section 5.1.[11] □

[10] Hierarchies are also called laminar families in the literature.

[11] We remark that (3) is not employed in showing (ii) ⇒ (i).

In the literature, (3) is called the *complementary slackness*.

As shown in the proof, this theorem is an immediate corollary of the well-known *intersection theorem* in discrete convex analysis (Theorem 8.17 of Murota (2003a)). However, it presents a familiar form for economists, as it is similar to Karush-Kuhn-Tucker's theorem.

Remark 1. Section 8.4 of Murota (2003a) and Section 6 of Murota (1998) develop a general Lagrange duality theory for an integer problem. As demonstrated there, under the M-convexity assumption, strong duality holds, and the solutions to the primal/dual problems can be translated into the saddle point of a Lagrange function. The present approach is distinguished from this result in that we explicitly refer to constraint functions and formulate the Lagrange function incorporating them. ∎

3. OPTIMAL TAXES/SUBSIDIES IN A JOB-MATCHING MODEL

This section applies the mathematical tool in Section 2 to a market model. We consider the model of job-market interventions due to Kojima et al. (2019), which accommodates hospital-intern markets as a representative example. While the authors do not assume quasi-linearity for doctors, we assume it for analytical purposes.

3.1. Underlying Framework

Let D denote the set of **doctors** and H the set of **hospitals**. A **matching** is represented by a function $\mu : D \to \bar{H}$, where $\bar{H} \equiv H \cup \{h_0\}$, and h_0 stands for an outside option; with a slight abuse of notation, we write $\mu(h) \equiv \{d \in D : \mu(d) = h\}$ for $h \in H$. Let $\Omega \equiv D \times H$. A **salary system** is a vector $s \in \mathbb{R}^\Omega$.

Each doctor d has a **valuation function** $v_d : \bar{H} \to \mathbb{Z}$. For $s \in \mathbb{R}^\Omega$, d's **utility function** is given by

$$v_d[s](h) = v_d(h) + s_{(d,h)} \text{ for all } h \in \bar{H},$$

where $s_{(d,h_0)}$ is defined to be 0. We define d's **indirect utility function** $V_d : \mathbb{R}^\Omega \to \mathbb{R}$ by

$$V_d(s) = \max_{h \in \bar{H}} v_d[s](h) \text{ for all } s \in \mathbb{R}^\Omega.$$

Each hospital h has a **revenue function** $v_h : 2^D \to \mathbb{Z}$. We define $h's$ **demand correspondence** $X_h : \mathbb{R}^\Omega \to 2^D$ by

$$X_h(s) = \left\{A \subseteq D : v_h(A) - \sum_{d \in A} s_{(d,h)} \geq v_h(A') - \sum_{d \in A'} s_{(d,h)} \text{ for all } A' \subseteq D\right\} \quad \text{for all } s \in \mathbb{R}^\Omega.$$

We make the following assumptions on $v_h(\cdot)$, where the latter one is due to Kelso and Crawford (1982):

- **Monotonicity:** for any $A, A' \subseteq D$ with $A \subseteq A'$, we have $v_h(A) \leq v_h(A')$.
- **Gross substitutes condition:** for any $s, s' \in \mathbb{R}^\Omega$ with $s \leq s'$ and $A \in X_h(s)$, there exists $A' \in X_h(s')$ such that $\{d \in A : s_{(d,h)} = s'_{(d,h)}\} \subseteq A'$.

Proposition 1 (Fujishige and Yang 2003). *$v_h(\cdot)$ satisfies the gross substitutes condition if and only if $v_h(\cdot)$ is $M^\natural$-concave.*[12]

In real job-matching markets, the government often imposes constraints on hospitals' hiring. Kojima et al. (2019) note that "[r]estrictions on the number of hires are commonplace, and these naturally exist in the form of interval constraints." An interval constraint consists of a floor constraint (minimum number of workers hired) and a ceiling constraint (maximum number of workers hired). Formally, for each $h \in H$, let $\underline{\delta}_h, \overline{\delta}_h \in \{0, 1, \ldots, |D|\}$ with $\underline{\delta}_h \leq \overline{\delta}_h$ denote the **floor constraint** and the **ceiling constraint**, respectively. We define h's **feasibility collection** $\mathcal{F}_h$ by

$$\mathcal{F}_h = \{A \subseteq D : \underline{\delta}_h \leq |A| \leq \overline{\delta}_h\}.$$

A **transfer policy** $t = (t_h)_{h \in H} \in \mathbb{Z}^H$ specifies the government transfer to each hospital h as t_h times the number of its hires; $t_h > 0$ indicates a subsidy, and $t_h < 0$ represents a tax. For $s \in \mathbb{R}^\Omega$ and $t \in \mathbb{R}^H$, h's **post-transfer profit**

[12] We identify $v_h(\cdot)$ with $\tilde{v}_h : \mathbb{Z}^D \to \mathbb{Z} \cup \{-\infty\}$ defined by $\tilde{v}_h(x) = v_h(\{d \in D : x_d = 1\})$ if $x \in \{0, 1\}^D$ and $\tilde{v}_h(x) = -\infty$ otherwise.

function is given by[13]

$$v_h[s,t](A) = v_h(A) + |A|t_h - \sum_{d \in A} s_{(d,h)} \text{ for all } A \subseteq D, \tag{4}$$

where the summation over the empty set is defined to be 0. We define h's **profit function** $V_h : \mathbb{R}^\Omega \times \mathbb{R}^H \to \mathbb{R}$ by

$$V_h(s,t) = \max_{A \subseteq D} v_h[s,t](A) \text{ for all } (s,t) \in \mathbb{R}^\Omega \times \mathbb{R}^H.$$

3.2. Equilibrium, Desirable Matching, and Optimal Transfer Policy

The government aims to realize a desirable matching in a market equilibrium, with desirability being tested in terms of feasibility and efficiency. We introduce additional concepts to discuss this goal formally.

A triple (μ, s, t) forms an **uncompelled competitive equilibrium** if

(i) for any $d \in D$, $v_d[s](\mu(d)) \geq v_d[s](h)$ for all $h \in \bar{H}$, and

(ii) for any $h \in H$, $v_h[s,t](\mu(h)) \geq v_h[s,t](A)$ for all $A \subseteq D$.

Kojima et al. (2019) introduce this equilibrium concept and prove its existence. Importantly, the hospitals are not compelled to obey the interval constraints in this definition. We call s an **equilibrium salary**.

We say that a matching μ is **feasible** if $\mu(h) \in \mathcal{F}_h$ for all $h \in H$. For a matching μ, we define the sum of valuations/revenues by

$$\mathbf{V}(\mu) = \sum_{d \in D} v_d(\mu(d)) + \sum_{h \in H} v_h(\mu(h)).$$

We say that μ is **constrained efficient** if μ is feasible and, for any feasible matching μ', it holds that $\mathbf{V}(\mu) \geq \mathbf{V}(\mu')$. The following lemma provides a sufficient condition for a matching in an equilibrium to be constrained efficient:

[13] If $v_h(\cdot)$ satisfies the gross substitutes condition, then $v_h[s,t](\cdot)$ also satisfies the condition. This claim can be verified by appeal to basic operations of $\mathrm{M}^\natural$-concave functions; see Theorem 6.15 of Murota (2003a). Kojima et al. (2020b) consider a general form of transfer policies and provide a necessary and sufficient condition under which a policy preserves the substitutes condition of the revenue function.

Lemma 1. *Let (μ, s, t) be an uncompelled competitive equilibrium such that μ is feasible. If the pair (μ, t) satisfies*

$$\left[t_h > 0 \Longrightarrow |\mu(h)| = \underline{\delta}_h\right] \text{ and } \left[t_h < 0 \Longrightarrow |\mu(h)| = \overline{\delta}_h\right], \tag{5}$$

then μ is constrained efficient.

Proof. See Section 5.2. □

Remark 2. The converse of Lemma 1 is not true. We provide a counterexample in Section S.2 of the Supplemental Material (Yokote 2020).

Intuitively, (5) states that the government intervention is minimal; it is allowed only if the resulting hires satisfy the constraints with equality. As suggested in standard market theory, the less government intervention there is, the more social welfare the market generates. This intuition is substantiated in Lemma 1. We say that t **induces μ to be constrained efficient** if (μ, t) satisfies (5).

A transfer policy t is **optimal** if there exists a salary system s and a matching μ such that the triple (μ, s, t) forms an uncompelled competitive equilibrium, μ is feasible, and t induces μ to be constrained efficient.

3.3. Main Theorems on Optimal Transfer Policies

We characterize the optimal transfer policies and the corresponding equilibrium salaries as the minimizers of a Lagrange function. Together with Kojima et al.'s (2019) result, this characterization establishes the existence of optimal transfer policies. For each $h \in H$, we define $G_h : \mathbb{Z} \to \mathbb{Z}$ by

$$G_h(t_t) = \begin{cases} -\underline{\delta}_h \cdot t_h & \text{if } t_h \geq 0, \\ -\overline{\delta}_h \cdot t_h & \text{if } t_h < 0. \end{cases} \tag{6}$$

Fig. 1 shows a graphical description of[14] $G_h(\cdot)$.

[14] $G_h(\cdot)$ is defined on $\mathbb{Z}$, but we draw a continuous function for presentational simplicity. Note that this function has a "convex" shape, which is utilized in the proof of Theorem 3.

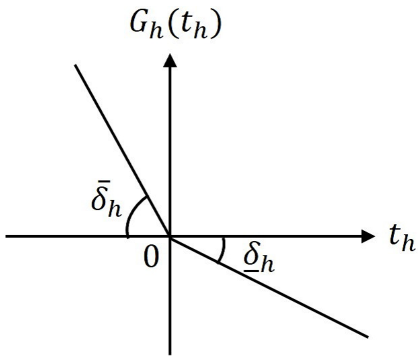


Fig. 1 Graphical description of $G_h(\cdot)$.

Set $G(t) \equiv \sum_{h \in H} G_h(t_h)$ for all $t \in \mathbb{Z}^H$. We define $F : \mathbb{Z}_+^\Omega \times \mathbb{Z}^H \to \mathbb{Z}$ by

$$F(s,t) = \sum_{d \in D} V_d(s) + \sum_{h \in H} V_h(s,t) + G(t) \text{ for all } (s,t) \in \mathbb{Z}_+^\Omega \times \mathbb{Z}^H. \quad (7)$$

Theorem 2 (Characterization of Optimal Transfer Policies and Equilibrium Salaries)**.** *Let $(s^*, t^*) \in \mathbb{Z}_+^\Omega \times \mathbb{Z}^H$. The following statements are equivalent:*

(i) (s^*, t^*) *is a solution to*

$$\min_{s \in \mathbb{Z}_+^\Omega} \min_{t \in \mathbb{Z}^H} F(s,t). \quad (8)$$

(ii) *There exists a matching μ such that the triple (μ, s^*, t^*) forms an uncompelled competitive equilibrium, μ is feasible, and t^* induces μ to be constrained efficient (i.e., (5) holds).*

Proof. See Section 5.3. □

Remark 3. Theorem 2 is silent about the existence of $(s^*, t^*) \in \mathbb{Z}_+^\Omega \times \mathbb{Z}^H$ satisfying (ii) (equivalently (i)), but this can be proven by combining existing results. First, as in the proof of Theorem 7 of Kojima et al. (2019), we consider a competitive equilibrium (a pair of a matching and a salary system) in the market where government-imposed feasibility collections are enforced (i.e., each hospital h has a revenue function $\tilde{v}_h(\cdot)$ such that $\tilde{v}_h(A) = v_h(A)$ if $A \in \mathcal{F}_h$ and $\tilde{v}_h(A) = -\infty$ otherwise). Then, we can find an integer-valued competitive salary system s^* and a corresponding matching[15] μ. Following the proof of Lemma 7 of Kojima et al. (2019), we can construct an integer-valued transfer policy t^* for which (5) holds. ∎

In the proof, we apply Theorem 1 to the problem of maximizing the sum of the agents' utilities/profits under the following constraints:

- each doctor is matched to at most one hospital, and
- each hospital's hires satisfy the floor/ceiling constraints.

As will be detailed in Section 5.3.2 (sketch of the proof), we construct three independent maximization problems, where the assumptions in Theorem 1 hold, and aggregate the Lagrange functions in each problem into a single function. The resulting function corresponds to (7).

In (7), the first and second summations represent the doctors' total utility and the hospitals' total profit, respectively, while $G(t)$ coincides with the government surplus at optimal solutions.[16] To see this point, let $h \in H$ and suppose that $t_h \geq 0$, i.e., the government subsidizes h. If h hires exactly $\underline{\delta}_h$ workers, then the government expenditure is equal to $-\underline{\delta}_h \cdot t_h = G_h(t_h)$. If h hires more than $\underline{\delta}_h$, then (5) implies that t_h is equal to 0. Namely, the government expenditure is $0 = G_h(t_h)$. A parallel argument holds for $t_h \leq 0$ (i.e., in the case of taxation).

Finally, we discuss the algorithmic aspect of minimizing $F(\cdot,\cdot)$.

Theorem 3 (L$^\natural$-Convexity of the Lagrange Function)**.** *$F(\cdot,\cdot)$ is an L$^\natural$-convex function.*

L$^\natural$-convex function minimization algorithms have been studied extensively in the literature, and their upper bounds on the number of iterations have been revealed; see Murota (2003b), Kolmogorov and Shioura (2009) or Murota and Shioura (2014).[17] As noted by Murota (2003b), complexity bounds are

[15] This statement can be proven in several ways. For example, following Candogan et al. (2016), we can translate the market into the M-convex submodular flow problem and utilize the existence of integer-valued potential (see Theorem 9.16 of Murota 2003a). Alternatively, we can adopt the techniques in Fujishige and Tamura (2007).

[16] This statement is formally proved in Claim 6 (see Section 5.3.1).

[17] Murota and Shioura (2014) demonstrate that the *exact* number of iterations can be obtained if the initial vector and the set of minimizers are known. Using this result, Murota et al. (2016) offer computational complexity results in an auction model. For a survey of algorithms, see Sections 7.3 and 8 of Murota (2016).

polynomial in the dimension of the variable (which is equal to $(|D| \times |H|) + |H|$ in our model) and the size of the effective domain. In practice, we can restrict the size of the effective domain by setting sufficiently high salaries and sufficiently low/high transfers. Together with Theorems 2 and 3, we conclude that an optimal transfer policy can be calculated via an efficient algorithm. Section S.3 of the Supplemental Material (Yokote 2020) provides an example of how this algorithm proceeds.

4. CONCLUDING REMARKS

As pointed out by Gul and Stacchetti (1999), the job-matching model subsumes the auction model as a special case. Applying our approach to the auction model, the resulting Lagrange function corresponds to the Lyapunov function of Ausubel (2006) and Sun and Yang (2009). Since existing auction algorithms for heterogeneous commodities can be recast into minimization of the Lyapunov function (see Murota et al. 2016), we can view existing auctions as a process for finding a saddle point of the Lagrange function. For a formal analysis, see Section S.4 of the Supplemental Material (Yokote 2020).

It would be possible to extend our approach to the trading network model under quasi-linearity due to Hatfield et al. (2013). Hatfield et al. (2019) prove that the full substitutes condition imposed on preferences is equivalent to $M^\natural$-concavity of the utility function,[18] which enables the application of discrete convex analysis.

5. THE APPENDICES

This section contains all the proofs.

[18] To be precise, the agents have gross substitutes and complements (see Sun and Yang 2009) or a *twisted* $M^\natural$-concave function (see Ikebe and Tamura 2015 and Section 3.5 of Murota 2016), which is essentially equivalent to an $M^\natural$-concave function.

5.1. Proof of Theorem 1

5.1.1. Preliminaries

For $x \in \mathbb{Z}^K$ and $A \subseteq K$, we define $x(A) = \sum_{k \in A} x_k$. For $X \subseteq \mathbb{R}^K$, we define the **indicator function of X**, $\delta_X : \mathbb{R}^K \to \mathbb{Z} \cup \{-\infty\}$, by

$$\delta_X(x) = \begin{cases} 0 & \text{if } x \in X, \\ -\infty & \text{otherwise.} \end{cases}$$

We say that $X \subseteq \mathbb{Z}^K$ with $X \neq \emptyset$ is an **M$^\natural$-convex set** if $\delta_X(\cdot)$ is an M$^\natural$-concave function. We say that $X \subseteq \mathbb{Z}^K$ with $X \neq \emptyset$ is an **L$^\natural$-convex set** if $\delta_X(\cdot)$ is an L$^\natural$-concave function.

For $f : \mathbb{Z}^K \to \mathbb{Z} \cup \{-\infty\}$ and $x^* \in \mathrm{dom} f$, we define the **supergradient of f at x^*** (in $\mathbb{Z}$) by

$$\partial_{\mathbb{Z}} f(x^*) = \{\hat{x} \in \mathbb{Z}^K : f(x^*) + \hat{x} \cdot (x - x^*) \geq f(x) \text{ for all } x \in \mathbb{Z}^K\}.$$

For $f : \mathbb{R}^K \to \mathbb{R} \cup \{-\infty\}$ and $x^* \in \mathrm{dom} f$, we define the **supergradient of f at x^*** (in $\mathbb{R}$) by

$$\partial_{\mathbb{R}} f(x^*) = \{\hat{x} \in \mathbb{R}^K : f(x^*) + \hat{x} \cdot (x - x^*) \geq f(x) \text{ for all } x \in \mathbb{R}^K\}.$$

Let $\mathcal{K} \subseteq 2^K \backslash \{\emptyset\}$. We define the **cone generated by** $(\mathbb{1}^A)_{A \in \mathcal{K}}$ (in $\mathbb{Z}$) by

$$cone_{\mathbb{Z}}(\mathcal{K}) = \left\{x \in \mathbb{Z}^K : x = \sum_{A \in \mathcal{K}} \lambda_A \cdot \mathbb{1}^A \text{ for some } (\lambda_A)_{A \in \mathcal{K}} \in \mathbb{Z}_+^{\mathcal{K}}\right\}.$$

We define the **cone generated by** $(\mathbb{1}^A)_{A \in \mathcal{K}}$ (in $\mathbb{R}$) by

$$cone_{\mathbb{R}}(\mathcal{K}) = \left\{x \in \mathbb{R}^K : x = \sum_{A \in \mathcal{K}} \lambda_A \cdot \mathbb{1}^A \text{ for some } (\lambda_A)_{A \in \mathcal{K}} \in \mathbb{R}_+^{\mathcal{K}}\right\}.$$

For $X \subseteq \mathbb{Z}^K$, let $\overline{X} \subseteq \mathbb{R}^K$ denote the **convex hull of X** and X° denote the **polar of X**, i.e., $X^\circ = \{y \in \mathbb{Z}^K : y \cdot x \leq 0 \text{ for all } x \in X\}$. For $X, Y \subseteq \mathbb{R}^K$, we define

$$-X = \{x \in \mathbb{R}^K : -x \in X\}, \ X - Y = \{x - y \in \mathbb{R}^K : x \in X, y \in Y\}.$$

The following theorem is essential to the proof.

Theorem 4 (Murota 2003a, Theorem 8.17 (M$^\natural$-concave intersection theorem))**.** For M$^\natural$-concave functions $f_1, f_2 : \mathbb{Z}^K \to \mathbb{Z} \cup \{-\infty\}$ and a point $x^* \in \mathrm{dom} f_1 \cap \mathrm{dom} f_2$, we have

$$f_1(x^*) + f_2(x^*) \leq f_1(x) + f_2(x) \text{ for all } x \in \mathbb{Z}^K$$

if and only if there exists $\hat{x} \in \mathbb{Z}^K$ such that

$$\hat{x} \in \partial_{\mathbb{Z}} f_1(x^*) \text{ and } -\hat{x} \in \partial_{\mathbb{Z}} f_2(x^*).$$

Theorem 5 (Murota 2003a, (5.8) (Convexity in intersection for L$^\natural$-convex sets))**.** For L$^\natural$-convex sets $X_1, X_2 \subseteq \mathbb{Z}^K$, we have

$$\overline{X_1} \cap \overline{X_2} \neq \emptyset \Longrightarrow X_1 \cap X_2 \neq \emptyset.$$

Claim 1. Let $\mathcal{K} \subseteq 2^K \backslash \{\emptyset\}$ with $\mathcal{K} \neq \emptyset$ be a hierarchy and $(\lambda_A)_{A \in \mathcal{K}} \in \mathbb{Z}^{\mathcal{K}}$. Then, the set

$$X = \{x \in \mathbb{Z}^K : x(A) \leq \lambda_A \text{ for all } A \in \mathcal{K}\} \tag{9}$$

is an M$^\natural$-convex set.

Proof. For each $A \in \mathcal{K}$, we define $\varphi_A : \mathbb{Z} \to \mathbb{Z} \cup \{-\infty\}$ by

$$\varphi_A(x) = \begin{cases} 0 & \text{if } x \leq \lambda_A, \\ -\infty & \text{otherwise.} \end{cases}$$

We define $\Phi : \mathbb{Z}^K \to \mathbb{Z} \cup \{-\infty\}$ by

$$\Phi(x) = \sum_{A \in \mathcal{K}} \varphi_A\big(x(A)\big) \text{ for all } x \in \mathbb{Z}^K.$$

Then, by Note 6.11 of Murota (2003a), $\Phi(\cdot)$ is an M$^\natural$-concave function. By Proposition 6.29 of Murota (2003a), the set of maximizers of $\Phi(\cdot)$ is an M$^\natural$-convex set. As X is the set of maximizers of $\Phi(\cdot)$, we obtain the desired condition. □

Remark 4. To develop some intuition for Claim 1, we provide a 2-dimentional case example. Let $K = \{k_1, k_2\}$ and $\mathcal{K} = \big\{\{k_1\}, \{k_2\}, \{k_1, k_2\}\big\}$. Set $\lambda_{\{k_1\}} = 5, \lambda_{\{k_2\}} = 6, \lambda_{\{k_1,k_2\}} = 8$. Then, the set X defined by (9) is described by the shaded area in Fig. 2.

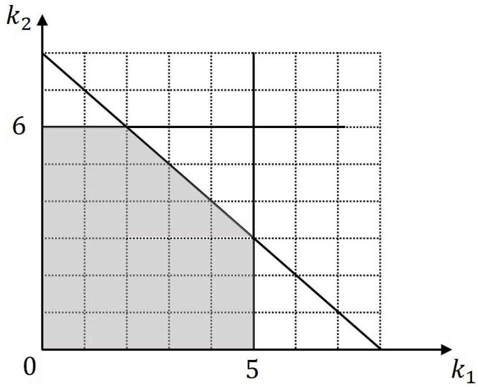


Fig. 2 Graphical descrption of X.

One easily verifies that this set is M$^\natural$-convex.

Claim 1 can also be proven by using the transformation of M$^\natural$-convexity by a network. A tree structure associated with a laminar family forms a network flow and the upper bounds $(\lambda_A)_{A \in \mathcal{K}}$ can be incorporated into the costs of flow. Then, Theorem 9.26 of Murota (2003a) establishes that X is an M$^\natural$-convex set.[19] ∎

Claim 2. Let $\mathcal{K} \subseteq 2^K \backslash \{\emptyset\}$ with $\mathcal{K} \neq \emptyset$ be a hierarchy. Then, the set

$$cone_{\mathbb{R}}(\mathcal{K}) \cap \mathbb{Z}^K$$

is an L$^\natural$-convex set.

Proof. This theorem follows from Claim 1 and the discrete conjugacy theorem (see Murota 2003a, Theorem 8.12). □

Claim 3. Let $\mathcal{K} \subseteq 2^K \backslash \{\emptyset\}$. Then,

$$\overline{cone_{\mathbb{Z}}(\mathcal{K})} = cone_{\mathbb{R}}(\mathcal{K}).$$

Proof. If $\mathcal{K} = \emptyset$, then the claim trivially holds. Suppose that $\mathcal{K} \neq \emptyset$. For each $\lambda \in \mathbb{Z}$ with $\lambda > 0$, let Δ^λ denote the λ-dimensional unit simplex.

Proof of $\subseteq$: Let $x \in \overline{cone_{\mathbb{Z}}(\mathcal{K})}$. By Carathéodory's theorem, there exist $x_j \in cone_{\mathbb{Z}}(\mathcal{K})$, $j = 1, \ldots, |K| + 1$, and $\alpha \in \Delta^{|K|}$ such that $x = \sum_j \alpha_j x_j$. For

[19] The author thanks a referee for pointing out this proof.

each $x_j \in cone_{\mathbb{Z}}(\mathcal{K})$, there exists $(\lambda_{A,j})_{A\in\mathcal{K}} \in \mathbb{Z}_+^{\mathcal{K}}$ such that $x_j = \sum_{A\in\mathcal{K}} \lambda_{A,j} \mathbb{1}^A$. Then,

$$x = \sum_{j=1}^{|K|+1} \alpha_j \sum_{A\in\mathcal{K}} \lambda_{A,j} \mathbb{1}^A = \sum_{A\in\mathcal{K}} \sum_{j=1}^{|K|+1} (\alpha_j \cdot \lambda_{A,j}) \mathbb{1}^A.$$

Hence, $x \in cone_{\mathbb{R}}(\mathcal{K})$.

Proof of $\supseteq$: Let $x \in cone_{\mathbb{R}}(\mathcal{K})$. Then, there exists $(\lambda_A)_{A\in\mathcal{K}} \in \mathbb{R}_+^{\mathcal{K}}$ such that $x = \sum_{A\in\mathcal{K}} \lambda_A \chi_A$. By $\mathbb{R}_+^{\mathcal{K}} = \overline{\mathbb{Z}_+^{\mathcal{K}}}$ and Carathéodory's theorem, there exist $(z_{A,j})_{A\in\mathcal{K}} \in \mathbb{Z}_+^{\mathcal{K}}$, $j = 1, \ldots, |\mathcal{K}| + 1$, and $\alpha \in \Delta^{|\mathcal{K}|}$ such that $\lambda_A = \sum_j \alpha_j z_{A,j}$ for all $A \in \mathcal{K}$. Then,

$$x = \sum_{A\in\mathcal{K}} \Big(\sum_{j=1}^{|\mathcal{K}|+1} \alpha_j z_{A,j} \Big) \mathbb{1}^A = \sum_{j=1}^{|\mathcal{K}|+1} \alpha_j \sum_{A\in\mathcal{K}} z_{A,j} \mathbb{1}^A.$$

Hence, $x \in \overline{cone_{\mathbb{Z}}(\mathcal{K})}$. □

Claim 4. For any hierarchy $\mathcal{K} \subseteq 2^K \backslash \{\emptyset\}$,

$$cone_{\mathbb{R}}(\mathcal{K}) \cap \mathbb{Z}^K = cone_{\mathbb{Z}}(\mathcal{K}).$$

Proof. One easily verifies that $\supseteq$ holds. We prove $\subseteq$ by induction on $|\mathcal{K}|$.

Induction base: Suppose $|\mathcal{K}| = 1$. Let $x \in cone_{\mathbb{R}}(\mathcal{K}) \cap \mathbb{Z}^K$. Assuming $\mathcal{K} = \{A\}$, $x = \lambda_A \cdot \mathbb{1}^A$ for some $\lambda_A \in \mathbb{R}_+$. Since $x \in \mathbb{Z}^K$, we have $\lambda_A \in \mathbb{Z}_+$. Hence, $x \in cone_{\mathbb{Z}}(\mathcal{K})$.

Induction step: Suppose the claim holds for $|\mathcal{K}| = t$ and we prove the claim for $|\mathcal{K}| = t + 1$, where $t \geq 1$.

Let $x \in cone_{\mathbb{R}}(\mathcal{K}) \cap \mathbb{Z}^K$ and $A' \in \mathcal{K}$. Then, there exists $(\lambda_A)_{A\in\mathcal{K}} \in \mathbb{R}_+^{\mathcal{K}}$ such that

$$x = \sum_{A\in\mathcal{K}\backslash\{A'\}} \lambda_A \cdot \mathbb{1}^A + \lambda_{A'} \cdot \mathbb{1}^{A'}.$$

This implies that

$$cone_{\mathbb{R}}(\mathcal{K}\backslash\{A'\}) \cap \big(\{x\} - cone_{\mathbb{R}}(\{A'\})\big) \neq \emptyset. \tag{10}$$

By Claim 3,

$$cone_{\mathbb{R}}(\mathcal{K}\backslash\{A'\}) = \overline{cone_{\mathbb{Z}}(\mathcal{K}\backslash\{A'\})}. \tag{11}$$

We also have

$$\begin{aligned}\{x\} - cone_{\mathbb{R}}(\{A'\}) &= \{x\} - \overline{cone_{\mathbb{Z}}(\{A'\})}\\ &= \overline{\{x\} - cone_{\mathbb{Z}}(\{A'\})},\end{aligned} \tag{12}$$

where the first equality follows from Claim 3 and the second equality follows from Proposition 3.17(4) of Murota (2003a).

By (10)-(12),

$$\overline{cone_{\mathbb{Z}}(\mathcal{K}\backslash\{A'\})} \cap \overline{\{x\} - cone_{\mathbb{Z}}(\{A'\})} \neq \emptyset.$$

As $\mathcal{K}\backslash\{A'\}$ is a hierarchy, by the induction hypothesis and Claim 2, it holds that $cone_{\mathbb{Z}}(\mathcal{K}\backslash\{A'\})$ is L$^\natural$-convex. One easily verifies that $\{x\} - cone_{\mathbb{Z}}(\{A'\})$ is also L$^\natural$-convex. By Theorem 5,

$$cone_{\mathbb{Z}}(\mathcal{K}\backslash\{A'\}) \cap \Big(\{x\} - cone_{\mathbb{Z}}(\{A'\})\Big) \neq \emptyset.$$

This implies that $x \in cone_{\mathbb{Z}}(\mathcal{K})$. □

5.1.2. Proof of (ii) ⇒ (i)

We mimic the proof of Karush-Kuhn-Tucker's theorem (for continuous settings) by Tiel (1984, p.103).

By the latter inequality in (1), $L(x^*, \lambda^*) \leq L(x^*, \lambda)$ for all $\lambda \in \mathbb{Z}^q_+$. Together with (2),

$$\sum_{j=1}^{q} \lambda_j^* g_j(x^*) \leq \sum_{j=1}^{q} \lambda_j g_j(x^*) \text{ for all } \lambda \in \mathbb{Z}^q_+,$$

$$0 \leq \sum_{j=1}^{q} (\lambda_j - \lambda_j^*) g_j(x^*) \text{ for all } \lambda \in \mathbb{Z}^q_+. \tag{13}$$

Since (13) holds for all $\lambda \in \mathbb{Z}^q_+$,

$$g_j(x^*) \geq 0 \text{ for all } j = 1, \ldots, q. \tag{14}$$

Letting $\lambda = \mathbf{0}$ in (13),

$$\sum_{j=1}^{q} \lambda_j^* g_j(x^*) \leq 0.$$

Combining this inequality with (14) yields

$$\sum_{j=1}^{q} \lambda_j^* g_j(x^*) = 0.$$

This equation and the former inequality in (1) imply

$$f(x) + \sum_{j=1}^{q} \lambda_j^* g_j(x) \le f(x^*) \text{ for all } x \in \mathbb{Z}^K.$$

This means that $f(x) \le f(x^*)$ whenever $g_j(x) \ge 0$ for all $j = 1, \ldots, q$. Together with (14), we obtain the desired condition.

5.1.3. *Proof of (i)* $\Rightarrow$ *(ii)*

By assumption, for all $j = 1, \ldots, q$, there exist $a_j \in \mathbb{Z}$ and $A_j \subseteq K$ with $A_j \neq \emptyset$ such that

$$g_j(x) = a_j - x(A_j) \text{ for all } x \in \mathbb{Z}^K.$$

Our purpose is to find $\lambda^* \in \mathbb{Z}_+^q$ such that (x^*, λ^*) satisfies (1). Suppose $A_j = A_{j'}$ for some $j, j' \in \{1, \ldots, q\}$ with $a_j \le a_{j'}$. Then, $g_{j'}(\cdot)$ is a redundant constraint function. In the proof below, we can ignore such j' by letting $\lambda_{j'}^* = 0$. Hence, w.l.o.g., we assume

$$A_j \neq A_{j'} \text{ for all } j, j' \in \{1, \ldots, q\}.$$

Set $C = \{x \in \mathbb{Z}^K : g_j(x) \ge 0 \text{ for all } j = 1, \ldots, q\}$. By Claim 1, $\delta_C(\cdot)$ is an M$^\natural$-concave function. Since x^* is a solution to the maximization problem under constraints,

$$f(x^*) + \delta_C(x^*) \ge f(x) + \delta_C(x) \text{ for all } x \in \mathbb{Z}^K.$$

By Theorem 4, there exists $\hat{x} \in \mathbb{Z}^K$ such that

$$\begin{aligned} \hat{x} &\in \partial_{\mathbb{Z}} f(x^*), \\ -\hat{x} &\in \partial_{\mathbb{Z}} \delta_C(x^*). \end{aligned} \tag{15}$$

By (15) and the definition of a supergradient, $-\hat{x} \in \partial_{\mathbb{R}} \delta_{\overline{C}}(x^*)$. As recognized in the literature on convex analysis (see, for example, Rockafellar (1970), Section 23), $-\partial_{\mathbb{R}} \delta_{\overline{C}}(x^*)$ is the normal cone to $\overline{C}$ at x^*. i.e.,

$$-\partial_{\mathbb{R}} \delta_{\overline{C}}(x^*) = \{y \in \mathbb{R}^K : y \cdot (x - x^*) \leq 0 \text{ for all } x \in \overline{C}\}.$$

By Proposition 5.2.4 of Hiriart-Urruty and Lemaréchal (2001), the normal cone is the polar of the tangent cone. This fact and (15) imply $\hat{x} \in cone_{\mathbb{R}}(\mathcal{K})$, where

$$\mathcal{K} = \left\{A \subseteq K : A = A_j \text{ for some } j \in \{1, \ldots, q\} \text{ with } g_j(x^*) = 0\right\}.$$

Since $\mathcal{K}$ is a hierarchy, by Claim 4, $\hat{x} \in cone_{\mathbb{Z}}(\mathcal{K})$. Hence, there exists $(\lambda^*_A)_{A \in \mathcal{K}} \in \mathbb{Z}^{\mathcal{K}}_+$ such that

$$\hat{x} = \sum_{A \in \mathcal{K}} \lambda^*_A \mathbb{1}^A.$$

For each $j \in \{1, \ldots, q\}$ with $g_j(x^*) = 0$, set $\lambda^*_j = \lambda^*_{A_j}$ for $A_j \in \mathcal{K}$. For each $j \in \{1, \ldots, q\}$ with $g_j(x^*) > 0$, set $\lambda^*_j = 0$.

We prove that (x^*, λ^*) is a saddle point of $L(\cdot, \cdot)$ defined by (2). We first fix λ^* and regard $L(\cdot, \lambda^*)$ as a function on $\mathbb{R}^K$. Then,

$$\begin{aligned}
\partial_{\mathbb{Z}} L(x^*, \lambda^*) &= \partial_{\mathbb{Z}} \Big(f(x^*) + \sum_{j=1}^{q} \lambda^*_j g_j(x^*) \Big) \\
&= \partial_{\mathbb{Z}} f(x^*) - \Big\{ \sum_{j=1}^{q} \lambda^*_j \mathbb{1}^{A_j} \Big\} \\
&\ni \hat{x} - \hat{x} \\
&= 0,
\end{aligned}$$

where the second equality, decomposition of superdifferential, follows from the fact that $g_j(\cdot)$, $j = 1, \ldots, q$, are affine functions.[20] This means that $L(\cdot, \lambda^*)$ is maximized at x^*.

[20] More generally, decomposition of superdifferential holds for two M$^\natural$-concave functions; see Theorem 8.35 of Murota (2003a).

Next fix x^* and regard $L(x^*, \cdot)$ as a function on $\mathbb{Z}^q_+$. As x^* satisfies the constraints, $g_j(x^*) \geq 0$ for all $j = 1, \ldots, q$. Hence, for any $\lambda \in \mathbb{Z}^q_+$,

$$\sum_{j=1}^{q} \lambda_j g_j(x^*) \geq 0. \tag{16}$$

Moreover, by the construction of λ^*,

$$\sum_{j=1}^{q} \lambda^*_j g_j(x^*) = 0. \tag{17}$$

We conclude

$$L(x^*, \lambda) = f(x^*) + \sum_{j=1}^{q} \lambda_j g_j(x^*) \geq f(x^*) + \sum_{j=1}^{q} \lambda^*_j g_j(x^*) = L(x^*, \lambda^*) \quad \text{for all } \lambda \in \mathbb{Z}^q_+,$$

where the inequality follows from (16) and (17). Finally, the complementary slackness follows from the construction of λ^*. □

5.2. Proof of Lemma 1

We prove a claim.

Claim 5. *Let μ be a matching and $s \in \mathbb{Z}^\Omega_+$. Then,*

$$\begin{aligned} &\sum_{d \in D} v_d[s](\mu(d)) + \sum_{h \in H} v_h[s,t](\mu(h)) \\ &= \sum_{d \in D} v_d(\mu(d)) + \sum_{h \in H} v_h(\mu(h)) + \sum_{h \in H} |\mu(h)| \cdot t_h. \end{aligned}$$

Proof. Since μ is a matching,

$$\sum_{d \in D} s_{(d,h)} = \sum_{h \in H} \sum_{d \in \mu(h)} s_{(d,h)}. \tag{18}$$

Then,

$$
\begin{aligned}
&\sum_{d\in D} v_d[s](\mu(d)) + \sum_{h\in H} v_h[s,t](\mu(h)) \\
&= \sum_{d\in D}\left\{v_d(\mu(d)) + s_{(d,h)}\right\} + \sum_{h\in H}\left\{v_h(\mu(h)) - \sum_{d\in\mu(h)} s_{(d,h)} + |\mu(h)|\cdot t_h\right\} \\
&= \sum_{d\in D} v_d(\mu(d)) + \sum_{h\in H} v_h(\mu(h)) + \sum_{h\in H} |\mu(h)|\cdot t_h,
\end{aligned}
$$

where the second equality follows from (18). □

We start the proof of Lemma 1. Let (μ, s, t) be an uncompelled competitive equilibrium such that μ is feasible and the pair (μ, t) satisfies (5). Let μ' be an arbitrary feasible matching. Then,

$$
\begin{aligned}
&\sum_{d\in D} v_d(\mu(d)) + \sum_{h\in H} v_h(\mu(h)) \\
&= \sum_{d\in D} v_d[s](\mu(d)) + \sum_{h\in H} v_h[s,t](\mu(h)) - \sum_{h\in H} |\mu(h)|\cdot t_h \\
&\geq \sum_{d\in D} v_d[s](\mu'(d)) + \sum_{h\in H} v_h[s,t](\mu'(h)) - \sum_{h\in H} |\mu(h)|\cdot t_h \\
&= \sum_{d\in D} v_d(\mu'(d)) + \sum_{h\in H} v_h(\mu'(h)) + \sum_{h\in H} |\mu'(h)|\cdot t_h - \sum_{h\in H} |\mu(h)|\cdot t_h \\
&= \sum_{d\in D} v_d(\mu'(d)) + \sum_{h\in H} v_h(\mu'(h)) \\
&+ \sum_{h\in H: t_h>0}\Big(|\mu'(h)| - |\mu(h)|\Big)\cdot t_h + \sum_{h\in H: t_h<0}\Big(|\mu'(h)| - |\mu(h)|\Big)\cdot t_h,
\end{aligned}
\tag{19}
$$

where the first equality follows from Claim 5, the first inequality follows from the fact that (μ, s, t) is an uncompelled competitive equilibrium, and the second equality follows from Claim 5. By (5) for (μ, t) and feasibility of μ', for any $h \in H$,

$$
\begin{aligned}
t_h > 0 &\Longrightarrow |\mu(h)| = \underline{\delta}_h \leq |\mu'(h)| \Longrightarrow \Big(|\mu'(h)| - |\mu(h)|\Big)\cdot t_h \geq 0, \\
t_h < 0 &\Longrightarrow |\mu(h)| = \overline{\delta}_h \geq |\mu'(h)| \Longrightarrow \Big(|\mu'(h)| - |\mu(h)|\Big)\cdot t_h \geq 0.
\end{aligned}
$$

It follows that

$$\sum_{h \in H: t_h > 0} \Big(|\mu'(h)| - |\mu(h)| \Big) \cdot t_h + \sum_{h \in H: t_h < 0} \Big(|\mu'(h)| - |\mu(h)| \Big) \cdot t_h \geq 0.$$

Together with (19), we conclude

$$\begin{aligned} \mathbf{V}(\mu) &= \sum_{d \in D} v_d(\mu(d)) + \sum_{h \in H} v_h(\mu(h)) \\ &\geq \sum_{d \in D} v_d(\mu'(d)) + \sum_{h \in H} v_h(\mu'(h)) \\ &= \mathbf{V}(\mu'). \end{aligned}$$

□

5.3. Proof of Theorem 2

5.3.1. Proof of (ii) ⇒ (i)

We prove a claim.

Claim 6. *Suppose that for a matching μ and $t \in \mathbb{Z}^H$,* (5) holds. *Then, for any $h \in H$,*

$$-|\mu(h)| \cdot t_h = G_h(t_h).$$

Proof. Let $h \in H$ be arbitrarily chosen. We consider three cases.
Case 1: If $t_h = 0$, then $-|\mu(h)| \cdot t_h = 0 = G_h(t_h)$.
Case 2: If $t_h > 0$, (5) implies $|\mu(h)| = \underline{\delta}_h$. Then, $-|\mu(h)| \cdot t_h = -\underline{\delta}_h \cdot t_h = G_h(t_h)$.
Case 3: If $t_h < 0$, (5) implies $|\mu(h)| = \overline{\delta}_h$. Then, $-|\mu(h)| \cdot t_h = -\overline{\delta}_h \cdot t_h = G_h(t_h)$. □

We start the proof of (ii) ⇒ (i). Let $(s^*, t^*) \in \mathbb{Z}_+^{\Omega} \times \mathbb{Z}^H$. Suppose that there exists a matching μ that satisfies the conditions in the statement of (ii). By Theorem 3, which we establish later, $L(\cdot, \cdot)$ is an L$^\natural$-convex function. By Theorem 7.14 of Murota (2003a) (known as the *L-optimality criterion*), it suffices to prove that, for any $\Psi \subseteq \Omega$ and $A \subseteq H$,

$$F(s^*, t^*) \leq F(s^* + \mathbb{1}^{\Psi}, t^* + \mathbb{1}^{A}), \tag{20}$$

$$F(s^*, t^*) \leq F(s^* - \mathbb{1}^{\Psi}, t^* - \mathbb{1}^{A}). \tag{21}$$

Let $\Psi \subseteq \Omega$ and $A \subseteq H$ be arbitrarily chosen. We prove (20) and skip the proof of (21) which can be obtained analogously.

The left-hand side of (20) is rewritten as follows:

$$\begin{aligned} F(s^*, t^*) &= \sum_{d \in D} V_d(s^*) + \sum_{h \in H} V_h(s^*, t^*) + G(t^*) \\ &= \sum_{d \in D} v_d[s^*](\mu(d)) + \sum_{h \in H} v_h[s^*, t^*](\mu(h)) + G(t^*) \\ &= \sum_{d \in D} v_d(\mu(d)) + \sum_{h \in H} v_h(\mu(h)) + \sum_{h \in H} |\mu(h)| \cdot t_h^* + \sum_{h \in H} G_h(t_h^*), \end{aligned} \tag{22}$$

where the second equality follows from the fact that (μ, s^*, t^*) is an uncompelled competitive equilibrium and the third equality follows from Claim 5. By Claim 6,

$$\sum_{h \in H} |\mu(h)| \cdot t_h^* + \sum_{h \in H} G_h(t_h^*) = 0. \tag{23}$$

Substituting (23) into (22) we obtain

$$F(s^*, t^*) = \sum_{d \in D} v_d(\mu(d)) + \sum_{h \in H} v_h(\mu(h)). \tag{24}$$

The right-hand side of (20) is rewritten as follows:

$$\begin{aligned} & F(s^* + \mathbb{1}^{\Psi}, t^* + \mathbb{1}^A) \\ &= \sum_{d \in D} V_d(s^* + \mathbb{1}^{\Psi}) + \sum_{h \in H} V_h(s^* + \mathbb{1}^{\Psi}, t^* + \mathbb{1}^A) + G(t^* + \mathbb{1}^A) \\ &\geq \sum_{d \in D} v_d[s^* + \mathbb{1}^{\Psi}](\mu(d)) + \sum_{h \in H} v_h[s^* + \mathbb{1}^{\Psi}, t^* + \mathbb{1}^A](\mu(h)) + G(t^* + \mathbb{1}^A) \\ &= \sum_{d \in D} v_d(\mu(d)) + \sum_{h \in H} v_h(\mu(h)) + \sum_{h \in H} |\mu(h)| \cdot (t_h^* + \mathbb{1}_h^A) + \sum_{h \in H} G_h(t_h^* + \mathbb{1}_h^A), \end{aligned}$$

where the inequality follows from the definition of indirect utility functions and the second equality follows from Claim 5. Together with (24), in order to prove (20), it suffices to prove that

$$\sum_{h \in H} |\mu(h)| \cdot (t_h^* + \mathbb{1}_h^A) + \sum_{h \in H} G_h(t_h^* + \mathbb{1}_h^A) \geq 0.$$

Let $h \in H$ be arbitrarily chosen. To prove the above inequality, it suffices to prove that

$$|\mu(h)| \cdot (t_h^* + \mathbb{1}_h^A) + G_h(t_h^* + \mathbb{1}_h^A) \geq 0. \tag{25}$$

Proof of (25): If $h \notin A$, then $\mathbb{1}_h^A = 0$. By Claim 6, the inequality holds with equality. Suppose that $h \in A$. We consider two cases.
Case 1: Suppose $t_h^* < 0$. By (5),

$$|\mu(h)| = \overline{\delta}_h. \tag{26}$$

By the definition of $G_h(\cdot)$,

$$G_h(t_h^*) - G_h(t_h^* + \mathbb{1}_h^A) = \overline{\delta}_h. \tag{27}$$

Then,

$$\begin{aligned}
|\mu(h)| \cdot (t_h^* + \mathbb{1}_h^A) + G_h(t_h^* + \mathbb{1}_h^A) &= -G_h(t_h^*) + |\mu(h)| \cdot \mathbb{1}_h^A + G_h(t_h^* + \mathbb{1}_h^A) \\
&= -\overline{\delta}_h + |\mu(h)| \cdot \mathbb{1}_h^A \\
&= -\overline{\delta}_h + \overline{\delta}_h \\
&= 0,
\end{aligned}$$

where the first equality follows from Claim 6, the second equality follows from (27), and the third equality follows from (26).
Case 2: Suppose $t_h^* \geq 0$. By the definition of $G_h(\cdot)$,

$$G_h(t_h^*) - G_h(t_h^* + \mathbb{1}_h^A) = \underline{\delta}_h. \tag{28}$$

Then,

$$\begin{aligned}
|\mu(h)| \cdot (t_h^* + \mathbb{1}_h^A) + G_h(t_h^* + \mathbb{1}_h^A) &= -G_h(t_h^*) + |\mu(h)| \cdot \mathbb{1}_h^A + G_h(t_h^* + \mathbb{1}_h^A) \\
&= -\underline{\delta}_h + |\mu(h)| \cdot \mathbb{1}_h^A \\
&\geq -\underline{\delta}_h + \underline{\delta}_h \\
&= 0,
\end{aligned}$$

where the first equality follows from Claim 6, the second equality follows from (28), and the inequality follows from feasibility of μ. □

5.3.2. Sketch of the proof of (i) ⇒ (ii)

We consider three maximization problems. In each problem, the objective functions are essentially the same; we maximize the sum of the agents' utilities/profits over possible matchings. The constraint functions vary across problems and are dealt with separately in order to guarantee the hierarchical structure.

Let $N = D \cup H$. Consider the domain $\{-1, 0, 1\}^{N\times\Omega}$. For $x \in \{-1, 0, 1\}^{N\times\Omega}$, $x_{(d,(d,h))} = 1$ (or -1) is intended to mean that d is matched to h. Fig. 3 visualizes the constraints in each problem.

In the first maximization problem, we set the constraint functions to represent the doctors' capacity constraints, i.e., each doctor is matched to at most one hospital (see Fig. 3 (a)). We require that the sum of the values in each shaded area be no greater than 0. If the sum is equal to 0, then we can realize a matching; d is matched to h if the value of $(d, (d, h))$ is -1, which entails that the value of $(h, (d, h))$ is 1 and h is matched to d. It is possible that the sum is equal to -1, but this case can be circumvented at optimal solutions by utilizing the complementary slackness and monotonicity of revenue functions (see the argument after (48) in the formal proof).

The second problem deals with the floor constraints (see Fig. 3 (b)), and the third deals with the ceiling constraints (see Fig. 3 (c)). These two can be handled analogously by exchanging -1 and $+1$ denoting the matching structure.

It is clear from Fig. 3 that, in each independent problem, the constraint structures form a hierarchy (they are mutually disjoint), and hence, Theorem 1 can be applied. We translate a saddle point into the min-max form of the Lagrange function. This transformation enables us to aggregate the results in each maximization problem into a single function.

5.3.3. Preliminaries for the proof of (i) ⇒ (ii)

Let $s \in \mathbb{Z}_+^\Omega$. For each $d \in D$, we define $\tilde{v}_d[s] : \mathbb{Z}^\Omega \to \mathbb{Z} \cup \{-\infty\}$ by

$$\tilde{v}_d[s](x) = \begin{cases} v_d[s](h) & \text{if } x \in \{0,1\}^\Omega \text{ and there exists } h \in H \text{ such that} \\ & x_{(d,h)} = 1,\ x_{(d,h')} = 0 \text{ for all } h' \neq h, \text{ and} \\ & x_{(d',h'')} = 0 \text{ for all } d' \neq d \text{ and } h'' \in H, \\ v_d[s](h_0) & \text{if } x = \mathbf{0}, \\ -\infty & \text{otherwise.} \end{cases}$$

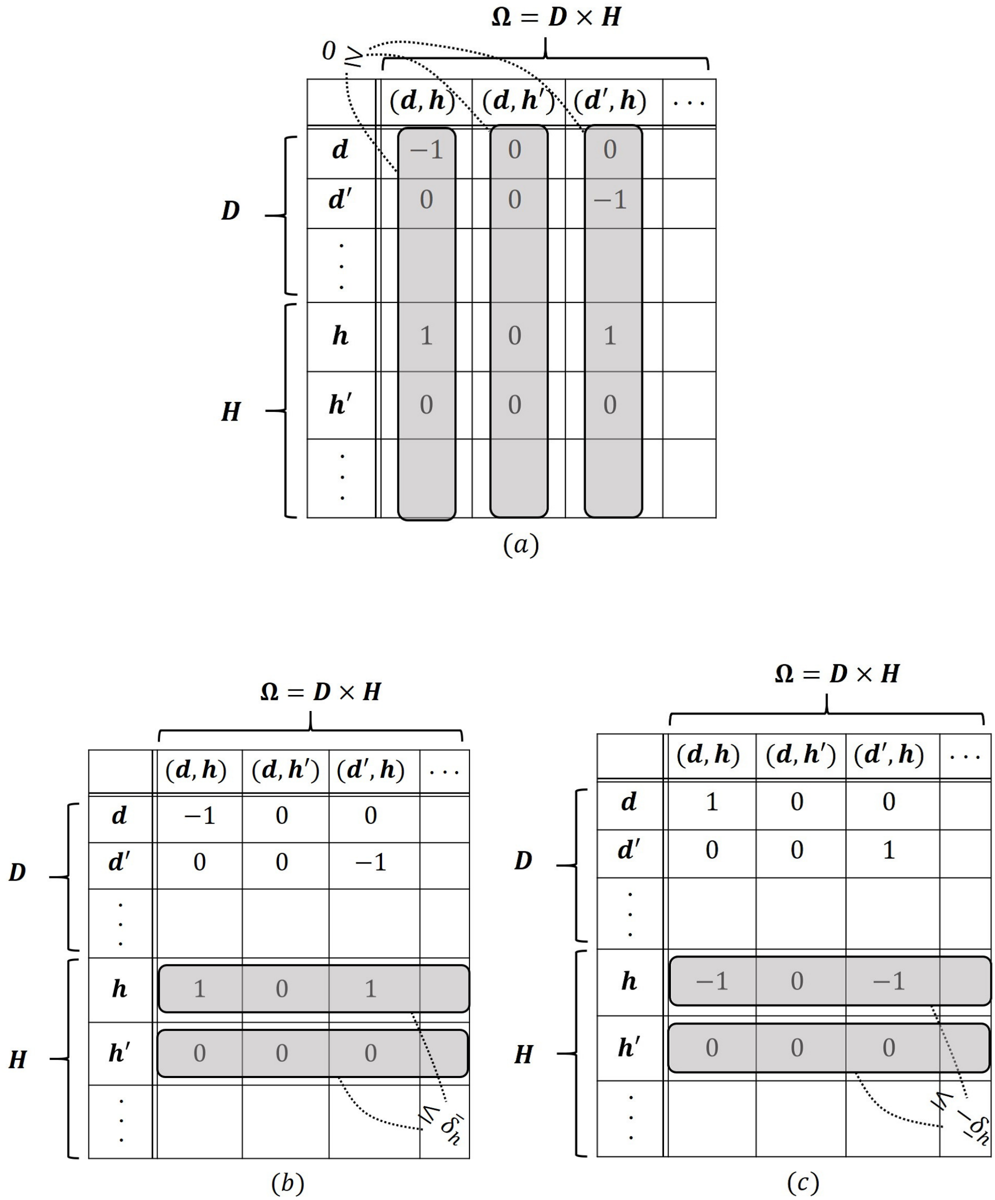


Fig. 3 Constraints in each of the three maximization problems.

Since the effective domain of $\tilde{v}_d[s](\cdot)$ consists of unit vectors, $\tilde{v}_d[s](\cdot)$ is $\mathrm{M}^\natural$-concave.

Let $s \in \mathbb{Z}_+^\Omega$ and $t \in \mathbb{Z}^H$. For each $h \in H$, we define $\tilde{v}_h[s,t] : \mathbb{Z}^\Omega \to \mathbb{Z} \cup \{-\infty\}$ by

$$\tilde{v}_h[s,t](x) = \begin{cases} v_h[s,t](\{d \in D : x_{(d,h)} = 1\}) & \text{if } x \in \{0,1\}^\Omega \text{ and} \\ & x_{(d,h')} = 0 \text{ for all } d \in D \text{ and } h' \neq h, \\ -\infty & \text{otherwise.} \end{cases}$$

By Proposition 1, $v_h(\cdot)$ is $\mathrm{M}^\natural$-concave. By Theorem 6.15 of Murota (2003a), $v_h[s,t](\cdot)$ is $\mathrm{M}^\natural$-concave. Hence, $\tilde{v}_h[s,t](\cdot)$ is also $\mathrm{M}^\natural$-concave.

For $x \in \{-1,0,1\}^{N\times\Omega}$ and $i \in N$, let $x|_i$ denote the projection of x on $\{-1,0,1\}^{\{i\}\times\Omega}$; we often identify $\{-1,0,1\}^{\{i\}\times\Omega}$ with $\{-1,0,1\}^\Omega$. For $x \in \{-1,0,1\}^{N\times\Omega}$, let x_D and x_H denote the projections of x on $\{-1,0,1\}^{D\times\Omega}$ and $\{-1,0,1\}^{H\times\Omega}$, respectively.

We formally describe the three maximization problems discussed in Section 5.3.2.

Maximization problem 1: We fix $\underline{t} \in \mathbb{Z}_+^H$ and $\bar{t} \in \mathbb{Z}_+^H$ throughout this problem. We define $f^1 : \mathbb{Z}^{N\times\Omega} \to \mathbb{Z} \cup \{-\infty\}$ by

$$f^1(x) = \sum_{d\in D} \tilde{v}_d[\mathbf{0}](-x|_d) + \sum_{h\in H} \tilde{v}_h[\mathbf{0}, \underline{t} - \bar{t}](x|_h) + G(-\bar{t}) + G(\underline{t}) \text{ for all } x \in \mathbb{Z}^{N\times\Omega}.$$

One can verify that, if $\tilde{v}_d[\mathbf{0}](\cdot)$ and $\tilde{v}_h[\mathbf{0}, \underline{t} - \bar{t}](\cdot)$ are $\mathrm{M}^\natural$-concave, then $f^1(\cdot)$ is also $\mathrm{M}^\natural$-concave.[21] Note that $\operatorname{dom} f^1 \subseteq \{-1,0,1\}^{N\times\Omega}$.

For each $\omega \in \Omega$, we define

$$g_\omega(x) = -x \cdot \mathbb{1}^{N\times\{\omega\}} \text{ for all } x \in \mathbb{Z}^{N\times\Omega}.$$

We consider the following problem:

$$\max_{x\in\{-1,0,1\}^{N\times\Omega}} f^1(x) \text{ subject to } g_\omega(x) \geq 0 \text{ for all } \omega \in \Omega. \tag{M1}$$

[21] Fujishige and Tamura (2007) utilize this preservation of $\mathrm{M}^\natural$-concavity in the analysis of two-sided markets (see equation (15) therein).

Let x^* be a solution to (M1). Then, by Theorem 1, there exists $s^* \in \mathbb{Z}_+^\Omega$ such that[22]

$$\begin{aligned}
&L(x^*, s^*) \\
&= \min_{s\in\mathbb{Z}_+^\Omega} \max_{x\in\{-1,0,1\}^{N\times\Omega}} L(x,s) \\
&= \min_{s\in\mathbb{Z}_+^\Omega} \max_{x\in\{-1,0,1\}^{N\times\Omega}} \Big\{ \sum_{d\in D} \tilde{v}_d[\mathbf{0}](-x|_d) + \sum_{h\in H} \tilde{v}_h[\mathbf{0}, \underline{t} - \bar{t}](x|_h) \\
&+ G(-\bar{t}) + G(\underline{t}) + \sum_{\omega\in\Omega} s_\omega g_\omega(x) \Big\} \\
&= \min_{s\in\mathbb{Z}_+^\Omega} \max_{x\in\{-1,0,1\}^{N\times\Omega}} \Big\{ \sum_{d\in D} \tilde{v}_d[\mathbf{0}](-x|_d) + \sum_{h\in H} \tilde{v}_h[\mathbf{0}, \underline{t} - \bar{t}](x|_h) \\
&+ G(-\bar{t}) + G(\underline{t}) - \sum_{\omega\in\Omega} s_\omega \sum_{i\in N} (x|_i)_\omega \Big\} \\
&= \min_{s\in\mathbb{Z}_+^\Omega} \max_{x\in\{-1,0,1\}^{N\times\Omega}} \Big\{ \sum_{d\in D} \big(\tilde{v}_d[\mathbf{0}](-x|_d) - s\cdot x|_d\big) + \sum_{h\in H} \big(\tilde{v}_h[\mathbf{0}, \underline{t} - \bar{t}](x|_h) - s\cdot x|_h\big) \\
&+ G(-\bar{t}) + G(\underline{t}) \Big\} \\
&= \min_{s\in\mathbb{Z}_+^\Omega} \max_{x\in\{0,1\}^{N\times\Omega}} \Big\{ \sum_{d\in D} \tilde{v}_d[s](x|_d) + \sum_{h\in H} \tilde{v}_h[s, \underline{t} - \bar{t}](x|_h) + G(-\bar{t}) + G(\underline{t}) \Big\}.
\end{aligned}$$

Note that, to derive the last equality, we change the domain for choosing $x|_d$ from $\{0,-1\}^{\{d\}\times\Omega}$ to $\{0,1\}^{\{d\}\times\Omega}$. Together with the complementary slackness, we obtain the following claim:

Claim 7. *Let $\underline{t} \in \mathbb{Z}_+^H$ and $\bar{t} \in \mathbb{Z}_+^H$. Then, for $x^* \in \{0,1\}^{N\times\Omega}$, the following statements are equivalent:*

- *$(-x_D^*, x_H^*)$ is a solution to (M1).*
- *There exists $s^* \in \mathbb{Z}_+^\Omega$ such that (x^*, s^*) is a solution to*

  $$\min_{s\in\mathbb{Z}_+^\Omega} \max_{x\in\{0,1\}^{N\times\Omega}} \Big\{ \sum_{d\in D} \tilde{v}_d[s](x|_d) + \sum_{h\in H} \tilde{v}_h[s, \underline{t} - \bar{t}](x|_h) + G(-\bar{t}) + G(\underline{t}) \Big\}.$$

  Moreover, s^* satisfies

$$s_\omega^* \cdot g_\omega(x^*) = 0 \text{ for all } \omega \in \Omega. \tag{29}$$

[22] Below the first equality follows because $L(x^*, s^*) = \min_s L(x^*, s) \le \max_x \min_s L(x,s) \le \min_s \max_x L(x,s) \le \max_x L(x, s^*) = L(x^*, s^*)$.

Maximization problem 2: We fix $s \in \mathbb{Z}_+^\Omega$ and $\underline{t} \in \mathbb{Z}_+^H$ throughout this problem. We define $f^2 : \mathbb{Z}^{N\times\Omega} \to \mathbb{Z} \cup \{-\infty\}$ by

$$f^2(x) = \sum_{d\in D} \tilde{v}_d[s](x|_d) + \sum_{h\in H} \tilde{v}_h[s,\underline{t}](x|_h) + G(\underline{t}) \text{ for all } x \in \mathbb{Z}^{N\times\Omega}.$$

For each $h \in H$, we define

$$\overline{g}_h(x) = \overline{\delta}_h - x \cdot \mathbb{1}^{\{h\}\times\Omega} \text{ for all } x \in \mathbb{Z}^{N\times\Omega}.$$

We consider the following problem:

$$\max_{x\in\{0,1\}^\Omega} f^2(x) \text{ subject to } \overline{g}_h(x) \geq 0 \text{ for all } h \in H. \tag{M2}$$

Let x^* be a solution to (M2). Then, by Theorem 1, there exists $\overline{t}^*$ such that

$$\begin{aligned}
&L(x^*,\overline{t}^*)\\
&= \min_{\overline{t}\in\mathbb{Z}_+^H} \max_{x\in\{0,1\}^{N\times\Omega}} L(x,\overline{t})\\
&= \min_{\overline{t}\in\mathbb{Z}_+^H} \max_{x\in\{0,1\}^{N\times\Omega}} \Big\{\sum_{d\in D} \tilde{v}_d[s](x|_d) + \sum_{h\in H} \tilde{v}_h[s,\underline{t}](x|_h) + G(\underline{t}) + \sum_{h\in H} \overline{t}_h \overline{g}_h(x)\Big\}\\
&= \min_{\overline{t}\in\mathbb{Z}_+^H} \max_{x\in\{0,1\}^{N\times\Omega}} \Big\{\sum_{d\in D} \tilde{v}_d[s](x|_d) + \sum_{h\in H} \big(\tilde{v}_h[s,\underline{t}](x|_h) - \overline{t}_h \cdot x \cdot \mathbb{1}^{\{h\}\times\Omega}\big)\\
&+ G(\underline{t}) + \sum_{h\in H} \overline{\delta}_h \cdot \overline{t}_h\Big\}\\
&= \min_{\overline{t}\in\mathbb{Z}_+^H} \max_{x\in\{0,1\}^{N\times\Omega}} \Big\{\sum_{d\in D} \tilde{v}_d[s](x|_d) + \sum_{h\in H} \tilde{v}_h[s,\underline{t}-\overline{t}](x|_h) + G(\underline{t}) + G(-\overline{t})\Big\}.
\end{aligned}$$

Together with the complementary slackness, we obtain the following claim:

Claim 8. *Let $s \in \mathbb{Z}_+^\Omega$ and $\underline{t} \in \mathbb{Z}_+^H$. Then, for $x^* \in \{0,1\}^{N\times\Omega}$, the following statements are equivalent:*

- *x^* is a solution to (M2).*
- *There exists $\overline{t}^*$ such that $(x^*,\overline{t}^*)$ is a solution to*

  $$\min_{\overline{t}\in\mathbb{Z}_+^H} \max_{x\in\{0,1\}^{N\times\Omega}} \Big\{\sum_{d\in D} \tilde{v}_d[s](x|_d) + \sum_{h\in H} \tilde{v}_h[s,\underline{t}-\overline{t}](x|_h) + G(\underline{t}) + G(-\overline{t})\Big\}.$$

  Moreover, $\overline{t}^*$ satisfies

  $$\overline{t}_h^* \cdot \overline{g}_h(x^*) = 0 \text{ for all } h \in H. \tag{30}$$

Maximization problem 3: We fix $s \in \mathbb{Z}_+^{\Omega}$ and $\bar{t} \in \mathbb{Z}_+^{H}$ throughout this problem. We define $f^3 : \mathbb{Z}^{N\times\Omega} \to \mathbb{Z} \cup \{-\infty\}$ by[23]

$$f^3(x) = \sum_{d\in D} \tilde{v}_d[s](x|_d) + \sum_{h\in H} \tilde{v}_h[s, -\bar{t}](-x|_h) + G(-\bar{t}) \text{ for all } x \in \mathbb{Z}^{N\times\Omega}.$$

For each $h \in H$, we define

$$\underline{g}_h(x) = -\underline{\delta}_h - x \cdot \mathbb{1}^{\{h\}\times\Omega} \text{ for all } x \in \mathbb{Z}^{N\times\Omega}.$$

We consider the following problem:

$$\max_{x\in\{-1,0,1\}^{N\times\Omega}} f^3(x) \text{ subject to } \underline{g}_h(x) \geq 0 \text{ for all } h \in H. \tag{M3}$$

Let x^* be a solution to (M3). Then, by Theorem 1, there exists $\underline{t}^*$ such that

$$\begin{aligned}
&L(x^*, \underline{t}^*) \\
&= \min_{\underline{t}\in\mathbb{Z}_+^H} \max_{x\in\{-1,0,1\}^{N\times\Omega}} L(x, \underline{t}) \\
&= \min_{\underline{t}\in\mathbb{Z}_+^H} \max_{x\in\{-1,0,1\}^{N\times\Omega}} \Big\{ \sum_{d\in D} \tilde{v}_d[s](x|_d) + \sum_{h\in H} \tilde{v}_h[s, -\bar{t}](-x|_h) \\
&+ G(-\bar{t}) + \sum_{h\in H} \underline{t}_h \underline{g}_h(x) \Big\} \\
&= \min_{\underline{t}\in\mathbb{Z}_+^H} \max_{x\in\{-1,0,1\}^{N\times\Omega}} \Big\{ \sum_{d\in D} \tilde{v}_d[s](x|_d) + \sum_{h\in H} \big(\tilde{v}_h[s, -\bar{t}](-x|_h) - \underline{t}_h \cdot x \cdot \mathbb{1}^{\{h\}\times\Omega} \big) \\
&+ G(-\bar{t}) - \sum_{h\in H} \underline{\delta}_h \cdot \underline{t}_h \Big\} \\
&= \min_{\underline{t}\in\mathbb{Z}_+^H} \max_{x\in\{0,1\}^{N\times\Omega}} \Big\{ \sum_{d\in D} \tilde{v}_d[s](x|_d) + \sum_{h\in H} \tilde{v}_h[s, \underline{t} - \bar{t}](x|_h) + G(-\bar{t}) + G(\underline{t}) \Big\}.
\end{aligned}$$

Note that, to derive the last equality, we change the domain for choosing $x|_h$ from $\{0, -1\}^{\{h\}\times\Omega}$ to $\{0, 1\}^{\{h\}\times\Omega}$. Together with the complementary slackness, we obtain the following claim:

Claim 9. *Let $s \in \mathbb{Z}_+^{\Omega}$ and $\bar{t} \in \mathbb{Z}_+^{H}$. Then, for $x^* \in \{0, 1\}^{N\times\Omega}$, the following statements are equivalent:*

---

[23] It is known that if a function $f(x)$ is M$^\natural$-concave in x, then $f(-x)$ is also M$^\natural$-concave in x; see Theorem 6.15 of Murota (2003a). Hence, $f^3(x)$ is M$^\natural$-concave in x.

- *$(x^*_D, -x^*_H)$ is a solution to (M3).*
- *There exists $\underline{t}^* \in \mathbb{Z}^H_+$ such that $(x^*, \underline{t}^*)$ is a solution to*

$$\min_{\underline{t}\in\mathbb{Z}^H_+} \max_{x\in\{0,1\}^{N\times\Omega}} \Big\{ \sum_{d\in D} \tilde{v}_d[s](x|_d) + \sum_{h\in H} \tilde{v}_h[s, \underline{t} - \bar{t}](x|_h) + G(-\bar{t}) + G(\underline{t}) \Big\}.$$

Moreover, $\underline{t}^*$ satisfies

$$\underline{t}^*_h \cdot \underline{g}_h(x^*) = 0 \text{ for all } h \in H. \tag{31}$$

5.3.4. *Proof of (i) ⇒ (ii)*

We define $\hat{F} : \mathbb{Z}^H \times \mathbb{Z}^\Omega_+ \times \mathbb{Z}^{N\times\Omega} \to \mathbb{Z} \cup \{-\infty\}$ and $\check{F} : \mathbb{Z}^H_+ \times \mathbb{Z}^H_+ \times \mathbb{Z}^\Omega_+ \times \mathbb{Z}^{N\times\Omega} \to \mathbb{Z} \cup \{-\infty\}$ by

$$\begin{aligned}
\hat{F}(t, s, x) &= \sum_{d\in D} \tilde{v}_d[s](x|_d) + \sum_{h\in H} \tilde{v}_h[s, t](x|_h) + G(t) \\
&\quad \text{for all } (t, s, x) \in \mathbb{Z}^H \times \mathbb{Z}^\Omega_+ \times \mathbb{Z}^{N\times\Omega}, \\
\check{F}(\bar{t}, \underline{t}, s, x) &= \sum_{d\in D} \tilde{v}_d[s](x|_d) + \sum_{h\in H} \tilde{v}_h[s, \underline{t} - \bar{t}](x|_h) + G(-\bar{t}) + G(\underline{t}) \\
&\quad \text{for all } (\bar{t}, \underline{t}, s, x) \in \mathbb{Z}^H_+ \times \mathbb{Z}^H_+ \times \mathbb{Z}^\Omega_+ \times \mathbb{Z}^{N\times\Omega}.
\end{aligned}$$

We prove two claims.

Claim 10. *Suppose that there exists a solution to*

$$\min_{t\in\mathbb{Z}^H} \min_{s\in\mathbb{Z}^\Omega} \max_{x\in\{0,1\}^{N\times\Omega}} \hat{F}(t, s, x). \tag{32}$$

Then, there also exists a solution to

$$\min_{\bar{t}\in\mathbb{Z}^H_+} \min_{\underline{t}\in\mathbb{Z}^H_+} \min_{s\in\mathbb{Z}^\Omega} \max_{x\in\{0,1\}^{N\times\Omega}} \check{F}(\bar{t}, \underline{t}, s, x), \tag{33}$$

and (32)=(33). In particular, for a solution (t^*, s^*, x^*) to (32), there exist $\bar{t}^* \in \mathbb{Z}^H_+$ and $\underline{t}^* \in \mathbb{Z}^H_+$ such that $\underline{t}^* - \bar{t}^* = t^*$ and $(\bar{t}^*, \underline{t}^*, s^*, x^*)$ is a solution to (33).

Proof. **Proof of (32) ≤ (33).** Let $(\bar{t}, \underline{t}, s) \in \mathbb{Z}_+^H \times \mathbb{Z}_+^H \times \mathbb{Z}_+^\Omega$ be arbitrarily chosen. It suffices to prove that there exists $t \in \mathbb{Z}^H$ such that

$$\max_{x \in \{0,1\}^{N \times \Omega}} \hat{F}(t, s, x) \leq \max_{x \in \{0,1\}^{N \times \Omega}} \check{F}(\bar{t}, \underline{t}, s, x). \tag{34}$$

Suppose that there exists $h' \in H$ with $\bar{t}_{h'} > 0$ and $\underline{t}_{h'} > 0$. We define $\bar{t}^* \in \mathbb{Z}_+^H$ and $\underline{t}^* \in \mathbb{Z}_+^H$ as follows:

$$\bar{t}^*_{h'} = \bar{t}_{h'} - 1,\ \underline{t}^*_{h'} = \underline{t}_{h'} - 1,\ \bar{t}^* = \left((\bar{t}_h)_{h \neq h'}, \bar{t}^*_{h'}\right),\ \underline{t}^* = \left((\underline{t}_h)_{h \neq h'}, \underline{t}^*_{h'}\right).$$

Then,

$$G(-\bar{t}^*) + G(\underline{t}^*) = \left\{G(-\bar{t}) - \bar{\delta}_{h'}\right\} + \left\{G(\underline{t}) + \underline{\delta}_{h'}\right\} \leq G(-\bar{t}) + G(\underline{t}),$$

where the inequality follows from $\underline{\delta}_{h'} \leq \bar{\delta}_{h'}$. Moreover, since $\underline{t} - \bar{t} = \underline{t}^* - \bar{t}^*$, we obtain

$$\check{F}(\bar{t}^*, \underline{t}^*, s, x) \leq \check{F}(\bar{t}, \underline{t}, s, x) \text{ for all } x \in \{0,1\}^{N \times \Omega}.$$

This inequality implies

$$\max_{x \in \{0,1\}^{N \times \Omega}} \check{F}(\bar{t}^*, \underline{t}^*, s, x) \leq \max_{x \in \{0,1\}^{N \times \Omega}} \check{F}(\bar{t}, \underline{t}, s, x).$$

By repeating the above procedure for all $h \in H$ with $\bar{t}_h > 0$ and $\underline{t}_h > 0$, we can construct $\bar{t}^{**} \in \mathbb{Z}_+^H$ and $\underline{t}^{**} \in \mathbb{Z}_+^H$ such that

$$\max_{x \in \{0,1\}^{N \times \Omega}} \check{F}(\bar{t}^{**}, \underline{t}^{**}, s, x) \leq \max_{x \in \{0,1\}^{N \times \Omega}} \check{F}(\bar{t}, \underline{t}, s, x), \text{ and} \tag{35}$$

$$\text{for any } h \in H,\ \bar{t}^{**}_h = 0 \text{ or } \underline{t}^{**}_h = 0 \text{ holds.} \tag{36}$$

By (36), for any $h \in H$,

$$G_h(-\bar{t}^{**}_h) + G_h(\underline{t}^{**}_h) = \begin{cases} G_h(-\bar{t}^{**}_h) + 0 & \text{if } \bar{t}^{**}_h > 0, \\ 0 + G_h(\underline{t}^{**}_h) & \text{if } \bar{t}^{**}_h = 0. \end{cases} \tag{37}$$

We define $t \in \mathbb{Z}^H$ as follows: for each $h \in H$,

$$t_h = \begin{cases} -\bar{t}^{**}_h & \text{if } \bar{t}^{**}_h > 0, \\ \underline{t}^{**}_h & \text{if } \bar{t}^{**}_h = 0. \end{cases} \tag{38}$$

Then, (37) and (38) imply that $G_h(-\bar{t}_h^{**}) + G_h(\underline{t}_h^{**}) = G_h(t_h)$. Taking the sum over $h \in H$, we obtain $G(-\bar{t}^{**}) + G(\underline{t}^{**}) = G(t)$. Moreover, by (38) and (36), for any $h \in H$,

$$\begin{aligned} t_h &= \begin{cases} 0 - \bar{t}_h^{**} & \text{if } \bar{t}_h^{**} > 0, \\ \underline{t}_h^{**} - 0 & \text{if } \bar{t}_h^{**} = 0, \end{cases} \\ &= \begin{cases} \underline{t}_h^{**} - \bar{t}_h^{**} & \text{if } \bar{t}_h^{**} > 0, \\ \underline{t}_h^{**} - \bar{t}_h^{**} & \text{if } \bar{t}_h^{**} = 0, \end{cases} \\ &= \underline{t}_h^{**} - \bar{t}_h^{**}. \end{aligned}$$

It follows that $t = \underline{t}^{**} - \bar{t}^{**}$. Together with $G(-\bar{t}^{**}) + G(\underline{t}^{**}) = G(t)$,

$$\hat{F}(t, s, x) = \check{F}(\underline{t}^{**}, \bar{t}^{**}, s, x) \text{ for all } x \in \{0,1\}^{N\times\Omega}.$$

This equation implies

$$\max_{x\in\{0,1\}^{N\times\Omega}} \hat{F}(t, s, x) = \max_{x\in\{0,1\}^{N\times\Omega}} \check{F}(\underline{t}^{**}, \bar{t}^{**}, s, x).$$

Together with (35), we obtain (34).

Proof of (32) ≥ (33): Let (t^*, s^*, x^*) be a solution to (32). We define $\bar{t}^* = (\bar{t}_h^*)_{h\in H}$ and $\underline{t}^* = (\underline{t}_h^*)_{h\in H}$ as follows: for each $h \in H$,

$$\bar{t}_h^* = \begin{cases} -t_h^* & \text{if } t_h^* \leq 0, \\ 0 & \text{if } t_h^* > 0, \end{cases} \quad \underline{t}_h^* = \begin{cases} 0 & \text{if } t_h^* \leq 0, \\ t_h^* & \text{if } t_h^* > 0. \end{cases} \tag{39}$$

Then, for any $h \in H$,

$$G_h(-\bar{t}_h^*) = \begin{cases} G_h(t_h^*) & \text{if } t_h^* \leq 0, \\ 0 & \text{if } t_h^* > 0, \end{cases} \quad G_h(\underline{t}_h^*) = \begin{cases} 0 & \text{if } t_h^* \leq 0, \\ G_h(t_h^*) & \text{if } t_h^* > 0. \end{cases}$$

These equations imply $G_h(-\bar{t}_h^*) + G_h(\underline{t}_h^*) = G_h(t_h^*)$. Taking the sum over $h \in H$, we obtain $G(-\bar{t}^*) + G(\underline{t}^*) = G(t^*)$. Moreover, since $\underline{t}^* - \bar{t}^* = t^*$ by (39),

$$\hat{F}(t^*, s^*, x) = \check{F}(\bar{t}^*, \underline{t}^*, s^*, x) \text{ for all } x \in \{0,1\}^{N\times\Omega}. \tag{40}$$

This equation implies

$$(32) = \hat{F}(t^*, s^*, x^*) = \max_{x \in \{0,1\}^{N \times \Omega}} \hat{F}(t^*, s^*, x) = \max_{x \in \{0,1\}^{N \times \Omega}} \check{F}(\bar{t}^*, \underline{t}^*, s^*, x) \geq (33),$$

where the inequality holds because (33) takes the minimum value over $(\bar{t}, \underline{t}, s) \in \mathbb{Z}_+^H \times \mathbb{Z}_+^H \times \mathbb{Z}_+^\Omega$.

As proven in the first half of the proof, (33) $\geq$ (32). Together with the above inequality, (33) $=$ (32). Moreover, by (40), $(\bar{t}^*, \underline{t}^*, s^*, x^*)$ is a solution to (33). Finally, as mentioned above, (39) implies $\underline{t}^* - \bar{t}^* = t^*$ and hence the last part of the statement follows. □

Claim 11. *Let* $(s, t) \in \mathbb{Z}_+^\Omega \times \mathbb{Z}^H$. *Then,*

$$\max_{x \in \{0,1\}^{N \times \Omega}} \hat{F}(t, s, x) = F(s, t) \equiv \sum_{d \in D} V_d(s) + \sum_{h \in H} V_h(s, t) + G(t).$$

Proof. This equation immediately follows from the definition of indirect utility functions. □

We start the proof of (i) $\Rightarrow$ (ii). Let (s^*, t^*) be a solution to (8). Let x^* be a solution to

$$\max_{x \in \{0,1\}^{N \times \Omega}} \hat{F}(t^*, s^*, x). \tag{41}$$

Then,

$$\begin{aligned}
\hat{F}(t^*, s^*, x^*) &= \max_{x \in \{0,1\}^{N \times \Omega}} \hat{F}(t^*, s^*, x) \\
&= F(s^*, t^*) \\
&= \min_{t \in \mathbb{Z}^H} \min_{s \in \mathbb{Z}_+^\Omega} F(s, t) \\
&= \min_{t \in \mathbb{Z}^H} \min_{s \in \mathbb{Z}_+^\Omega} \max_{x \in \{0,1\}^{N \times \Omega}} \hat{F}(t, s, x),
\end{aligned}$$

where the scond and fourth equalities follow from Claim 11. It follows that (t^*, s^*, x^*) is a solution to (32). By Claim 10, setting $\bar{t}^*$ and $\underline{t}^*$ as in (39), $(\bar{t}^*, \underline{t}^*, s^*, x^*)$ is a solution to (33) and satisfies

$$\underline{t}^* - \bar{t}^* = t^*. \tag{42}$$

Then,[24]

$$
\begin{aligned}
& \check{F}(\bar{t}^*, \underline{t}^*, s^*, x^*) && (43)\\
= & \min_{\bar{t}\in\mathbb{Z}_+^H} \min_{\underline{t}\in\mathbb{Z}_+^H} \min_{s\in\mathbb{Z}_+^\Omega} \max_{x\in\{0,1\}^{N\times\Omega}} \check{F}(\bar{t}, \underline{t}, s, x) && (44)\\
= & \min_{s\in\mathbb{Z}_+^\Omega} \max_{x\in\{0,1\}^{N\times\Omega}} \check{F}(\bar{t}^*, \underline{t}^*, s, x) && (45)\\
= & \min_{\bar{t}\in\mathbb{Z}_+^H} \max_{x\in\{0,1\}^{N\times\Omega}} \check{F}(\bar{t}, \underline{t}^*, s^*, x) && (46)\\
= & \min_{\underline{t}\in\mathbb{Z}_+^H} \max_{x\in\{0,1\}^{N\times\Omega}} \check{F}(\bar{t}^*, \underline{t}, s^*, x) && (47)\\
= & \max_{x\in\{0,1\}^{N\times\Omega}} \check{F}(\bar{t}^*, \underline{t}^*, s^*, x). && (48)
\end{aligned}
$$

By (43)=(45) and Claim 7, $(-x_D^*, x_H^*)$ is a solution to (M1). Suppose that $g_\omega(-x_D^*, x_H^*) > 0$ for some $\omega \in \Omega$ and let $(d', h') \equiv \omega$. Then, $x^*_{(d',(d',h'))} = 1$ and $x^*_{(h',(d',h'))} = 0$. By (29), $s^*_{(d',h')} = 0$. Together with monotonicity of $v_{h'}(\cdot)$, replacing the $(h', (d', h'))$-th coordinate of x^* with 1 yields a solution to (41). Hence, w.l.o.g., we assume that $(-x_D^*, x_H^*)$ is a solution to (M1) satisfying all the constraints with equality. Namely, for each $\omega \in \Omega$,

$$0 = g_\omega(-x_D^*, x_H^*) = (-x_D^*, x_H^*) \cdot \mathbb{1}^{N\times\{\omega\}} = -\sum_{d\in D}(x^*|_d)_\omega + \sum_{h\in H}(x^*|_h)_\omega.$$

Then, defining $\mu : D \to H$ by

$$\mu(d) = h \iff x^*_{(d,(d,h))} = 1 \text{ for all } d \in D \text{ and } h \in H,$$

μ forms a matching.

By (43)=(46) and Claim 8, x^* is a solution to (M2). Hence, x^* satisfies the ceiling constraints. By (43)=(47) and Claim 9, $(x_D^*, -x_H^*)$ is a solution to (M3). Hence, x^* satisfies the floor constraints. It follows that μ is feasible. Moreover, by (42) and (43)=(48), x^* maximizes every agent's utilities/profits under the salary system s^* and the transfer system t^*. We conclude that (μ, s^*, t^*) forms an uncompelled competitive equilibrium.

[24] In the transformation below, one can verify the following equations: (44)=(45), (44)=(46), (44)=(47), and (44)=(48).

It remains to prove that (5) holds. If $t^*_h < 0$, since $\bar{t}^*_h$ is constructed by (39), we obtain $\bar{t}^*_h > 0$. By (30), for any $h \in H$,

$$\begin{aligned} \bar{g}_h(x^*) = \bar{\delta}_h - x^* \cdot \mathbb{1}^{\{h\}\times\Omega} = 0, \\ |\{d \in D : x^*_{(h,(d,h))} = 1\}| = \bar{\delta}_h, \\ |\mu(h)| = \bar{\delta}_h. \end{aligned}$$

We can deal with the case of $t^*_h > 0$ in a paralell manner by using (31). □

5.3.5. Remark on the proof of Theorem 2

Referees of the journal raised a concern about the length of the proof of Theorem 2, especially that of the implication (i) ⇒ (ii), and proposed an alternative short proof.[25] Here we explain their ideas, as well as the advantage of our proof.

A referee pointed out that the proof can be shortened by using the "L-convex intersection theorem," which states the optimality condition of the sum of two L$^\natural$-convex functions.[26] Consider two functions g_1 and g_2 defined over salaries s and transfers t such that

$$g_1(s,t) = \sum_{d \in D} V_d(s) + G(t), \;\; g_2(s,t) = \sum_{h \in H} V_h(s,t).$$

Then, $F(s,t) = g_1(s,t) + g_2(s,t)$. Applying the L-convex intersection theorem to the two functions, we can prove the existence of a pair of salary/transfer systems satisfying the conditions stated in (ii).[27]

We fully admit that the referee's proof is shorter and easier to understand. Meanwhile, our proof appears to be more informative about how to construct

[25] The author deeply appreciates the two referees for sending an alternative proof.

[26] The L-convex intersection theorem is a cousin of the M-convex intersection theorem (see Theorem 8.17 of Murota 2003a), replacing M$^\natural$-convex functions in the latter theorem with L$^\natural$-convex functions. This theorem is an immediate corollary of the L-separation theorem (see Theorem 8.16 of Murota 2003a).

[27] Another referee suggested adopting the techniques in Ausubel (2006) and Sun and Yang (2009). In their proofs, given a minimizer p^* of the Lyapunov function, any efficient allocation is supported by p^* as an equilibrium (a related claim is formerly established in Lemma 6 of Gul and Stacchetti (1999)). A parallel claim holds in our model: given a minimizer (s^*, t^*), any constrained efficient matching is supported by (s^*, t^*) as an equilibrium (this claim can be verified from the choice of x^* after Claim 11 in the proof). It might be possible to prove this claim without relying on the saddle-point approach.

the function $L(\cdot)$ (and the corresponding algorithm). As will be detailed in the proof, $L(\cdot)$ is derived by setting constraints in a visually simple way (see Fig. 3 in Section 5.3.2) and applying the discrete saddle-point theorem. Providing a recipe seems to be important in extending our approach to other socio-economic problems, such as those mentioned in Section 1.

5.4. Proof of Theorem 3

Let $d \in D$ and $h \in H$ be arbitrarily chosen. As the sum of L$^\natural$-convex functions is L$^\natural$-convex (see Theorem 7.11 of Murota 2003a), it suffices to prove the following:[28]

$$V_d : \mathbb{Z}_+^\Omega \to \mathbb{Z} \text{ is L}^\natural\text{-convex.} \tag{49}$$

$$V_h : \mathbb{Z}_+^\Omega \times \mathbb{Z}^H \to \mathbb{Z} \text{ is L}^\natural\text{-convex.} \tag{50}$$

$$G_h : \mathbb{Z}^H \to \mathbb{Z} \text{ is L}^\natural\text{-convex.} \tag{51}$$

To prove (49) and (50), it suffices to prove that $V_d : \mathbb{Z}^\Omega \to \mathbb{Z}$ and $V_h : \mathbb{Z}^\Omega \times \mathbb{Z}^H \to \mathbb{Z}$ are L$^\natural$-convex, because restricting the domain to non-negative vectors preserves L$^\natural$-convexity (see Theorem 7.11 of Murota (2003a)).

Proof of (49): We define $\hat{v}_d : \mathbb{Z}^\Omega \to \mathbb{Z} \cup \{-\infty\}$ by $\hat{v}_d(x) = \tilde{v}_d[\mathbf{0}](x)$ for all $x \in \mathbb{Z}^\Omega$ (recall the definition of $\tilde{v}_d[\mathbf{0}](\cdot)$ in Section 5.3.3). As dom $\hat{v}_d \subseteq \left\{x \in \{0,1\}^\Omega : \sum_{\omega\in\Omega} x_\omega \leq 1\right\}$, $-\hat{v}_d(\cdot)$ is M$^\natural$-convex. As $V_d(\cdot)$ is a convex conjugate function of $-\hat{v}_d(\cdot)$, by the discrete conjugacy theorem (see Theorem 8.12 of Murota (2003a)), $V_d(\cdot)$ is L$^\natural$-convex.

Proof of (50): We define $\hat{v}_h : \mathbb{Z}^\Omega \to \mathbb{Z} \cup \{-\infty\}$ by $\hat{v}_h(x) = \tilde{v}_h[\mathbf{0},\mathbf{0}](x)$ for all $x \in \mathbb{Z}^\Omega$ (recall the definition of $\tilde{v}_h[\mathbf{0},\mathbf{0}](\cdot)$ in Section 5.3.3). We define $\bar{v}_h : \mathbb{Z}^\Omega \times \mathbb{Z} \to \mathbb{Z} \cup \{-\infty\}$ by

$$\bar{v}_h(x,r) = \begin{cases} \hat{v}_h(x) & \text{if } r = -\sum_{\omega\in\Omega} x_\omega, \\ -\infty & \text{otherwise.} \end{cases}$$

Since $\hat{v}_h(\cdot)$ is M$^\natural$-concave, $\bar{v}_h(\cdot,\cdot)$ satisfies a discrete concavity called *M-concavity* (see (6.4) of Murota (2003a)). Its concave conjugate function $\bar{v}_h^0(\cdot,\cdot)$

[28] As the domains of $V_d(\cdot)$ and $V_h(\cdot)$ are different, we cannot take their sum directly. However, this problem can be circumvented by extending the domain of $V_d(\cdot)$ to $\tilde{V}_d(\cdot,\cdot)$ defined by $\tilde{V}_d(s,t) = V_d(s)$ for all $(s,t) \in \mathbb{Z}_+^\Omega \times \mathbb{Z}^H$. One easily verifies that L$^\natural$-convexity of $V_d(\cdot)$ is preserved in $\tilde{V}_d(\cdot,\cdot)$. Similarly, we can extend the domain of $G_h(\cdot)$.

is given as follows: for any $(s, t_h) \in \mathbb{Z}^{\Omega} \times \mathbb{Z}$,

$$\begin{aligned}
\bar{v}_h^0(s, t_h) &= -\sup_{(x,r)\in\mathbb{Z}^{\Omega}\times\mathbb{Z}} \left\{\bar{v}_h(x, r) - (s, t_h) \cdot (x, r)\right\} \\
&= -\sup_{x\in\mathbb{Z}^{\Omega}} \left\{\hat{v}_h(x) - s \cdot x - t_h \cdot \left(-\sum_{\omega\in\Omega} x_\omega\right)\right\} \\
&= -\sup_{x\in\mathbb{Z}^{\Omega}} \left\{\hat{v}_h(x) - s \cdot x + t_h \cdot \sum_{\omega\in\Omega} x_\omega\right\} \\
&= -V_h(s, t).
\end{aligned}$$

By the discrete conjugacy theorem (Theorem 8.12 of Murota (2003a)), $V_h(\cdot,\cdot)$ is L-convex. As L-convexity is stronger than L$^\natural$-convexity, $V_h(\cdot,\cdot)$ is L$^\natural$-convex.
Proof of (51): As described in Fig. 1, $G_h(\cdot)$ satisfies

$$G_h(t_h - 1) + G_h(t_h + 1) \geq 2G_h(t_h) \text{ for all } t_h \in \mathbb{Z},$$

which is equivalent to L$^\natural$-convexity (and also to M$^\natural$-convexity) in the 1-dimensional case; see Section 1.2 of Murota (2003a) for a detailed discussion. □

References

Ausubel, L.M., 2006. An efficient dynamic auction for heterogeneous commodities. American Economic Review 96, 602–629.

Baumol, W.J., Oates, W.E., 1971. The use of standards and prices for protection of the environment. Swedish Journal of Economics 73, 42–54.

Burch, I., Gilchrist, J., 2018. Survey of global activity to phase out internal combustion engine vehicles. Center of Climate Protection: Santa Rosa, CA, USA .

Candogan, O., Epitropou, M., Vohra, R.V., 2016. Competitive equilibrium and trading networks: A network flow approach, in: Proceedings of the 2016 ACM Conference on Economics and Computation, ACM. pp. 701–702.

Dupuy, A., Galichon, A., Jaffe, S., Kominers, S.D., 2020. Taxation in matching markets. International Economic Review, forthcoming .

Fujishige, S., Tamura, A., 2007. A two-sided discrete-concave market with possibly bounded side payments: An approach by discrete convex analysis. Mathematics of Operations Research 32, 136–155.

Fujishige, S., Yang, Z., 2003. A note on Kelso and Crawford's gross substitutes condition. Mathematics of Operations Research 28(3), 463–469.

Gul, F., Stacchetti, E., 1999. Walrasian equilibrium with gross substitutes. Journal of Economic Theory 87(1), 95–124.

Hatfield, J.W., Kominers, S., Nichifor, A., Ostrovsky, M., Westkamp, A., 2013. Stability and competitive equilibrium in trading networks. Journal of Political Economy 121(5), 966–1005.

Hatfield, J.W., Kominers, S.D., Nichifor, A., Ostrovsky, M., Westkamp, A., 2019. Full substitutability. Theoretical Economics 14, 1535–1590.

Hiriart-Urruty, J.B., Lemaréchal, C., 2001. Fundamentals of convex analysis. Springer.

Ikebe, Y.T., Tamura, A., 2015. Stability in supply chain networks: An approach by discrete convex analysis. Journal of the Operations Research Society of Japan 58, 271–290.

Kamada, Y., Kojima, F., 2017. Recent developments in matching with constraints. American Economic Review 107, 200–204.

Kamada, Y., Kojima, F., 2020. Accommodating various policy goals in matching with constraints. Japanese Economic Review 71, 101–133.

Kelso, A., Crawford, V., 1982. Job matching, coalition formation, and gross substitutes. Econometrica , 1483–1504.

Kojima, F., Sun, N., Yu, N.N., 2019. Job matching under constraints. Working Paper (version: July 29, 2019) .

Kojima, F., Sun, N., Yu, N.N., 2020a. Job matching under constraints. American Economic Review, forthcoming .

Kojima, F., Sun, N., Yu, N.N., 2020b. Job matching with subsidy and taxation. Working Paper .

Kojima, F., Tamura, A., Yokoo, M., 2018. Designing matching mechanisms under constraints: An approach from discrete convex analysis. Journal of Economic Theory 176, 803–833.

Kolmogorov, V., Shioura, A., 2009. New algorithms for convex cost tension problem with application to computer vision. Discrete Optimization 6, 378–393.

Murota, K., 1998. Discrete convex analysis. Mathematical Programming 83, 313–371.

Murota, K., 2003a. Discrete convex analysis. Society for Industrial and Applied Mathematics.

Murota, K., 2003b. On steepest descent algorithms for discrete convex functions. SIAM Journal on Optimization 14, 699–707.

Murota, K., 2016. Discrete convex analysis: A tool for economics and game theory. Journal of Mechanism and Institution Design 1, 151–273.

Murota, K., Shioura, A., 2014. Exact bounds for steepest descent algorithms of l-convex function minimization. Operations Research Letters 42, 361–366.

Murota, K., Shioura, A., Yang, Z., 2016. Time bounds for iterative auctions: a unified approach by discrete convex analysis. Discrete Optimization 19, 36–62.

Parry, I.W., Walls, M., Harrington, W., 2007. Automobile externalities and policies. Journal of Economic Literature 45, 373–399.

Rockafellar, R.T., 1970. Convex analysis. Princeton University Press.

Sun, N., Yang, Z., 2009. A double-track adjustment process for discrete markets with substitutes and complements. Econometrica 77, 933–952.

Tiel, J.v., 1984. Convex analysis: An introductory text. John Wiley.

Yokote, K., 2020. Supplemental material for "On optimal taxes and subsidies: A discrete saddle-point theorem with application to job matching under constraints". Available at https://sites.google.com/site/kojiyokote/.

Journal of Mechanism and Institution Design
ISSN: 2399-844X (Print), 2399-8458 (Online)
DOI:10.22574/jmid.2020.12.003

ON A CLASS OF LINEAR-STATE DIFFERENTIAL GAMES WITH SUBGAME INDIVIDUALLY RATIONAL AND TIME CONSISTENT BARGAINING SOLUTIONS

Simon Hoof

Paderborn University, Germany
simon.hoof@upb.de

ABSTRACT

We consider n-person pure bargaining games in which the space of feasible payoffs is constructed via a normal form differential game. At the beginning of the game the agents bargain over strategies to be played over an infinite time horizon. An initial cooperative solution (a strategy tuple) is called subgame individually rational (SIR) if it remains individually rational throughout the entire game and time consistent (TC) if renegotiating it at a later time instant yields the original solution. For a class of linear-state differential games we show that any solution which is individually rational at the beginning of the game satisfies SIR and TC if the space of admissible cooperative strategies is restricted to constants. We discuss an application from environmental economics.

Keywords: Differential games, bargaining solutions, time consistency.

JEL Classification Numbers: C61, C71, C78.

1. INTRODUCTION

Suppose there are two countries that pollute the environment by emitting pollutants and further assume that there is a production technology that

I am indebted to Claus-Jochen Haake, three anonymous referees, and participants of the 9th Murat Sertel Workshop for valuable comments. This work was partially supported by the German Research Foundation (DFG) within the Collaborative Research Center "On-The-Fly Computing" (SFB 901) under the project number 160364472-SFB901/3.

maps pollutant emissions into output and output is increasing in emissions. Hence, each country has an incentive to pollute the environment. The total pollution stock, however, is assumed to be a common bad and payoffs are thus decreasing in the total stock of pollution. Total pollution further comes with a stock externality. If one neighboring country increases the rate of emissions, then it harms the other one, because the pollution stock increases. Clearly, there are some underlying dynamics in the sense that current emissions increase the total stock of pollution over time. The problem at hand can be modeled as a differential game in which the countries are agents, emission rates are the respective action variable of each country and the pollution stock is the state variable (E. J. Dockner & Van Long, 1993). Considering the noncooperative equilibrium as a solution of the game, it turns out that it is not efficient, because both countries are overemitting with respect to the social optimum and thus both countries could gain by coordinating their actions. The two countries face a bargaining situation in the sense of Nash (1950b, p. 155):

> "A TWO-PERSON bargaining situation involves two individuals who have the opportunity to collaborate for mutual benefit in more than one way."

Liu (1973) linked cooperative bargaining games to differential games. He studied a class of differential games in normal form in which strategies are assumed to be determined by the Nash bargaining solution. Let us assume that the countries determine their current emission rate via feedback strategies, i.e., current emissions are a function of the current pollution stock. At the beginning of the dynamic game the countries bargain over the strategies to be played during the game, while the noncooperative equilibrium payoff serves as the disagreement point. They would come to an agreement if it was overall individually rational, i.e., the payoffs under cooperation exceed the noncooperative equilibrium payoffs at the beginning of the game. A bargaining solution thus determines the cooperative strategies to be played during the game. Haurie (1976) pointd out two potential reasons for the breakdown of an initial agreement:

(i) If there is an agent whose current noncooperative outside option payoff dominates the current cooperative payoff, then this agent should abandon the agreement and play noncooperatively in the remaining game.

(ii) If the agents are about to renegotiate the initial agreement at a later time instant, it is not granted that they agree on the original solution.

The first property is called subgame individually rationality (SIR) and the second one time consistency (TC). In the literature on cooperation in differential games it is usually assumed that the agents jointly maximize the weighted sum of payoffs. The resulting strategies are the cooperative strategies to be played during the entire game and they satisfy the TC property by construction. Cooperation is assumed to last if the gains of cooperation exceed the noncooperative equilibrium payoff for each agent and at each time instant. One further distinguishes between games with transferable (Petrosyan & Zaccour, 2018) and nontransferable utility (Yeung & Petrosyan, 2018). If payoffs are transferable the designer tries to construct a SIR payoff distribution procedure. If payoffs are nontransferable, then the designer tries to adjust the weights in the joint maximization program such that the resulting payoffs are SIR.

The present approach deviates from the standard approach in the sense that the cooperative strategies are not determined by a joint payoff maximization program, but via a bargaining solution. The goal of the paper is to characterize a variety of linear-state infinite horizon differential games[1] and show how to construct bargaining solutions that satisfy (i) and (ii). The advantage of linear-state games is their analytical tractability and the coincidence of open loop and feedback equilibria (E. J. Dockner et al., 2000, Ch. 7.2).

Taking the noncooperative equilibrium payoffs as given (disagreement point), we induce a cooperative n-person pure bargaining game by considering all strategies that payoff dominate the disagreement point. It is shown that SIR and TC bargaining solutions exist if we restrict the set of admissible cooperative strategies to constants. At a first glance it seems that the restriction of the strategy space to constants is rather limiting as we neglect the case in which strategies are time and state dependent functions. We will show, however, that the class of games under consideration exhibits a subgame perfect Nash equilibrium in constant actions. That is, even though we allow for strategies that depend on the current state and thus adjust for the evolution of the state over time, it is an equilibrium strategy to fix a constant action during the entire game.

The rest of this article is organized as follows. Section 2 discusses the relation between bargaining and differential games. Section 3 examines disagreement points. Section 4 focuses on linear-state games. Section 5

[1] The same class of games is considered in Jørgensen et al. (2003). They are also concerned with subgame individually rational solutions, but the cooperative strategies are determined by maximizing the joint sum of payoffs.

presents some applications. Section 6 concludes.

2. LINKING BARGAINING AND DIFFERENTIAL GAMES

Let $N = \{1, 2, \ldots, n\}$ denote the set of agents. A bargaining problem $(S, \boldsymbol{d})$ is described by a set $S \subseteq \mathbb{R}^n$ of feasible payoffs and a given disagreement point $\boldsymbol{d} \in S$. Let $\mathcal{B}$ denote the set of all bargaining problems. A bargaining solution is a mapping $\phi : \mathcal{B} \to \mathbb{R}^n$ that assigns to each problem $(S, \boldsymbol{d}) \in \mathcal{B}$ a unique point in the set of feasible payoffs $\phi(S, \boldsymbol{d}) \in S$. In the present paper the set of feasible payoffs is constructed via a differential game. Therefore we first introduce the differential game terminology and then show how a differential game can be used to induce a bargaining game.

Let $T \subseteq \mathbb{R}$ denote the time space, $X \subseteq \mathbb{R}$ the set of feasible states and $A_i \subseteq \mathbb{R}$ the set of admissible actions for each $i \in N$. Here we consider stationary (time invariant) infinite horizon games and thus fix $T = [0, \infty)$. At each point in time $t \in T$, each agent $i \in N$ executes an action $a_i(t) \in A_i$. Let $A = \bigtimes_{i \in N} A_i$ denote the jointly admissible action space. A profile of action trajectories $\boldsymbol{a} : T \to A$ determines the state trajectory $x : T \to X$ via the law of motion $f : X \times A \to X$. The transition function f is a stationary differential equation

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = f(x(t), \boldsymbol{a}(t)) \tag{1}$$

which governs the evolution of the state over time and we assume that the function $f(x, \boldsymbol{a})$ is continuously differentiable in x and $\boldsymbol{a}$.

Technically, we distinguish between actions and strategies. The former is an element of a set, while the latter is a function. A stationary feedback strategy is a function $\sigma_i : X \to A_i$ that determines the current action. An agent thus observes $x(t)$ and sets $a_i(t)$ according to $\sigma_i(x(t))$. Let us formally introduce the set of admissible strategies by

$$\Sigma_i = \{\sigma_i : X \to A_i \mid \sigma_i(x) \text{ is Lipschitz continuous on } X\}.$$

The conditions imposed on the state equation f and strategy spaces $\{\Sigma_i\}_{i \in N}$ imply that the solution $(x(t))_{t \geq 0}$ of the differential equation (1) exists and is unique as well as continuous (Başar & Olsder, 1999, Thm. 5.1).

We further need to introduce payoffs (utility). At any position $(t, x(t)) \in T \times X$ the payoff functional $u_i : T \times X \times \bigtimes_{i \in N} \Sigma_i \to \mathbb{R}$ of any agent $i \in N$ is given by the discounted stream of instantaneous payoffs

$$u_i(t, x(t), \boldsymbol{\sigma}) = \int_t^{\infty} e^{-r_i(s-t)} G_i(x(s), \boldsymbol{\sigma}(x(s)))\mathrm{d}s. \tag{2}$$

Here, $r_i > 0$ is the time preferences rate and $G_i : X \times A \to \mathbb{R}$ the instantaneous utility function. The functions $\{G_i(x, \boldsymbol{a})\}_{i \in N}$ are assumed to be continuously differentiable in x and $\boldsymbol{a}$. Since the setup is stationary the payoff functional u_i does not depend on the current time t, but only on the current state $x(t)$, i.e., $u_i(x(t), \boldsymbol{\sigma}) = u_i(t', x(t), \boldsymbol{\sigma})$ for all $t' \in T$ and we thus suppress t as an explicit argument (Caputo, 2005, Thm. 19.2). For notational convenience we denote the current state with $x = x(t)$. We denote the class of all stationary infinite horizon games by

$$\mathcal{G} = \{\Gamma(x) \mid \Gamma(x) = \langle N, (\Sigma_i)_{i \in N}, (u_i(x, \cdot))_{i \in N} \rangle\}.$$

We make two assumptions about jointly admissible strategy profiles $\boldsymbol{\sigma} \in \bigtimes_{i \in N} \Sigma_i$. A strategy profile $\boldsymbol{\sigma}$ is admissible if the state stays in the state space and the utility functional u_i is finite for all i. Therefore consider the parametrized solution of (1) originating from any position $(t, x) \in T \times X$

$$x(s) = y(s; t, x, \boldsymbol{\sigma}) = x + \int_t^s f(x(k), \boldsymbol{\sigma}(x(k)))\mathrm{d}k.$$

Now we define a set $\Sigma \subseteq \bigtimes_{i \in N} \Sigma_i$ of jointly admissible strategies by

$$\Sigma = \left\{ \boldsymbol{\sigma} \in \bigtimes_{i \in N} \Sigma_i \;\middle|\; \forall (t, x) \in T \times X : y(s; t, x, \boldsymbol{\sigma}) \in X \; \forall s \geq t, \right.$$
$$\left. \forall x \in X : \max_{i \in N} \{|u_i(x, \boldsymbol{\sigma})|\} < \infty \right\}.$$

Given a game $\Gamma(x) \in \mathcal{G}$ as a primitive we define the set of feasible payoffs by considering all payoff allocations $\boldsymbol{u}(x, \boldsymbol{\sigma}) = (u_i(x, \boldsymbol{\sigma}))_{i \in N}$ under admissible strategies such that $S(x) \subseteq \mathbb{R}^n$ is given by

$$S(x) = \{\boldsymbol{u}(x, \boldsymbol{\sigma}) \in \mathbb{R}^n \mid \boldsymbol{\sigma} \in \Sigma\}.$$

To properly define a bargaining problem we need to fix a state dependent disagreement point $\boldsymbol{d}(x) \in S(x)$. For now we assume that $\overline{\boldsymbol{\sigma}} \in \Sigma$ is a given disagreement strategy that induces for all $x \in X$ the disagreement point $\boldsymbol{d}(x) = \boldsymbol{u}(x, \overline{\boldsymbol{\sigma}})$. In Section 3 we argue that the noncooperative equilibrium payoff is a reasonable disagreement point.

Definition 1 (Bargaining Problem)**.** *The pair $(S(x), \boldsymbol{d}(x))$ is a bargaining problem.*

Usually, a bargaining solution is a point in the payoff space $S(x)$. Here, however, the agents bargain over strategies that determine the payoffs. Therefore we consider bargaining solutions as functions in the strategy space Σ. Let $\mathcal{B} = \{(S(x), \boldsymbol{d}(x)) \mid x \in X\}$ denote the set of all bargaining problems.

Definition 2 (Bargaining Solution)**.** *A bargaining solution is a correspondence $\phi : \mathcal{B} \rightrightarrows \Sigma$ that assigns to each problem $(S(x), \boldsymbol{d}(x)) \in \mathcal{B}$ a set $\phi(S(x), \boldsymbol{d}(x)) \subseteq \Sigma$ of admissible strategy profiles.*

We assume that the bargaining solution is the set of maximizers of a continuous function $H : \mathbb{R}^n \to \mathbb{R}$. The arguments of H are the cooperation dividends

$$E_i(x, \boldsymbol{\sigma}) = u_i(x, \boldsymbol{\sigma}) - d_i(x).$$

With $\boldsymbol{E}(x, \boldsymbol{\sigma}) = (E_i(x, \boldsymbol{\sigma}))_{i \in N}$ the bargaining solution is then given by

$$\phi(S(x), \boldsymbol{d}(x)) = \arg\max_{\boldsymbol{\sigma} \in \Sigma} H(\boldsymbol{E}(x, \boldsymbol{\sigma})).$$

An element of $\phi(S(x), \boldsymbol{d}(x))$ is termed a cooperative strategy and is denoted by $\boldsymbol{\sigma}^*_x$. One should note that the set $\phi(S(x), \boldsymbol{d}(x))$ may differ with respect to x and a cooperative strategy thus generally depends on the current state x. Let us briefly illustrate the idea. Assume that the agents rely on the Nash bargaining solution to determine a cooperative strategy at $(t, x) = (0, x_0)$ and let us further assume that there exist two maximizers

$$\left\{\boldsymbol{\sigma}^{*,1}_{x_0}, \boldsymbol{\sigma}^{*,2}_{x_0}\right\} = \arg\max_{\boldsymbol{\sigma} \in \Sigma} \prod_{i \in N} E_i(x_0, \boldsymbol{\sigma}).$$

Now the agents reevaluate the problem at $(t, x) = (1, x_1) \neq (0, x_0)$ and end up with a unique maximizer

$$\left\{\boldsymbol{\sigma}^{*}_{x_1}\right\} = \arg\max_{\boldsymbol{\sigma} \in \Sigma} \prod_{i \in N} E_i(x_1, \boldsymbol{\sigma}).$$

Assume that $\boldsymbol{\sigma}^{*,1}_{x_0} = \boldsymbol{\sigma}^{*}_{x_1}$ and $\boldsymbol{\sigma}^{*,2}_{x_0} \neq \boldsymbol{\sigma}^{*}_{x_1}$ hold, then the first cooperative strategy $\boldsymbol{\sigma}^{*,1}_{x_0}$ is time consistent at x_1 and the second one $\boldsymbol{\sigma}^{*,2}_{x_0}$ is not. The goal of the article is to identify bargaining solutions on a restricted domain of games that are time consistent over the entire state space X and which happen to be also individually rational.

Axiomatic bargaining theory (see e.g. Peters, 1992) studies solutions that can be uniquely identified by a number of axioms. Here we do not intend to pin down a unique bargaining solution by some suitable axiomatization, but to check whether there exists an initial cooperative strategy $\boldsymbol{\sigma}^*_{x_0} \in \phi(S(x_0), \boldsymbol{d}(x_0))$ that (i) remains individually rational throughout the game and (ii) is robust with respect to renegotiations.

We distinguish between a weak and strong notion of (i) and (ii). A cooperative strategy is weak if (i) and (ii) only hold along the cooperative state trajectory $x^*(s) = y(s; 0, x_0, \boldsymbol{\sigma}^*_{x_0})$ and strong if (i) and (ii) hold along any admissible state trajectory $x(s) = y(s; 0, x_0, \boldsymbol{\sigma})$, $\boldsymbol{\sigma} \in \Sigma$. We define the differential game concepts rigorously.

Definition 3 (Overall Individual Rationality)**.** *A solution ϕ satisfies Overall Individual Rationality if there exists a cooperative strategy $\boldsymbol{\sigma}^*_{x_0} \in \phi(S(x_0), \boldsymbol{d}(x_0))$ such that*

$$\min_{i \in N} \{E_i(x_0, \boldsymbol{\sigma}^*_{x_0})\} \geq 0. \tag{OIR}$$

A bargaining solution satisfies OIR if each agent is not harmed by cooperating at the beginning of the game. This property is a global necessary condition for cooperation, because otherwise there would be an agent who does not want to cooperate at all. As time goes by it might be the case that cooperation under the initial agreement lacks payoff dominance with respect to the noncooperative outside option. We thus also introduce the subgame notion of individual rationality.

Definition 4 (Subgame Individual Rationality)**.** *A solution ϕ satisfies Weak/Strong Subgame Individual Rationality if there exists a cooperative strategy $\boldsymbol{\sigma}^*_{x_0} \in \phi(S(x_0), \boldsymbol{d}(x_0))$ such that*

$$\min_{i \in N} \{E_i(x^*(s), \boldsymbol{\sigma}^*_{x_0})\} \geq 0 \quad \forall s \geq 0, \tag{WSIR}$$

$$\min_{i \in N} \{E_i(x, \boldsymbol{\sigma}^*_{x_0})\} \geq 0 \quad \forall x \in X. \tag{SSIR}$$

The strong version is robust in the sense that even if the cooperative path $x^*(s)$ is perturbed to $x(s) = x^*(s) + \varepsilon$, $\varepsilon \neq 0$, it always pays off to stick to the initial agreement. In order to satisfy time consistency we require that there exists an initial cooperative strategy that solves any bargaining problem occurring at a later point in time.

Definition 5 (Time Consistency)**.** *A solution ϕ satisfies Weak/Strong Time Consistency if*

$$\bigcap_{s \geq 0} \phi(S(x^*(s)), \boldsymbol{d}(x^*(s))) \neq \emptyset, \tag{WTC}$$

$$\bigcap_{x \in X} \phi(S(x), \boldsymbol{d}(x)) \neq \emptyset. \tag{STC}$$

Clearly, SSIR implies WSIR implies OIR and STC implies WTC. Apparently, the reverse does not hold in general.

In the literature on cooperative differential games (see e.g. Jørgensen & Zaccour, 2002; Zaccour, 2008) there are different terminologies for the previously introduced concepts. We follow Başar (1989) in calling the renegotiation condition (ii) time consistency. It is a property that generally applies to optimal control problems as well as differential games. The strong version corresponds to subgame perfectness in standard game theory lingo and is also called Markov perfect equilibrium or feedback equilibrium. In cooperative differential games it is usually assumed that the agents jointly maximize the sum of individual payoffs in order to determine the cooperative strategies to be played throughout the game. When solving this optimal control problem via a dynamic programming approach strong time consistency is automatically given. The weak version of subgame individual rationality (WSIR) is then just called time consistency and following Kaitala & Pohjola (1990) the strong version (SSIR) is called agreeability. Here, however, we do not assume that the agents jointly maximize their payoffs, but that they bargain over strategies. We therefore explicitly distinguish between the consistency (renegotiating) and the individual rationality condition.

3. DISAGREEMENT POINT

Clearly, the bargaining solution $\phi(S(x), \boldsymbol{d}(x))$ depends on the disagreement point $\boldsymbol{d}(x) = \boldsymbol{u}(x, \overline{\boldsymbol{\sigma}})$ and thus on the disagreement strategies $\overline{\boldsymbol{\sigma}} \in \Sigma$. Here the disagreement point is not given, but constructed from the underlying game $\Gamma(x)$. There is no objective rationale that yields the "correct" disagreement point, but we would like to argue that the Nash (1950a, 1951) equilibrium payoff is the natural choice for pure bargaining games. A Nash equilibrium is a strategy tuple such that no agent has an incentive to unilaterally change her strategy. It is straightforward to define the solution concept for differential games (Starr & Ho, 1969a,b). Given the feedback information structure we must make sure, however, that no agent has an incentive to change her strategy (a function) for different states $x \in X$.

Definition 6 (Equilibrium)**.** *The n-tuple $\overline{\boldsymbol{\sigma}} \in \Sigma$ is a subgame perfect Nash equilibrium for $\Gamma(x) \in \mathcal{G}$ if for all agents $i \in N$ and positions $x \in X$ the following inequalities hold:*

$$u_i(x, \overline{\boldsymbol{\sigma}}) \geq u_i(x, \sigma_i, \overline{\boldsymbol{\sigma}}_{-i}) \quad \forall \sigma_i \in \Sigma_i$$

where $\overline{\boldsymbol{\sigma}}_{-i} = (\overline{\sigma}_j)_{j \in N \setminus \{i\}}$ are the equilibrium strategies of the opponents.

For any $x \in X$ the disagreement point $\boldsymbol{d}(x) \in \mathbb{R}^n$ is then given by the *current* payoff-to-go under the noncooperative equilibrium strategies, i.e., $d_i(x) = u_i(x, \overline{\boldsymbol{\sigma}})$ for all $i \in N$.

Let us elaborate why the Nash equilibrium payoff is the natural disagreement point. Pure bargaining games are a special kind of cooperative game in the sense that either the grand coalition forms and jointly realizes the payoffs $\boldsymbol{u}(x, \boldsymbol{\sigma}_x^*) \in S(x)$ or there exists an agent who disagrees with $\boldsymbol{\sigma}_x^*$ and then every agent $i \in N$ gets $d_i(x)$. We therefore do not allow for the formation of any coalition $C \subset N$, but the singleton $C = \{i\}$. At the beginning of the game $(t, x) = (0, x_0)$ the agents can thus either jointly agree on playing $\boldsymbol{\sigma}_{x_0}^* \in \Sigma$ or they cannot come to an agreement. Hence in case of disagreement the agents face a fully noncooperative game (n singletons) as of which the Nash equilibrium is the rationale solution concept with payoffs $\boldsymbol{d}(x_0)$. That is, they should bargain over gains compared to the outside option which is playing a noncooperative game. Let us assume the agents settled on $\boldsymbol{\sigma}_{x_0}^*$ at the beginning of the game, but suppose they reopen negotiations at any position $(t, x) \in (0, \infty) \times X$. They then face the same situation as in the beginning of the game, but the outside option is rather the *current* Nash equilibrium payoff at position x and thus $\boldsymbol{d}(x)$.

By invoking dynamic programming techniques, one can show that the equilibrium strategies solve a system of coupled differential equations. The following Lemma is fundamental when characterizing the equilibrium of a differential game (see e.g. E. Dockner & Wagener, 2014, Thm. 1).

Lemma 1. *The n-tuple* $\overline{\boldsymbol{\sigma}} \in \Sigma$ *is a subgame perfect Nash equilibrium for* $\Gamma(x) \in \mathcal{G}$ *if there exist n functions* $\{d_i : X \to \mathbb{R}\}_{i \in N}$ *that are continuously differentiable in x and solve the following system of coupled Hamilton-Jacobi-Bellman (HJB) equations:*

$$\begin{aligned} r_i d_i(x) &= \max_{a_i \in A_i} \{G_i(x, a_i, \overline{\boldsymbol{\sigma}}_{-i}(x)) + d_i'(x) f(x, a_i, \overline{\boldsymbol{\sigma}}_{-i}(x))\} \\ &= G_i(x, \overline{\boldsymbol{\sigma}}(x)) + d_i'(x) f(x, \overline{\boldsymbol{\sigma}}(x)). \end{aligned} \tag{3}$$

Further, the transversality condition $\lim_{t\to\infty} e^{-r_i t} d_i(x) = 0$ *must be satisfied for all agents* $i \in N$ *and all feasible states* $x \in X$.

The function $d_i(\cdot)$ is termed value function. We apply Lemma 1 in the next section in order to establish the existence of an equilibrium in constant strategies.

4. LINEAR-STATE GAMES

In the present paper we restrict the domain of games under consideration to a class $\mathcal{G}_{\text{Lin}} \subset \mathcal{G}$ of linear-state games and therefore consider state equations and instantaneous payoff functions that satisfy the following functional form:

$$f(x, \boldsymbol{a}) = h(\boldsymbol{a}) - \delta x, \tag{4}$$

$$G_i(x, \boldsymbol{a}) = g_i(\boldsymbol{a}) + \eta_i x, \tag{5}$$

where $\delta, \eta_i \in \mathbb{R}$ denote parameters. Finally, we assume that for all different $i, j \in N$ the state equation as well as the payoffs are independent of the actions of distinct players, i.e.,

$$\frac{\partial^2 h(\boldsymbol{a})}{\partial a_j \partial a_i} = 0 \quad \text{and} \quad \frac{\partial^2 g_i(\boldsymbol{a})}{\partial a_j \partial a_i} = 0. \tag{6}$$

There is thus no multiplicative interaction between the actions of the players. Let us define the class $\mathcal{G}_{\text{Lin}}$ of linear-state games by

$$\mathcal{G}_{\text{Lin}} = \{\Gamma(x) \in \mathcal{G} \mid \text{Eqs. (4), (5) and (6) hold}\}.$$

Proposition 1. *For $\Gamma(x) \in \mathcal{G}_{\text{Lin}}$ there exists an equilibrium in constant strategies $\overline{\boldsymbol{k}} \in A \subseteq \mathbb{R}^n$ with associated linear value functions $\{d_i(x) = \alpha_i x + \beta_i\}_{i \in N}$. The constants α_i and β_i are given by*

$$\alpha_i = \frac{\eta_i}{r_i + \delta}, \tag{7}$$

$$\beta_i = \frac{1}{r_i}\left[g_i(\overline{\boldsymbol{k}}) + \alpha_i h(\overline{\boldsymbol{k}})\right]. \tag{8}$$

The equilibrium strategies $\overline{\sigma}_i(x) = \overline{k}_i$ are given by the root of the first order condition for $i \in N$

$$L_i(\overline{k}_i, \alpha_i) = 0$$

where $L_i(\cdot, \cdot)$ denotes the derivative of the right hand side of the HJB equation with respect to the action a_i

$$L_i(a_i, \alpha_i) = \frac{\partial\left[g_i(a_i, \overline{\boldsymbol{\sigma}}_{-i}(x)) + \alpha_i h(a_i, \overline{\boldsymbol{\sigma}}_{-i}(x))\right]}{\partial a_i}(a_i, \alpha_i). \tag{9}$$

Proof. We guess a functional form for the value function $d_i(x) = \alpha_i x + \beta_i$ where $\alpha_i, \beta_i \in \mathbb{R}$ are constants. We show that $\{d_i(x)\}_{i \in N}$ satisfy the HJB equations of Lemma 1. Consider the function $L_i(\cdot, \cdot)$ as defined by (9). Due to (6) $L_i(\cdot, \cdot)$ does not depend on the equilibrium strategies $\overline{\sigma}_j(x)$ of any other agent $j \in N \setminus \{i\}$. For all $i \in N$ let $\overline{k}_i$ denote the root of the first order condition (9), i.e., $\overline{k}_i$ solves $L_i(\overline{k}_i, \alpha_i) = 0$. Since $\overline{k}_i$ only depends on α_i it is a constant. Collect the maximizers in $\overline{\boldsymbol{k}} = (\overline{k}_i)_{i \in N}$. The maximized HJB equation (3) then reads

$$r_i\left(\alpha_i x + \beta_i\right) = g_i(\overline{\boldsymbol{k}}) + \eta_i x + \alpha_i\left(h(\overline{\boldsymbol{k}}) - \delta x\right).$$

The maximized HJB equation is satisfied for all $x \in X$ if α_i and β_i satisfy (7) and (8) respectively. □

We can now state the main result. We use the information about the noncooperative equilibrium strategies and restrict the domain of admissible cooperative strategies to constants. Then we are able to show that all OIR solutions satisfy SSIR and STC.

Proposition 2. *If the space of admissible cooperative strategies is restricted to constants* $\Sigma_i^* = \{\sigma_i(x) = k_i \mid k_i \in A_i\} \subset \Sigma_i$*, then all* OIR *bargaining solutions satisfy* STC *and* SSIR.

Proof. Let $\boldsymbol{k} \in A$ and for $s \geq t$ let $y(s;t,x,\boldsymbol{k})$ denote the feasible solutions of (4) conditioned on the current position (t,x)

$$y(s;t,x,\boldsymbol{k}) = \frac{1}{\delta}\left[h(\boldsymbol{k}) - e^{-\delta(s-t)}(h(\boldsymbol{k}) - \delta x)\right].$$

For any $x \in X$ the feasible current payoff-to-go is then given by

$$\begin{aligned} u_i(x,\boldsymbol{k}) = &\int_t^\infty e^{-r_i(s-t)}\left[g_i(\boldsymbol{k}) + \eta_i y(s;t,x,\boldsymbol{k})\right] ds \\ =&\frac{\eta_i}{r_i+\delta}x + \frac{1}{r_i}\left[g_i(\boldsymbol{k}) + \frac{\eta_i}{r_i+\delta}h(\boldsymbol{k})\right] \\ =&\alpha_i x + \frac{1}{r_i}\left[g_i(\boldsymbol{k}) + \alpha_i h(\boldsymbol{k})\right]. \end{aligned} \tag{10}$$

When considering constant actions $\sigma_i(x) = k_i$ for all $i \in N$, then the slope of any feasible payoff functional $\partial u_i(x,\boldsymbol{k})/\partial x = \alpha_i$ is equal to the slope of the noncooperative equilibrium payoff $d_i'(x)$. Since we fix $d_i(x)$ as the outside option the potential gains of cooperating from $(t,x) \in [0,\infty) \times X$ onwards become

$$E_i(\boldsymbol{k}) = u_i(x,\boldsymbol{k}) - d_i(x) = \frac{1}{r_i}\left[g_i(\boldsymbol{k}) - g_i(\overline{\boldsymbol{k}}) + \alpha_i\left(h(\boldsymbol{k}) - h(\overline{\boldsymbol{k}})\right)\right]. \tag{11}$$

At $t = 0$ an OIR solution solves

$$\begin{aligned} &\boldsymbol{k}^* \in \arg\max_{\boldsymbol{k} \in A} H(\boldsymbol{E}(\boldsymbol{k})) \\ \text{s.t.} \quad &\min_{i \in N}\{E_i(\boldsymbol{k})\} \geq 0. \end{aligned}$$

As $\boldsymbol{k}^*$ does not depend on the state SSIR and STC are satisfied for all $x \in X$. □

One should note that the result is not constructive. Restricting the set of admissible cooperative strategies to constants is not sufficient, because the bargaining solution could still vary with respect to the state. The result is entirely driven by the fact that the cooperation dividends $\boldsymbol{E}(\boldsymbol{k})$ do not depend on x. If this was not the case, i.e., the cooperation dividends do depend on the state $\boldsymbol{E}(x, \boldsymbol{k})$, then the bargaining solution $\boldsymbol{k}_x^* \in \arg\max_{\boldsymbol{k}\in A} H(\boldsymbol{E}(x, \boldsymbol{k}))$ still varies over the state space.

State invariance of the cooperative dividends $\boldsymbol{E}(\boldsymbol{k})$ has an interesting implication, because then the set of individually rational actions

$$A^{\mathrm{IR}} = \left\{ \boldsymbol{k} \in A \;\middle|\; \min_{i\in N}\{E_i(\boldsymbol{k})\} \geq 0 \right\}$$

is also invariant with respect to x. The set of feasible payoffs $S(x)$ still depends on the state x, though. Note, however, that any bargaining problem $(S(x), \boldsymbol{d}(x))$ is only a linearly shifted variant of the initial problem

$$(S(x), \boldsymbol{d}(x)) = (S(x_0) + \boldsymbol{\alpha} \cdot (x - x_0), \boldsymbol{d}(x_0) + \boldsymbol{\alpha} \cdot (x - x_0))$$

where we should recall that $\boldsymbol{\alpha} = (\alpha_i)_{i\in N}$ are the coefficients of the value functions $d_i(x) = \alpha_i x + \beta_i$. Since we are only interested in OIR solutions – rational agents bargain over additional surplus compared to the disagreement point – we may define a normalized problem by

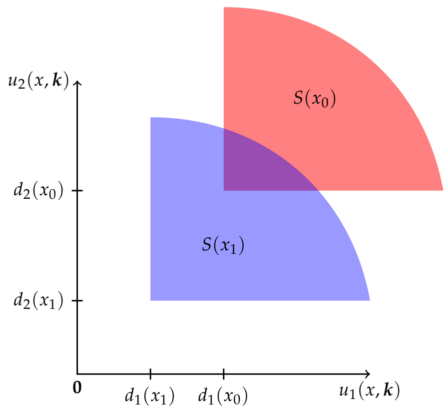
$$S^{\mathrm{IR}} = S(x) - \boldsymbol{d}(x) = \left\{ \boldsymbol{E}(\boldsymbol{k}) \mid \boldsymbol{k} \in A^{\mathrm{IR}} \right\} \subseteq \mathbb{R}_+^n.$$

As S^{IR} does not depend on x, all normalized problems coincide. Figure 1 illustrates the relationship. In Figure 1a we consider two problems $(S(x_0), \boldsymbol{d}(x_0))$ and $(S(x_1), \boldsymbol{d}(x_1))$ for some different states $x_0, x_1 \in X$. Even though the original problems do not coincide in the payoff space $S(x_0) \neq S(x_1)$, they are equivalent in terms of current surplus $S(x_0) - \boldsymbol{d}(x_0) = S(x_1) - \boldsymbol{d}(x_1)$ (cf. Figure 1b).

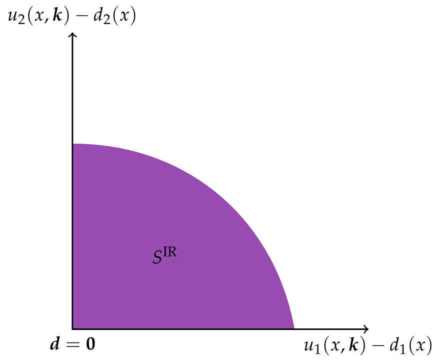
There is a straightforward policy implication: if a linear-state model approximates a real world phenomenon sufficiently well and there are agents who want to jointly benefit by cooperation, then the policymaker (designer) should only allow constant strategies such that no agent has an incentive to deviate from the initial cooperative agreement as times goes by.

5. APPLICATION

We discuss an application drawn from environmental economics (E. J. Dockner & Van Long, 1993). There are two countries that pollute the air. Let



(a) original problem $(S(x), \boldsymbol{d}(x))$



(b) normalized problem $(S^{\mathrm{IR}}, \mathbf{0})$

Figure 1: Normalization $S^{\mathrm{IR}} = S(x) - \boldsymbol{d}(x)$

$k_i \in A_i = \mathbb{R}_{++}$ denote the constant emission rate and $x \in X = \mathbb{R}_+$ the stock of pollution. A country gets some payoff from current pollution via $g_i(k_i) = \ln(k_i)$, but has to take into account a stock externality described by η_i. As total payoffs are given by $G_i(x, k_i) = \ln(k_i) + \eta_i x$, we assume that the externality parameter is negative $\eta_i < 0$ and payoffs thus decrease in the total stock of pollution. Let us further assume that the pollution stock increases linearly in emissions $h(\boldsymbol{k}) = k_1 + k_2$ and that pollution is naturally absorbed with $\delta > 0$ such that the state equation reads $f(x, \boldsymbol{k}) = k_1 + k_2 - \delta x$. The utility under constant strategies is given by (10)

$$u_i(x, k_i, k_{-i}) = \alpha_i x + \frac{1}{r_i}\left[\ln(k_i) + \alpha_i\left(k_i + k_{-i}\right)\right].$$

It is straightforward to verify that there exists a noncooperative equilibrium characterized by the action

$$\overline{k}_i = \underset{k_i \in A_i}{\arg\max}\left\{u_i(x, k_i, \overline{k}_{-i})\right\} = -\frac{1}{\alpha_i} \overset{(7)}{=} -\frac{r_i + \delta}{\eta_i} > 0.$$

One should verify that

$$d_i(x) = u_i(x, \overline{\boldsymbol{k}}) = \alpha_i x + \frac{1}{r_i}\left[\ln\left(-\frac{1}{\alpha_i}\right) - \alpha_i\left(\frac{1}{\alpha_i} + \frac{1}{\alpha_{-i}}\right)\right]$$

in fact solves the general HJB equation

$$r_i d_i(x) = \max_{a_i \in A} \{\ln(a_i) + \eta_i x + d_i'(x)(a_i + \overline{a}_{-i} - \delta x)\}.$$

The maximized HJB equation reads

$$r_i d_i(x) = \ln\left(-\frac{1}{d_i'(x)}\right) + \eta_i x - d_i'(x)\left(\frac{1}{d_i'(x)} + \frac{1}{d_{-i}'(x)} + \delta x\right).$$

Substituting $d_i'(x) = \alpha_i$ and dividing by r_i yields the claimed result

$$d_i(x) = \underbrace{\frac{(\eta_i - \alpha_i \delta)}{r_i}}_{=\alpha_i} x + \frac{1}{r_i}\left[\ln\left(-\frac{1}{\alpha_i}\right) - \alpha_i\left(\frac{1}{\alpha_i} + \frac{1}{\alpha_{-i}}\right)\right].$$

The cooperation dividend is then given by (11)

$$E_i(\boldsymbol{k}) = \frac{1}{r_i}\left[\ln\left(\frac{k_i}{\overline{k}_i}\right) - \frac{1}{\overline{k}_i}\sum_{j=1}^{2}\left(k_j - \overline{k}_j\right)\right]$$

where we use $\alpha_i = -1/\overline{k}_i$. We consider the bargaining problem

$$(S, \mathbf{0}) = (\{(E_1(\boldsymbol{k}), E_2(\boldsymbol{k})) \mid \boldsymbol{k} \in A\}, \mathbf{0}).$$

First, we show that there are conceptional analogies to an Edgeworth box. To be more precise: the set of individually rational cooperative actions $A^{\mathrm{IR}} \subset A$ has the shape of a lens, the noncooperative equilibrium strategies $\overline{\boldsymbol{k}}$ anchor the zero indifference curves, and the set of Pareto efficient actions – denoted by $A^{\mathrm{P}} \subset A$ – corresponds to the contract curve.

We solve $E_i(\boldsymbol{k}) = 0$ for k_{-i} and denote the solution by $\hat{k}_{-i} : A_i \to A_{-i}$ such that $E_i(k_i, \hat{k}_{-i}(k_i)) = 0$ for all $k_i \in A_i$. We can explicitly solve for $\hat{k}_{-i}(k_i)$ with

$$\hat{k}_{-i}(k_i) = \overline{k}_i \ln\left(\frac{k_i}{\overline{k}_i}\right) - k_i + \sum_{j=1}^{2} \overline{k}_j.$$

Consider the graphs of the zero indifference curves through the point $\overline{\boldsymbol{k}}$

$$\mathscr{G}(\hat{k}_2) = \{(k_1, \hat{k}_2(k_1)) \mid k_1 \in A_1\} \subset A,$$
$$\mathscr{G}(\hat{k}_1) = \{(\hat{k}_1(k_2), k_2) \mid k_2 \in A_2\} \subset A.$$

Now we make the following observation: the functions $\hat{k}_{-i}(k_i)$ are strictly concave $\hat{k}''_{-i}(k_i) = -\overline{k}_i/k_i^2 < 0$. The curve $\mathscr{G}(\hat{k}_1)$ is then strictly *convex* in the A space. The curves are tangent $\mathscr{G}(\hat{k}_2) \cap \mathscr{G}(\hat{k}_1) = \{\overline{\boldsymbol{k}}\}$ for the trivial problem $A^{\mathrm{IR}} = \{\overline{\boldsymbol{k}}\}$, but they must intersect twice $|\mathscr{G}(\hat{k}_2) \cap \mathscr{G}(\hat{k}_1)| = 2$ for a nontrivial problem $|A^{\mathrm{IR}}| > 1$ and thus describe a lens.

We can further construct a curve of Pareto efficient action pairs (contract curve) and determine the resulting Pareto frontier $\partial S \subset S$. Define the set of efficient actions by

$$A^{\mathrm{P}} = \left\{ \boldsymbol{k} \in A \;\middle|\; \frac{\partial E_1(\boldsymbol{k})/\partial k_1}{\partial E_1(\boldsymbol{k})/\partial k_2} = \frac{\partial E_2(\boldsymbol{k})/\partial k_1}{\partial E_2(\boldsymbol{k})/\partial k_2} \right\}.$$

The contract curve $\tilde{k}_2 : (0, \overline{k}_1) \to (0, \overline{k}_2)$ given by

$$\tilde{k}_2(k_1) = -\frac{\overline{k}_2}{\overline{k}_1} k_1 + \overline{k}_2$$

solves

$$\frac{\partial E_1(k_1, \tilde{k}_2(k_1))/\partial k_1}{\partial E_1(k_1, \tilde{k}_2(k_1))/\partial k_2} = \frac{\partial E_2(k_1, \tilde{k}_2(k_1))/\partial k_1}{\partial E_2(k_1, \tilde{k}_2(k_1))/\partial k_2}.$$

The efficient actions are then given by $A^{\mathrm{P}} = \{(k_1, \tilde{k}_2(k_1)) \mid k_1 \in (0, \overline{k}_1)\}$. There is a relationship between the contract curve and the Pareto frontier ∂S in the sense that each point on the contract curve defines a point on the Pareto frontier such that

$$\partial S = \left\{ (E_1(\boldsymbol{k}), E_2(\boldsymbol{k})) \mid \boldsymbol{k} \in A^{\mathrm{P}} \right\}.$$

From an economic point of view there are two axioms that should be shared by all bargaining solutions, namely: individual rationality as well as group rationality (efficiency). We thus pick a solution from the set $A^{\mathrm{IR,P}} = A^{\mathrm{IR}} \cap A^{\mathrm{P}}$.

Let us illustrate the results with a numerical example. Consider the following parametrization $(r_1, r_2, \eta_1, \eta_2, \delta) = (0.10, 0.15, -1, -1, 1)$. The equilibrium actions are directly given by $\overline{\boldsymbol{k}} = (1.10, 1.15)$. We are going to solve for the Egalitarian $\boldsymbol{k}^E$ and Utilitarian solution $\boldsymbol{k}^U$ respectively defined by

$$\boldsymbol{k}^E = \underset{\boldsymbol{k} \in A^{\mathrm{IR}}}{\arg\max} \left\{ \min\{E_1(\boldsymbol{k}), E_2(\boldsymbol{k})\} \right\},$$

$$\boldsymbol{k}^U = \underset{\boldsymbol{k} \in A^{\mathrm{IR}}}{\arg\max} \{E_1(\boldsymbol{k}) + E_2(\boldsymbol{k})\}.$$

The two solutions are polar cases describing full egalitarianism

$$E_1(\boldsymbol{k}^E) = E_2(\boldsymbol{k}^E)$$

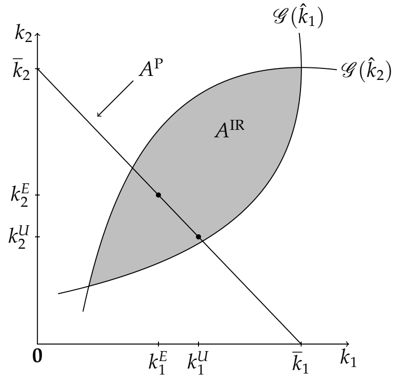
and maximal efficiency in the sense of surplus maximization

$$E_1(\boldsymbol{k}^U) + E_2(\boldsymbol{k}^U) \geq E_1(\boldsymbol{k}) + E_2(\boldsymbol{k}) \quad \forall \boldsymbol{k} \in A^{\mathrm{IR}}.$$

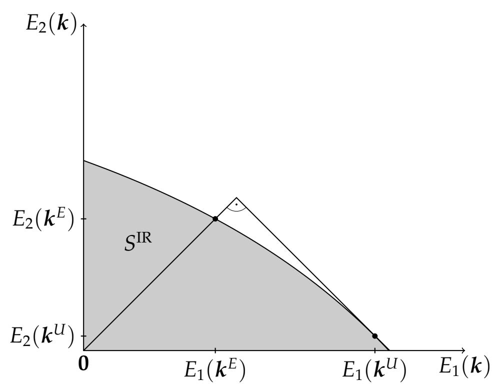
Since both solutions are located on the contract curve $\tilde{k}_2(k_1)$, it is straightforward to calculate them. The action $k_1^E = 0.5047$ solves $E_1(k_1^E, \tilde{k}_2(k_1^E)) = E_2(k_1^E, \tilde{k}_2(k_1^E))$ such that the Egalitarian solution is given by $\boldsymbol{k}^E = (k_1^E, \tilde{k}_2(k_1^E)) = (0.5047, 0.6224)$. For the Utilitarian solution one simply solves

$$k_1^U = 0.6717 = \arg\max_{k_1>0}\{E_1(k_1, \tilde{k}_2(k_1)) + E_2(k_1, \tilde{k}_2(k_1))\}.$$

The Utilitarian solution is then given by $\boldsymbol{k}^U = (k_1^U, \tilde{k}_2(k_1^U)) = (0.6717, 0.4478)$.



(a) action space



(b) payoff space

Figure 2: Egalitarian and Utilitarian bargaining solutions

Figure 2 illustrates the solutions in the action space A (cf. Figure 2a) and IR payoff space S^{IR} (cf. Figure 2b). We find that the emission rate of country 1 is less under egalitarianism than under utilitarianism $k_1^E < k_1^U$ and vice versa for country 2. Apparently the results are driven by the heterogeneous discount rates $r_1 < r_2$, because country 1 can be considered more patient than country 2.

Since country 2 discounts future payoffs more heavily than country 1, it would be also able to emit more pollutants in order to equalize the surplus.

On the other hand, if we were only concerned about total surplus maximization, then country 1 is favored, because it does not discount future payoffs that heavily. In Figure 2b we find the geometric analogon of the analytic results. The Egalitarian solution is given by the intersection of the diagonal and the Pareto frontier, while the Utilitarian solution is given by the tangential point of the line with slope -1 and the Pareto frontier.

6. CONCLUSION

We addressed the question of identifying subgame individually rational (SIR) and time consistent (TC) bargaining solutions for a variety of linear-state differential games. For these kind of games there exists a subgame perfect equilibrium in constant strategies. We showed that all overall individually rational bargaining solutions satisfy SIR and TC, if one restricts the space of admissible cooperative strategies to constants. The approach of using cooperative strategies that are functionally equivalent to given equilibrium strategies should be applicable more broadly to design SIR and TC cooperative solutions. In subsequent work we hope to generalize the results for the rich class of linear-quadratic games. There seems to be an inverse relationship, however, between the complexity of the model and the bargaining solutions that satisfy SIR and TC. In an accompanying paper (Hoof, 2020) we showed for a nonlinear model of dynamic cake eating that a bargaining solution must maximize some linear-homogenous function in order to satisfy SIR and TC. Generally, we plan to pin down classes of games such that bargaining solutions can be uniquely identified by SIR and/or TC and that those two properties thus serve as axioms in dynamic models.

References

Başar, T. (1989). Time Consistency and Robustness of Equilibria in Non-Cooperative Dynamic Games. In F. van der Ploeg & A. de Zeeuw (Eds.), *Dynamic policy games in economics* (Vol. 181, pp. 9 – 54). Elsevier.

Başar, T., & Olsder, G. J. (1999). *Dynamic Noncooperative Game Theory* (2nd ed.). SIAM.

Caputo, M. R. (2005). *Foundations of Dynamic Economic Analysis*. Cambridge University Press.

Dockner, E., & Wagener, F. (2014). Markov perfect Nash equilibria in models with a single capital stock. *Economic Theory*, *56*(3), 585 – 625.

Dockner, E. J., Jørgensen, S., Long, N. V., & Sorger, G. (2000). *Differential Games in Economics and Management Science*. Cambridge University Press.

Dockner, E. J., & Van Long, N. (1993). International Pollution Control: Cooperative versus Noncooperative Strategies. *Journal of Environmental Economics and Management*, *25*(1), 13 – 29.

Haurie, A. (1976). A note on nonzero-sum differential games with bargaining solution. *Journal of Optimization Theory and Applications*, *18*(1), 31 – 39.

Hoof, S. (2020). *A pure bargaining game of dynamic cake eating*. (mimeo)

Jørgensen, S., Martín-Herrán, G., & Zaccour, G. (2003). Agreeability and Time Consistency in Linear-State Differential Games. *Journal of Optimization Theory and Applications*, *119*(1), 49 – 63.

Jørgensen, S., & Zaccour, G. (2002). Time Consistency in Cooperative Differential Games. In G. Zaccour (Ed.), *Decision & control in management science* (pp. 349 – 366). Springer.

Kaitala, V., & Pohjola, M. (1990). Economic Development and Agreeable Redistribution in Capitalism: Efficient Game Equilibria in a Two-Class Neoclassical Growth Model. *International Economic Review*, *31*(2), 421 – 438.

Liu, P. T. (1973). Nonzero-sum differential games with bargaining solutions. *Journal of Optimization Theory and Applications*, *11*(3), 284 – 292.

Nash, J. F., Jr. (1950a). Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, *36*(1), 48 – 49.

Nash, J. F., Jr. (1950b). The Bargaining Problem. *Econometrica*, *18*(2).

Nash, J. F., Jr. (1951). Non-Cooperative Games. *Annals of Mathematics*, *54*(2), 286 – 295.

Peters, H. J. (1992). *Axiomatic Bargaining Game Theory* (Vol. 9). Springer.

Petrosyan, L. A., & Zaccour, G. (2018). Cooperative Differential Games with Transferable Payoffs. In T. Basar & G. Zaccour (Eds.), *Handbook of dynamic game theory* (pp. 595 – 632). Springer.

Starr, A. W., & Ho, Y. C. (1969a). Further Properties of Nonzero-Sum Differential Games. *Journal of Optimization Theory and Applications*, *3*(4), 207 – 219.

Starr, A. W., & Ho, Y. C. (1969b). Non-Zero Sum Differential Games. *Journal of Optimization Theory and Applications*, *3*(3), 184 – 206.

Yeung, D. W. K., & Petrosyan, L. A. (2018). Nontransferable Utility Cooperative Dynamic Games. In T. Başar & G. Zaccour (Eds.), *Handbook of dynamic game theory* (pp. 633 – 670). Springer.

Journal of Mechanism and Institution Design
ISSN: 2399-844X (Print), 2399-8458 (Online)
DOI:10.22574/jmid.2020.12.004

MATCHING WITH COMPATIBILITY CONSTRAINTS: THE CASE OF THE CANADIAN MEDICAL RESIDENCY MATCH

Muhammad Maaz

McMaster University, Canada
University of Toronto, Canada
m.maaz@mail.utoronto.ca

Anastasios Papanastasiou

McMaster University, Canada
papanasa@mcmaster.ca

ABSTRACT

The Canadian medical residency match has received considerable attention in the medical community as several students go unmatched every year. Simultaneously, multiple residency positions go unfilled, largely in Quebec, the Francophone province of Canada. In Canada, positions are designated with a language restriction, a phenomenon that has not been described previously in the matching literature. We develop a model of matching with compatibility constraints, where, based on a dual-valued characteristic, a subset of students is incompatible with a subset of hospitals, and show how such constraints lead to inefficiency. We derive a lower bound for the number of Anglophone and Francophone residency positions such that every student is matched for all instances of (a form of) preferences. Our analysis suggests that to guarantee a stable match for every student, a number of positions at least equal to the population of bilingual students must be left unfilled.

Keywords: Two-sided matching, CaRMS, matching with constraints.

JEL Classification Numbers: C78, D82.

We thank two anonymous referees for their helpful comments that led to substantial improvements of the paper. The authors are responsible for any remaining errors.

1. INTRODUCTION

The seminal paper by Gale & Shapley (1962) introduced the deferred acceptance (DA) algorithm as a mechanism for establishing stable matchings in two-sided matching problems. Since then, applications of DA have flourished such as the well-known medical residency match. This application was motivated by Roth's observation that the National Resident Matching Program (NRMP) in the United States, which is responsible for allocating medical school graduates to their post-graduate training (also called a residency), had independently arrived at the Gale-Shapley DA algorithm (Roth, 1984, 2003). In 1999, the DA algorithm was modified to include the ability for student couples to apply to match together. This modified algorithm is called the Roth-Peranson algorithm (Roth & Peranson, 1999), and was adopted in many other countries, including Canada (Canadian Residency Matching Service, 2019). Since then, matching theory has remained a fertile field, both theoretically and practically, with the question of real-world *constraints* inspiring much of the matching work in the 21st century.

In Canada, medical students apply to be matched to postgraduate training (also called a residency) at a Canadian hospital through the Canadian Residency Matching Service (CaRMS) (Canadian Residency Matching Service, 2019), which uses a version of the DA algorithm.[1] The unique constraint that exists is that some positions are designated for French-speaking students in order to provide French services to the public. This is due to the status of French as the second official language of Canada (Esman, 1982). While this guarantees equal status for French and English in federal jurisprudence, some provinces also give French special status. The province of New Brunswick, for example, is officially bilingual, while the province of Quebec, Canada's largest province, is officially unilingually French (Esman, 1982). As well, French is often taught as a second language in English-speaking provinces like Ontario, while English is also taught in Francophone provinces (Esman, 1982).

[1] The CaRMS actually runs four different matches (2019): 1. R-1: This is what graduating or graduated medical students apply to for their postgraduate training. 2. MSM: Medicine Subspecialty Match. This is for residents currently in an internal medicine program seeking to enter subspeciality training. 3. FM/EM: Family Medicine/Emergency Medicine. This is for residents who are currently in or have completed family medicine training and wish to pursue further training in emergency medicine. 4. PSM: Pediatric Subspecialty Match. This is for residents currently in a pediatric residency program who wish to pursue subspecialty training. In this paper, when we talk about the residency match, we are referring to the R-1 match.

According to CaRMS data (2019), in the 2019 R-1 match, 103 out of 2984 Canadian medical graduates (CMG's) went unmatched - meaning that 96.5% did indeed obtain a residency position. While comparing favorably to other residency matching clearinghouses - for example, in the US, 79.6% of applicants to the NRMP were matched in 2019 (National Resident Matching Program, 2019) - much attention in Canada has been drawn to the issue of unmatched medical residents. The Canadian Medical Association (2019) has increasingly been sounding the alarm over the issue of unmatched medical students, as the number of unmatched CMGs steadily increasing every year. Other professional organizations, like the Association of Faculties of Medicine of Canada (2018), abbreviated as the AFMC, have been lobbying the government to respond as well (provincial governments are responsible for funding residency positions). It is worth noting that unmatched medical students cannot practice medicine, despite nearly a decade in school, and are often left with little in terms of job prospects (Association of Faculties of Medicine of Canada, 2018).

In the Canadian medical literature, much discussion has been ongoing as to what to do about the CaRMS. Wilson & Bordman (2017), in a commentary in the Canadian Medical Association Journal, the preeminent general medical journal in Canada, declared that the CaRMS was "broken", citing the fact that 68 graduates went unmatched, while 64 residency positions were unfilled (including 56 in family medicine in the province of Quebec). This commentary attracted attention and responses in the subsequent months, including those from doctors, deans of medical schools, the CaRMS itself, and impassioned personal anecdotes from unmatched graduates (Sorokopud-Jones, 2018; Yeung, 2018; Willett, 2017; Silverberg & Purdy, 2018; Persad, 2018a,b; Moineau, 2018). News media have picked up on the problem of unmatched residents in recent years as well, with considerable coverage surrounding the tragic suicide of Dr. Robert Chu who went unmatched despite attempting to do so twice (Warsh, 2017). The frustration over the CaRMS has even spilled into some professional associations which staged demonstrations outside the Ontario provincial legislature (Association of Faculties of Medicine of Canada, 2018).

Wilson and Bordman's commentary, as well as match data analysis by the AFMC, demonstrated there was a seeming disconnect between the two sides of the matching market. There are more positions than graduates (Association of Faculties of Medicine of Canada, 2018), which at first glance is a favorable situation. Again, comparing with the United States, there are indeed fewer positions than students in the NRMP, so the sub-100% match rate is perhaps easily explained away by that disparity (National Resident Matching Program, 2019). However, in Canada, there are approximately 102 positions for every 100 medical graduates (Association of Faculties of Medicine of Canada, 2018). In addition, it seems that unfilled residency positions tend to largely be in Quebec (Wilson & Bordman, 2017), and Quebec graduates match to other provinces more than other provinces' students match to Quebec (Association of Faculties of Medicine of Canada, 2018). All in all, the plight of the unmatched is one of the most important issues facing the Canadian medical community today.

We now review some related literature. Observations of "undesirable" (from a policymaker's perspective) matches yielded by current matching algorithms led to work on possible modifications to the basic DA algorithm. This is not a new problem. As far back as fifty years ago, McVitie & Wilson (1970) studied the stable marriage problem with unequal sets of men and women. Clearly, by the Pigeonhole Principle (Lakins, 2016), some elements will remain unmatched. McVitie & Wilson (1970) proved the Rural Hospital Theorem, which states that unmatched participants in one stable matching are unmatched in all stable matchings. This result was later restated by Roth (1986) as, in the resident-hospital matching market, "any hospital that fails to fill all of its positions in some stable outcome will not only fill the same number of positions at any other stable outcome, but will fill them with exactly the same residents." The theorem was termed the Rural Hospital Theorem on the basis that rural hospitals tend to have greater difficulty filling their residency positions as they are seen as less desirable than urban ones. From these early results, we can see that the idea of imbalances and disparities arising in matching markets is not new.

The aforementioned urban-rural disparity was observed in the data in countries that used centralized clearinghouses for their medical residents, and some countries became proactive in attempting to manipulate the matching algorithm in order to correct the imbalance. Kamada & Kojima (2010, 2012) studied the Japanese medical residency match, which uses the student-proposing DA. In response to public pressure about the lack of rural doctors, the Japanese government instituted regional quotas based on prefectures (government districts) (Kamada & Kojima, 2010), the idea being to set caps on how many residents may work in urban prefectures. Kamada & Kojima (2010) demonstrated that such tampering with the DA algorithm results in inefficiency and possible instability, as well as a lower match rate (fewer doctors overall receive positions). They proposed a *flexible deferred acceptance* algorithm that results in stability and respects regional quotas, and show, through simulations, that while this still yields a lower match rate than normal DA, it does fill more positions than the Japanese implementation of regional quota DA (Kamada & Kojima, 2012).

The opposite problem of setting floor constraints instead of ceiling constraints is seemingly less tractable. Kamada & Kojima (2010, 2012) point out that floor constraints are likely much harder to use. For example, if no student wants to be matched to a specific region, then individual rationality would be compromised, and even with an individually rational matching, stability is not guaranteed (Kamada & Kojima, 2015). Recent work in the computer science literature has found that checking the mere existence of a feasible matching with floor constraints is $\mathcal{NP}$-complete (Goto et al., 2016). It remains unclear whether such constraints are tractable, and what the definitions of concepts like individual rationality and stability would be in such situations (Goto et al., 2016).

Our paper's contribution is thus twofold. From an economic theory point of view, we study a novel situation that has not been described in other well-studied matching markets in the literature. While there is a growing literature on introducing constraints into matching problems, these papers focus on other constraints, such as quotas. The situation described above in Canada, where due to language designations, a subset of students is incompatible with a subset of residency positions, has not been treated by other papers, to the authors' knowledge. Secondly, with regards to real world applications, given the intense scrutiny around the Canadian residency match, this paper aims to build a theoretical basis that can explain how and why the much-derided outcomes described above have arisen. On this basis, possible solutions to the problems affecting the CaRMS can be developed. This paper therefore serves as an extension of the theory of matching as well as an analysis of the CaRMS match.

2. MODEL

2.1. Preliminaries

As per Roth & Sotomayor (1992), our hospital-residents model is a four-tuple $\langle H, I, q, P\rangle$:

- H is a finite set of hospitals.[2]
- I is a finite set of students. The sets H and I are disjoint.
- q is a vector of hospital capacities: q_h for $h \in H$ gives the capacity of hospital h.
- P is a collection of preference relations, such that:
  - For each $i \in I$, P_i denotes the preferences of student i over $H \cup \{\emptyset\}$, hence we derive the strict preference relation $\succ_i$; so, $h_1 \succ_i h_2$ means that student i strictly prefers hospital h_1 to h_2.
  - For each $h \in H$, P_h denotes the preferences of hospital h over $I \cup \{\emptyset\}$, hence, as with the students, we derive the strict preference relation $\succ_h$, which is defined similarly.[3]

Student i is said to be *acceptable* to hospital h if $i \succ_h \emptyset$, and hospital h is acceptable to student i if $h \succ_i \emptyset$. Alternatively, we will use throughout this paper the terminology that student i **applies to**, or is an applicant of, hospital h, if $h \succ_i \emptyset$.[4]

Note that, since hospitals have a capacity of more than one, in reality they would have preferences between sets of students, not necessarily individual students. However, we will assume that hospitals have *responsive preferences*, meaning that replacing a less-preferred student with a more-preferred one, or filling a vacancy with an acceptable student makes it better off (Roth & Sotomayor, 1992).

A **matching** is a function $\mu : H \cup I \to \mathcal{P}(H \cup I)$ such that (Roth & Sotomayor, 1992):

---

2 Note this is purely semantics. Medical professionals may protest that in Canada it is actually universities that "host" residency positions, and have affiliations with hospitals which is where the resident would actually practice. This is true, however we are using "hospitals" as this is the standard terminology used in the matching literature.

3 If for some student i, $\emptyset \succ_i h$, then i prefers being unmatched to being matched to h. If for some hospital h, $\emptyset \succ_h i$, then h prefers keeping some position unfilled rather than being matched to i.

4 This language of *applying* is from the real-world set-up of the residency match, where, when medical students seek residencies, they go through an application process entailing sending a CV, reference letters, and participating in an interview. At the end of the process, students submit a ranking to the CaRMS (or whichever centralized matching system) of the hospitals they applied to, and similarly hospitals rank the students that submitted applications to them according to the strength of their applications.

1. No hospital exceeds its quota, with some positions possibly left unfilled: $\mu(h) \subseteq I \cup \{\emptyset\}$ such that $|\mu(h)| \leq q_h$ for all $h \in H$,

2. each student is matched to at most one hospital or not at all: $\mu(i) \subseteq H \cup \{\emptyset\}$ such that $|\mu(i)| \leq 1$ for all $i \in I$,

3. student i is matched to hospital h if and only if hospital h is matched to a set containing student i: $i \in \mu(h) \iff \mu(i) = \{h\}$ for all $h \in H$ and $i \in I$.

We call a pair $(h, i) \in H \times I$ a **blocking pair** if i and h are both acceptable to each other, and *both* of the following two conditions hold (Roth & Sotomayor, 1992):

1. $h \succ_i \mu(i)$, and,

2. either $i \succ_h i'$ for some $i' \in \mu(h)$, or, $|\mu(h)| < q_h$ and $i \succ_h \emptyset$

From the concept of a blocking pair we can define one of the central concepts in matching theory: stability. A matching μ is **stable** if and only if there do not exist any blocking pairs under μ (Gale & Shapley, 1962).

2.2. Deferred acceptance algorithm

The current CaRMS configuration uses the Roth-Peranson algorithm, which is the student-proposing deferred acceptance algorithm (Roth & Peranson, 1999). As well, this is the algorithm that we analyze in the context of matching residents to residencies throughout this paper. The **student-proposing deferred acceptance (DA) algorithm** is defined as follows (Roth & Sotomayor, 1992):

Step 1. Each student i proposes to its most preferred hospital. A hospital h receiving more than q_h proposals shortlists its q_h most preferred students according to its preferences P_h, and rejects the rest, while a hospital h receiving less than q_h proposals shortlists all of its proposals.

Step k. Any student i who was rejected at step $k-1$ proposes to the hospital it prefers the most among the hospitals it applied to (i.e. hospitals it finds acceptable) that hasn't rejected it yet. At each step, each hospital h takes the q_h top students from its shortlist and its proposers, and rejects the others.

The algorithm terminates when there are no more rejections. At termination, the matching is given by the shortlists of the hospitals in the most recent step.

The algorithm also gives a stable matching if the hospitals propose (Roth & Sotomayor, 1992), although this can be a different matching than the one given by the student-proposing version. Note that it is possible for there to be stable matchings other than the one yielded by the DA algorithm (Gale & Shapley, 1962).

2.3. Introducing compatibility constraints

We build upon the basic model in Section 2.1. Our motivation for this model comes from the CaRMS language constraints. Namely, every student can be designated as either Anglophone, Francophone, or both (ie. bilingual). On the other hand, the set of hospitals can be partitioned into two disjoint sets on the basis of language as well.[5] Student i and hospital h are **compatible** only if they share the same language characteristic, and are incompatible otherwise. Thus, in our formulation, an English-speaking student can apply only to Anglophone hospitals, and French-speaking students can apply only to Francophone hospitals, and bilingual students can apply to both Anglophone and Francophone hospitals.[6] In addition, hospitals rank only the students that apply to them.[7]

We can generalize the idea of such language incompatibilities to any sort of incompatibility based on some arbitrary two-valued characteristic. In general, we define a **matching with compatibility constraints problem** as a standard hospital-residents model as per Section 2.1 with the following additional constraints:

[5] There is, of course, the situation where one hospital can have some Anglophone positions and some Francophone positions. However, we can simply imagine this hospital as two different hospitals, one containing all the Anglophone positions, and one containing all the Francophone positions. Therefore, the set of hospitals can always be partitioned into two disjoint sets: English and French.

[6] This is the same as having French-speaking-only students prefer no match over a match with Anglophone hospitals, and vice versa for English-speaking-only students.

[7] Note that, in reality, it is the hospitals who impose such restrictions - for example, a hospital restricts its positions to French speakers. It does not necessarily follow that English-speaking students will not apply to Francophone hospitals. However, Irving et al. (2008) have shown that one can assume without loss of generality that preferences are *consistent* in two-sided matching problems, meaning that for hospital h and student i, $h \succ_i \emptyset$ if and only if $i \succ_h \emptyset$. Therefore, it follows that though these language restrictions are exogenously imposed by the hospitals, we can safely say that the students also do not apply to hospitals which would find them unacceptable due to language constraints.

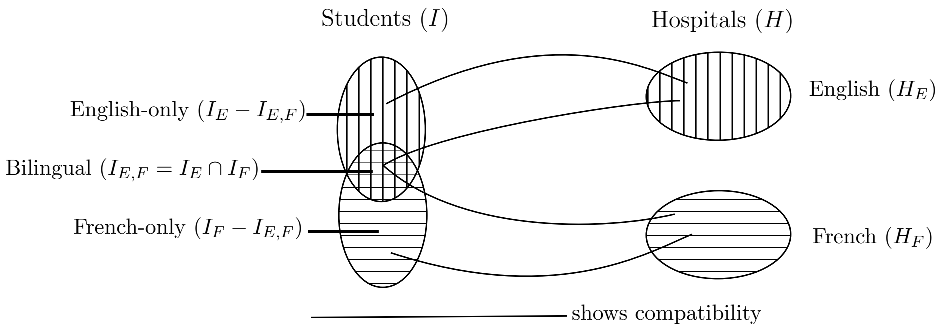


Figure 1: Schematic of matching with compatibility constraints applied to the Anglophone(E)/Francophone(F) constraints in the CaRMS

- There is a two-valued characteristic $C = \{c_1, c_2\}$.
- Each student $i \in I$ has the characteristic c_1, c_2, or both. Let the set of students with characteristic c_1 be denoted as I_1, and the set of students c_2 be denoted I_2, such that $I = I_1 \cup I_2$. Let the intersection of these sets $I_1 \cap I_2$ be denoted $I_{1,2}$. We restrict the sets $I_1 - I_{1,2}$ and $I_2 - I_{1,2}$ (the sets of students who only have c_1 and who only have c_2, respectively) to be non-empty.
- There is a partition of hospitals H into two disjoint sets H_1 and H_2, which correspond to the characteristics c_1 and c_2.
- A student-hospital pair (h, i) is compatible if they share the same characteristic, and incompatible otherwise. A student is unacceptable to a hospital and a hospital is unacceptable to a student if they do not share the same characteristic. Hence, a student can only apply to compatible hospitals, however they may not apply to all. See Figure 1 for a representation.

We now apply this terminology in the context of our example. Our characteristic set is $C = \{E, F\}$, where E is the English-speaking characteristic, and F denotes the French-speaking characteristic. English-speaking-only students $I_E - I_{E,F}$ are incompatible with the Francophone hospitals H_F, while the French-speaking-only students $I_F - I_{E,F}$ are incompatible with the Anglophone hospitals H_E. This is shown in Figure 1.

3. RESULTS

3.1. Stability

Stability is an important consideration in matching markets. As Roth & Sotomayor (1992) have shown, instability often leads to a collapse of matching markets. In order to demonstrate stability, we can show that the matching with compatibility constraints is an instance of the stable marriage with incomplete preferences problem (SMI problem). First introduced by Gale & Sotomayor (1985), an SMI problem is a one-to-one matching problem where preferences are not complete, meaning that some hospital-student pairs are not mutually acceptable. The following lemma will help us to establish stability.

Lemma 1. *The hospital-residents problem with compatibility constraints is an instance of the SMI problem.*

Proof. Let S be a finite set of residency positions. For every hospital $h \in H$ with quota q_h, construct q_h copies of h, each copy with the same preference relation as h, and each copy with capacity of 1. Place these copies in S. Rewrite the preference relations of every student $i \in I$ by replacing every hospital $h \in P_i$ with a list of the elements of S that were derived from h, arbitrarily breaking ties to maintain strict preferences. Now, the many-to-one sided matching problem between I and H has been translated into a one-to-one matching problem between I and S; i.e. it is a stable marriage problem. Due to compatibility constraints, preferences are incomplete (i.e., some hospital-student pairs are not mutually acceptable, per the definition of Irving et al. (2008)). Therefore, it is a stable marriage problem with incomplete preferences. □

This result allows us to immediately establish stability, as follows.

Corollary 1. *With compatibility constraints, DA yields a stable matching.*

Proof. Gale & Sotomayor (1985) showed that the DA algorithm yields a stable matching for the SMI problem. Combining this result with Lemma 1 completes the proof. □

Therefore, we have shown that even when compatibility constraints are introduced as per Section 2.3, the DA algorithm still finds a stable matching.

3.2. Existence of unmatched students in stable matchings

As touched upon in the introduction of the paper, a key issue in the CaRMS is that some students go unmatched, despite more residency positions than students. As well, many positions also go unfilled, largely in Quebec. With our matching with compatibility constraints framework, we can demonstrate that such a result is theoretically possible with the following motivating example.

Consider a case where there are one English-speaking-only student, one bilingual student, and one French-speaking-only student. At first glance it seems that one should only need three positions, since there are only three students, say 2 Anglophone and 1 Francophone positions. But, the problem with this is that if the bilingual student places the Francophone position as first in his preferences, and likewise the Francophone position does so to the bilingual student, they will be matched after running student-proposing DA. This leaves the French-only student without a position. On the other hand, if there are 1 Anglophone and 2 Francophone positions, then the bilingual student could out-compete the English-speaking-only student analogous to the above case, leaving the English-speaking-only student without a position.

This contrasts with the well-known result that when there are as many students as residency positions, and preferences are complete, then there are no unmatched students and no unfilled positions after running DA (Roth & Sotomayor, 1992).

We can look further at the case where there are *more* residency positions than students. For example, in the CaRMS, there are about 102 positions for every 100 students (Association of Faculties of Medicine of Canada, 2018). Observe that in the example where the bilingual student ranks the Anglophone position first, and vice versa the Anglophone position ranks it first, then adding further Francophone positions does nothing to help the overall match rate, as the English-speaking-only student is still left without a position - and indeed leaves those Francophone positions unfilled. This mirrors the current situation in the CaRMS where English-speaking-only students seem to be bear the brunt of the unmatched issue, while Francophone positions go unfilled.

However, now observe what would happen if there were 2 Anglophone and 2 Francophone positions. Then, even if the bilingual student gets matched to an Anglophone position, there is still one left over for the English-speaking-only student. Similarly, they cannot compete the French-speaking-only student out of a position because there is still one position left over for the French-speaking only student. This example provides the motivation for the following section.

3.3. Establishing an I-saturating stable matching

An I-saturating matching is defined as a matching in which, for all $i \in I$, $\mu(i) \neq \{\emptyset\}$ (Gibbons, 1985). Thus an I-saturating stable matching is a matching that is also *stable*.

As the illustrative example above showed, it is insufficient to set the number of positions equal to the number of students. Consideration must be given to the number of Anglophone and Francophone positions individually. As well, the role of preferences is important. For example, with 2 Francophone and 1 Anglophone position, if the bilingual student is matched to the Francophone position then no student will go unmatched. However, the issue is that a social planner choosing how many hospital positions to have (which mimics the situation in Canada well, as funding for residency positions comes from the government) does not know a priori how the students will rank the hospitals nor how hospitals will rank students.[8] If only the very limited information of how many there are in each class is known, how many residency positions should be allocated, such that every student obtains a position *no matter* what ends up transpiring during the residency application process? In the vein of the illustrative example, we will establish a necessary and sufficient condition such that no student is unmatched in all possibilities of (a form of) preferences.

First, we introduce a new definition for preference completeness. If every student finds all their respective compatible hospitals acceptable, and vice versa, i.e., every hospital finds all of their respective compatible students acceptable, then we say that preferences are **compatibility-wise complete**, or CW-complete. To use the language of Irving et al. (2008) and others who study the SMI problem, *complete preferences* would be satisfied if every hospital-student pair is acceptable to both hospital and student; due to compatibility constraints, this is generally not possible. However, CW-complete preferences are, in that sense, *as complete* as preferences can be under these constraints.

Next, we introduce some additional notation to make the statement easier

[8] There are a host of factors that contribute to how hospitals rank applicants, including marks, reference letters, academic publications, and community service (Lakoff et al., 2020). Similarly, there are a host of factors that contribute to how students rank hospitals, including prestige, reputation in a particular medical field (for example, students interested in trauma medicine would like to go to premier trauma centers), family, and cost of living (Dow et al., 2020).

to read. Let the set of English-speaking-only students be E, the set of French-speaking-only students be F, and the set of bilingual students be B, with sizes e, f, and b, respectively. These are all subsets of I, and we label their elements as: $E = \{i_1^E, i_2^E ... i_e^E\}$, $F = \{i_1^F, i_2^F ... i_f^F\}$, and $B = \{i_1^B, i_2^B ... i_b^B\}$. Let the set of Anglophone hospitals be X and the set of Francophone hospitals be Y, with total quotas x and y, respectively. We restrict $e, f, b, x, y > 0$. Let the set of all possible CW-complete preferences be $\mathbb{P}$. Then, we can show the following result.

Theorem 1. *Every stable matching is I-saturating in all instances of CW-complete preferences if and only if $x \geq e + b$ and $y \geq f + b$. Formally: $(\forall P \in \mathbb{P})$(every student has a position in all stable matchings)$\Leftrightarrow$ $(x \geq e + b) \wedge (y \geq f + b)$.*

Proof. We first prove the 'if' part of the statement: every stable matching is I-saturating in all instances of CW-complete preferences if $x \geq e + b$ and $y \geq f + b$.

Suppose to the contrary that student i does not have a position in some stable matching. Let the number of students with the same characteristic as i (note this includes bilingual students), including i, be k. Note that if i is bilingual, then k is the number of students that share at least one characteristic with i, and so k is the number of English-speaking-only, French-speaking-only, and bilingual students - i.e. *all* students.

As preferences are CW-complete and the matching is stable, student i does not form a blocking pair with any of its compatible hospitals. By assumption, the number of positions with i's characteristic is weakly greater than k, and because students cannot occupy more than one position, at most $k - 1$ positions with i's characteristic are filled, and at least one position compatible with student i is left unfilled. By CW-completeness, student i and the hospital with that unfilled position form a blocking pair. Thus, the matching is not stable, which is a contradiction. Therefore, every student is matched in all stable matchings for all instances of CW-complete preferences.

Next, we prove the 'only if' part of the statement. Consider its contrapositive: $(x < e + b) \vee (y < f + b) \Rightarrow (\exists P \in \mathbb{P})$(there exists a stable matching at which some student is unmatched). It suffices to show the existence of such a matching μ for some P, so we will use a constructive proof.

First, consider the case where $x < e + b$. Consider an instance P in which:

- Every Anglophone hospital h: $i_m^B \succ_h i_n^E$ for all $m \leq b$ and for all $n \leq e$.

- Also, every Anglophone hospital h: $i^E_m \succ_h i^E_n$ for all $m < n$.

- Every bilingual student i: $h_x \succ_i h_y$ for all $h_x \in X$ and for all $h_y \in Y$.

We need to show that some student is left unmatched in some stable matching. We show that student i^E_e is left unmatched in *every* stable matching. Suppose not, so there is a stable matching μ with $x < e + b$ in which i^E_e is assigned to an Anglophone hospital h. From the definition of stability, h does not form a blocking pair with any bilingual or English-speaking-only student. Under the above preferences, this is only possible if all English-speaking and bilingual students are matched to some other Anglophone hospital that they prefer to h. This implies that μ assigns $e+b$ students to $x < e+b$ Anglophone positions, an impossibility. Thus, student i^E_e is left unmatched in every stable matching.

Along the same lines, we can show that when $y < f + b$ some students go unmatched in some stable matching for some P. For example, consider a set of CW-complete preferences in which:

- Every Francophone hospital: $i^B_m \succ i^F_n$ for all $m \leq b$ and for all $n \leq f$.

- Also, every Francophone hospital: $i^F_m \succ i^F_n$ for all $m < n$.

- Every bilingual student: $h_y \succ h_x$ for all $h_x \in X$ and for all $h_y \in Y$.

Following the same steps as in the first case, we obtain that student i^F_f is left unmatched in *every* stable matching, which implies the required result. This completes the proof for the 'only if' part of the statement.

□

Theorem 1 shows that $(x \geq e+b) \wedge (y \geq f+b)$ is needed for every stable matching to be I-saturating. The next result establishes that this condition is also needed for the existence of an I-saturating matching in every instance of preferences.[9]

Lemma 2. *If there exists a stable matching in which every student is matched* $\forall P \in \mathbb{P}$, *then* $(x \geq e+b) \wedge (y \geq f+b)$.

[9] We thank an anonymous referee for constructive suggestions on structuring the proof of Lemma 3.2.

Proof. We need to show $(\forall P \in \mathbb{P})$ (there is some stable matching in which every student is matched) $\implies (x \geq e + b) \wedge (y \geq f + b)$, equivalently, $(\nexists P \in \mathbb{P})$ (there is some student unmatched in every stable matching) $\implies (x \geq e + b) \wedge (y \geq f + b)$. It suffices to prove the contrapositive $(x < e + b) \vee (y < f + b) \implies (\exists P \in \mathbb{P})$(there is some student unmatched in every stable matching). The result follows directly from the second part of our proof of Theorem 1, by observing that the argument is established for every stable matching. □

In plainer words, Lemma 2 means that, in order to guarantee every student a match in all CW-complete preference possibilities, the number of Anglophone positions needs to be at least equal to the number of English-speaking students (including bilingual students) and the number of Francophone positions needs to be at least equal to the number of French-speaking students (including bilingual students). For example, if we have 5 students (2 English-speaking-only, 2 French-speaking-only, and 1 bilingual), then in order to ensure that every student is matched (assuming CW-completeness), no matter what the preferences are, we would actually need 6 positions (3 Anglophone and 3 Francophone) instead of, as we might think at first glance, 5 positions for 5 students.

Such a requirement on positions would also guarantee, by Theorem 1, that every student is matched in every stable matching, when preferences are CW-complete. Since the student-proposing DA algorithm specifically gives the student-optimal stable matching (Roth & Sotomayor, 1992), this result also applies to the special case of the CaRMS, a fact which might be useful in policymaking.

From Theorem 1 and Lemma 2 we obtain the following result:

Corollary 2. *If there exists a stable matching under which every student is matched for all CW-complete preferences, at least as many positions as the number of bilingual students are left unfilled in every stable matching.*

Proof. If the assumption in the statement holds, by Lemma 2 $x \geq e + b$ and $y \geq f + b$ and by Theorem 1 exactly $f + e + b$ students are matched, and so $x + y - e - f - b \geq b$ positions are left unfilled, in every stable matching. □

In general, from conditions $x \geq e + b$ and $y \geq f + b$ in Theorem 1 and Lemma 2, the total number of positions $x + y \geq e + f + 2b$ is greater than the number of students, $e + f + b$ and some positions must remain unfilled.[10] Corollary 2 shows that at least one position for every bilingual student is left unfilled, and demonstrates the inefficiency of introducing compatibility constraints in matching markets, in the sense that if language restrictions are lifted, all positions are filled and no students are left unmatched in every stable matching with complete preferences.[11] Our condition implies that there is an inherent trade-off for the policymaker deciding how many residency positions to fund: setting the number of residency positions in accordance with the lower bound of Theorem 1 would mean that every student is matched, but would also mean some positions would be unfilled, which could be a waste of resources. The policymaker must therefore consider these two opposing goals: guaranteeing a match for every student, or filling every residency position.

4. CONCLUDING REMARKS

In this paper we developed the matching with compatibility constraints model, where a dual-valued characteristic causes a subset of students to be incompatible with a subset of hospitals, in order to investigate the phenomenon of language restrictions in the Canadian medical residency match. This is, to our best knowledge, the first paper to investigate this unique feature of the Canadian residency match and use it to explain its present problems under the lens of standard two-sided matching theory. Notably, we investigated theoretically how this could lead to the current issue of unmatched students and unfilled positions observed in the CaRMS. We showed that even when there are more residencies than students, as is the case in Canada, it is not guaranteed that every student is able to obtain a position.

We defined a weaker form of preference completeness, called compatibility-wise completeness, or CW-completeness, which is as complete as preferences can be under compatibility constraints. We then showed that when we assume CW-completeness (i.e. all English-speaking students apply to all Anglophone

[10] Except for the degenerate case when $I_E \cap I_F = \emptyset$, meaning $b = 0$, we effectively have two separate standard hospital-residents problems: one between the English students and hospitals, and one between the French students and hospitals. Then, it suffices to have the number of Anglophone positions equal to the number of English students, and similarly for the number of Francophone positions equal to the number of French students.

[11] Assuming that every student prefers to be matched to their least preferred hospital over going unmatched, and every hospital prefers to hire its least preferred student than leaving a position empty.

residencies), then we can guarantee every student obtaining a position by having the number of Anglophone positions equal to the number of English-speaking students and the number of Francophone positions equal to the number of French-speaking students. Interestingly, the total required number of positions to guarantee this is greater than the number of students - which contrasts with the result in standard matching models stating that under complete preference relations, having positions equal in number to the students guarantees a match for everyone. Unfortunately, even given this guarantee, we cannot solve the problem of unfilled residency positions. Rather surprisingly, the number of bilingual students leads to an increase in inefficiency, in the sense of unfilled positions in Corollary 2.

The real-world applicability of this prescription may be limited as preferences in the real world are not likely to be CW-complete. There are significant logistical hurdles that applicants to residency positions must pass through for each application, including reference letters and interviews. Due to this, medical students in the CaRMS do not rank all hospitals with whom they are compatible. Taking into this account, the number of required residency positions to guarantee that every student matches is likely to be larger, albeit by an unknown amount, than the requirement under CW-complete preferences.

Our model certainly has implications for the CaRMS and the analysis of the current issues that have received so much attention in the medical community. Its generalized formulation in terms of arbitrary two-valued characteristics allows it to be applied to any variant of one-to-one and many-to-one matching situations. For example, in a marriage market, it could be used to analyze the effect of the existence of religious preferences. Future theoretical work could take this framework in numerous directions. In addition, it would be interesting to see how the framework applied empirically to, for instance, the study of the CaRMS. It would be interesting to see how varying the number of Anglophone and Francophone positions affects the match rate by simulating the CaRMS. We leave it to future theoreticians and empiricists to build upon the results laid out in this paper.

References

Association of Faculties of Medicine of Canada. (2018, Jan). *Reducing the number of unmatched Canadian medical graduates: A way forward* (Tech. Rep.). Association of Faculties of Medicine of Canada.

Canadian Medical Association. (2019, Apr). *CaRMS update: 40% fewer students go unmatched over last year.* https://www.cma.ca/carms-update-40-fewer-students-go-unmatched-over-last-year.

Canadian Residency Matching Service. (2019, Dec). *Home.* https://www.carms.ca/.

Dow, T., McGuire, C., Crawley, E., & Davies, D. (2020). Application rates to surgical residency programs in Canada. *Canadian Medical Education Journal*, *11*(3), e92–e100.

Esman, M. J. (1982). The politics of official bilingualism in Canada. *Political Science Quarterly*, *97*(2), 233–253.

Gale, D., & Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, *69*(1), 9–15.

Gale, D., & Sotomayor, M. (1985). Some remarks on the stable matching problem. *Discrete Applied Mathematics*, *11*(3), 223–232.

Gibbons, A. (1985). *Algorithmic graph theory*. Cambridge university press.

Goto, M., Iwasaki, A., Kawasaki, Y., Kurata, R., Yasuda, Y., & Yokoo, M. (2016). Strategyproof matching with regional minimum and maximum quotas. *Artificial intelligence*, *235*, 40–57.

Irving, R. W., Manlove, D. F., & Scott, S. (2008). The stable marriage problem with master preference lists. *Discrete Applied Mathematics*, *156*(15), 2959–2977.

Kamada, Y., & Kojima, F. (2010). Improving efficiency in matching markets with regional caps: The case of the Japan residency matching program. *Discussion Papers, Stanford Institute for Economic Policy Research*, *1*.

Kamada, Y., & Kojima, F. (2012). Stability and strategy-proofness for matching with constraints: A problem in the Japanese medical match and its solution. *American Economic Review*, *102*(3), 366–70.

Kamada, Y., & Kojima, F. (2015). Efficient matching under distributional constraints: Theory and applications. *American Economic Review*, *105*(1), 67–99.

Lakins, T. J. (2016). *The tools of mathematical reasoning*. American Mathematical Soc.

Lakoff, J., Howse, K., Cofie, N., Heeneman, S., & Dalgarno, N. (2020). Analysis of factors affecting Canadian medical students' success in the residency match. *Canadian Medical Education Journal*, *11*(3), e43–e55.

McVitie, D. G., & Wilson, L. B. (1970). Stable marriage assignment for unequal sets. *BIT Numerical Mathematics*, *10*(3), 295–309.

Moineau, G. (2018). Association of faculties of medicine of Canada response to: The unmatched by Dr. Amit Persad. *Canadian Medical Education Journall*, *9*(2), e95–e96.

National Resident Matching Program. (2019, Sep). *2019 main residency match by* *the numbers*. National Resident Matching Program http://www.nrmp.org/new-2019-main-residency-match-by-the-numbers.

Persad, A. (2018a). The overall culture of residency selection needs fixing. *Canadian Medical Association Journal*, *190*(14), e443–e443.

Persad, A. (2018b). The unmatched. *Canadian Medical Education Journal*, *9*(2), e89–e92.

Roth, A. E. (1984). The evolution of the labor market for medical interns and residents: A case study in game theory. *Journal of Political Economy*, *92*(6), 991–1016.

Roth, A. E. (1986). On the allocation of residents to rural hospitals: a general property of two-sided matching markets. *Econometrica: Journal of the Econometric Society*, 425–427.

Roth, A. E. (2003). The origins, history, and design of the resident match. *Jama*, *289*(7), 909–912.

Roth, A. E., & Peranson, E. (1999). The redesign of the matching market for american physicians: Some engineering aspects of economic design. *American economic review*, *89*(4), 748–780.

Roth, A. E., & Sotomayor, M. (1992). Two-sided matching. *Handbook of game theory with economic applications*, *1*, 485–541.

Silverberg, S. L., & Purdy, K. M. (2018). Unmatched Canadian medical graduates. *Canadian Medical Association Journal*, *190*(4), e118–e118.

Sorokopud-Jones, M. (2018). The increase in unmatched Canadian medical graduates: Who is to blame and should we be concerned? *University of Manitoba Journal of Medicine*, *1*(1).

Warsh, F. (2017, Jun). Robert Chu's suicide sends a message medicine cannot ignore. *Toronto Star*. Retrieved from https://www.thestar.com/opinion/commentary/2017/06/20/robert-chus-suicide-sends-a-message-medicine-cannot-ignore.html

Willett, J. (2017, Dec). What to do about the Canadian residency matching process [letter]. *Canadian Medical Association Journal*. Retrieved from https://www.cmaj.ca/content/what-do-about-canadian-residency-matching-process

Wilson, C. R., & Bordman, Z. N. (2017). What to do about the Canadian resident matching service. *Canadian Medical Association Journal*, *189*(47), e1436–1447.

Yeung, E. Y. (2018, Sep). Canadian Resident Matching Service: are we trying to reinvent the wheel? [letter]. *Canadian Medical Association Journal*. Retrieved from https://www.cmaj.ca/content/canadian-resident-matching-service-are-we-trying-reinvent-wheel

Journal of Mechanism and Institution Design
ISSN: 2399-844X (Print), 2399-8458 (Online)
DOI:10.22574/jmid.2020.12.005

RIGHTS AND RENTS IN LOCAL COMMONS

Sjur Didrik Flåm

University of Bergen, Norway
sjur.flaam@uib.no

ABSTRACT

Motivated by management problems in national fisheries, we examine management of renewable resources in local or regional commons. This paper suggests that property rights, or lack thereof, be replaced by well-defined user rights. It shows that the use of commons can be conditioned, paid for, or valued, via market mechanisms. To that end, direct deals and double auctions are expedient. Either institution can distribute, restore and secure resource rent. Either can also focalize debates as to which assignments, regulations or taxation of rights might be fair or legitimate.

Keywords: Rights, rent, commons.

JEL Classification Numbers: C71, D33, D51, Q22.

MSC Classification: 46N10, 91A80.

1. INTRODUCTION

Many commons are local or regional. Yet their governance or use is often difficult or disputed - and especially so after disruptive changes. This paper aims to address what arrangements might prove efficient, fair and stable.

For a simple example, imagine an old street with no driveways or garages, where in modern times, too many cars are parked both legally or illegally. Local welfare may improve then by giving regular, specified parking permits to residents only (Fisher, 2017).

My sincere thanks go to the four anonymous referees, and my colleagues P. Bakka, I. Gaasland, R. Hannesson and B. Sandvik for their helpful feedback, and the Informatics department for excellent facilities, and Røwdes Fond for its financial support.

This example of localized commons showcases typical characteristics. Neither use nor value of the commons is static. Indeed, either feature is much affected by changes in exogenous conditions. These facts motivate this paper to consider similar settings and inquire: *can permits - and rental use of these - mitigate dissipation of welfare? May well-defined user rights entail fair and legitimate sharing of surplus*?

Implicitly, these questions reflect the necessity of *collective action* (Hughes, 1976; Olson, 1965) - or of environmental regulation and *law* (Fisher, 2017; Standing, 2019).[1] As usual, large parts of law must be applied. In particular, the shared use of local commons ought to account for at least three defining aspects. First, throughout history, access has often depended on proximity, customary privileges or tacit conventions. Second, insiders seldom play on a level playing field with outsiders. Third, rights are not fully or approximately amended by changed abundance or use of the resources in question.

Against this backdrop, the paper considers restricted access to circumscribed commons - alongside rationed use of valuable resources therein. For an important and motivating instance, imagine a coastal region or nation that controls and determines the total quotas fishermen can catch, within its exclusive waters, during the season.[2] Suppose those quotas be split according to fixed rules among legitimate parties. More precisely, suppose that qualified agents are allotted specific shares.

The privileges thus handed out are valuable allowances to "catch" specified amounts of various "species". Such allowances come as *user rights*, dated and with time-limited validity. Those considered here are local in nature and entail *no* property rights. Hence they are neither heritable nor permanently transferable (Moxnes, 2012). Yet, in the short or medium term, licensed users can ask or bid for such rights on efficient markets. That is, in the main, suppose rights are *rented* or *leased*, but *not* owned.

This assumption echoes Ostrom (1990) on good governance in that common resources need not be privatized. To wit, *empirical* studies find that agent-based exchanges and negotiations can promote efficiency and resolve conflicts; (see Chap. 3 in Ostrom, 1990). Such findings invite a search for *theoretical* underpinnings, and they motivate this paper to inquire: *Might market-like mechanisms - operated by agents themselves - restore and safeguard substantial*

[1] Thereby the permits acquire economic and legal status.

[2] Straddling stocks, migrating across high seas and between diverse economic zones, do not fit the frames of this paper (Ekerhovd et al., 2021). By contrast, sharing of local water does.

parts of the resource rent? *Can the distribution of realized rent also be fair and legitimate* (Young, 1994)?

Each question has twin faces. One, regards *institutions* - as fits the resurgence of *political economics* (Bowles, 2004; Hodgson, 2000), (Ostrom, 1990). The other face regards *solution concepts* - as fits *mathematical economics* (Debreu, 1954, 1959), here just invoking simple convex analysis (Mordukhovich & Nam, 2013). What combines the two faces is disequilibrium *processes*, mediated by stable institutions which facilitate convergence to - or emergence of - steady-state solutions. The rules remain fairly fixed, yet interactions are dynamic, open and evolutionary.

The paper is planned as follows. Section 2 provides preliminaries and briefly describes the generic agent who can use, or wants to use, the commons and resources that come with it. Section 3 depicts diverse stake-holders. Section 4 formalizes the allocation of aggregate quotas. Section 5 considers efficiency of allocations - best characterized and implemented by shadow prices. The latter define - and implicitly divide - total resource rent. Section 6 captures how shadow pricing can comply with competitive equilibrium and induce corresponding behavior (Flåm, 2020). Section 7 considers whether direct deals or double auctions can bring stakeholders to such equilibrium. Section 8 concludes.

The paper addresses the above questions asked by analysts of institutions and mechanisms, economists concerned with resource management and markets - as well as scholars of environmental law. Alongside these are computer scientists who study agent-based models. Elements of convex analysis provide the main tools.

2. PRELIMINARIES

Let X denote a real vector space. For simplicity, X has a finite dimension. Hence $X = R^S$ for some set S denoting a finite list of ”species”.[3]

Any vector $(x_s) = x \in X$ is seen as a quantified bundle of diverse natural resources, all stemming - or taken out - from some local or regional commons. By assumption, rights regarding such bundles are perfectly added, divided and transferred - without friction or transaction costs - among various identifiable and legitimate parties.

[3] $x \in X$ can code a *contingent claim*, written on scenarios or states $s \in S$, revealed ex post. Then, if S is infinite, X cannot be Euclidean. Extension is possible.

It unburdens discourse and notation to use coordinate-free expressions. Accordingly, the shorthand $x^* \in X^*$ stands for a linear functional $\hat{x} \in X \mapsto x^*\hat{x} := x^*(\hat{x}) \in R$, seen here as an *arbitrage-free price* regime.[4] Given any *gross payoff* function

$$\pi : X \mapsto R \cup \{-\infty\}, \tag{1}$$

such a price regime $x^* \in X^*$ is declared a *supgradient* of π at $x \in X$, as is signalled by writing

$$x^* \in \partial\pi(x), \quad \text{iff } x \in \arg\max \{\pi(\hat{x}) - x^*\hat{x} : \hat{x} \in X\} \quad \text{with } \pi(x) \text{ finite.} \tag{2}$$

Thus, $x^* \in \partial\pi(x)$ iff the *conjugate function*

$$x^* \in X^* \mapsto \pi^*(x^*) := \sup \{\pi(\hat{x}) - x^*\hat{x} : \hat{x} \in X\} \tag{3}$$

attains a finite, maximal value when $\hat{x} = x$. For interpretation: $x^* \in \partial\pi(x)$ iff the finite, price-taking *profit* $\pi^*(x^*)$ equals gross *revenues* $\pi(x)$ less cost x^*x of *input*:

$$\pi^*(x^*) = \pi(x) - x^*x.$$

Some remarks on objectives follow. One may return to these - or skip them.

Any gross payoff (1) should be *proper*, meaning that the *effective domain*

$$dom\pi := \{\hat{x} \in X : \pi(\hat{x}) \in R\} =: \pi^{-1}(R) \tag{4}$$

is non-empty. Typically, the set $dom\pi$ is closed convex - often bounded whence compact - and most frequently part of R^S_+. To focus arguments, the implicit but rigid *constraint* $x \in dom\pi$ is neither repeated nor spelled out. In short, only feasible points $x \in dom\pi$ are considered, usually belonging to R^S_+.

(*On lack of smoothness*). When payoff π is finite near x and differentiable at that point, the optimality condition (2) takes the customary form $x^* = \pi'(x)$. Boundary choice in $dom\pi$, or lack of smoothness there, cannot be ignored though. So, in the sequel, it is not necessary that payoff π be a differentiable function.

(*Linear instances*). For an important example, if an agent holds resource bundle $x \in X$, acts under a linear criterion $y \in Y \mapsto y^*y \in R$, and uses a

[4] For instance, when $X = R^S$, with S finite, $x^*x = \sum_{s \in S} x^*_s x_s$. *No arbitrage* means that it is impossible to gain anything just by trading user rights. Consequently, pricing x^* must be linear.

resource-consuming linear technology $0 \leq y \in Y \mapsto Ay \in X$, he may aim at best payoff

$$\pi(x) := \sup\{y^*y : x \geq Ay \ \ \& \ \ y \geq 0\}. \tag{5}$$

Then, boundary or extreme point solutions y become the rule. Consequently, π is *closed*,[5] *concave* and piece-wise linear, but not always differentiable or finite.[6] Moreover, when $\pi(x)$ is finite,

$$\partial\pi(x) = \arg\min\ \{x^*x : A^*x^* \geq y^* \ \ \& \ \ x^* \geq 0\}$$

where $A^* : X^* \to Y^*$ denotes the transpose of A. $\diamond$

The generic agent pursues improvement or maximization of his own payoff π (1). For analytic or technical reasons, it is convenient - in fact, desirable - that any such function, as in (5), be closed concave. Further, for modelling and narrative, π should also reflect *transferable utility*. How might these properties emerge? That query is addressed next.

For argument, let $\mathcal{X} := R\times X$ denote the augmented space of pairs $\chi = (r, x)$, each composed of monetary reserve $r \in R$ and user right $x \in X$. Suppose the agent at hand is already entitled to user right $\bar{x} \in X$. He has a *preference relation* $\succsim$ on the larger space $\mathcal{X}$.[7] Let $\succ$ denote *strict preference*. Given any $\chi = (r, x) \in \mathcal{X}$ in the domain of $\succsim$, since more money is strictly desirable, suppose

$$\begin{cases} \hat{r} > r & \Longrightarrow \quad (\hat{r}, x) \succ (r, x), \text{ and} \\ (\hat{r}, \hat{x}) \succ (r, x) & \Longrightarrow \quad (\hat{r} - \varepsilon, \hat{x}) \succ (r, x) \text{ for small enough } \varepsilon > 0. \end{cases} \tag{6}$$

When $\bar{\chi} \in \mathcal{X}$, let

$$\{\succsim \bar{\chi}\} := \{\chi \in \mathcal{X} : \chi \succsim \bar{\chi}\} \tag{7}$$

denote the upper level, *preferred set*. Now, if the agent already holds an ”initial position” $\bar{\chi} := (\bar{r}, \bar{x}) \in \mathcal{X}$, he will pay no more money than

$$\pi(x\,|\bar{\chi}) := \pi(x) := \sup\{r \in R : (-r, x) + \bar{\chi} \succsim \bar{\chi}\}, \tag{8}$$

[5] A function $\pi : X \to R \cup \{-\infty\}$ is *closed* (or *upper semicontinuous*) iff each upper level set $\{x : \pi(x) \geq r\}, r \in R$, is closed.

[6] For this instance, see Flåm & Gramstad (2012) on direct exchange in linear economies.

[7] The order relation $\succsim$ is reflexive and transitive, but not necessarily complete. It might be represented by a *utility function* $U : \mathcal{X} \to R \cup \{-\infty\}$. The latter should then be closed and quasi-concave.

for additional quota $x \in X$. Such threshold payment reflects his reservation or indifference level, conditioned by his actual endowment, holding or "wealth" $\bar{\chi}$. The function π, so defined (8), extends from X to the augmented space $\mathcal{X} = R \times X$ by

$$\pi(\chi\,|\bar{\chi}) := \pi(\chi) := \sup\{r \in R : (-r,0) + \chi + \bar{\chi} \gtrsim \bar{\chi}\}.$$

Moreover, it is additively separable in money with constant slope 1 :

$$\pi((r,0)+\chi) \in R \Longrightarrow \pi((r,0)+\chi) = r + \pi(\chi). \tag{9}$$

So, seen as a "commodity" which affects preferences (6), money trades at unit price 1. Consequently,

$$(r^*,x^*) = \chi^* \in \partial\pi(\chi) \;\; \textit{with}\; \chi = (r,x) \Longleftrightarrow r^* = 1 \;\&\; x^* \in \partial\pi(x).$$

It follows straightforwardly:

Proposition 2.1 (On Indifference Payments). *If the preferred set* (7) *comes closed convex, the indifferent payment function* π (8) *is closed concave.*[8] *Since that function is quasi-linear in money with unit slope* (9)*, the agent acts as though "utility" is transferable.* □

3. THE AGENTS

Fixed here is a finite ensemble I, $\#I \geq 2$, of economic agents.[9] Each member $i \in I$, upon facing - and planning for - the upcoming season, considers *four* features of his role:

∗ *first*, he holds a *right* $\bar{x}_i \in X$ to use renewable resources found in the commons,
∗ *second,* he himself chooses to *take out* some resource bundle $x_i \in X$ from there,
∗ *third,* his resulting revenue or gross payoff is determined then by a proper function

$$x_i \in X \mapsto \pi_i(x_i) \in R \cup \{-\infty\}, \tag{10}$$

interpreted, when $x_i \in dom\pi_i$ (4), as market income less the cost of all other factors than in bundle x_i,

---

8 For simpler notation, the dependence of π on $\bar{x}$ is not emphasized. Clearly, if the entitlement changes, so does π.

9 The setting is a closed club; nobody enters or leaves.

$*$ *fourth*, he values his *excess take-out* $x_i - \bar{x}_i$ by some common price $x^* = (x^*_s) \in X^*$. Thus, he buys (sells) rights to take species $s \in S$ iff $x^*_s(x_{is} - \bar{x}_{is}) > 0$ (resp. < 0).

In short, agent i aims at maximization of his idiosyncratic, economic criterion

$$x_i \in X \mapsto \pi_i(x_i) - x^*(x_i - \bar{x}_i) \in R \cup \{-\infty\}\,. \tag{11}$$

(11) singles out net transaction $x_i - \bar{x}_i$ of natural input factors, valued by common price x^*.

Three types of agents merit special mention. A *first*, called a *pure user* of resources, has no established or historical rights in the commons: his $\bar{x}_i = 0$. Yet, by money-based transactions, he acquires some user right $x_i \neq 0$. For example, the agent in question could be a competent fisherman, already owning fishing gear and vessel, but a priori no permit. He doesn't necessarily dwell in communities adjacent to the commons. Geographically, he might be an outsider, hence given *no* user right for free.

A *second* type $i \in I$, called a *pure holder* of rights, has $\bar{x}_i \neq 0$ but chooses $x_i = 0$. He has no capacity - or no motivation - to use own rights himself. Instead, for suitable payments, he transfers $\bar{x}_i$ directly to others. Alternatively, he could put his rights up for rent, fully or partly, at various platforms for auctions or direct deals (Krishna, 2010). At those venues - or on the opposite side of markets - other parties bid for use of the rights.[10] A pure holder could represent a commune, county or region which borders to - or contains - the commons.[11]

Third, between the said two extremes types, there are others - often many - each of which has $\bar{x}_i$ and x_i both non-zero. Of particular notice are those with $\bar{x}_i = x_i \neq 0$. For interpretation, these are local fishermen - active in that walk of life by long standing and tradition.

In short, three types come forward here: *pure users* versus *pure holders* of rights, and besides or between them, parties who *act in both capacities*. All operate within regulated, stable frames - one season after the other. Considered here is just one generic season.

[10] Modern versions of such platforms are computerized and accessible via the internet.

[11] It might be an agency, institution or syndicate.

4. SEASONAL QUOTAS

Suppose some collective agency or directorate decides - up front, before the season - the aggregate quota $\bar{x}_I = (\bar{x}_{Is}) \in X = R^S$ to be taken, during the season, from diverse stocks $s \in S$ (Clarke, 1976).[12] Presumably, that decision reflects appropriate concerns with sustainability and welfare - alongside competence in resource economics and management.[13]

Thereby, agent $i \in I$ is entitled to take - within specified places and times - his "fraction" $\bar{x}_i = \varphi_i(\bar{x}_I) \in X$ from the prescribed, total allowance $\bar{x}_I$, $\sum_{i \in I} \varphi_i(\bar{x}_I) = \bar{x}_I$. The sharing rules $\varphi_i : X \to X$, $i \in I$, are presumed fixed and *constitutional* in nature - continuous from one season to the next.[14] What member $i \in I$ thus receives is a short-term, non-heritable *user right*. Suppose such rights can be renewed for free - say, annually or periodically[15] - among qualified recipients.

Some remarks conclude this section. All can be skipped; none are essential.

(*On dynamics and stochasticity*). Cost, growth, prices or technologies can change from one season to the next. Thus, *between seasons* the system is dynamic and most likely stochastic. For instance, responding to environmental impacts, the resource manager may impose a seasonal moratorium on catch of different species. When the season starts, everybody perfectly understands the particular framing of the upcoming season.

(*On stable frames*). Interaction is governed by time-invariant, separated com-

[12] The "agency" might be autonomous or independent. Presumably, it is benevolent, legitimate and respected. Other bodies - or the concerned agents themselves - ought to control compliance with rules and penalize violations.

[13] Stock dynamics affect seasonal quotas. Their evolution is not considered here. For this, see Bjørndal & Conrad (1987); Clarke (1976); Gordon (1954); Hanley et al. (2007).

[14] By contrast, the overall abundance $\bar{x}_I$ may evolve from one season to the next. Innovations of various sorts, exogenous shocks and random factors drive changes in $\bar{x}_I$. Considered here is just *one* representative season. By hypothesis, the "constitution" $i \mapsto \varphi_i(\cdot)$ may accommodate - but hardly be overthrown by - novel regulations, states or technologies.

[15] The paper side-steps these issues, simply presuming that rules $i \mapsto \varphi_i(\cdot)$ are given. Economic historians illuminate these issues (Libecap, 1986). It's not precluded that some rules $\varphi_i(\cdot)$ derive from *long-term property rights*. To value these - and maybe tax the proceeds - it appears important that *short-term user rights* be traded on well functioning markets. Suppliers could come forward fairly few - say, as municipalities, regions or the state. In any case, concerns with efficiency, fairness or taxation lend justification to establishment of markets.

petencies and institutions - featuring a resource-managing ”directorate,” fixed user rights, efficient operation of auctions, direct deals or markets - and, of course, presence of ”police” and possible penalties.

(*On strategic behavior*). Instead of individual payoff $\pi_i(x_i)$, which precludes externalities, the more general format $\pi_i(x_i, x_{-i})$, with $x_{-i} := (x_j)_{j \neq i}$, would allow congestion (or overcrowding) - and thereby some efficiency loss - in commons or markets, say of Cournot type; see Flåm (2016). This extended optic distracts, though, from the main purpose here: to identify and distribute (or restore) the fully efficient, pure resource rent. Asymmetric information - a main query in game theory (Goswami et al., 2014) - appears to have minor or negligible impacts in the present setting.

(*On irreversible investments* (Clarke et al., 1979)). In several industries, notably fisheries, use of natural resources requires heavy, highly specialized investment. Then, short-term user rights do not easily square with procurement of expensive, long-lived equipment. *Three* features may partly mitigate the problems. *First*, agents might enjoy perfect foresight - as is frequently assumed in market theory. *Second*, the said equipment is often used in other or similar contexts as well. *Third*, there might be - or eventually emerge - a rental market for “gear and vessels.”

(*On claimants to resource rent*). *Three* parties could collect resource rent:
first, pure holders of rights, leasing these to active users,
second, active users of own rights,
third, the public sector by levying tax on quantities taken out.

Competition may eventually take pure profit away from pure users. Unlike the Malthus-Ricardo theory on income distribution between workers, merchants, and land-owners, here is scarcely any room for gentry (Dorfman, 1989) - be it coastal or landed. The basic reason is that there are no property rights in the commons.

(*On matching of agents*). It is conceivable - albeit not likely - that agents form a disjoint union $F \cup W = I$ of two non-empty, finite sets. Each $f \in F$ is a *firm* which needs input $x_f \in X$ but lacks user right: $\bar{x}_f = 0$. By contrast, each $w \in W$ holds the right to use *wealth* $\bar{x}_w \in X$, but lacks capacity. Then, for stable matchings (f, w), see Fujishige & Yang (2017) and references therein. See also Ex. 259.3 in Osborne & Rubinstein (1994).

5. EFFICIENCY, SHADOW PRICES, AND SHARING

The setting is one of perfectly transferable payoffs, quotas and rights. Consequently, the convoluted criterion

$$x_I \in X \mapsto \pi_I(x_I) := \sup\left\{\sum_{i\in I}\pi_i(x_i) : \sum_{i\in I} x_i = x_I\right\} \tag{12}$$

makes good sense - and especially so for total abundance $x_I = \bar{x}_I$. Henceforth suppose that $\pi_I(\bar{x}_I)$ be finite.

An allocation (x_i) which solves (12) for $x_I = \bar{x}_I$ is declared *efficient*. Call any *supgradient* $x^* \in \partial\pi_I(\bar{x}_I)$ (2) a *shadow price*. Such prices support efficiency:

Proposition 5.1 (On Efficiency, Shadow Prices, and Equal Margins). *Whenever (x_i) is an efficient allocation of the aggregate $\bar{x}_I$, any shadow price $x^* \in \partial\pi_I(\bar{x}_I)$ entails common margins in that*

$$\partial\pi_I(\bar{x}_I) \subseteq \cap_{i\in I}\partial\pi_i(x_i). \tag{13}$$

Conversely, provided $\bar{x}_I = \sum_{i\in I} x_i$, the turned-around inclusion

$$\partial\pi_I(\bar{x}_I) \supseteq \cap_{i\in I}\partial\pi_i(x_i).$$

also holds. If moreover, $\cap_{i\in I}\partial\pi_i(x_i)$ is non-empty, then (x_i) is efficient.

Proof. If $x^* \in \partial\pi_I(\bar{x}_I)$, and (x_i) solves (12), then $\sum_{i\in I}\hat{x}_i = \hat{x}_I$ implies

$$\sum_{i\in I}\pi_i(\hat{x}_i) \le \pi_I(\hat{x}_I) \le \pi_I(\bar{x}_I) + x^*(\hat{x}_I - \bar{x}_I) = \sum_{i\in I}\left[\pi_i(x_i) + x^*(\hat{x}_i - x_i)\right].$$

In this string, posit $\hat{x}_j = x_j$ for each $j \ne i$ to get

$$\pi_i(\hat{x}_i) - x^*\hat{x}_i \le \pi_i(x_i) - x^*x_i \ \text{ for all } \hat{x}_i \in X.$$

Since $i \in I$ is arbitrary, this family of inequalities, alongside (2), implies $x^* \in \partial\pi_i(x_i)\ \forall i \in I$.

Conversely, suppose $x^* \in \cap_{i\in I}\partial\pi_i(x_i)$ and $\sum_{i\in I} x_i = \bar{x}_I$. Since $\pi_i(\hat{x}_i) \le \pi_i(x_i) + x^*(\hat{x}_i - x_i)\ \forall\hat{x}_i \in X$, $\forall i \in I$, summation across I yields

$$\sum_{i\in I}\pi_i(\hat{x}_i) \le \sum_{i\in I}\pi_i(x_i) + x^*\sum_{i\in I}(\hat{x}_i - x_i).$$

In the last inequality, let $\sum_{i\in I} \hat{x}_i = \bar{x}_I$ to see that allocation (x_i) solves (12). Further, the instance $\sum_{i\in I} \hat{x}_i = \hat{x}_I$ entails $\pi_I(\hat{x}_I) \leq \pi_I(\bar{x}_I) + x^*(\hat{x}_I - \bar{x}_I)\ \forall \hat{x}_I \in X$. Hence $x^* \in \partial\pi_I(\bar{x}_I)$. □

Proposition 5.1 invokes no assumptions - apart, of course, from availability of suitable allocations (x_i) and prices x^*. Existence of such "items" is briefly addressed next so as to "reduce" or play down the role of convexity:

Proposition 5.2 (On the Existence of Price-Supported Pareto Optimum). *Suppose that each criterion π_i (10) be proper and sup-compact.*[16] *Then the optimal value $\pi_I(\bar{x}_I)$ (12) is attained.*

If convolution π_I (12) is bounded below near $\bar{x}_I$ and coincides at that point with its closed concave envelope, then $\partial\pi_I(\bar{x}_I)$ is non-empty.

Proof (*sketched*). One argues directly that π_I (12) inherits concavity from its underlying terms π_i - and likewise for having π_I closed. By compactness, the latter property secures that $\pi_I(\bar{x}_I)$ is attained. Finally, any closed concave function, bounded below near $\bar{x}_I$, has non-empty supdifferential $\partial\pi_I(\bar{x}_I)$ (Zălinescu, 2002). □

A running example accommodates just *one* species: $\#S = 1$ so that $X = R$. Suppose agent $i \in I$ has *catch capacity* $\kappa_i \geq 0$ and earns gross revenue $r_i \geq 0$ per unit caught. Then, $\pi_i(x_i) = r_i x_i$ if $x_i \in [0, \kappa_i] = dom\pi_i$ (4) - and $-\infty$ elsewhere. So,

$$\partial\pi_i(x_i) = \begin{cases} (+\infty, r_i] & \text{if } x_i = 0, \\ r_i & \text{if } 0 < x_i < \kappa_i, \\ [r_i, -\infty) & \text{if } x_i = \kappa_i, \end{cases} \tag{14}$$

and $\partial\pi_i(x_i)$ is empty otherwise. Let $I := \{1, \ldots, \#I\}$ and suppose $r_1 > r_2 > \cdots$. For specified *aggregate quota* $q := \bar{x}_I > 0$, define $i(q)$ to be the smallest index $i \in I$, if any, such that

$$\sum \{\kappa_i : i \leq i(q)\} \geq q.$$

Provided $i(q)$ is well defined, the efficient allocation (x_i) has $x_i = \kappa_i$ for each intra-marginal $i < i(q)$. The marginal agent $i(q)$ gets what remains of the total quota. For simplicity in argument, suppose his allotment is strictly interior to

[16] A function π_i (10) is *sup-compact* iff each upper level set $\{\pi_i \geq r\}$, $r \in R$, comes compact.

the interval $[0, \kappa_{i(q)}]$. Every extra-marginal agent $i > i(q)$ gets $x_i = 0$. Thus, each agent i, but at most one, ends up with boundary choice $x_i \in \{0, \kappa_i\}$.

By Proposition 5.1 and (14), user rights trade at unit price $x^* = r_{i(q)} \in \cap_{i \in I} \partial \pi_i(x_i)$. Agent i earns profit

$$\pi_i(x_i) - x^*(x_i - \bar{x}_i) = \begin{cases} r_i \kappa_i - r_{i(q)}(\kappa_i - \bar{x}_i) & \text{if } i < i(q), \\ r_i x_i - r_{i(q)}(x_i - \bar{x}_i) & \text{if } i = i(q), \\ r_{i(q)} \bar{x}_i & \text{if } i > i(q). \end{cases}$$

In the unlikely event that $\sum_{i \in I} \kappa_i < q := \bar{x}_I$, user rights are worthless: $x^* = 0$. $\diamond$

Price-supported allocations are Pareto optimal. They also stand out on other grounds - as is described in the next section.

6. EQUILIBRIUM AND WELFARE

Proposition 5.1 revolves around coincidence of marginal valuations. That feature - combined with price-taking and optimizing behavior - characterizes competitive equilibria:

Definition 6.1 (Competitive Equilibrium in User Rights). *An allocation cum price pair* $[(x_i), x^*] \in X^I \times X^*$ *constitutes a* ***price-taking equilibrium*** *iff*

$$\begin{cases} \textit{quota markets clear}: & \sum_{i \in I} x_i = \bar{x}_I, \\ \textit{with best choices}: & \pi_i(x_i) - x^* x_i = \max\{\pi_i - x^*\} \in R \ \forall i \in I. \end{cases}$$

Proposition 6.1 (On Competitive Equilibrium, Profit and Sharing of Resource Rent). *If agent $i \in I$ regards the price $x^* \in X^*$ of quotas as exogenous, he may aim at the highest profit*

$$\pi_i^*(x^*) := \sup\{\pi_i(\hat{x}_i) - x^* \hat{x}_i : \hat{x}_i \in X\}. \tag{15}$$

Provided $\pi_i(0) \geq 0$, that profit is non-negative. Moreover, for any shadow price $x^* \in \partial \pi_I(\bar{x}_I)$ and allocation (x_i) which solves (12) with $x_I = \bar{x}_I$, agent i can take home the highest total payment

$$\Pi_i(x^*) := \pi_i^*(x^*) + x^* \bar{x}_i = [\pi_i(x_i) - x^* x_i] + x^* \bar{x}_i, \tag{16}$$

composed of pure profit $\pi_i^*(x^*) \geq 0$, derived from production, plus resource rent $\rho_i := x^* \bar{x}_i$ from own entitlement. In particular, a pure resource user i just *collects production profit $\pi_i^*(x^*)$. At the other extreme, a pure right holder i merely receives his part $\rho_i = x^*\bar{x}_i$ of the total resource rent $\rho_I := x^*\bar{x}_I$.*[17]

Proof. (15) is a matter of definition and price-taking behavior. Clearly $\pi_i^*(x^*) \geq \pi_i(0) - x^*0 \geq 0$. When (x_i) solves (12) for $x_I = \bar{x}_I$, and $x^* \in \partial\pi_I(\bar{x}_I)$, it follows from (13) that $x^* \in \partial\pi_i(x_i)$. Hence maximal profit (15) is attained with input x_i (2). So, net value (16) follows. □

As is well known, competitive equilibrium connects to Pareto efficiency via two fundamental welfare theorems (Luenberger, 1995). Those connections are applicable here. At the same time, incentives and stability come to the fore. To wit, consider a *transferable-payoff, cooperative game with player set I* in which any non-empty *coalition* $\mathcal{I} \subseteq I$ could aim for a joint payoff

$$\pi_{\mathcal{I}}(\bar{x}_{\mathcal{I}}) := \sup\left\{\sum_{i\in\mathcal{I}}\pi_i(x_i) : \sum_{i\in\mathcal{I}} x_i = \bar{x}_{\mathcal{I}} := \sum_{i\in\mathcal{I}}\bar{x}_i\right\}.$$

Then, a payoff profile $i \in I \mapsto \Pi_i \in R$ belongs to the *core* (Osborne & Rubinstein, 1994) iff

$$\sum_{i\in\mathcal{I}}\Pi_i \geq \pi_{\mathcal{I}}(\bar{x}_{\mathcal{I}}) \;\; \forall\mathcal{I} \subseteq I \text{ with equality for the grand coalition } \mathcal{I} = I. \quad (17)$$

Proposition 6.2 (On Price-Supported Payoff-Sharing and Core Solutions). *For any $x^* \in X^*$ the payment profile $i \in I \mapsto \Pi_i := \pi_i^*(x^*) + x^*\bar{x}_i$ satisfies the inequalities in* (17). *Moreover, equality holds for the grand coalition $\mathcal{I} = I$ iff $x^* \in \partial\pi_I(\bar{x}_I)$. Thus, any shadow price x^*, alongside any best allocation (x_i) of $\bar{x}_I$ across I, implements a core solution in which agent i gets $\Pi_i = \Pi_i(x^*)$* (16).

Conversely, if x_i solves (15) *and $\sum_{i\in I} x_i = \bar{x}_I$, then (x_i) is a best allocation of $\bar{x}_I$. Further $x^* \in \partial\pi_I(\bar{x}_I)$, and the profile $i \mapsto \Pi_i(x^*)$* (16) *is "in the core"* (17).

Proof. For any non-empty $\mathcal{I} \subseteq I$ and price $x^* \in X^*$, inequality (17) holds

[17] Some goal or *golden rule* (Clarke, 1976) might define a steady yield vector $\bar{x}_I \in X$ alongside a *shadow price* $x^* \in \partial\pi_I(\bar{x}_I)$. Regarding the components of $\bar{x}_I$ as material dividends, furnished by nature, $x^*\bar{x}_I$ equals the total *resource rent*. Absent perfect markets, resource pricing is difficult (Neher, 1990). Formula $x^*\bar{x}_I$ fits, however, competitive economies in which production factors are paid according to their marginal values.

because

$$\sum_{i\in\mathcal{I}} \Pi_i(x^*) \geq \sup\left\{\sum_{i\in\mathcal{I}} [\pi_i(x_i) + x^*(\bar{x}_i - x_i)] : (x_i) \in X^{\mathcal{I}}\right\}$$
$$\geq \sup\left\{\sum_{i\in\mathcal{I}} \pi_i(x_i) + x^*(\bar{x}_i - x_i)] : \sum_{i\in\mathcal{I}} x_i = \bar{x}_{\mathcal{I}}\right\} = \pi_{\mathcal{I}}(\bar{x}_{\mathcal{I}}).$$

In particular, $\sum_{i\in I} \Pi_i(x^*) \geq \pi_I(\bar{x}_I)$. For the converse of the last inequality, note that $x^* \in \partial\pi_I(\bar{x}_I)$ and $\sum_{i\in I} x_i = x_I$ imply $\sum_{i\in I} \pi_i(x_i) \leq \pi_I(x_I) \leq \pi_I(\bar{x}_I) + x^*(x_I - \bar{x}_I)$, whence

$$\sum_{i\in I} [\pi_i(x_i) + x^*(\bar{x}_i - x_i)] \leq \pi_I(\bar{x}_I)\ \forall [x_i] \in X^I. \tag{18}$$

In (18) take supremum over profiles (x_i) to see that $\sum_{i\in I} \Pi_i(x^*) \leq \pi_I(\bar{x}_I)$. Inequality (18) also implies that $x^* \in \partial\pi_I(\bar{x}_I)$. Finally, Proposition 5.1 justifies the claims about attainment of core payments. □

Remarks (*on competitive equilibrium*). The solution concept in Def. 6.1, is *Walrasian* (Debreu, 1959). It implies, strictly speaking, that all trade takes place *once*, *in* equilibrium - at common, clearing prices. This framing begs *two* reservations. *First*, most likely - since data may change from one season to another - equilibrium becomes contingent, depending on exogenous factors. This issue is addressed later by emphasizing that Debreu's concept of *valuation equilibrium* (Debreu, 1954) appears more fitting ; see remark after Proposition 7.4.

Second, the Walrasian setting precludes that rights be traded *out of* equilibrium, prior to the season, maybe repeatedly and at "personal" prices - as will be considered next.

7. RIGHTS TRADED

How might equilibrium come about? In what manner could clearing prices emerge?[18] To address these questions, this section considers two market mechanisms, namely: *direct deals* and *double auctions* - in that order.

[18] Walrasian tâtonnement might illuminate matters, but isn't practical; see Flåm (2020) and the agent-based approach (Bowles, 2004).

Direct deals (Flåm, 2014) are decentralized and iterative. For a backdrop, invoking shadow prices $x^* \in \cap_{i\in I}\partial\pi_i(x_i)$, Proposition 5.1 captures and describes efficient outcomes. Otherwise, if $\cap_{i\in I}\partial\pi_i(x_i)$ is empty, then already $\cap_{i\in\mathcal{I}}\partial\pi_i(x_i) = \varnothing$ for some subset $\mathcal{I} \subseteq I$, comprising at least two agents. In this case, resource valuations differ among members of $\mathcal{I}$; that is, some *bid-ask spread* prevails across $\mathcal{I}$.

That concept, is formally defined next and generalized - beyond price difference in *one* good - to multi-commodity settings. Recall that the *directional derivative* of a concave $\pi_i(\cdot)$ at x_i - with $\pi_i(x_i)$ finite - in direction d_i is given by

$$\pi_i'(x_i; d_i) = \lim_{r\to 0^+} \frac{\pi_i(x_i + rd_i) - \pi_i(x_i)}{r}.$$

Hereafter, for each $i \in I$, suppose $\pi_i'(x_i; d_i)$ is closed in d_i. Then,

$$\pi_i'(x_i; d_i) = \inf\left\{x_i^* d_i : x_i^* \in \partial\pi_i(x_i)\right\}. \tag{19}$$

Definition 7.1 (Bid-Ask Spread). *Let $\|\cdot\|$ denote the norm on X. An agent ensemble $\mathcal{I} \subseteq I$, $\#\mathcal{I} \geq 2$, which actually holds a profile $i \in \mathcal{I} \mapsto x_i \in X$, displays* ***bid-ask spread***

$$\mathfrak{S}_{\mathcal{I}}(\mathbf{x}) := \max_{(d_i)} \left\{ \sum_{i\in\mathcal{I}} \pi_i'(x_i; d_i) : \sum_{i\in\mathcal{I}} d_i = 0, \ \sum_{i\in\mathcal{I}} \|d_i\|^2 \leq 1 \right\}. \tag{20}$$

Remark (*on bid-ask spread*). Suppose (19) is applicable. Then, invoke Theorem 1.86 in (Penot, 2013) to replace the resulting max inf format (20) with inf max. Thus,

$$\mathfrak{S}_{\mathcal{I}}(\mathbf{x}) = \inf_{(x_i^*)} \max_{(d_i)} \left\{ \sum_{i\in\mathcal{I}} x_i^* d_i : x_i^* \in \partial\pi_i(x_i) \ \& \ \sum_{i\in\mathcal{I}} d_i = 0, \ \sum_{i\in\mathcal{I}} \|d_i\|^2 \leq 1 \right\}. \tag{21}$$

Here, the operation $\max_{(d_i)}$ is easily performed. Indeed, using shorthand $\bar{x}^* := \sum_{i\in\mathcal{I}} x_i^*/\#\mathcal{I}$ for the average of supgradients $x_i^* \in \partial\pi_i(x_i), i \in \mathcal{I}$, when some $x_i^* \neq \bar{x}^*$, yields

$$\mathfrak{S}_{\mathcal{I}}(\mathbf{x}) = \min\left\{ [\sum_{i\in\mathcal{I}} \left\|x_i^* - \bar{x}^*\right\|^2]^{-1/2} \sum_{i\in\mathcal{I}} x_i^*(x_i^* - \bar{x}^*) : x_i^* \in \partial\pi_i(x) \right\}. \ \diamond \tag{22}$$

As indicated, a positive spread signals that agents' valuations diverge:

Proposition 7.2 (On Bid-Ask Spreads). *Bid-ask spread* (20) *is non-negative*: $\mathfrak{S}_{\mathcal{I}}(\mathbf{x}) \geq 0$. *Moreover, with each* $\partial\pi_i(x_i)$ *non-empty compact, then*

$$\mathfrak{S}_{\mathcal{I}}(\mathbf{x}) = 0 \Longleftrightarrow \cap_{i\in\mathcal{I}}\partial\pi_i(x_i) \neq \varnothing.$$

Proof. The simple fact that each $\pi_i'(x_i; 0) = 0$ implies $\mathfrak{S}_{\mathcal{I}}(\mathbf{x}) \geq 0$. Further, for any $x^* \in \cap_{i\in\mathcal{I}}\partial\pi_i(x_i)$,

$$\mathfrak{S}_{\mathcal{I}}(\mathbf{x}) \leq \max\left\{\sum_{i\in\mathcal{I}} x^* d_i : \sum_{i\in\mathcal{I}} d_i = 0, \ \sum_{i\in\mathcal{I}} \|d_i\|^2 \leq 1\right\} = 0,$$

in which case, $\mathfrak{S}_{\mathcal{I}}(\mathbf{x}) = 0$. Conversely, when each $\partial\pi_i(x_i)$ is non-empty and compact but $\cap_{i\in\mathcal{I}}\partial\pi_i(x_i) = \varnothing$, use a non-zero vector $\mathbf{d} = (d_i) \in X^{\mathcal{I}}$ to separate the compact convex product set $\Pi_{i\in\mathcal{I}}\partial\pi_i(x_i)$ strictly from the "diagonal" $\{(x_i^*) : \text{all } x_i^* \text{ are equal}\}$. Then, $\sum_{i\in\mathcal{I}} d_i = 0$, and one may take $\sum_{i\in\mathcal{I}} \|d_i\|^2 = 1$. Finally, by choosing the appropriate sign of $\mathbf{d}$, it follows that $\mathfrak{S}_{\mathcal{I}}(\mathbf{x}) > 0$. □

In the running example (see Sect. 5), when $x_1 \in [0, \kappa_1)$ and $x_2 \in (0, \kappa_2]$, agent ensemble $\mathcal{I} = \{1, 2\}$ has spread $\mathfrak{S}_{\mathcal{I}}(\mathbf{x}) = r_1 - r_2 > 0$. Further, the overall spread $\mathfrak{S}_I(\mathbf{x})$ is nil iff $r_{i(q)} = \cap_{i\in I}\partial\pi_i(x_i)$. ◇

When $\mathfrak{S}_{\mathcal{I}}(\mathbf{x}) > 0$, members of $\mathcal{I}$ have good reasons to transact among themselves. Clearly, direct deals within $\mathcal{I}$ are purely redistributive. So, with no loss of generality, agent $i \in \mathcal{I}$ gets an updated user right

$$x_i^{+1} := x_i + \sigma d_i \ \text{ with } \sum_{i\in\mathcal{I}} d_i = 0 \ \text{ and } \sum_{i\in\mathcal{I}} \|d_i\|^2 \leq 1 \tag{23}$$

for some *step-size* $\sigma \geq 0$. It's assumed that $\mathbf{x}^{+1} := (x_i^{+1})$, so defined with $x_i^{+1} = x_i$ for $i \notin \mathcal{I}$, remains feasible, meaning $x_i^{+1} \in dom\pi_i$ - that is, $\pi_i(x_i^{+1}) \in R$ for $i \in \mathcal{I}$.

Assumption (On Payoff Improvements). *Whenever agents* $i \in \mathcal{I} \subseteq I$, $\#\mathcal{I} \geq 2$, *are trading user rights among themselves, redistribution* (23) *entails total payoff improvement*

$$\Delta\pi_{\mathcal{I}}(\mathbf{x}) := \sum_{i\in\mathcal{I}} [\pi_i(x_i^{+1}) - \pi_i(x_i)] \geq \sigma\mathfrak{S}_{\mathcal{I}}(\mathbf{x}) \tag{24}$$

for some "step-size" $\sigma > 0$, *depending on the stage counter* k *- or the number of encounters, so far, at which* $\mathcal{I}$ *traded.*

Repeated direct deals are modelled next as a discrete-time process (23) - in the nature of an **algorithm:**

Start, with $x_i = \bar{x}_i$ for each $i \in I$.
Activate, by some protocol or matching mechanism, an agent ensemble $\mathcal{I} \subseteq I$, $\#\mathcal{I} \geq 2$.
If $\mathfrak{S}_{\mathcal{I}}(\mathbf{x}) > 0$, its members trade among themselves so that (23) yields (24).
Continue to **Activate** some agent ensemble until convergence.

To repeat: at discrete stages $k = 0, 1, \ldots$user right x_i^k for $i \in \mathcal{I}^k \subseteq I$ is updated by

$$x_i^{k+1} := x_i^k + \sigma^k d_i^k \ \text{ with } \sum_{i \in \mathcal{I}^k} d_i^k = 0 \ \text{ and } \sum_{i \in \mathcal{I}^k} \left\| d_i^k \right\|^2 \leq 1$$

Henceforth suppose step-sizes dwindle, meaning $\sigma^k \to 0^+$.

The resulting process depends, of course, on which ensemble $\mathcal{I}^k \subseteq I$ of agents trade at stage k. In particular, it appears prudent that ensembles come forward almost periodically:

Proposition 7.3 (On Eventual Disappearance of Bid-Ask Spreads). *Let the members of some ensemble* $\mathcal{I} \subseteq I$, $\#\mathcal{I} \geq 2$, *trade repeatedly among themselves* (23)*, with no less than a fixed, finite lapse between consecutive stages. Suppose step-sizes dwindle in that* $\sigma^k \to 0^+$, *but* ***not too fast****, meaning that for any subsequence* $\mathbf{x}^k$, $k \in K$, *along which* $\mathcal{I}$ *trades,*

$$\mathbf{x} = \lim_{k \in K} \mathbf{x}^k \ \ \& \ \ \mathfrak{S}_{\mathcal{I}}(\mathbf{x}) > 0 \Longrightarrow \sum_{k \in K} \sigma^k = +\infty. \tag{25}$$

Then, for any subsequential limit $\mathbf{x} = \lim_{k \in K} \mathbf{x}^k$, *it holds* $\mathfrak{S}_{\mathcal{I}}(\mathbf{x}) = 0$.

Proof. Consider any subsequential limit $\mathbf{x} = \lim_{k \in K} \mathbf{x}^k$. Since $\sigma^k \to 0^+$, and the ensemble $\mathcal{I}$ comes on stage almost cyclically, it brings no loss of generality to assume that the members of $\mathcal{I}$ trade (among themselves) at *all* stages $k \in K$. Now, if $\mathfrak{S}_{\mathcal{I}}(\mathbf{x}) > 0$, invoke (25) to see that

$$+\infty > \pi_I(\bar{x}) - \sum_{i \in I} \pi_i(\bar{x}_i) \geq \lim_{k \to +\infty} \sum_{i \in I} \pi_i(x_i^{k+1}) - \sum_{i \in I} \pi_i(\bar{x}_i)$$

$$\geq \sum_{k \in K} \sum_{i \in \mathcal{I}} [\pi_i(x_i^{k+1}) - \pi_i(x_i^k)] \geq \sum_{k \in K} \Delta\pi_{\mathcal{I}}(\mathbf{x}^k) \geq \sum_{k \in K} \sigma^k \mathfrak{S}_{\mathcal{I}}(\mathbf{x}) = +\infty. \quad \square$$

Declare a family $\mathbb{I}$ of subsets from I *decisive* if

$$\mathfrak{S}_{\mathcal{I}}(\mathbf{x}) = 0 \; \forall \mathcal{I} \in \mathbb{I} \Rightarrow \mathfrak{S}_I(\mathbf{x}) = 0. \tag{26}$$

Two examples: (I) If some $i \in I$ has differentiable payoff π_i and $\mathbb{I}$ comprises all agent pairs $\{i, j \neq i\}$, the family $\mathbb{I}$ is decisive. This setting indicates that some ”smooth” agent i serves as a hub for trade.
(II) Suppose $\mathbb{I}$ comprises all ensembles $\mathcal{I} \subseteq I$ with $\#\mathcal{I} = \#S + 1$, and let each be $\partial\pi_i(x_i)$ compact. Then, since $X = R^S$, by Helly's theorem (Valentine, 1964), $\mathbb{I}$ becomes decisive. $\diamond$

Proposition 6.3 immediately implies:

Corollary to Propostion 7.3 (On Competitive Equilibria as Cluster Points). *If the members of each ensemble $\mathcal{I}$, belonging to a decisive family $\mathbb{I}$ (26), trade at least once between themselves within any lapse of some prescribed length, then for every accumulation point $\mathbf{x}$ of traded holdings $\mathbf{x}^k$, $k = 0, 1, \ldots$ it holds $\mathfrak{S}_I(\mathbf{x}) = 0$. If moreover, all $\partial\pi_i(x_i)$ are non-empty compact at such a point $\mathbf{x} = (x_i)$, there exists a price $x^* \in X^*$ such that $[(x_i), x^*]$ is a competitive equilibrium.*

Direct trade unfolds as a discrete-time process. Most exchanges happen *off* equilibrium, and limits depend on sequential matchings of agents.

Considered next is a markedly different institution, called *double auction*. Trade takes place indirectly at such venues, once, and only *in* equilibrium - chiefly depending on endowments:

The double auction (Flåm, 2021) is a one-shot affair. Each participant $i \in I$ anonymously, “simultaneously” and silently submits his payoff function $\pi_i(\cdot)$ - say, “in closed form”, in “a sealed envelope” - to a common auctioneer. Thereby, qua “bidder”, agent i commits to “buy” whatever user right $x_i \in X$ for whatever payment $\leq \pi_i(x_i)$.

The auctioneer convolutes the received functions $\pi_i(\cdot)$, $i \in I$, so as to solve (12), when $x_I = \bar{x}_I$, for an optimal allocation (x_i), supported by some shadow price $x^* \in \partial\pi_I(\bar{x}_I)$. Thereafter, he allots user right x_i to agent i for net payment $x^*(x_i - \bar{x}_i)$. Thus, *the auction house clears*. Aggregate demand equals total supply: $\sum_{i\in I} x_i = \bar{x}_I$. The auctioneer merely redistributes seasonal endowments $\bar{x}_i$ and "charges" payments $x^*(x_i - \bar{x}_i)$ for "excess use".[19] Proposition 5.1 immediately entails:

Proposition 7.4 (Redistribution via a Double Auction). *Suppose no party $i \in I$ behaves strategically but rather submits his function $\pi_i(\cdot)$ to the auctioneer as is. Also suppose the latter solves instance $\pi_I(\bar{x}_I)$ (12) for an efficient allocation (x_i), supported by a shadow price $x^* \in \partial\pi_I(\bar{x}_I)$. Thereby, the auctioneer implements a competitive equilibrium $[(x_i), x^*]$ in user rights.* □

Remarks: (*On double auctions*). As modelled, the double auction relates to *Walrasian tâtonnement*. It is markedly different though. The auctioneer of Walras pushes a price process so as to equilibrate intended demand with supply. The double auction is no process; it convolutes price-quantity curves to have one-shot clearing. Such institutions weaken the customary criticism of the Walrasian paradigm.

However, since payoff $\pi_i(\cdot\,|\bar{\chi}_i)$ going to any party i depends on his endowment $\bar{\chi}_i = (\bar{r}_i, \bar{x}_i)$, there are non-standard *wealth effects*. Specializing Definition 6.1 allows such effects to be incorporated: Following Debreu (Debreu, 1954), a *valuation equilibrium* $[x^*, (x_i)]$ prevails iff

$$\begin{cases} \textit{markets clear}: & \sum_{i\in I} x_i = \bar{x}_I, \\ \textit{with no pure profit}: & \pi_i(x_i) - x^* x_i = \max\{\pi_i - x^*\} = 0 \ \forall i \in I. \end{cases}$$

Then, for any agent $i \in I$, if he is a *consumer*, $\hat{x}_i \succ_i x_i \Longrightarrow x^*\hat{x}_i > x^* x_i$. If he is a *producer*, $x^*\hat{x}_i > x^* x_i \Longrightarrow \hat{x}_i$ is infeasible; see Flåm (2021). With this solution concept, i takes home *pure resource rent* $x^*\bar{x}_i$ and no more.

Notice though, that the customary distinction between consumers and producers does not apply well in commons.

[19] By assumption, the auctioneer takes no fees and discriminates nobody.

8. CONCLUDING REMARKS

Management of local commons and resources brings up important concerns (Ostrom, 1990). Many revolve around how to "get institutions right." This paper advocates stable rules and separation of powers. Three major points stand out:

First, an independent bio-economic agency ought, qua *quantity adjustor*, to care about the aggregate take and the intertemporal allocation of resources, thereby mitigating negative externalities and blocking *tragedies of the commons* (Hardin, 1968).[20]

Second, for efficiency[21] and fairness, the paper recommends that user rights should replace or supplement property rights.[22] Under fair sharing, a rental market might enhance efficiency and eventually meet with common acceptance.

Third, there is no escape from strict rules, tight control and suitable penalties.

References

Bjørndal, T., & Conrad, J. M. (1987). The dynamics of an open access fishery. *Canadian Journal of Economics*, *20*, 74–85.

Bowles, S. (2004). *Microeconomics*. Princeton University Press.

Clarke, C. W. (1976). *Mathematical bioeconomics*. J. Wiley & Sons, New York.

Clarke, C. W., Clarke, F. H., & Munro, G. R. (1979). The optimal exploitation of renewable resource stocks: Problems of irreversible investment. *Econometrica*, *47*(1), 25–47.

Debreu, G. (1954). Valuation equilibrium and Pareto optimum. *Proceedings of the National Academy of Sciences*, *47*(1), 588–92.

Debreu, G. (1959). *Theory of value*. Yale University Press, New Haven.

Dorfman, R. (1989). Thomas Robert Malthus and David Ricardo. *Journal of Economic Perspectives*, *3*(3), 153–64.

Ekerhovd, N.-A., Flåm, S. D., & Steinshamn, S. I. (2021). On shared use of straddling stocks. *European Journal of Operations Research*, *forthcoming*.

[20] One can hardly expect much constructive *logic of collective action* (Olson, 1965) among parties who have more of opposed than of common interest.

[21] It requires no justification to emphasize the welfare gains generated by efficient use of scarce production factors (Kaldor, 1939).

[22] Shifting parts of resource rent - away from users of rights - towards permanent holders of such items, is apt to provoke contestation.

Fisher, E. (2017). *Environmental law*. Oxford University Press.

Flåm, S. D. (2014). Bilateral exchange and competitive equilibrium. *Set-Valued and Variational Analysis*, *24*, 1–11.

Flåm, S. D. (2016). Noncooperative games, coupling constraints, and partial efficiency. *Bulletin of Economic Theory*, *4*, 213–29.

Flåm, S. D. (2020). Emergence of price-taking behavior. *Economic Theory*, *70*, 847–70.

Flåm, S. D. (2021). Towards competitive equilibrium by double auctions. *Pure and Applied Functional Analysis*, *forthcoming*.

Flåm, S. D., & Gramstad, K. (2012). Direct exchange in linear economies. *International Game Theory Review*, *14*(4), 1–18.

Fujishige, S., & Yang, Z. (2017). On a spontaneous decentralized market process. *Journal of Mechanism and Institution Design*, *2*, 1–37.

Gordon, H. S. (1954). The economic theory of a common property resource: The fishery. *Journal of Political Economy*, *62*, 124–42.

Goswami, M. P., Mitra, M., & Sen, A. (2014). Strategy proofness and Pareto efficiency in quasilinear exchange economies. *Theoretical Economics*, *9*, 361–81.

Hanley, N., Shogren, J. F., & White, B. (2007). *Environmental economics* (2nd ed.). Palgrave, New York.

Hardin, G. (1968). The tragedy of the commons. *Science*, *162*, 143–48.

Hodgson, G. M. (2000). What is the essence of institutional economics? *Journal of Economic Issues*, *34*, 317–29.

Hughes, J. R. T. (1976). *Social control in the colonial economy*. University of Virgina Press, Charlottesville.

Kaldor, N. (1939). Welfare propositions of economics and interpersonal comparisons of utility. *Economic Journal*, *49*, 549–52.

Krishna, V. (2010). *Auction theory*. Academic Press, Amsterdam.

Libecap, G. D. (1986). Property rights in economic history: Implications for research. *Explorations in Economic History*, *23*, 227–52.

Luenberger, D. G. (1995). *Microeconomic theory*. McGraw-Hill, New York.

Mordukhovich, B. S., & Nam, N. M. (2013). *An easy path to convex analysis and applications*. Morgan & Claypool Publishers.

Moxnes, E. (2012). Individual transferable quotas versus auctioned seasonal quotas: An experimental investigation. *Marine Policy*, *36*, 339–49.

Neher, P. A. (1990). *Natural resource economics, conservation and exploitation*. Cambridge University Press.

Olson, M. (1965). *The logic of collective action, public goods, and the theory of groups*. Harvard University Press.

Osborne, M. J., & Rubinstein, A. (1994). *A course in game theory*. MIT Press.

Ostrom, E. (1990). *Governing the commons*. Cambridge University Press.
Penot, J.-P. (2013). *Calculus without derivatives*. Springer Verlag, Berlin.
Standing, G. (2019). *Plunder of the commons*. Pelican Books, Random House, UK.
Valentine, F. A. (1964). *Convex sets*. Mc-Graw-Hill, New York.
Young, P. (1994). *Equity*. Princeton University Press.
Zălinescu, C. (2002). *Convex analysis in general vector spaces*. World Scientific, London.

How to prepare a paper for submission

Before submitting a paper to **this Journal**, the authors are advised to follow closely the following instructions to prepare the paper.

1. Papers submitted to this Journal must be unpublished and original work that is neither under review elsewhere nor will be submitted elsewhere for publication without withdrawing from this Journal.

2. This Journal requires that all results (experimental, empirical and computational) be replicable. All underlying data necessary to replicate results must be made available to the Journal.

3. Papers must be submitted in electronic form (preferably as pdf files) and the size of the text font must be at least 12 point. Each figure and table must be included on the relevant page of the paper and should not be collected at the end of the paper. The list of references should appear after any appendices, as the last part of the paper.

4. There is no restriction on the number of pages of any submitted manuscript. We seek to process any first submission within 3 months. However, very long manuscripts may take considerably more time to review.

5. Each submitted paper should have an abstract of no more than 150 words containing no mathematical formulas, complete with no more than 3 suggested keywords and JEL codes. We also encourage authors to make the introduction of their submitted articles understandable to the widest audience possible.

6. The first page of each submitted paper should contain every author's e-mail address, phone number and affiliation.

7. The editors or the publishing Society will not hold any responsibility for views expressed by authors in this Journal.

How to submit a paper

1. To submit a paper, we encourage the author(s) of the paper to become member(s) of the **Society for Promotion of Mechanism and Institution Design**. To join the Society, please go to the Society website: `http://www.mechanism-design.org/`.

   Currently, the annual membership rates are £60 (GBP) for regular members and £40 (GBP) for students. Concessions are available for members from low-income countries.

2. Papers should be submitted electronically to the **Journal of Mechanism and Institution Design** through the above website.

Aims & Scope of the Journal

The Journal of Mechanism and Institution Design aims to publish original articles that deal with the issues of designing, improving, analysing and testing economic, financial, political, or social mechanisms and institutions. It welcomes theoretical, empirical, experimental, historical and practical studies. It seeks creative, interesting, rigorous, and logical research and strives for clarity of thought and expression. We particularly encourage less experienced researchers such as recent PhD graduates to submit their work to the Journal and are sympathetic towards those papers that are novel and innovative but which have been unsuccessful elsewhere. We hope that the published articles will be interesting and valuable to a broad audience from the areas of economics, finance, politics, law, computer science, management, history, mathematics, government, and related disciplines. The journal is an open-access, independent, peer-reviewed, non-profit, English-language journal with the purpose of disseminating and sharing the latest knowledge and understanding of the subject widely.