Journal of Mechanism and Institution Design

Special Issue in Honour of Vincent P. Crawford

Guest Editor: Alexander Teytelboym

CONTENTS

Volume 9, Issue 1, December 2024
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
Universität Klagenfurt, Austria
Southwestern University of Economics and Finance, China.

Cover & Logo Artwork @ Jasmine Yang
LaTeX Editor & Journal Design @ Paul Schweinzer (using 'confproc')

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

A Letter from the Editors

WE are very, very proud to present this new and special volume of the Journal as a festschrift in honor of Vincent P. Crawford to celebrate his fundamental contributions to economic theory, game theory, and our Society for the Promotion of Mechanism and Institution Design. We are grateful to the many colleagues and especially the guest editor Alexander Teytelboym who made this special issue possible.

The special issue arose from the Conference on Mechanism and Institution Design held July 8-12, 2024, in Budapest, Hungary. The conference was organized by Péter Biró and hosted by Corvinus University of Budapest. It featured more than 300 participants, four one-hour keynote lectures by Paul Milgrom, Roger Myerson, Al Roth, and Eva Tardos, a one-hour special lecture by Vince Crawford, and three 45-minute plenary lectures by Nagore Iriberri (University of the Basque Country), Kristof Madarasz (London School of Economics), and Alex Teytelboym (University of Oxford). These sessions and the 284 contributed talks made the conference a fantastic event with, to us, this summer's most vibrant, intellectually stimulating, and fun vibe.

As with each new issue, we want to thank all associate editors, referees, and supporters for their contributions to the Journal. Everyone involved in the scientific selection and technical production of the Journal is volunteering their capabilities, time, and effort. As a result, our Journal of Mechanism and Institution Design can publish high-quality research free of charge to both authors and readers, allowing for free and open access to the public. We hope the community sharing this vision will grow and become stronger, tackling the most pressing design issues of our time.

The dual goals of this Journal lie in the publication of high-quality papers in the wide area of mechanism and institution design and in promoting our research in the public interest. Supporting these purposes, we hope to publish more special issues in the future, celebrating scholars who have made pioneering scientific breakthroughs in the field of mechanism and institution design. We are thrilled that Vince Crawford allowed us to make an exciting and superb start. In the words of our special issue guest editor: "Here's to you, Commodore Crawford!"

Paul Schweinzer & Zaifu Yang, Klagenfurt & York, December 16, 2024.

Journal of Mechanism and Institution Design
ISSN: 2399-844X (Print), 2399-8458 (Online)
DOI:10.22574/jmid.2024.12.001

A FESTSCHRIFT FOR VINCE CRAWFORD

Alexander Teytelboym

University of Oxford, United Kingdom
alexander.teytelboym@economics.ox.ac.uk

1. INTRODUCTION

The five papers in this festschrift that celebrates the 75th birthday of Vincent P. Crawford (henceforth "Vince") attempt to reflect a mere fraction of his monumental contributions to economic theory. This special issue of the *Journal of Mechanism and Institution Design* touches on just three of Vince's earlier research interests—social choice, information economics, and matching markets—which led to unforgettable papers pored over by generations of first-year graduate students as well as by more seasoned connoisseurs. The issue only scratches the surface of the many areas in which Vince has made long-lasting contributions—for example, his work has also been fundamental in bargaining and arbitration, learning, contract theory, evolutionary game theory, and, more recently, level-k thinking. But Vince deserves multiple festschrifts and this is only the first one.

Some of Vince's earliest published work was on social choice[1] and specifically looked at the possibility for *self-administered procedures* (rather than mechanisms run by the social planner) to get to Pareto-efficient and "fair" allocations. The poignant first paper in the issue reflects Vince's early interest in social choice. "Kenneth Arrow's Last Theorem" by Paul Milgrom describes

I am very grateful to Zaifu Yang for inviting me to be a guest editor for this special issue and to Péter Biró for organizing the 2024 Conference on Mechanism and Institution Design where we celebrated Vince's birthday. I want to thank Scott Kominers, Al Roth, Paul Schweinzer for their comments and suggestions for this note. Finally, I want to record a huge debt of gratitude to Vince, my doctoral advisor, for his unwavering support and inspiration from the first day I met him in 2010.

[1] See, for example, Crawford (1977), which appeared in Vince's PhD thesis, as well as Crawford (1979) and Crawford (1980).

and formalizes Arrow's *final* contribution to economics: the impossibility of non-dictatorial and efficient collective decision-making across generations. Arrow's key observation was that efficient policies in the far future must give higher weight to the most patient agents who end up dictating societal decisions.

The second paper in the issue, "Bayesian Bullshit", is a delightful take on Vince's most cited and probably most revered paper, which introduced the sender-receiver game and provided the seminal strategic analysis of cheap talk about private information (Crawford & Sobel 1982). Sajan Srivastava, Tymofiy Mylovanov, and Rakesh Vohra consider a variation on Crawford and Sobel's model in which the sender is potentially a "bullshitter" who doesn't care about the truth at all (formally, the sender does not know their type when they send a message, and moreover has state-independent preferences). The authors show that depending on the bullshitter's preferences, more or less information can be transmitted in equilibrium and the (informed) sender and the receiver can be either made better or worse off.[2]

The remaining papers in the issue are related to two of Vince's papers that beautifully linked general equilibrium theory with the theory of matching markets introduced by Gale & Shapley (1962). In response to these two papers coauthored with Elsie Knoer and Lex Kelso respectively, Al Roth inscribed Vince's copy of Roth & Sotomayor (1990), "For Vince, who discovered that this stuff was economics."

In his groundbreaking 1981 paper with Knoer, Vince considered a model of two-sided matching (between, say, workers and firms) in which agents care about salaries in addition to the quality of the match.[3] Crawford & Knoer (1981) showed that an adapted Gale-Shapley algorithm corresponds to a very natural dynamic salary-adjustment process, reminiscent of an English auction, which converges to the core and competitive equilibrium matching most favoured by the firms, who make offers in the process. In his even more influential 1982 paper with Kelso (which Vince considered at the time "just an extension" of the 1981 paper!), Vince extended the model to allow firms to hire multiple workers (which was possible under the original model as long as preferences were additively separable). Moreover, he highlighted the importance of the "gross substitutes" condition which guaranteed the existence, structure and convergence to a competitive equilibrium in an economy with indivisible goods.

[2] In the spirit of the title of the paper, the authors produced an AI-generated podcast of the paper.

[3] Much later, Crawford (2008) argued that the efficiency of National Resident Matching Program could, in fact, be improved by allowing for flexible salaries.

Kelso & Crawford (1982) showed that under substitutability making offers worker by worker (or good by good in other settings) suffices for the process to converge to an efficient, equilibrium outcome; that is, despite the productivity interactions, with substitutability a combinatorial auction is not needed for reaching an efficient allocation. Thanks to Kelso & Crawford (1982), the powerful gross substitutability condition is now central in the literatures on matching markets (Hatfield & Milgrom 2005), dynamic auctions (Ausubel 2006), and pseudomarkets with indivisible goods (Gul et al. 2024; Nguyen & Vohra 2024). The gross substitutability condition also played a key role in the design of the original US spectrum auction (Milgrom 2000) and its modern variants appear in the designs of many matching markets (see, e.g., Sönmez & Switzer 2013 for an application and Hatfield & Kominers 2016 for a survey).

Inspired by Vince's work, Alfred Galichon, Larry Samuelson, and Lucas Vernet have recently developed a new class of "equilibrium flow" models that generalizes bipartite matching models, minimum-cost flow problems and even hedonic pricing models. In Galichon et al. (2022), they introduce a special case of the (weak) gross substitutes condition for correspondences, which they term "unified gross substitutes"; this condition has attractive monotone comparative statics properties for equilibrium flow models. In their paper "The Existence of Equilibrium Flows," in this issue, they establish the key existence result for equilibrium flow models. Essence, after all, does require existence.

Another marvellous result in Kelso & Crawford (1982) is the "entry comparative static": adding an extra agent to one side of a two-sided matching market makes all the agents on the other side of market weakly better off (Crawford 1991, among others, further generalized this result, and many generalizations have followed since). In his paper for the issue, Scott Duke Kominers provides an elegant alternative proof of the entry comparative static that exploits the more recent "respect for improvements" property of Balinski & Sönmez (1999), i.e., the fact that making one agent unambiguously more highly ranked in their potential match partners' preferences improves that agent's outcome under deferred acceptance.

The final paper in the issue is a nod to the importance that Vince's work has had for (what is now called) market design and especially the design of dynamic auctions. Zaifu Yang and Jingsheng Yu propose an ascending "menu auction" à la Bernheim & Whinston (1986) in which bidders can have complex preferences (allowing for complements) as well as budget constraints—neither of which were permitted in Kelso and Crawford's original setting since they preclude

the existence of a competitive equilibrium with indivisible goods. Remarkably, Yang and Yu's auction produces a strongly Pareto-efficient outcome in the core even though competitive prices might not exist.

Vince's work inspires every generation of scholars that encounter it, and the papers in this issue are a testament to an enduring power of Vince's ideas. Scholarship, mentorship and friendship aren't ships that always sail together, but in his spectacular career Vince has had command of all three of them. Here's to you, Commodore Crawford!

References

Ausubel, L. (2006). An efficient dynamic auction for heterogeneous commodities. *American Economic Review*, *96*(3), 602–29.

Balinski, M., & Sönmez, T. (1999). A tale of two mechanisms: Student placement. *Journal of Economic Theory*, *84*(1), 73–94.

Bernheim, B., & Whinston, M. (1986). Menu auctions, resource allocation, and economic influence. *Quarterly Journal of Economics*, *101*(1), 1–31.

Crawford, V. P. (1977). A game of fair division. *Review of Economic Studies*, *44*(3), 235–47.

Crawford, V. P. (1979). A procedure for generating pareto-efficient egalitarian-equivalent allocations. *Econometrica*, *47*(1), 49–60.

Crawford, V. P. (1980). A self-administered solution of the bargaining problem. *Review of Economic Studies*, *47*(2), 385–92.

Crawford, V. P. (1991). Comparative statics in matching markets. *Journal of Economic Theory*, *54*(2), 389–400.

Crawford, V. P. (2008). The flexible-salary match: A proposal to increase the salary flexibility of the national resident matching program. *Journal of Economic Behavior & Organization*, *66*(2), 149–60.

Crawford, V. P., & Knoer, E. M. (1981). Job matching with heterogeneous firms and workers. *Econometrica*, *49*(2), 437–50.

Crawford, V. P., & Sobel, J. (1982). Strategic information transmission. *Econometrica*, *50*(6), 1431–51.

Gale, D., & Shapley, L. S. (1962). College admissions and the stability of marriage. *American Mathematical Monthly*, *69*(1), 9-15.

Galichon, A., Samuelson, L., & Vernet, L. (2022). Monotone comparative statics for equilibrium problems. *arXiv preprint arXiv:2207.06731 [econ.TH]*.

Gul, F., Pesendorfer, W., & Zhang, M. (2024). Efficient allocation of indivisible goods in pseudomarkets with constraints. *Journal of Political Economy*, *132*(11),

3708–36.

Hatfield, J. W., & Kominers, S. D. (2016). Hidden substitutes. In *Proceedings of the 16th ACM Conference on Electronic Commerce.*

Hatfield, J. W., & Milgrom, P. (2005). Matching with contracts. *American Economic Review*, *95*(4), 913–35.

Kelso, A. S., Jr., & Crawford, V. P. (1982). Job matching, coalition formation, and gross substitutes. *Econometrica*, *50*(6), 1483–1504.

Milgrom, P. (2000). Putting auction theory to work: Ascending auctions with package bidding. *Journal of Political Economy*, *108*(2), 245–72.

Nguyen, T., & Vohra, R. (2024). (Near-)substitute preferences and equilibria with indivisibilities. *Journal of Political Economy*, *132*(12), 4122–54.

Roth, A. E., & Sotomayor, M. A. O. (1990). *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. United Kingdom: Cambridge University Press.

Sönmez, T., & Switzer, T. (2013). Matching with (branch-of-choice) contracts at the United States Military Academy. *Econometrica*, *81*(2), 451–88.

Journal of Mechanism and Institution Design
ISSN: 2399-844X (Print), 2399-8458 (Online)
DOI:10.22574/jmid.2024.12.002

KENNETH ARROW'S LAST THEOREM

Paul Milgrom

Stanford University, USA
milgrom@stanford.edu

ABSTRACT

In Kenneth Arrow's last week of life at age 95, he reported that "I began my research career with an impossibility theorem. If I had time now, my last theorem would be an impossibility theorem about social choice for environmental policy." This paper completes the formalization, proof, and discussion of the theorem that Arrow then described.

Keywords: Impossibility theorem, environmental economics.

JEL Classification Numbers: D63, Q00.

1. INTRODUCTION

In his final days, Kenneth Arrow's interest in economic scholarship never waned. When I visited him, he wanted to discuss two topics: income equality and climate change. In one conversation, he proposed an impossibility theorem that applies especially to collective choice for environmental policies. Below, I introduce a complete mathematical model based on his description.

The finite set of agents $n \in N$ consume environmental amenities over time periods $t = 0, 1, 2 \ldots$. We may interpret each agent as a clan or dynasty, which cares for its living and future members. The *state* of the system at time t is

I gratefully acknowledge support from the National Science Foundation, grant SES1947514. This report is also based in part upon work supported by the National Science Foundation under Grant No. DMS-1928930 and by the Alfred P. Sloan Foundation under grant G-2021-16778, while the author was in residence at the Simons Laufer Mathematical Sciences Institute (formerly MSRI) in Berkeley, California, during the Fall 2023 semester. I thank Marion Ott for her careful reading, Michael Crystal for his research assistance, Ilya Segal for pointing out the related paper by Jackson and Yariv, and a referee for suggesting improvements to the proof and its interpretation.

a positive integer $s_t \in S$, where S may be finite or infinite. In each period, the collective observes the state and takes an *action* $a_t \in \{1,\ldots,A\}$, which governs the next state according to the transition probabilities $p(s_{t+1}|s_t,a_t)$.

A *policy* is a function that describes a pure or mixed action $\pi(s)$ for each state s. In each period, agent n earns flow utility $u_n(s_t,a_t) \in [0,1]$ that depends on the realized state and action. Let Π denote the set of policies. Each agent evaluates its future utility starting in state s as the discounted sum of future flow utilities, $U_{nt}^{\pi}(s) = \mathbb{E}_{\pi}\left[\sum_{\tau=t}^{\infty} \delta_n^{\tau-t} u_n(s_\tau, \pi(s_\tau)) \mid s_t = s\right]$, in which the policy determines the probability distribution over future states and actions and δ_n is n's discount factor. A policy π is Pareto dominated in state s by another policy or a mixture of policies π' if for all agents n, $U_{n0}^{\pi}(s) \leq U_{n0}^{\pi'}(s)$ and for some n, the inequality is strict. A policy π is *efficient* if it is not Pareto dominated in any state.

Arrow proposed a single axiom to govern the collective choice: the chosen policy should be efficient. Using this one axiom, the simplest version of Arrow's Last Theorem is the following.

Theorem. Suppose that $p(\cdot|\cdot) > 0$ and that for all $n \neq m$, $\delta_n \neq \delta_m$. Then for any efficient policy π, there is an m (the "dictator") such that in every state, π is a most preferred policy for m.

Proof. Let $U_t^{\pi}(s) = (U_t^{\pi}(s))_{n\in N}$ denote the payoff vector beginning at time t in state s with policy π and let $F(s) \in [0,1]^N$ denote the set of payoff vectors corresponding to feasible policies. Since policies may be randomized, F is convex and for any efficient policy π^*, $U_0^{\pi^*}(s)$ is on the boundary of $F(s)$. So $U_0^{\pi^*}(s)$ it is supported by a hyperplane. This means that for some nonzero vector $\lambda_s \in R_{+^N}$ with $\sum_n \lambda_{ns} = 1$ and all $u \in F(s)$, $\lambda \cdot u \leq \lambda_s \cdot U_0^{\pi^*}$. In terms of policies, this inequality means:

$$(\forall s \in S)\pi^* \in \arg\max_{\pi\in\Pi} \mathbb{E}_{\pi}[\sum_{\tau=0}^{\infty} \sum_{n\in N} \lambda_{ns}\delta_n^{\tau} u_n(s_\tau, \pi(s_\tau) \mid s_0 = s].$$

Fix any state s and let $m := \arg\max_{n:\lambda_{ns}>0} \delta_n$, that is, m is the agent with the highest discount factor among *influential* agents (agents with positive welfare weights). Then, π^* must also solve the continuation problem of maximizing the payoff after the first period given that a transition to state $s_1 = s'$ has occurred.

$$\pi^* \in \arg\max_{\pi\in\Pi} \sum_{s'} p(s' \mid s, \pi^*(s))\delta_m \mathbb{E}_{\pi}[\sum_{\tau=1}^{\infty} \sum_{n\in N} \frac{\delta_n}{\delta_m}\lambda_{ns}\delta_n^{\tau-1} u_n(s_\tau, \pi(s_\tau) \mid s_1 = s'].$$

So,

$$\pi^* \in \arg\max_{\pi\in\Pi} \mathbb{E}_\pi[\sum_{\tau=0}^{\infty}\sum_{n\in N}\left(\lambda_{ns}\frac{\delta_n}{\delta_m}\right)\delta_n^\tau u_n(s_\tau,\pi(s_\tau)\mid s_0=s].$$

Comparing this to the original maximization problem, the coefficients $\lambda_{ns}\frac{\delta_n}{\delta_m}$ have replaced the coefficients λ_{ns} of the original problem. Iterating this coefficient replacement step t times leads to:

$$\pi^* \in \arg\max_{\pi\in\Pi} \mathbb{E}_\pi[\sum_{\tau=0}^{\infty}\sum_{n\in N}\left(\lambda_{ns}\left(\frac{\delta_n}{\delta_m}\right)^t\right)\delta_n^\tau u_n(s_\tau,\pi(s_\tau)\mid s_0=s].$$

If $n\neq m$, then $\delta_n<\delta_m$. So, for all $n\neq m$, $\lim_{t\to\infty}\lambda_{ns}\left(\frac{\delta_n}{\delta_m}\right)^t=0$, leading to:

$$\pi^* \in \arg\max_{\pi\in\Pi} \mathbb{E}_\pi[\sum_{\tau=0}^{\infty}\sum_{n\in N}\lambda_{ms}\delta_m^\tau u_m(s_\tau,\pi(s_\tau)\mid s_0=s].$$

So, π^* is a most preferred policy for m. □

1.1. Discussion

The essence of the proof is to note that if π^* is an efficient policy starting in some state s today, that implies a maximization using welfare weights for agents that must be unchanging over time. However, efficiency requires that decisions in the distant future must give relatively greater weight to the preferences of more patient individuals, at least among agents with non-zero weights. Those two things can be consistent only if there is just one agent with a non-zero weight.

The dictatorship conclusion in Arrow's Last Theorem conjures memories of his famous Impossibility Theorem, but this new theorem differs in some important ways. The original Impossibility Theorem asserts that any *social welfare function* satisfying three reasonable axioms must reflect the preferences of just one agent, who is called the dictator. Similarly, Arrow's Last Theorem introduces an axiom – efficiency – and concludes that the socially chosen policy must be a most preferred policy of a single agent. However, the conclusion of Arrow's Last Theorem is different from his first impossibility theorem because the definitions of dictatorship are different: the agent that dictates the policy for one value profile in Arrow's Last Theorem may not be a dictator for the social welfare function.

To illustrate the difference, consider the social welfare function that works as follows. First, any given preference profile is mapped into an ordered list of agents. Then, the first-ranked agent identifies her most preferred policy or policies. If there are multiple most preferred policies, then the second-ranked agent picks her most preferred policies among those, and so on in a process of serial dictatorship until some policy is chosen. For *every* function that maps preference profiles to an ordered list of agents, the social welfare function represented by this two-step construction selects an efficient policy and for every profile, that policy is the most preferred one for some agent, but in contrast to Arrow's original impossibility theorem, different profiles may identify different dictators.

Arrow's Last Theorem is to examine a property of a policy rather than of a social welfare functions. Consequently, my preferred interpretation is that it asserts that, when its conditions apply, it is impossible for a policy to efficiently *compromise* the interests of agents with different time preferences.[1] One may wonder: what rules out the possibility that the policy chooses a preferred action for less patient agents at early dates and for more patient agents at later dates? The answer is that the assumption that transition probabilities $p(\cdot \mid \cdot)$ are all strictly positive rules out such a policy because it makes it impossible to keep track of the date t by encoding it in the state.

If some transitions were allowed to have zero probability, we could construct the state space to be the union of two non-communicating sets of states, which allows that an efficient policy can have different dictators in the two non-communicating sets. Or, it could be possible that the transitions ensure that no state can occur more than once. For example, let $S_t = \{(t-1)S + 1, \ldots, Kt\}$ and suppose that that each state in S_t can transition only to a state in S_{t+1}. That allows the state to encode the date and allows efficient policies that increasingly favor more patient agents at later dates. Restricting transition probabilities to be positive both eliminates those possibilities. It ensures the unity of the model (just one set of communicating states) and enforces a perspective that Arrow championed: that the date should be irrelevant for setting or evaluating environmental policy beginning today.

Finally, the theorem also assumes that different agents have different rates of time discount. Omitting that assumption, a similar proof implies that any efficient policy maximizes a weighted sum of utilities of a *dictatorial group*

[1] A similar point was made by Jackson & Yariv (2015), although their formulation and framing of the problem were different.

of agents who all have the same discount factor.

Kenneth Arrow told me that this result would be his final theorem, and throughout Arrow's career, he encouraged intelligent debate. On his behalf, I have completed the formalization and proof of his theorem and tried to provide the kind of challenging remarks that he would have welcomed.

References

Jackson, M., & Yariv, L. (2015). Collective dynamic choice: The necessity of time inconsistency. *American Economic Journal: Microeconomics*, *7*(4), 150–78.

Journal of Mechanism and Institution Design
ISSN: 2399-844X (Print), 2399-8458 (Online)
DOI:10.22574/jmid.2024.12.003

BAYESIAN BULLSHIT

Sajan Srivastava

University of Pennsylvania, USA
sajansr@sas.upenn.edu

Tymofiy Mylovanov

University of Pittsburgh, USA
tym10@pitt.edu

Rakesh Vohra

University of Pennsylvania, USA
rvohra@seas.upenn.edu

ABSTRACT

A bullshitter neither knows nor cares about the truth, and therefore, it has been asserted, is more pernicious than a liar. We examine this assertion within the standard model of cheap talk communication where a bullshitter is modeled as an uninformed Sender. We show that in some circumstances, uncertainty about whether the Sender is informed or not can increase the welfare of the Receiver.

Keywords: Bayesian persuasion, information design, cheap talk.

JEL Classification Numbers: C6, D82, D83.

1. INTRODUCTION

> Political language is designed to make lies sound truthful and murder respectable, and to give an appearance of solidity to pure wind.

George Orwell

We thank Juuso Toikka and an anonymous referee for valuable suggestions.

2. INTRODUCTION

A LIAR knows the truth and consciously seeks to conceal it. Liars are absent in the canonical economic model of communication: cheap talk. As Sobel (2020) points out, lying requires a common understanding of what words mean, which is not a feature of cheap talk models. To account for lying, Sobel (2020) adds the concept of a common language to a standard model of cheap talk. This allows one to interpret every equilibrium type-action distribution as involving lies of omission. The central finding is that fully anticipated lies don't interfere with the exchange of information.

To lie is not the same as to bullshit because a bullshitter, unlike a liar, neither knows nor cares about the truth. A bullshitter does care about concealing both their ignorance of and interest in the truth.

The Oxford English Dictionary defines bullshit as 'to talk nonsense or rubbish' and 'to bluff one's way through something by talking nonsense'. According to Partridge et al. (2006), the word entered into general use during the first world war to describe a focus on the inessential (spit & polish) over the essential (combat readiness). Frankfurt (2009) argues that bullshit is now widespread because the opportunities to engage in it have expanded:

> "....the production of bullshit is stimulated whenever a person's obligations or opportunities to speak about some topic are more excessive than his knowledge of the facts that are relevant to that topic."

He also asserts that it is more dangerous than lying.

> "When an honest man speaks, he says only what he believes to be true; and for the liar, it is correspondingly indispensable that he considers his statements to be false. For the bullshitter, however, all bets are off ... He does not reject the authority of the truth, as the liar does, and oppose himself to it. He pays no attention to it at all. By virtue of this, bullshit is a greater enemy of truth than lies are."

Bullshitting does not fit into the cheap talk model because the Sender is no more informed than the Receiver. This lacunae is filled by presuming the existence of an uninformed Sender. We use it to examine Frankfurt's assertion that bullshitting is the 'greater enemy'. His assertion relies on a skepticism about the Receiver's ability to identify bullshit (Frankfurt 2009):

> "... most people are rather confident of their ability to recognise bullshit and to avoid being taken in by it"

Just as the Receiver in cheap talk accounts for the fact that an informed Sender will obscure the truth, why can't they do the same with a bullshitter? If so, is it obvious that they are worse off compared to the case where the Sender is informed?

We answer this question in a standard cheap talk environment. There is an underlying one dimensional state of nature unknown to the Receiver who wishes to choose an action that matches the state. The Receiver faces a Sender who may or may not be informed about the underlying state. The probability that the Sender is informed is p. An informed Sender's preferences differ from that of the Receiver and depend upon the state. They wish the Receiver to choose an action equal to the state plus a commonly known *positive bias*. As the Sender's preferences differ from the Receiver's, they have an interest in obscuring the truth.

As remarked earlier, to account for lying one must have a common language. In the standard cheap talk model, messages correspond to subsets of the state space and these subsets will be the common language. Thus, if the informed sender says S, where S is some subset of the state space, this is understood to mean that the Sender is saying that the underlying state is in S. If the Sender says S and the state is outside S, this is an outright lie. If the state is in S but S is not a singleton, this is a lie of omission. The non-babbling equilibria of the standard model of cheap talk correspond to lies of omission and we focus on these. This makes the informed Sender the 'liar' in our model.

With probability $1-p$, the Sender is uninformed and has preferences that are state-independent. The uninformed Sender is the bullshitter in our model because they are ignorant of the underlying state (i.e. the truth) and, as will be seen, have no incentive to acquire information about it. The uninformed Sender wishes to be perceived as informed, and this is hard wired into the model by assuming $p > 0$. We consider two possibilities. The uninformed Sender prefers higher to lower actions or the converse. One can interpret the uninformed Sender as adding noise to the message sent by the informed Sender. Blume et al. (2007) considers the impact of noise in the standard cheap talk model by assuming every message has a chance of being uninformative. In our case, *some* but not all messages have a chance of being uninformative. This is because there are some messages an uninformed Sender never sends, making some of the informed Sender's messages noise free.

The possibility of an uninformed Sender ($p < 1$) changes equilibrium outcomes (modulo selection) in the cheap talk game by *changing* the behavior of the informed Sender rather than through direct manipulation of the Receiver's beliefs. Thus, our model highlights the *indirect* effect an uninformed Sender has on dialog by influencing the communication strategies of the informed Sender. Whether the Receiver is the worse for it, compared to the case $p = 1$, *depends* on the commonly known preference of the uninformed Sender.

If the uninformed Sender prefers higher to lower actions, the Receiver is worse off compared to the case when $p = 1$. An informed Sender is also ex-ante worse off. In this case bullshit dilutes informative communication, consistent with Frankfurt's lament about the prevalence of bullshit.[1]

If, however, the uninformed Sender prefers lower to higher actions, the Receiver's expected payoff can be *higher* compared to the case when $p = 1$ provided p is not too small. The informed Sender with positive bias is ex-ante always better off provided the bias is not too small. This contrasts with Frankfurter's unqualified assertion that bullshitting is worse than lying. It highlights that whether bullshitting has a negative effect or not depends upon the uninformed Sender's preferences relative to those of the informed Sender.

One might conjecture that an uninformed Sender is no different than an informed Sender with a very large bias. If true, it would mean that the uninformed Sender's ignorance of the underlying state plays no role. We show this not to be case. The equilibrium messaging strategy of an informed but highly biased Sender *differs* from that of an uninformed Sender.

We also consider the possibility that the preferences of the uninformed Sender may be unknown. Specifically, the uninformed Sender may prefer higher actions or lower actions. We find that if both possibilities are equally likely, the Receiver can be better off, provided that the likelihood of facing an uninformed Sender is sufficiently small.

Our model also highlights another way in which a bullshitter can affect communication. The standard Crawford & Sobel (1982) cheap talk model supports equilibria where the Sender can partition the message space into finer intervals. This increases the precision of information transmitted and leads to higher payoffs for both Receiver and informed Sender. This is no longer true in the presence of a bullshitter. In particular, an equilibrium with three messages as opposed to two, only leads to higher payoffs if p is sufficiently

[1] Many find it hard to follow Wittgenstein's injunction: "Whereof one cannot speak, thereof one must be silent."

large. Thus, the presence of a bullshitter reduces the incentives to use a larger message space.

The remainder of this paper is organized as follows. The next section reviews the basic cheap talk model of Crawford & Sobel (1982). The subsequent section investigates what happens in this model when the possibility of an uninformed Sender being present is allowed. The next section discusses some variations on the base model to highlight the robustness of our conclusions. The final section draws comparisons between bullshit in a cheap talk environment and in a Bayesian persuasion environment.

3. CHEAP TALK MODEL

We recall the textbook cheap talk communication game between an informed Sender with known positive bias and an uninformed Receiver who must choose an action (Crawford & Sobel 1982).

1. The Sender's type is the underlying state, $\theta \in [0,1]$.
2. The Receiver is ignorant of the state but believes it to be drawn uniformly from $[0,1]$. This belief is known to the Sender.
3. Receiver's action space is $A = [0,1]$.
4. Receiver's payoff from action $a \in [0,1]$ is $-(a-\theta)^2$.
5. Sender's payoff from action $a \in [0,1]$ is $-(a-(\theta+b))^2$ where $b > 0$ is the bias.
6. Sender's message space is $[0,1]$.

The cheap talk game admits multiple equilibria. One reason is that the meaning of messages is arbitrary. Given any equilibrium, one can generate another equilibrium by changing the interpretation of messages.

Rather than assuming the Sender's strategy is a signaling rule (a function from type into message), we follow Gordon (2010) in assuming the informed Sender's strategy is an increasing set of thresholds. Adjacent thresholds identify an interval and distinct types of the informed Sender send a common message if and only if their types are in the same interval. This formulation eliminates message indeterminacy. It also means that the informed Sender never lies, but they don't disclose the whole truth either.

It is straightforward to see that an equilibrium of the cheap talk game generates a type-action distribution if and only if it is an equilibrium of the threshold game. Crawford & Sobel (1982) show that there is a finite upper bound,

n^*, on the number of thresholds. Furthermore, for each integer k between 1 and n^* there is a choice of b such that it is a unique equilibrium outcome with k thresholds. For economy of exposition, we focus primarily on equilibria with a single threshold. Section 5.3, discusses what happens when this assumption is relaxed. In the appendices we explore other variations of this model.[2]

Suppose there exists a threshold $t \in [0,1]$ and the Sender sends the message B (for bottom) if state $\theta \in [0,t]$ and the message T (for top) otherwise. In the first case the Receiver chooses action $0.5t$ to maximize their expected payoff. In the other case, the Receiver's optimal action choice is $0.5(1+t)$. At an equilibrium threshold t the following indifference condition for the informed Sender must hold:

$$t+b-0.5t = 0.5(1+t)-(t+b) \;\Rightarrow\; t = 0.5-2b.$$

We need $b \leq 0.25$ to ensure this is well defined and this will be assumed throughout.

When the possibility of an uninformed Sender is introduced, the equilibrium threshold t changes. We are interested in how this change affects the Receiver and Sender's expected payoffs. This is summarized in the next two Lemmas which are standard.

Lemma 3.1. *The Receiver's expected payoff as a function of t, denoted $\rho(t)$, is increasing for $t \leq 0.5$, and decreasing otherwise.*

Proof.

$$\begin{aligned}\rho(t) &= -t \times t^{-1}\int_0^t (0.5t-x)^2dx - (1-t)\times(1-t)^{-1}\int_t^1 (0.5(1+t)-x)^2dx \\ &= -\int_0^t (0.5t-x)^2dx - \int_t^1 (0.5(1+t)-x)^2dx \\ &= \frac{(t-1)^3-t^3}{12}.\end{aligned}$$

Now, $\rho(t)$ achieves a maximum in $[0,1]$ at $t=0.5$ (where the derivative is zero). □

The informed Sender's payoff will depend upon the underlying state. Hence, we focus on the informed Sender's ex-ante expected payoff.

[2] For more on equilibrium selection in this context see Gordon et al. (2021).

Lemma 3.2. *The Sender's ex-ante expected payoff as a function of t, denoted $\sigma(t)$ is increasing for $t \leq 0.5$ and decreasing otherwise.*

Proof.

$$\begin{aligned}\sigma(t) &= -t \times t^{-1} \int_0^t (0.5t - b - x)^2 dx - (1-t) \\ &\quad \times (1-t)^{-1} \int_t^1 (0.5(1+t) - b - x)^2 dx \\ &= -\int_0^t (0.5t - b - x)^2 dx - \int_t^1 (0.5(1+t) - b - x)^2 dx \\ &= \frac{(t-1)^3 - t^3}{12} - b^2.\end{aligned}$$

The derivative of $\sigma(t)$ with respect to t is $-0.5t + 0.25$. The second derivative is -0.5 which is negative. Therefore $\sigma(t)$ achieves a maximum at $t = 0.5$. □

Under cheap talk $t = 0.5 - 2b$. In this case, the Sender's ex-ante expected payoff is

$$\begin{aligned}&\frac{(0.5 - 2b - 1)^3 - (0.5 - 2b)^3}{12} - b^2 \\ &= \frac{16.5b^3 + 10.25 - 19.25b + 10.25b^2 - 12b^2 - 9}{12}.\end{aligned}$$

As observed already in Crawford & Sobel (1982), for any $t \in [0,1]$, that $\rho(t) - \sigma(t) = b^2$. Therefore, the informed Sender's and the Receiver's payoffs vary with t in the same way.

If the informed Sender can commit to a choice of t before the state is realized, i.e., engage in Bayesian persuasion as in Kamenica & Gentzkow (2011), then, $\sigma(t)$ would represent the informed Sender's payoff at threshold t. From the proof of Lemma 3.2 we see that the threshold that maximizes the Sender's ex-ante expected payoff does not depend on her bias.

4. THE UNINFORMED SENDER

We introduce an uninformed Sender who prefers higher to lower actions *independent* of the state (in Section 4.1, we consider the reverse). In a sense the

uninformed and informed Sender are 'pulling' in the same direction, as each would choose a higher action than the Receiver would.

The Receiver does not know if the Sender they are facing is informed or not. They believe that the Sender is informed with probability p. When $p = 1$, we recover the standard cheap talk model, and when $p = 0$, there is no possibility of valuable information being transmitted to the Receiver and, thus, we only have a "babbling" equilibrium in which the Receiver's best response to any message is to choose the action that corresponds to the mean of the distribution from which the state is drawn; in our model with $\theta \sim U[0,1]$, this entails setting $a = 0.5$. Therefore, in what follows we assume that $p \in (0,1)$.

The uninformed Sender considers the messages that an informed Sender will send with positive probability and picks from among these. It is easy to see that the uninformed Sender always sends the message T (see Section 5.1). Even if the uninformed Sender were to learn the state but the Receiver is unaware of this, the previously uninformed Sender will always send message T. Hence, the uninformed Sender gains no benefit from knowing the state.

One might imagine that an uninformed Sender is no different than an informed Sender with a very large positive bias (exceeding 0.25). Section 5.1 shows this is not the case because in some instances the high bias informed Sender will send message B.

The Receiver accounts for the possibility that the Sender is uninformed when interpreting message T. This causes the informed Sender to adjust her equilibrium threshold t. Our first result shows that the Receiver and the informed Sender are ex-ante worse off compared to the case $p = 1$.

The intuition for this is straightforward. Given the focus on two message equilibria, to determine the effect of a bullshitter, it is enough to look at the indifferent informed Sender in equilibrium. The equilibrium without bullshitters is a partition in which all Senders with types in $[0, t^*)$ send the message B and all Senders with types in $[t^*, 1]$ send message T. Given $b > 0$, the indifferent type t^* satisfies $t^* < 0.5$. The introduction of a bullshitter will change the indifferent type to $\bar{t}$, say. If $\bar{t} < t^*$, the equilibrium provides less information (the partition is more asymmetric), while if it is higher than t^* then the equilibrium is more informative, as long as $|\bar{t} - 0.5| < |t^* - 0.5|$. This is because partitions that lead to lower expected variance lead to a higher payoff. Equal sized partitions always have lower variance than unequal ones (see Goltsman et al. 2009).

Henceforth, we denote by IS the event that the Sender is informed and de-

note by US the event that Sender is uninformed. We use T to denote message T as well as the event that message T was sent. Similarly with B.

Theorem 4.1. *Suppose the uninformed Sender prefers higher to lower actions. Let t^* be the equilibrium threshold in a non-babbling equilibrium. Then, the following are true.*

1. $t^* \leq 0.5 - 2b$.
2. *The ex-ante payoff of an informed Sender with bias $b \leq 0.25$ and the Receiver is lower than their ex-ante payoff when $p = 1$.*

4.1. Preference for Lower Actions

Now, suppose the uninformed Sender prefers **lower** to higher actions independent of the state. Now, the uninformed Sender and informed Sender are "pulling" in different directions. In Appendix A, we show that the Receiver's expected action is increasing in the probability that the uninformed Sender sends message T. Therefore, the uninformed Sender with preference for *lower* actions will always send message B. The Receiver takes this into account when interpreting message B and in turn this will cause the informed Sender to adjust their equilibrium threshold t.

Theorem 4.2. *Suppose the uninformed Sender prefers lower to higher actions. Then, there is a non-babbling equilibrium where the equilibrium threshold exceeds $0.5 - 2b$.*

Whether the informed Sender or Receiver is better off will depend on whether this threshold is smaller or larger than 0.5. From the proof of Theorem 4.2 in Appendix A, the equilibrium threshold t^* satisfies:

$$t^* = \frac{-(3+4(b-1)p) + \sqrt{16b^2p^2 - 8(b+1)p + 9}}{4p}.$$

When $b = 0$, it is straightforward to verify that $t^* > 0.5$ for all $p \in (0,1)$. Therefore, by Lemma 3.2, the informed Sender will be ex-ante worse off compared to $p = 1$. By continuity it follows that for some small $b > 0$, the possibility of a bullshitter lowers the informed Sender's ex-ante expected payoff.

5. ROBUSTNESS

In this section we consider several variations of our model to verify the robustness of our main qualitative insights. Section 5.1 replaces the uninformed Sender with a highly biased informed Sender. Section 5.2 introduces an additional uninformed Sender with opposing preferences. Section 5.3 examines equilibria when the informed Sender is permitted to partition the message space into more intervals.

5.1. High bias vs Bullshit

Suppose now both types of Senders are informed but differ in their bias and we exclude the possibility that they can send cheap talk messages about their biases. With probability p, the informed Sender has low bias and this event is denoted IS^L. With probability $1-p$, the informed Sender has high bias and this event is denoted IS^H. We compare the value of the information sent to the Receiver in this case with the case in which the Receiver faces either a low-biased informed Sender or an uninformed Sender with preference for higher actions.

Denote by $t_l \in [0,1]$ the equilibrium threshold used by IS^L to determine her message. The corresponding threshold for IS^H is denoted $t_h \in [0,1]$. Without loss of generality, we assume $t_h > t_l$. The Receiver's best response upon seeing message T is given by:

$$a_T = 0.5\left(\frac{p(t_l^2 - t_h^2) + t_h^2 - 1}{p(t_l - t_h) + t_h - 1}\right).$$

The Receiver's best response upon seeing message B is:

$$a_B = 0.5\left(\frac{p(t_l^2 - t_h^2) + t_h^2}{p(t_l - t_h) + t_h}\right).$$

Now, t_l must satisfy the following indifference condition:

$$\left(0.5\left(\frac{p({t_l}^2 - {t_h}^2) + {t_h}^2 - 1}{p(t_l - t_h) + t_h - 1}\right) - (t_l + b_l)\right)^2 = \left((t_l + b_l) - 0.5\left(\frac{p({t_l}^2 - {t_h}^2) + {t_h}^2}{p(t_l - t_h) + t_h}\right)\right)^2$$

Similarly, t_h must satisfy the following indifference condition:

$$\left(0.5\left(\frac{p({t_l}^2-{t_h}^2)+{t_h}^2-1}{p(t_l-t_h)+t_h-1}\right)-(t_h+b_h)\right)^2$$
$$=\left((t_h+b_h)-0.5\left(\frac{p({t_l}^2-{t_h}^2)+{t_h}^2}{p(t_l-t_h)+t_h}\right)\right)^2$$

Interior solutions to the above are unavailable in closed form. However, we show below that these equilibrium conditions imply that a highly biased informed Sender may behave in a way distinct from an uninformed Sender.

When p is close to 0.5, there exist many (b_l, b_h) pairs that support an interior equilibrium. However, b_h and b_l must be close to zero. Thus, for some p there may be no pure strategy two message cheap talk equilibrium with a highly biased informed Sender. If there is a mixed strategy equilibrium, it would mean that the informed Sender with high bias must send message B with positive probability. Therefore, a highly biased informed Sender may behave differently from an uninformed Sender with a preference for higher actions. In particular, a highly biased informed Sender may inhibit communication altogether.

Lemma 5.1. *In a binary message space, there is a choice of $b_l < 0.25$ and b_h such that an informed Sender with high bias is distinguishable from an uninformed Sender.*

Proof. Suppose there is an equilibrium where $a_T \neq a_B$, i.e., some information is transmitted. The equilibrium threshold t of an informed Sender with bias b must satisfy the following indifference condition:

$$(a_T-(t+b))^2=((t+b)-a_B)^2$$
$$\Rightarrow[(a_T-(t+b))+(t+b-a_B)][(a_T-(t+b))-(t+b-a_B)]=0$$
$$\Rightarrow[a_T-a_B][a_T+a_B-2b-2t]=0.$$

Given $a_T \neq a_B$ (to exclude babbling), it follows that $a_T+a_B-2b-2t=0$.

At an equilibrium we need the indifference condition for both the high *and* low bias informed Sender to be satisfied, i.e.:

$$a_T+a_B-2b_h-2t_h=0,\ a_T+a_B-2b_l-2t_l=0.$$

Here t_h and t_l are the equilibrium thresholds for the high and low bias informed Sender respectively. Combining these equations, we obtain:

$$-2b_h + 2b_l - 2t_h + 2t_l \;\Rightarrow\; t_h - t_l = b_h - b_l.$$

Suppose b_h and b_l take on values such that $t_h, t_l \in (0,1)$. Then, the high-bias informed Sender will send message B with positive probability. Recall that the uninformed Sender always sends message T. □

5.2. Two Uninformed Senders

Here we consider the possibility that the preferences of the uninformed Sender are unknown. We model this by assuming the possibility of two uninformed Senders in addition to an informed Sender. There is an uninformed Sender who prefers higher actions, denoted by US^H, and an uninformed Sender who prefers lower actions, denoted by US^L. Nature selects one from among US^H, US^L or the informed Sender with probabilities p_1, p_2, and $(1-p_1-p_2)$, respectively. We use IS to denote the event that nature selects an informed Sender.

The informed Sender sends a state-dependent message, while US^H and US^L send their respective messages without regard to the state. We determine whether the informed Sender and the Receiver, denoted R, are better or worse off than in the case of no uninformed Senders (i.e. $p_1 = p_2 = 0$).

As before, $t \in (0,1)$ denotes the threshold which determines which message IS sends. IS sends message B when $\theta \in [0,t)$, and message T when $\theta \in [t,1]$. The strategy of the uninformed senders is, possibly, mixed. US^H sends message T with probability α and B with probability $(1-\alpha)$ for all $\theta \in [0,1]$, and US^L sends message T with probability β and B with probability $(1-\beta)$ for all $\theta \in \Theta$. Once the Sender is chosen and the message sent, R observes it, but not the identity of the Sender.

R observes α, β, p_1, and p_2, using these to inform her beliefs on the Sender's type and, thus, θ. Using these beliefs, R chooses an action to maximize her expected payoff. R's best response to message T is given by:

$$a_T = 0.5\left(\frac{p_1(\alpha+t^2-1)+p_2(\beta+t^2-1)-t^2+1}{p_1(\alpha+t-1)+p_2(\beta+t-1)-t+1}\right).$$

R's best response to message B is then given by:

$$a_B = 0.5\left(\frac{p_1(\alpha+t^2-1)+p_2(\beta+t^2-1)-t^2}{p_1(\alpha+t-1)+p_2(\beta+t-1)-t}\right).$$

US^H will choose α to solve $\max_{\alpha\in[0,1]}[a_T\alpha + a_B(1-\alpha)]$. In Appendix B, we show that the optimal solution is $\alpha = 1$. Thus, US^H will only send message T. Similarly, US^L will only send message B.

The informed Sender's threshold t must satisfy the following indifference condition:

$$-((t+b)-a_B)^2 = -(a_T-(t+b))^2.$$

Focusing on the non-babbling case:

$$\begin{aligned} &0.5\left(\frac{p_1t^2+p_2(t^2-1)-t^2+1}{p_1t+p_2(t-1)-t+1}\right)-t-b \\ &= t+b-0.5\left(\frac{p_1t^2+p_2(t^2-1)-t^2}{p_1t+p_2(t-1)-t}\right). \end{aligned} \tag{1}$$

The solution to equation (1) is not monotone in either p_1 or p_2 but it does decrease with b.

Lemma 5.2. *Suppose $p_1 = p_2 = p_u \in (0, 0.5)$. Then, in a non-babbling equilibrium, ex-ante, the informed Sender is always better off.*

In the presence of two uninformed Senders ($p_1, p_2 > 0$), one can support non-babbling equilibria in which $b \in (\frac{1}{4}, \frac{1}{2})$. Recall that in the standard cheap talk setting, such equilibria exist only when $b < \frac{1}{4}$.

Unsurprisingly, the informed Sender's payoff decreases monotonically in b. When the informed Sender is *highly biased* (i.e., $b > 0.25$), the presence of uninformed Senders makes the informed Sender strictly better off for any value of $p_u \in (0, \frac{1}{2})$.

Turning to the other extreme, when $b = 0$, the presence of the uninformed Senders does not reduce the ex-ante expected payoff of the informed Sender for any value of $p_u > 0$. Hence, there will be a threshold bias level, denoted b^* depending on p_u, where for all $b < b^*$, the informed Sender is no worse off under $p_u > 0$ compared to $p_u = 0$. As p_u increases, the informed Sender's payoff peaks and then decreases, eventually falling below the standard cheap talk payoff. Essentially, an informed Sender with sufficiently large bias achieves higher utility levels when there is a positive but small prevalence of bullshit. Note that in this setting bullshit is, on average, neutral; the expected bias of the uninformed message is 0.

In the presence of a single uninformed Sender, the informed Sender's message is less informative. However, two uninformed Sender's with opposing

preferences– which is essentially "noise" in the model – can benefit the informed Sender and the Receiver! This may appear unsurprising given Blume et al. (2007) which allows for noisy communication in the standard cheap talk model. In that paper the Receiver with some probability $\lambda \in [0,1]$ receives a state independent random message from a continuous distribution. They show for λ sufficiently small, there exists an equilibrium that is Pareto superior to all noiseless cheap talk equilibria. This result does not imply the claim we make here because Blume et al. (2007) relies on flexibility in the choice of the number of messages. In their example 1, for instance, the Pareto optimal equilibrium in the noiseless cheap talk model has two messages, while their Pareto superior equilibrium with noise has three messages. In our case we hold the number of messages *fixed* at two.

5.3. The Enlarged Message Space

Crawford & Sobel (1982) and Gordon (2010) show that, for sufficiently small bias, there exist cheap talk equilibria that support more than two messages. In the standard cheap talk model (where $p = 1$), both the informed Sender and the Receiver are better off with a larger message space. Here we examine whether the same is true in the presence of an uninformed Sender. We restrict ourselves to a three message equilibrium where the uninformed sender does not mix.

The informed Sender uses thresholds $t_1 \in [0,1]$ and $t_2 \in [0,1]$ where $t_2 > t_1$ to determine which message to send. Let the informed Sender send message B when $\theta \in [0,t_1]$, message M when $\theta \in [t_1,t_2]$, and message T when $\theta \in [t_2,1]$. In Appendix C, we show that there is an equilibrium where the uninformed Sender sends only message T.

In the absence of an uninformed Sender, a three message equilibrium exists provided the following condition on the informed Sender's bias is satisfied (Crawford & Sobel 1982):

$$2 \leq \frac{1}{2}\left(\sqrt{1+\frac{2}{b}}-1\right) \Rightarrow b \leq \frac{1}{12}.$$

When the uninformed Sender is present with probability $1-p$, a two message equilibrium is supported when

$$b \leq \frac{2-\sqrt{4-3p}}{12} \text{ or } b \geq \frac{2+\sqrt{4-3p}}{12}.$$

Hence, the value of b under which a three message equilibrium exists under cheap talk *and* in the presence of an uninformed Sender must satisfy $b \leq \left(2-\sqrt{4-3p}\right)/12$. For such a b, the possible values of p consistent with a three message equilibrium in the presence of an uninformed Sender must be at least as large as $16b(1-3b)$.

We have determined numerically that for $b = 1/20$ there is a threshold such that for p below it, the informed Sender's non-babbling equilibrium payoff with three messages is *below* their payoff in the two message non-babbling equilibrium. This is illustrated in Figure 1. A closed form expression for this threshold is unavailable. The possible values of p are plotted on the horizontal while the vertical axes corresponds to the expected payoff of the informed Sender.

The orange curve shows how the expected payoff of the informed Sender in a three message equilibrium changes with p. The blue curve shows how the expected payoff of the informed Sender in a two message equilibrium changes with p. Both curves are plotted only for the range of p under which a three message non-babbling equilibrium exists, i.e. $p \geq 16b(1-3b)$. The red region corresponds to the values of p for which both non-babbling two and three message equilibria exist but where a two message equilibrium is preferable to the informed Sender.

To understand why this is happening, we examined numerically the case $b = 1/20$. The resulting partition induced by the three message equilibrium is very uneven, increasing expected variance. This effect offsets the benefits of being able to more closely tailor the message to the realized state.

In pure cheap talk, a larger message space allows the informed Sender to transmit more information to the Receiver. However, in the presence of an uninformed Sender, enlarging the message space need not increase the informed Sender's or Receiver's expected payoff. For $p \geq 16b(1-3b)$ the informed sender's expected payoff under both a two and three message equilibrium increases in p. However, there exists exactly one $p_s \in (16b(1-3b), 1)$ below which the informed Sender's expected payoff in a three message equilibrium is strictly below that in a two message equilibrium. For $p > p_s$ the converse is true. Now, the Receiver's expected payoffs in both the two and three message equilibria also increase in p. There exists a $p_r \in (16b(1-3b), 1)$ in which the Receiver's expected payoff in the three message equilibria is below that of the two message equilibrium for $p < p_r$. It appears that $p_s > p_r$ for all $b \in [0, 1/12]$. Additionally, it seems that p_s and p_r each increase in b, while

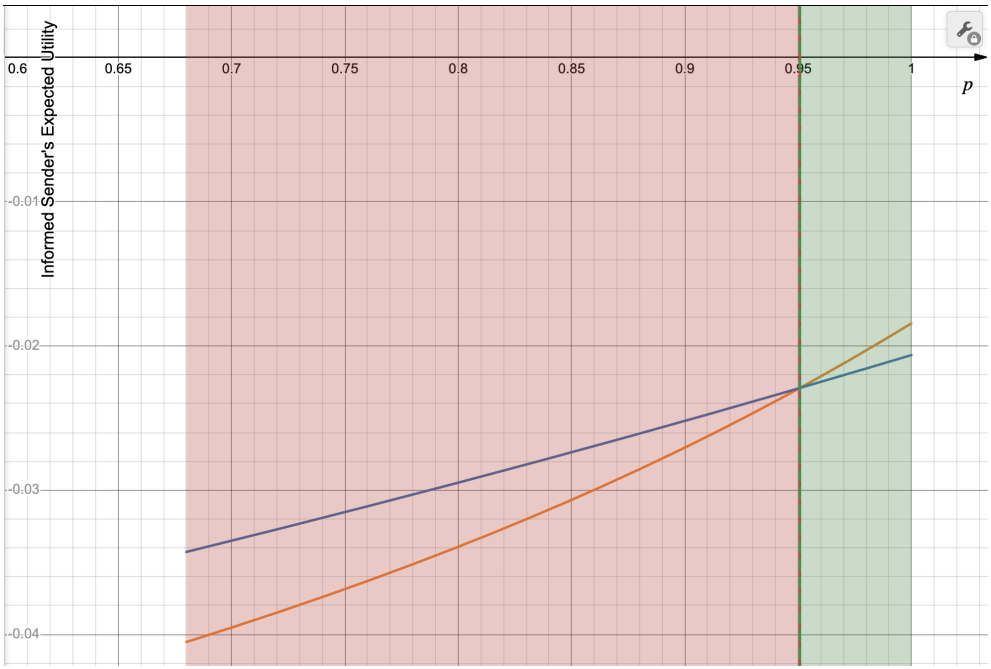

Figure 1: IS's ex ante expected utility as function of p

$p_s - p_r$ decreases in b.

Thus, the presence of an uninformed Sender harms communication in two ways. First, it may cause the informed Sender and Receiver to prefer a smaller message space. This limits the *amount* of information the informed Sender is able to convey to the Receiver. Second, the informed Sender's equilibrium thresholds change, which changes the *type* of information the informed Sender is able to convey to the Receiver.

6. CHEAP TALK VS. BAYESIAN PERSUASION

It is natural to ask what would happen if the informed Sender could commit, before learning the state, to a disclosure policy. This is the Bayesian persuasion setting of Kamenica & Gentzkow (2011). Persuasion differs from cheap talk in that the informed Sender's bias influences the disclosure policy but not the inferences that the Receiver draws.

In this section, we allow the informed Sender with bias $b > 0$ to influence the Receiver's action by committing to a single threshold prior to learning

the state and promising the Receiver to faithfully send one message when the state is above the threshold and a different message when the state is below the threshold. As before, the uninformed Sender prefers higher actions to lower ones independent of the state.

It is easy to see that the informed Sender will do better when they can commit to a message strategy. It is less clear that the Receiver will benefit in the presence of an uninformed Sender. As before, we restrict all Senders to one of two messages. The uninformed Sender sends message T with probability 1. The Receiver's best response to each message remains as before:

$$a_T = 0.5\left(\frac{pt^2-1}{pt-1}\right),\ a_B = 0.5t.$$

The informed Sender chooses her threshold t *before* the state is realized so as to maximize her expected ex ante utility:

$$\begin{aligned}\max_t E[u_{IS}] &= \max_{t\in[0,1]} E[u_{IS}|T]P(T) + E[u_{IS}|B]P(B)\\ &= \max_{t\in[0,1]} -\frac{pt(3t^2-6t+4)-1}{12(pt-1)} - b^2.\end{aligned}$$

The first-order optimality condition for an interior optimal is:

$$2pt^3 - 2pt^2 - 3t^2 + 4t - 1 = 0.$$

The optimal choice of t, denoted t^* satisfies

$$t^* = \frac{3-\sqrt{9-8p}}{4p}.$$

The expected payoff of the Receiver is

$$\frac{(4p+3\sqrt{9-8p}-9)(9-8p)}{96p^2}.$$

The expected payoff of the informed Sender is

$$\frac{(4p+3\sqrt{9-8p}-9)(9-8p)}{96p^2} - b^2.$$

Lemma 6.1. *If the uninformed Sender has a preference for higher actions, the informed Sender and Receiver are strictly better off under persuasion than cheap talk.*

Proof. If not, the equilibrium payoff of the Receiver under cheap talk exceeds that of its payoff under persuasion. In other words,

$$\begin{aligned}&\frac{27(\sqrt{8p(2b^2p+b-1)+9}-3)}{96p^2}\\&-\frac{4p(12b^2p+(3b+6)\sqrt{8p(2b^2p+b-1)+9}}{96p^2}\\&+\frac{8p-27)}{96p^2}>\frac{(4p+3\sqrt{9-8p}-9)(9-8p)}{96p^2}.\end{aligned}$$

This implies that

$$\begin{aligned}&27\left(\sqrt{8p(2b^2p+b-1)+9}-3\right)\\&-4p\left(12b^2p+(3b+6)\sqrt{8p(2b^2p+b-1)+9}+8p-27\right)>\\&\left(4p+3\sqrt{9-8p}-9\right)(9-8p).\end{aligned}$$

Therefore $p<0$ which cannot be. □

References

Blume, A., Board, O. J., & Kawamura, K. (2007). Noisy talk. *Theoretical Economics*, *2*(4), 395–440.

Crawford, V. P., & Sobel, J. (1982). Strategic information transmission. *Econometrica*, *50*(6), 1431–1451.

Frankfurt, H. G. (2009). *On Bullshit*. Princeton University Press.

Goltsman, M., Hörner, J., Pavlov, G., & Squintani, F. (2009). Mediation, arbitration and negotiation. *Journal of Economic Theory*, *144*(4), 1397-1420.

Gordon, S. (2010). *On infinite cheap talk equilibria* (Tech. Rep.). University of Montreal.

Gordon, S., Kartik, N., yu Lo, M. P., Olszewski, W., & Sobel, J. (2021). *Effective communication in cheap talk games* (Tech. Rep.). Columbia University.

Kamenica, E., & Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, *101*(6), 2590-2615. doi: 10.1257/aer.101.6.2590

Partridge, E., Dalzell, T., Victor, T., & Partridge, E. (2006). *The New Partridge Dictionary of Slang and Unconventional English*. Routledge.

Sobel, J. (2020). Lying and deception in games. *Journal of Political Economy*, *128*(3), 907-947.

Appendix A

Suppose two types of Senders: an informed one and an uninformed one who prefers higher actions. In this section we derive the equilibrium actions of each Sender and the Receiver.

Nature draws the state θ uniformly at random from $[0,1]$ and selects the type of Sender (informed or uninformed), with probabilities p and $(1-p)$, respectively with $p \in (0,1)$. If the uninformed is selected, the uninformed Sender sends message T with probability α and message B with probability $(1-\alpha)$ for all $\theta \in \Theta$. If the informed Sender is selected, the informed Sender chooses a threshold $t \in [0,1]$ and sends message B when $\theta \in [0,t)$, and T when $\theta \in [t,1]$.

Lemma 6.2. *Let a_T denote the Receiver's best response upon seeing message T. Then,*

$$a_T = 0.5\left(\frac{p(\alpha+t^2-1)-\alpha}{p(\alpha+t-1)-\alpha}\right).$$

Proof. The probability that the Receiver sees message T is $p(1-t)+(1-p)\alpha$. Given message T, the probability that the Sender is informed is

$$\frac{p(1-t)}{p(1-t)+(1-p)\alpha}.$$

The probability that the Sender is uninformed is

$$\frac{(1-p)\alpha}{p(1-t)+(1-p)\alpha}.$$

The Receiver's expected payoff from choosing action a is given by:

$$E[u_R|T] = E[u_R|IS,T]P(IS|T) + E[u_R|US,T]P(US|T) =$$

$$\frac{p(1-t)}{p(1-t)+(1-p)\alpha}\int_t^1 -\frac{(a-\theta)^2}{1-t}d\theta + \frac{(1-p)\alpha}{p(1-t)+(1-p)\alpha)}\int_0^1 -(a-\theta)^2 d\theta =$$

$$-\frac{p\int_t^1 (a^2-2a\theta+\theta^2)\,d\theta}{p(1-t)+(1-p)\alpha} - \frac{(1-p)\alpha\int_0^1 (a^2-2a\theta+\theta^2)\,d\theta}{p(1-t)+(1-p)\alpha}$$

To determine the Receiver's optimal choice of a we determine when the derivative of (A) is zero:

$$\partial \left[-\frac{p\int_t^1 (a^2 - 2a\theta + \theta^2)\, d\theta}{p(1-t) + (1-p)\alpha} - \frac{(1-p)\alpha \int_0^1 (a^2 - 2a\theta + \theta^2)\, d\theta}{p(1-t) + (1-p)\alpha} \right] \Big/ \partial a = 0$$

$$\Rightarrow -2 \left(\frac{p\int_t^1 (a-\theta)\, d\theta}{p(1-t) + (1-p)\alpha} + \frac{(1-p)\alpha \int_0^1 (a-\theta)\, d\theta}{p(1-t) + (1-p)\alpha} \right) = 0$$

$$\Rightarrow \frac{p\left[a\theta - 0.5\theta^2\right]_t^1}{p(1-t) + (1-p)\alpha} + \frac{(1-p)\alpha\left[a\theta - 0.5\theta^2\right]_0^1}{p(1-t) + (1-p)\alpha} = 0$$

$$\Rightarrow a_T = 0.5 \left(\frac{p(\alpha + t^2 - 1) - \alpha}{p(\alpha + t - 1) - \alpha} \right). \qquad \square$$

Lemma 6.3. *Let a_B denote the Receiver's best response upon seeing message B. Then,*

$$a_B = 0.5 \left(\frac{p(\alpha + t^2 - 1) - \alpha + 1}{p(\alpha + t - 1) - \alpha + 1} \right).$$

Proof. The probability the Receiver sees message B is $pt + (1-p)(1-\alpha)$. Given that the message is B, the probability that the Sender is informed

$$\frac{pt}{pt + (1-p)(1-\alpha)}.$$

Given that the message is B, the probability that the Sender is uninformed is

$$\frac{(1-p)(1-\alpha)}{pt + (1-p)(1-\alpha)}.$$

The Receiver's expected payoff from choosing action a is given by:

$$E[u_R|B] = E[u_R|IS, B]P(IS|B) + E[u_R|US, B]P(US|B)$$

$$= \frac{pt}{pt + (1-p)(1-\alpha)} \int_0^t -(a-\theta)^2 \frac{1}{t} d\theta + \frac{(1-p)(1-\alpha)}{pt + (1-p)(1-\alpha)} \int_0^1 -(a-\theta)^2 d\theta$$

$$= -\frac{p\int_0^t (a^2 - 2a\theta + \theta^2)\, d\theta}{pt + (1-p)(1-\alpha)} - \frac{(1-p)(1-\alpha)\int_0^1 (a^2 - 2a\theta + \theta^2)\, d\theta}{pt + (1-p)(1-\alpha)}. \qquad (2)$$

To determine the Receiver's optimal choice of a we set the derivative of (2) with respect to a equal to zero:

$$\partial\left[-\frac{p\int_0^t(a^2-2a\theta+\theta^2)\,d\theta}{pt+(1-p)(1-\alpha)}-\frac{(1-p)(1-\alpha)\int_0^1(a^2-2a\theta+\theta^2)\,d\theta}{pt+(1-p)(1-\alpha)}\right]\Big/\partial a=0$$

$$\Rightarrow -2\left(\frac{p\int_0^t(a-\theta)\,d\theta}{pt+(1-p)(1-\alpha)}+\frac{(1-p)(1-\alpha)\int_0^1(a-\theta)\,d\theta}{pt+(1-p)(1-\alpha)}\right)=0$$

$$\Rightarrow \frac{p\left[a\theta-0.5\theta^2\right]_0^t}{pt+(1-p)(1-\alpha)}+\frac{(1-p)(1-\alpha)\left[a\theta-0.5\theta^2\right]_0^1}{pt+(1-p)(1-\alpha)}=0$$

$$\Rightarrow a_B=0.5\left(\frac{p(\alpha+t^2-1)-\alpha+1}{p(\alpha+t-1)-\alpha+1}\right).$$

□

Lemma 6.4. *The uninformed Sender's optimal choice of α is $\alpha=1$.*

Proof. The uninformed Sender's problem is:

$$\begin{aligned}\max_\alpha E[a] &= \max_\alpha[\alpha a_T+(1-\alpha)a_B]\\ &= \max_{\alpha\in[0,1]}\left(0.5\left(\frac{p(\alpha+t^2-1)-\alpha}{p(\alpha+t-1)-\alpha}\right)\times\alpha\right.\\ &\qquad\left.+0.5\left(\frac{p(\alpha+t^2-1)-\alpha+1}{p(\alpha+t-1)-\alpha+1}\right)\times(1-\alpha)\right)\end{aligned}$$

The first-order condition for an interior optimal yields the following:

$$\alpha=-\frac{p^2(t-1)-2p(t-1)+\sqrt{(p-1)^2(t-1)t}+t-1}{(p-1)^2}. \tag{3}$$

Given $t\in(0,1)$ and $p\in(0,1)$, equation (3) has no solution with $\alpha\in(0,1)$. Hence, the optimal value of α is either 0 or 1. Direct computation shows that it must be $\alpha=1$. □

Theorem 4.1

Proof. It is clear that in equilibrium the uninformed Sender will always send message T (a formal proof is in Appendix A). If t is the threshold used by the informed Sender,

the Receiver's best response to messages T and B as a function of t, are given by (see Appendix A):

$$a_T = 0.5\left(\frac{pt^2-1}{pt-1}\right),\ a_B = 0.5t.$$

The informed Sender's indifference condition is:

$$-\left((t+b)-a_B\right)^2 = -\left(a_T-(t+b)\right)^2.$$

This leads to two possibilities. The first is that

$$(t+b)-a_B+a_T-(t+b)=0 \ \Rightarrow a_B = a_T.$$

Hence,

$$pt^2-t = pt^2-1 \ \Rightarrow t=1,$$

which corresponds to a babbling equilibrium which we ignore. The second possibility is that

$$\begin{aligned}
(t+b)-a_B &= a_T-(t+b)\\
\Rightarrow t+b-0.5t &= 0.5\frac{pt^2-1}{pt-1}-t-b\\
\Rightarrow 0.5t+b &= 0.5\frac{pt^2-1}{pt-1}-t-b\\
\Rightarrow 1.5t+2b &= 0.5\frac{pt^2-1}{pt-1}\\
\Rightarrow (pt-1)(1.5t+2b) &= 0.5(pt^2-1)\\
\Rightarrow 1.5pt^2+2bpt-1.5t-2b &= 0.5(pt^2-1)\\
\Rightarrow pt^2+2bpt-1.5t-2b+0.5 &= 0\\
\Rightarrow 2pt^2+4bpt-3t-4b+1 &= 0\\
\Rightarrow 2pt^2+(4bp-3)t-4b+1 &= 0\\
\Rightarrow t = \frac{-(4bp-3)\pm\sqrt{16b^2p^2+8(b-1)p+9}}{4p}&.
\end{aligned}$$

For t to be well defined we need the term in the square root to be non-negative. The expression $16b^2p^2+8(b-1)p+9$ has a derivative with respect to p of $32b^2p-8(1-b)$ which is negative given $b^2 \leq 1/8$. Therefore, the expression achieves a minimum at $p=0$, so the term in the square root is non-negative and the roots are real.

We eliminate the root

$$\frac{-(4bp-3)+\sqrt{16b^2p^2+8(b-1)p+9}}{4p}$$

from consideration because it exceeds 1. To see why, suppose otherwise:

$$\frac{-(4bp-3)+\sqrt{16b^2p^2+8(b-1)p+9}}{4p}<1$$

$$\Rightarrow -(4bp-3)+\sqrt{16b^2p^2+8(b-1)p+9}<4p$$

$$\Rightarrow 16b^2p^2+8(b-1)p+9<16p^2(1+b)^2+9-24p(1+b)$$

$$\Rightarrow 8(b-1)p<16p^2+32p^2b-24p-24bp$$

$$\Rightarrow 1<p,$$

which cannot be.

Therefore, the informed Sender's equilibrium threshold, denoted t^*, satisfies:

$$t^*=\frac{3-4bp-\sqrt{16b^2p^2+8(b-1)p+9}}{4p}. \tag{4}$$

We verify that $t^*\leq 0.5-2b$:

$$\frac{3-4bp-\sqrt{16b^2p^2+8(b-1)p+9}}{4p}\leq 0.5-2b$$

$$\Rightarrow 3-4bp-\sqrt{16b^2p^2+8(b-1)p+9}\leq 2p-8bp$$

$$\Rightarrow 3+4bp-2p\leq \sqrt{16b^2p^2+8(b-1)p+}$$

$$\Rightarrow (3+4bp-2p)^2\leq 16b^2p^2+8(b-1)p+9$$

$$\Rightarrow 9+(4bp-2p)^2+6(4bp-2p)\leq 16b^2p^2+8(b-1)p+9$$

$$\Rightarrow 9+16b^2p^2+4p^2-16bp^2+24bp-12p\leq 16b^2p^2+8(b-1)p+9$$

$$\Rightarrow 4p^2-16bp^2+16bp-4p\leq 0$$

$$\Rightarrow 4p^2(1-4b)+4p(4b-1)\leq 0$$

$$\Rightarrow 4p(p-1)(1-4b)\leq 0,$$

which is clearly true.

We also verify that $t^* \geq 0$:

$$\frac{3-4bp-\sqrt{16b^2p^2+8(b-1)p+9}}{4p} \geq 0$$

$$\Rightarrow 3-4bp \geq \sqrt{16b^2p^2+8(b-1)p+9}$$

$$\Rightarrow 9+16b^2p^2-24bp \geq 16b^2p^2+8(b-1)p+9$$

$$\Rightarrow -24bp \geq +8(b-1)p$$

$$\Rightarrow 8p \geq 32bp$$

$$\Rightarrow 0.25 \geq b,$$

which is true by assumption.

Now, we compare the informed Sender's ex-ante payoff which is

$$-\int_0^{t^*}(\theta+b-0.5t^*)^2d\theta-\int_{t^*}^1(0.5\frac{p(t^*)^2-1}{pt^*-1}-\theta-b)^2d\theta,$$

with their ex-ante payoff when $p=1$ which is $\sigma(0.5-2b)$.

Consider the function $\phi(x)=\int_{t^*}^1(x-\theta-b)^2d\theta$. A straightforward calculation shows that it is convex and achieves a minimum at $x=0.5(1+t^*)+b$. Therefore, for $x \leq 0.5(1+t^*)+b$, $\phi(x)$ is decreasing. Now, $\frac{p(t^*)^2-1}{pt^*-1} \leq 1+t^*$. Hence,

$$-\int_0^{t^*}(\theta+b-0.5t)^2d\theta-\int_{t^*}^1(0.5\frac{p(t^*)^2-1}{pt-1}-\theta-b)^2d\theta$$

$$\leq -\int_0^{t^*}(\theta+b-0.5t)^2d\theta-\int_{t^*}^1(0.5(1+t^*)-\theta-b)^2d\theta$$

$$=\sigma(t^*) \leq \sigma(0.5-2b).$$

The last inequality follows from Lemma 3.2 and the fact that $t^* \leq 0.5-2b$.

The Receiver's expected payoff is

$$-\int_0^{t^*}(\theta-0.5t)^2d\theta-\int_{t^*}^1(0.5\frac{p(t^*)^2-1}{pt-1}-\theta)^2d\theta.$$

We will use Lemma 3.1 to compare it with their expected payoff when $p=1$ which is $\rho(0.5-2b)$. Consider the function $\mu(x)=\int_{t^*}^1(x-\theta)^2d\theta$. A straightforward calculation shows that it is convex and achieves a minimum at $x=0.5(1+t^*)$. Therefore, for

$x \leq 0.5(1+t^*)$, $\mu(x)$ is decreasing. As $\frac{p(t^*)^2-1}{pt^*-1} \leq 1+t^*$ it follows that the Receiver's expected payoff is bounded above by

$$-\int_0^{t^*}(\theta-0.5t)^2 d\theta - \int_{t^*}^1 (0.5(t^*+1)-\theta)^2 d\theta = \rho(t^*) \leq \rho(0.5-2b).$$

□

Theorem 4.2

Proof. Given that the uninformed Sender strictly prefers to send message B, the Receiver's best response – as functions of the informed Sender's threshold t – to messages T and B, respectively, are given by (see Appendix A):

$$a_T = 0.5(t+1),\ a_B = 0.5\left(\frac{p(t^2-1)+1}{p(t-1)+1}\right).$$

The informed Sender's indifference condition is:

$$-((t+b)-a_B)^2 = -(a_T-(t+b))^2.$$

There are two possibilities. The first is that

$$a_T = a_B \Rightarrow p(t+1)(t-1)+(t+1) = p(t^2-1)+1 \Rightarrow t=0.$$

This corresponds to a babbling equilibrium which we ignore. The second possibility is that

$$\begin{aligned}
t+b-a_B &= a_T - t - b \\
\Rightarrow t+b-0.5\left(\frac{p(t^2-1)+1}{p(t-1)+1}\right) &= 0.5(t+1)-t-b \\
\Rightarrow \left(\frac{p(t^2-1)+1}{p(t-1)+1}\right)+t+1 &= 4(t+b) \\
\Rightarrow p(t^2-1)+1+p(t-1)(t+1)+t+1 &= 4(t+b)[p(t-1)+1] \\
\Rightarrow 2p(t^2-1)+t+2 &= 4(t+b)[p(t-1)+1] \\
\Rightarrow 2p(t^2-1)+t+2 &= 4pt^2-4pt+4bpt-4bp+4(t+b) \\
\Rightarrow 2pt^2+(4(b-1)p+3)t-4bp+2p+4b-2 &= 0 \\
\Rightarrow 2pt^2+(4(b-1)p+3)t+2p(1-2b)+2(2b-1) &= 0 \\
\Rightarrow 2pt^2+(4(b-1)p+3)t+2p(1-2b)-2(1-2b) &= 0 \\
\Rightarrow 2pt^2+(4(b-1)p+3)t+2(p-1)(1-2b) &= 0
\end{aligned}$$

$$\Rightarrow 2pt^2 + (4(b-1)p+3)t - 2(1-p)(1-2b) = 0.$$

There are two roots:

$$t^* = \frac{-(3+4(b-1)p) \pm \sqrt{16b^2p^2 - 8(b+1)p+9}}{4p}.$$

Consider first

$$t^* = \frac{-(3+4(b-1)p) + \sqrt{16b^2p^2 - 8(b+1)p+9}}{4p}. \tag{5}$$

We verify that it is in $[0.5-2b, 1]$:

$$\frac{-(3+4(b-1)p) + \sqrt{16b^2p^2 - 8(b+1)p+9}}{4p} \leq 1$$

$$\Rightarrow -(3+4(b-1)p) + \sqrt{16b^2p^2 - 8(b+1)p+9} \leq 4p$$

$$\Rightarrow \sqrt{16b^2p^2 - 8(b+1)p+9} \leq 4p + (3+4(b-1)p)$$

$$\Rightarrow \sqrt{16b^2p^2 - 8(b+1)p+9} \leq 4p + 3 + 4bp - 4p$$

$$\Rightarrow \sqrt{16b^2p^2 - 8(b+1)p+9} \leq 3 + 4bp$$

$$\Rightarrow 16b^2p^2 - 8(b+1)p+9 \leq 9 + 16b^2p^2 + 24bp$$

$$\Rightarrow -8(b+1)p \leq 24bp,$$

which is certainly true.

Next, we show that this root exceeds $0.5-2b$. Suppose not, then

$$\frac{-(3+4(b-1)p) + \sqrt{16b^2p^2 - 8(b+1)p+9}}{4p} \leq 0.5-2b$$

$$\Rightarrow -(3+4(b-1)p) + \sqrt{16b^2p^2 - 8(b+1)p+9} \leq 2p - 8pb$$

$$\Rightarrow \sqrt{16b^2p^2 - 8(b+1)p+9} \leq 2p - 8pb + 3 + 4(b-1)p$$

$$\Rightarrow \sqrt{16b^2p^2 - 8(b+1)p+9} \leq 2p - 8pb + 3 + 4bp - 4p$$

$$\Rightarrow \sqrt{16b^2p^2 - 8(b+1)p+9} \leq -2p - 4pb + 3$$

$$\Rightarrow 16b^2p^2 - 8(b+1)p+9 \leq 9 + (2p+4pb)^2 - 6(2p+4pb)$$

$$\Rightarrow 16b^2p^2 - 8bp - 8p \leq (2p+4pb)^2 - 12p - 24bp$$

$$\Rightarrow 16b^2p^2 \leq (2p+4pb)^2 - 4p - 16bp$$

$$\Rightarrow 16b^2p^2 \leq 4p^2 + 16p^2b^2 + 16p^2b - 4p - 16bp$$

$$\Rightarrow 0 \leq 4p^2 + 16p^2b - 4p - 16bp$$

$$\Rightarrow 0 \leq 4p^2(1+4b) - 4p(1+4b) = 4p(p-1)(1+4b),$$

which is impossible given $p \leq 1$.

Now, we consider the other root

$$t^* = \frac{-(3+4(b-1)p) - \sqrt{16b^2p^2 - 8(b+1)p + 9}}{4p}. \tag{6}$$

This root can be eliminated because it is negative. To see why, suppose otherwise:

$$\frac{-(3+4(b-1)p) - \sqrt{16b^2p^2 - 8(b+1)p + 9}}{4p} \geq 0$$

$$\Rightarrow -(3+4(b-1)p) - \sqrt{16b^2p^2 - 8(b+1)p + 9} \geq 0$$

$$\Rightarrow -3 - 4(b-1)p \geq \sqrt{16b^2p^2 - 8(b+1)p + 9}$$

$$\Rightarrow -3 + 4(1-b)p \geq \sqrt{16b^2p^2 - 8(b+1)p + 9}.$$

Squaring both sides yields:

$$9 + 16(1-b)^2p^2 - 24(1-b)p \geq 16b^2p^2 - 8(b+1)p + 9$$

$$\Rightarrow 16(1-2b+b^2)p^2 - 24(1-b)p \geq 16b^2p^2 - 8(b+1)p$$

$$\Rightarrow 16p^2 - 32bp^2 + 16b^2p^2 - 24(1-b)p \geq 16b^2p^2 - 8(b+1)p$$

$$\Rightarrow 16p^2 - 32bp^2 - 24(1-b)p \geq -8(b+1)p$$

$$\Rightarrow 16p^2 - 32bp^2 \geq 24(1-b)p - 8(b+1)p$$

$$\Rightarrow 16p^2(1-2b) \geq 16p(1-2b)$$

which is a contradiction given $p \leq 1$. □

Appendix B

Suppose there are three types of Senders: an informed Sender, an uninformed Sender who prefers higher actions (US^H), and an uninformed Sender who prefers lower actions (US^L). The Receiver believes that the Sender is US^H with probability p_1, US^L with probability p_2, and informed with probability $(1-p_1-p_2)$.

The informed Sender sets threshold $t \in [0,1]$ and sends message B when $\theta \in [0,t)$, and T when $\theta \in [t,1]$. Suppose US^H sends message T with probability α and B with probability $(1-\alpha)$ for all $\theta \in \Theta$, and US^L sends message T with probability β and B with probability $(1-\beta)$ for all $\theta \in \Theta$.

Lemma 6.5. *Let a_T denote the Receiver's best response upon seeing message T. Then,*

$$a_T = 0.5\left(\frac{p_1(\alpha+t^2-1)+p_2(\beta+t^2-1)-t^2+1}{p_1(\alpha+t-1)+p_2(\beta+t-1)-t+1}\right).$$

Proof. The probability that the Receiver sees message T is $\alpha p_1+\beta p_2+(1-t)(1-p_1-p_2)$. Given that the message is T, the probability that the Sender is uninformed and prefers higher actions is

$$\frac{\alpha p_1}{\alpha p_1+\beta p_2+(1-t)(1-p_1-p_2)}.$$

Similarly, the probability that the Sender is uninformed and prefers lower actions is

$$\frac{\beta p_2}{\alpha p_1+\beta p_2+(1-t)(1-p_1-p_2)}.$$

Finally, given message T, the probability that the Sender is informed is

$$\frac{(1-t)(1-p_1-p_2)}{\alpha p_1+\beta p_2+(1-t)(1-p_1-p_2)}.$$

The Receiver's expected payoff from choosing action a is:

$$E[u_R|T] = E[u_R|US^H,T]P(US^H|T)+E[u_R|US^L,T]P(US^L|T)+E[u_R|IS,T]P(IS|T)$$

$$= \frac{\alpha p_1}{\alpha p_1+\beta p_2+(1-t)(1-p_1-p_2)}\int_0^1 -(a-\theta)^2 d\theta$$

$$+\frac{\beta p_2}{\alpha p_1+\beta p_2+(1-t)(1-p_1-p_2)}\int_0^1 -(a-\theta)^2 d\theta$$

$$+\frac{(1-t)(1-p_1-p_2)}{\alpha p_1+\beta p_2+(1-t)(1-p_1-p_2)}\int_t^1 \frac{-(a-\theta)^2}{1-t}\,d\theta$$

$$=-\frac{[\alpha p_1+\beta p_2]\int_0^1 (a-\theta)^2 d\theta}{\alpha p_1+\beta p_2+(1-t)(1-p_1-p_2)}-\frac{[1-p_1-p_2]\int_t^1 (a-\theta)^2\,d\theta}{\alpha p_1+\beta p_2+(1-t)(1-p_1-p_2)}. \quad (7)$$

To determine the Receiver's optimal choice of a we set the derivative of (7) with respect to a equal to zero:

$$-\frac{[\alpha p_1+\beta p_2]\int_0^1 2(a-\theta)d\theta}{\alpha p_1+\beta p_2+(1-t)(1-p_1-p_2)}-\frac{[1-p_1-p_2]\int_t^1 2(a-\theta)\,d\theta}{\alpha p_1+\beta p_2+(1-t)(1-p_1-p_2)}=0$$

$$\frac{(\alpha p_1+\beta p_2)\Big[a-0.5\Big]}{\alpha p_1+\beta p_2+(1-t)(1-p_1-p_2)}+\frac{(1-p_1-p_2)\Big[(a-0.5)-(at-0.5t^2)\Big]}{(\alpha p_1+\beta p_2+(1-t)(1-p_1-p_2))(1-t)}=0$$

$$\Rightarrow a_T=0.5\left(\frac{p_1(\alpha+t^2-1)+p_2(\beta+t^2-1)-t^2+1}{p_1(\alpha+t-1)+p_2(\beta+t-1)-t+1}\right). \qquad \square$$

Lemma 6.6. *Let a_T denote the Receiver's best response upon seeing message T. Then,*

$$a_B=0.5\left(\frac{p_1(\alpha+t^2-1)+p_2(\beta+t^2-1)-t^2}{p_1(\alpha+t-1)+p_2(\beta+t-1)-t}\right).$$

Proof. The probability the Receiver sees message B is $(1-\alpha)p_1+(1-\beta)p_2+t(1-p_1-p_2)$. Given that the message is B, the probability that the Sender is uninformed and prefers higher actions is

$$\frac{(1-\alpha)p_1}{(1-\alpha)p_1+(1-\beta)p_2+t(1-p_1-p_2)}.$$

Given that the message is B, the probability that the Sender is uninformed and prefers lower actions is

$$\frac{(1-\beta)p_2}{(1-\alpha)p_1+(1-\beta)p_2+t(1-p_1-p_2)}.$$

Finally, if the message is B, the probability that the Sender is informed is

$$\frac{t(1-p_1-p_2)}{(1-\alpha)p_1+(1-\beta)p_2+t(1-p_1-p_2)}.$$

The Receiver's expected payoff from choosing action a is $E[u_R|B] =$

$$E[u_R|US^H,B]P(US^H|B) + E[u_R|US^L,B]P(US^L|B) + E[u_R|IS,B]P(IS|B) = \tag{8}$$

$$-\frac{(1-\alpha)p_1\int_0^1(a-x)^2dx}{(1-\alpha)p_1+(1-\beta)p_2+t(1-p_1-p_2)} - \frac{(1-\beta)p_2\int_0^1(a-x)^2dx}{(1-\alpha)p_1+(1-\beta)p_2+t(1-p_1-p_2)}$$

$$-\frac{t(1-p_1-p_2)}{(1-\alpha)p_1+(1-\beta)p_2+t(1-p_1-p_2)}\int_0^t\frac{(a-\theta)^2}{t}\,d\theta =$$

$$-\frac{[(1-\alpha)p_1+(1-\beta)p_2]\int_0^1(a-\theta)^2d\theta}{(1-\alpha)p_1+(1-\beta)p_2+t(1-p_1-p_2)} - \frac{[1-p_1-p_2]\int_0^t(a-\theta)^2\,d\theta}{(1-\alpha)p_1+(1-\beta)p_2+t(1-p_1-p_2)}.$$

To determine the Receiver's optimal choice of a, we differentiate (8) with respect to a and set to zero:

$$-\frac{[(1-\alpha)p_1+(1-\beta)p_2]\int_0^1(2a-2\theta)d\theta}{(1-\alpha)p_1+(1-\beta)p_2+t(1-p_1-p_2)} - \frac{[1-p_1-p_2]\int_0^t(2a-2\theta)\,d\theta}{(1-\alpha)p_1+(1-\beta)p_2+t(1-p_1-p_2)} = 0$$

$$-2\left(\frac{[(1-\alpha)p_1+(1-\beta)p_2]\int_0^1(a-\theta)d\theta}{(1-\alpha)p_1+(1-\beta)p_2+t(1-p_1-p_2)} + \frac{[1-p_1-p_2]\int_0^t(a-\theta)\,d\theta}{(1-\alpha)p_1+(1-\beta)p_2+t(1-p_1-p_2)}\right) = 0$$

$$\frac{[(1-\alpha)p_1+(1-\beta)p_2]\int_0^1(a-\theta)d\theta}{(1-\alpha)p_1+(1-\beta)p_2+t(1-p_1-p_2)} + \frac{[1-p_1-p_2]\int_0^t(a-\theta)\,d\theta}{((1-\alpha)p_1+(1-\beta)p_2+t(1-p_1-p_2)} = 0$$

$$\frac{[(1-\alpha)p_1+(1-\beta)p_2]\left[a\theta-0.5\theta^2\right]_0^1}{(1-\alpha)p_1+(1-\beta)p_2+t(1-p_1-p_2)} + \frac{[1-p_1-p_2]\left[a\theta-0.5\theta^2\right]_0^t}{((1-\alpha)p_1+(1-\beta)p_2+t(1-p_1-p_2)} = 0$$

$$\frac{[(1-\alpha)p_1+(1-\beta)p_2]\left[a-0.5\right]}{(1-\alpha)p_1+(1-\beta)p_2+t(1-p_1-p_2)} + \frac{[1-p_1-p_2]\left[at-0.5t^2\right]}{((1-\alpha)p_1+(1-\beta)p_2+t(1-p_1-p_2)} = 0$$

$$\frac{[(1-\alpha)p_1+(1-\beta)p_2]\Big[a-0.5\Big]}{(1-\alpha)p_1+(1-\beta)p_2+t(1-p_1-p_2)}+\frac{[1-p_1-p_2]\Big[a-0.5t\Big]}{(1-\alpha)p_1+(1-\beta)p_2+t(1-p_1-p_2)}=0$$

$$\Rightarrow a_B=0.5\left(\frac{p_1(\alpha+t^2-1)+p_2(\beta+t^2-1)-t^2}{p_1(\alpha+t-1)+p_2(\beta+t-1)-t}\right). \qquad \square$$

The uninformed Sender who prefers higher actions must solve the following problem:

$$\max_\alpha E[a]=\max_\alpha[a_TP(T)+a_BP(B)]=\max_\alpha[a_T\alpha+a_B(1-\alpha)]$$

$$=\max_{\alpha\in[0,1]}\left(0.5\left(\frac{p_1(\alpha+t^2-1)+p_2(\beta+t^2-1)-t^2+1}{p_1(\alpha+t-1)+p_2(\beta+t-1)-t+1}\right)\times\alpha\right.$$
$$\left.+0.5\left(\frac{p_1(\alpha+t^2-1)+p_2(\beta+t^2-1)-t^2}{p_1(\alpha+t-1)+p_2(\beta+t-1)-t}\right)\times(1-\alpha)\right)$$

Lemma 6.7. *The uninformed Sender who prefers higher actions will choose* $\alpha=1$.

Proof. The first-order condition for an interior optimal is:

$$\begin{aligned}\alpha^* &= \frac{1}{(p_1-1){p_1}^2}\Big(-\big({p_1}^2(\beta^2{p_2}^2+2\beta p_1p_2t-\beta p_1p_2+2\beta p_2{}^2t\\
&-2\beta p_2{}^2-2\beta p_2t+\beta p_2+{p_1}^2t^2-{p_1}^2t+2p_1p_2t^2-3p_1p_2t+p_1p_2\\
&-2p_1t^2+2p_1t+{p_2}^2t^2-2{p_2}^2t+{p_2}^2-2p_2t^2+3p_2t-p_2+t^2-t)\big)^{\frac{1}{2}}\\
&-bp^2q-tp^3+p^3-p^2qt+p^2q+p^2t-p^2\Big).\end{aligned}$$

Given $t\in(0,1)$, $p_1\in(0,1)$, $p_2\in(0,1)$, and $\beta\in(0,1)$, this expression has no solution in $(0,1)$. Hence, at optimality $\alpha=0$ or $\alpha=1$.

Setting $\alpha=0$ yields the following expected payoff for US^H:

$$0.5\left(\frac{p_1(t^2-1)+p_2(\beta+t^2-1)-t^2}{p_1(t-1)+p_2(\beta+t-1)-t}\right).$$

Setting $\alpha=1$ yields the following expected payoff for US^H:

$$0.5\left(\frac{p_1t^2+p_2(\beta+t^2-1)-t^2+1}{p_1t+p_2(\beta+t-1)-t+1}\right).$$

With $t\in(0,1)$ and $p_1\in(0,1)$, $p_2\in(0,1)$, and $\beta\in(0,1)$, $\alpha=1$ yields a greater payoff for US^H. Thus, the uninformed Sender who prefers higher actions will always send message T. $\square$

By symmetry, the uninformed Sender who prefers lower actions will choose $\alpha = 0$, i.e., will always send message B. The informed Sender will choose threshold t so that

$$((t+b)-a_B)^2 = (a_T-(t+b))^2.$$

Lemma 5.2

Proof. From equation (1), we have:

$$0.5\left(\frac{p_u t^2+p_u(t^2-1)-t^2+1}{p_u t+p_u(t-1)-t+1}\right)+0.5\left(\frac{p_u t^2+p_u(t^2-1)-t^2}{p_u t+p_u(t-1)-t}\right)-2t=2b. \quad (9)$$

Denote the left hand side of equation (9) by $f(t)$. In equilibrium, $f(t^*)=2b$ and the derivative of $f(t)$ with respect to t is:

$$\frac{\partial f}{\partial t}=\frac{0.125p_u^2-0.125p_u}{(p_u t-0.5p_u-0.5t)^2}+\frac{0.125(p_u^3-1.5p_u^2+0.5p_u)}{(p_u-0.5)(p_u t-0.5p_u-0.5t+0.5)^2}-1$$

$$=\frac{p_u(p_u-1)}{8((p_u-0.5)t-0.5p_u)^2}+\frac{p_u(2p_u-1)(p_u-1)}{16(p_u-0.5)((p_u-0.5)t+0.5(1-p_u))^2}-1. \quad (10)$$

In the right hand side of equation (10) let

$$A=\frac{p_u(p_u-1)}{8((p_u-0.5)t-0.5p_u)^2}$$

and

$$C=\frac{p_u(2p_u-1)(p_u-1)}{16(p_u-0.5)((p_u-0.5)t+0.5(1-p_u))^2}.$$

Denote the respective numerators by N_A and N_C and the respective denominators by D_A and D_C. Recall that we must have $p_u \in (0,0.5)$ and $t \in (0,1)$.

Now, $N_A = p_u(p_u-1)$ is negative and $D_A = 8((p_u-0.5)t-0.5p_u)^2$ is positive. Hence, $A<0$. Also, $N_C = p_u(2p_u-1)(p_u-1)$ is positive as it is the product of two negative numbers. In addition, $D_C = 16(p_u-0.5)((p_u-0.5)t+0.5(1-p_u))^2$ is negative as it is the product of a positive and a negative number. Hence, $D_C<0$. Therefore, $C<0$.

As $\frac{\partial f}{\partial t}=A+C-1<0$ it follows that f is decreasing in t over the relevant domain. Denote the standard cheap talk threshold by $\bar{t}=0.5-2b$. Assume, for a contradiction, that $f(\bar{t})<f(t^*)$. This yields the following inequality:

$$f(\bar{t})=\frac{b(2b^2(p_u-0.5)^2+\frac{1}{8}p_u(p_u-1)-\frac{1}{32})}{b^2(p_u-0.5)^2-\frac{1}{64}}<2b=f(t^*). \quad (11)$$

We now show that the denominator on the left hand side of inequality (11) is negative. If not,

$$b^2(p_u-0.5)^2-\frac{1}{64}>0$$
$$\Rightarrow b^2(p_u-0.5)^2>\frac{1}{64}$$
$$\Rightarrow b(p_u-0.5)>\frac{1}{8}$$
$$\Rightarrow p_u>\frac{1}{8b}+\frac{1}{2}.$$

This cannot be as $p_u \in (0,0.5)$ and $b>0$.

Hence, if we multiply both sides of (11) by $b^2(p_u-0.5)^2-\frac{1}{64}$, the sign of the inequality reverses. Therefore, when $f(\bar{t})<f(t^*)$, the following holds:

$$b(2b^2(p_u-0.5)^2+\frac{1}{8}p_u(p_u-1)-\frac{1}{32})>2b(b^2(p_u-0.5)^2-\frac{1}{64})$$
$$\Rightarrow 2b^3(p_u-0.5)^2+\frac{1}{8}bp_u(p_u-1)-\frac{1}{32}b>2b^3(p_u-0.5)^2-\frac{1}{32}b$$
$$\Rightarrow \frac{1}{8}bp_u(p_u-1)>0.$$

Provided that $b>0$ this is equivalent to $p_u(p_u-1)>0$. Therefore, one of the following must be true:

1. Both p_u and (p_u-1) are negative.
2. Both p_u and (p_u-1) are positive.

Now p_u cannot be negative, nor can it be that (p_u-1) is positive, as $p \in (0,\frac{1}{2})$. Therefore, $f(\bar{t})<f(t^*)$ which is a contradiction. This means that $f(\bar{t})\geq f(t^*)$. Since $f(\cdot)$ is a decreasing function of t, $f(\bar{t})\geq f(t^*)$ implies that $\bar{t}\leq t^*$. Thus, the presence of the uninformed Sender raises the informed Sender's equilibrium threshold when compared to the standard cheap talk model. This implies that, ex ante, the informed Sender who faces two uninformed Senders with opposing preferences, sends message B with greater probability than in the cheap talk case.

Now, we evaluate $f(0.5)$ and show that it is less than $f(t^*)$. This will imply, $t^*<0.5$ and therefore, by Lemma 3.2, the informed Sender is better off. Now,

$$f(0.5)=0.5\left(\frac{0.25p_u-0.75p_u+0.75}{0.5p_u-0.5p_u+0.5}\right)+0.5\left(\frac{0.25p_u-0.75p_u-0.25}{0.5p_u+0.5p_u-0.5}\right)-1$$
$$=0.5\frac{-0.5p_u+0.75}{0.5}+0.5\frac{-0.5p_u-0.25}{-0.5}-1=0<2b=f(t^*). \quad \square$$

Appendix C

We consider the possibility of an equilibrium with multiple messages. Suppose there exists an equilibrium with n messages denoted by M_k for $k = 1,2,...,n$, with M_1 representing the bottom message and M_n representing the top message. The messages partition $[0,1]$ into n intervals, separated by $n-1$ thresholds. Denote by t_k the threshold that separates the intervals M_k and M_{k+1}. Thus, $t_1 \leq t_2 \leq \ldots \leq t_{n-1}$. Suppose the uninformed Sender with a preference for higher actions sends message M_k with probability q_k. Our goal is to identify the choice of q_k's that maximize the uninformed Sender's expected payoff.

Upon observing message M_1, the Receiver chooses action a to maximize her expected payoff:

$$E[u_R|M_1] = E[u_R|IS,M_1]P(IS|M_1) + E[u_R|US,M_1]P(US|M_1)$$

$$= \frac{pt_1}{pt_1 + (1-p)q_1} \int_0^{t_1} -\frac{(a-\theta)^2}{t_1} d\theta + \frac{(1-p)q_1}{pt_1 + (1-p)q_1} \int_0^1 -(a-\theta)^2 d\theta$$

$$\Rightarrow a_{M_1} = 0.5\left(\frac{p({t_1}^2 - q_1) + q_1}{p(t_1 - q_1) + q_1}\right).$$

More generally, upon seeing message M_i (for arbitrary integer $i = 2,3,...,n-1$), the Receiver chooses action a to maximize the following:

$$E[u_R|M_i] = E[u_R|IS,M_t]P(IS|M_t) + E[u_R|US,M_i]P(US|M_i)$$

$$= \frac{p(t_i - t_{i-1})}{p(t_i - t_{i-1}) + (1-p)q_i} \int_{t_{i-1}}^{t_i} -\frac{(a-\theta)^2}{t_i - t_{i-1}} d\theta + \frac{(1-p)q_i}{p(t_i - t_{i-1}) + (1-p)q_i} \int_0^1 -(a-\theta)^2 d\theta$$

$$\Rightarrow a_{M_i} = 0.5\left(\frac{p({t_i}^2 - t_{i-1}^2 - q_i) + q_i}{p(t_i - t_{i-1} - q_i) + q_i}\right).$$

Upon seeing message M_n, the Receiver chooses action a to maximize the following:

$$E[u_R|M_n] = E[u_R|IS,M_n]P(IS|M_n) + E[u_R|US,M_n]P(US|M_n)$$

$$= \frac{p(1 - t_{n-1})}{p(1 - t_{n-1}) + (1-p)q_n} \int_{t_{n-1}}^1 -\frac{(a-\theta)^2}{1 - t_{n-1}} d\theta + \frac{(1-p)q_n}{p(1 - t_{n-1}) + (1-p)q_n} \int_0^1 -(a-\theta)^2 d\theta$$

$$\Rightarrow a_{M_n} = 0.5\left(\frac{p(1 - t_{n-1}^2 - q_n) + q_n}{p(1 - t_{n-1} - q_n) + q_n}\right).$$

Thus, the expected action of the Receiver is given by:

$$E[a|s=US]=\sum_{t=1}^{n} q_t a_{M_t} = q_1 a_{M_1} + \sum_{t=2}^{n-1}[q_t a_{M_t}] + q_n a_{M_n}$$

$$= q_1\left(0.5\left(\frac{p({t_1}^2-q_1)+q_1}{p(t_1-q_1)+q_1}\right)\right) + \sum_{i=2}^{n-1}\left[q_i\left(0.5\left(\frac{p(t_i^2-t_{i-1}^2-q_i)+q_i}{p(t_i-t_{i-1}-q_i)+q_i}\right)\right)\right]$$

$$+q_n\left(0.5\left(\frac{p(1-t_{n-1}^2-q_n)+q_n}{p(1-t_{n-1}-q_n)+q_n}\right)\right).$$

Let

$$f_1(q_1)=q_1\left(\frac{p{t_1}^2+(1-p)q_1}{2(pt_1+(1-p)q_1)}\right),\ f_i(q_i)=q_i\left(\frac{p(t_i^2-t_{i-1}^2)+(1-p)q_i}{2(p(t_i-t_{i-1})+(1-p)q_1)}\right)$$

for $i=2,\ldots,n-2$ and

$$f_n(q_n)=f_n=q_n\left(\frac{p(1-{t_{n-1}}^2)+(1-p)q_n}{2(p(1-t_{n-1})+(1-p)q_n)}\right).$$

Then, $E[a|s=US]=\sum_{i=1}^{n-1} f_i(q_i)$. The uninformed Sender's problem is the following:

$$\max \sum_{i=1}^{n-1} f_i(q_i)$$

$$\text{s.t. } \sum_{i=1}^{n-1} q_i = 1;\ q_i \geq 0\ \forall i=1,\ldots,n-1$$

Let $(q_1^*,\ldots,q_{n-1}^*)$ be an optimal solution to this program. A direct computation shows the following:

- $t_i+t_{i-1}<1 \Rightarrow f_i(q_i)$ is convex in q_i;
- $t_i+t_{i-1}=1 \Rightarrow f_i(q_i)$ is linear in q_i;
- $t_i+t_{i-1}>1 \Rightarrow f_i(q_i)$ is concave in q_i.

Let $r-1$ be the largest index such that $t_{r-1}\leq 0.5$ then $f_1,\ldots,f_{r-1}$ are convex. Consider the following sub-problem:

$$\max \sum_{i=1}^{r-1} f_i(q_i)$$

$$\text{s.t.} \sum_{i=1}^{r-1} q_i = \sum_{i=1}^{r-1} q_i^*;\ q_i \geq 0\ \forall i = 1,\ldots,r-1.$$

As the function being optimized is convex, the maximum of the subproblem must occur at one of the extreme points. A straightforward computation shows that the optimal solution will be $q_{r-1} = \sum_{i=1}^{r-1} q_i^*$. In other words, $q_1 = 0$, i.e., the uninformed Sender never sends the lowest message.

Three Messages

Now, consider the possibility of an equilibrium with three messages, the informed Sender uses thresholds $t_1 \in [0,1]$ and $t_2 \in [0,1]$ where $t_2 > t_1$. The informed Sender sends message B when $\theta \in [0,t_1]$, M when $\theta \in [t_1,t_2]$, and T when $\theta \in [t_2,1]$. The uninformed Sender sends message T with probability α and M with probability $(1-\alpha)$. By the argument above the uninformed Sender *never* sends message B.

The uninformed Sender chooses α to maximize the expected action of the Receiver, since the uninformed Sender strictly prefers higher actions. Thus, the uninformed Sender will choose $\alpha \in [0,1]$ to maximize

$$uR := 0.5\left(\frac{p(\alpha+t_2^2-1)-\alpha}{p(\alpha+t_2-1)-\alpha}\right)\times\alpha + 0.5\left(\frac{p(\alpha+t_2^2-t_1^2-1)-\alpha+1}{p(\alpha+t_2-t_1-1)-\alpha+1}\right)\times(1-\alpha).$$

The first-order condition for optimality is:

$$\begin{aligned}\partial uR/\partial\alpha = 0 \Rightarrow \alpha^* = {} & 2p^2{t_1}^3t_2 - 2p^2{t_1}^3 - 2p^2{t_1}^2{t_2}^2 + 2p^2{t_1}^2 \\ & - \sqrt{-2p^2{t_1}^3t_2 + 2p^2{t_1}^3 + 2p^2{t_1}^2{t_2}^2}.\end{aligned} \tag{12}$$

Note that for some values of p, t_1 and t_2, it may be optimal to choose α strictly between zero and one.[3]

The informed Sender's indifference condition between messages T and M is:

$$0.5(\frac{p(\alpha+t_2^2-1)-\alpha}{p(\alpha+t_2-1)-\alpha}) - t_2 - b = t_2 + b - 0.5(\frac{p(\alpha+t_2^2-t_1^2-1)-\alpha+1}{p(\alpha+t_2-t_1-1)-\alpha+1}).$$

The indifference condition between messages M and B is:

$$0.5(\frac{p(\alpha+t_2^2-t_1^2-1)-\alpha+1}{p(\alpha+t_2-t_1-1)-\alpha+1}) - t_1 - b = t_1 + b - 0.5t_1$$

When $\alpha = 1$,

$$t_2 = \frac{5-36bp-\sqrt{144b^2p^2+24(b-1)p+25}}{12p}$$

[3] See the graph: https://www.desmos.com/calculator/rdv7j9qiwa.

and

$$t_1 = \frac{5 - 12bp - \sqrt{144b^2p^2 + 24(b-1)p + 25}}{6p}.$$

If we substitute this into equation (12) we see that there is no interior solution. Hence, the uninformed Sender's best response to the choice of thresholds above is $\alpha = 1$.

We now focus on the case $\alpha = 1$ to determine the upper limit on b to support a three message equilibrium. The Receiver's best response to each message is:

$$a_T = 0.5\left(\frac{pt_2{}^2 - 1}{pt_2 - 1}\right),\ a_M = 0.5(t_1 + t_2),\ a_B = 0.5t_1.$$

The informed Sender's thresholds (ignoring the babbling case) are determined by the following indifference conditions:

$$0.5\left(\frac{pt_2{}^2 - 1}{pt_2 - 1}\right) - (t_2 + b) = t_2 + b - 0.5(t_2 + t_1). \tag{13}$$

From equation (13) the equilibrium value of t_2 as a function of t_1, denoted $t_2^*(t_1)$, must satisfy:

$$t_2^*(t_1) = \frac{-\sqrt{p^2(t_1 - 4b)^2 + 8bp - 2p(t_1 + 4) + 9} - 4bp + pt_1 + 3}{4p}.$$

The second indifference condition is

$$0.5(t_2 + t_1) - (t_1 + b) = t_1 + b - 0.5t_1. \tag{14}$$

From equation (14) the equilibrium value of t_1 as a function of t_2, denoted $t_1^*(t_2)$, must satisfy:

$$t_1^*(t_2) = 0.5t_2 - 2b.$$

Solving these equations we find that the equilibrium thresholds satisfy the following:

$$t_1^*(p,b) = \frac{36bp - 72b - 7p + 12 - sq}{6(p-3)(2p-3)},$$
$$t_2^*(p,b) = \frac{12bp + 7p - 12 + sq}{6p(2p-3)}$$

for $sq = \sqrt{144b^2p^2 - 120bp^2 + 144bp - 24p^3 + 157p^2 - 276p + 144}$. Next, we determine the upper limit of b in order to support such an equilibrium.

In the absence of an uninformed Sender, a three message equilibrium exists provided the following condition on the informed Sender's bias is satisfied (Crawford & Sobel 1982):

$$2 \leq \frac{1}{2}\left(\sqrt{1 + \frac{2}{b}} - 1\right) \Rightarrow b \leq \frac{1}{12}.$$

When the uninformed Sender is present with probability $1-p$, a two message equilibrium is supported when $t_1 \geq 0$:

$$\frac{36bp-72b-7p+12-sq}{6(p-3)(2p-3)} \geq 0.$$

As $p \in (0,1)$ the denominator in the left hand side of the above is positive. Hence,

$$36bp-72b-7p+12-sq \geq 0$$

$$\Rightarrow 36bp-72b-7p+12 \geq sq$$

$$\Rightarrow (36bp-72b-7p+12)^2 \geq 144b^2p^2-120bp^2+144bp-24p^3+157p^2-276p+144$$

$$\Rightarrow 1152b^2p^2-384bp^2+1728bp-108p^2+108p-5184b^2p+5184b^2+1728b+24p^3 \geq 0$$

$$\Rightarrow 12(p-3)(2p-3)(p+48b^2-16b) \geq 0$$

$$\Rightarrow p+48b^2-16b \geq 0.$$

The roots of the quadratic on the left are $b = (2 \pm \sqrt{4-3p})/12$ and they are real for $p \in [0,1]$. Therefore, $b \leq (2-\sqrt{4-3p})/12$ or $b \geq (2+\sqrt{4-3p})/12$. Hence, the value of b under which a three message equilibrium exists under cheap talk *and* in the presence of an uninformed Sender must satisfy $b \leq (2-\sqrt{4-3p})/12$. For such a b, the possible values of p consistent with a three message equilibrium in the presence of an uninformed Sender must be at least as large as $16b(1-3b)$.

Appendix D

Suppose there exist two types of informed Senders: one with low bias b_l denoted (IS^L) and another with high bias b_h denoted (IS^H). Both Senders send a state-dependent message. We compare the value of the information sent to the Receiver in this case with that in which the Receiver faces either a low-biased informed Sender or an uninformed Sender with preferences for higher actions only.

IS^L sets threshold $t_l \in [0,1]$ to determine which message is sent. IS^L sends message B when $\theta \in [0,t_l)$, and T when $\theta \in [t_l,1]$. Let $t_H \in [0,1]$ be the threshold for IS^H to determine her binary message. Thus, IS^H sends message B when $\theta \in [0,t_h)$, and T when $\theta \in [t_h,1]$. Assume that the informed Sender is IS^L with probability p and IS^H with probability $1-p$.

The Receiver observes the message, but not the type of the informed Sender. The Receiver knows p, b_l, and b_h and uses these to inform her beliefs on the Sender's type and, thus, θ. Then, the Receiver chooses an action to maximize her expected payoff.

Receiver's Action in Response to T

The probability that the Receiver sees message T is:

$$P(T) = P(T|IS^L)P(IS^L) + P(T|IS^H)P(IS^H) = p(1-t_l) + (1-p)(1-t_h).$$

Given that the message is T, the probability that the sender is IS^L is given by:

$$P(IS^L|T) = \frac{P(IS^L)P(T|IS^L)}{P(T)} = \frac{p(1-t_l)}{p(1-t_l)+(1-p)(1-t_h)}.$$

Given that the message is T, the probability that the sender is IS^H is given by:

$$P(IS^H|T) = \frac{P(IS^H)P(T|IS^H)}{P(T)} = \frac{(1-p)(1-t_h)}{p(1-t_l)+(1-p)(1-t_h)}.$$

R's expected payoff from choosing action $a \in [0,1]$ is given by:

$$E[u_R|T] = E[u_R|IS^L,T]P(IS^L|T) + E[u_R|IS^H,T]P(IS^H|T)$$

$$= \frac{p(1-t_l)}{p(1-t_l)+(1-p)(1-t_h)} \int_{t_l}^{1} -\frac{(a-\theta)^2}{1-t_l} d\theta + \frac{(1-p)(1-t_h)}{p(1-t_l)+(1-p)(1-t_h)}$$

$$\int_{t_h}^{1} -\frac{(a-\theta)^2}{1-t_h} d\theta = -\frac{p\int_{t_l}^{1}(a^2-2a\theta+\theta^2)\,d\theta}{p(1-t_l)+(1-p)(1-t_h)} - \frac{(1-p)\int_{t_h}^{1}(a^2-2a\theta+\theta^2)\,d\theta}{p(1-t_l)+(1-p)(1-t_h)}.$$

To determine R's optimal choice of action, we invoke the first-order condition:

$$\partial \left[-\frac{p\int_{t_l}^{1}(a^2-2a\theta+\theta^2)\,d\theta}{p(1-t_l)+(1-p)(1-t_h)} - \frac{(1-p)\int_{t_h}^{1}(a^2-2a\theta+\theta^2)\,d\theta}{p(1-t_l)+(1-p)(1-t_h)} \right] \Big/ \partial a = 0$$

$$-2\left(\frac{p\int_{t_l}^{1}(a-\theta)\,d\theta}{p(1-t_l)+(1-p)(1-t_h)} + \frac{(1-p)\int_{t_h}^{1}(a-\theta)\,d\theta}{p(1-t_l)+(1-p)(1-t_h)} \right) = 0$$

$$\frac{p\left[a\theta - 0.5\theta^2\right]_{t_l}^{1}}{p(1-t_l)+(1-p)(1-t_h)} + \frac{(1-p)\left[a\theta - 0.5\theta^2\right]_{t_h}^{1}}{p(1-t_l)+(1-p)(1-t_h)} = 0$$

$$\frac{p(a-0.5-at_l+0.5{t_l}^2)+(1-p)(a-0.5-at_h+0.5{t_h}^2)}{p(1-t_l)+(1-p)(1-t_h)} = 0$$

$$a_T = 0.5\left(\frac{p({t_l}^2-{t_h}^2)+{t_h}^2-1}{p(t_l-t_h)+t_h-1}\right).$$

Receiver's Action in Response to B

The probability that the Receiver sees message B is given by:

$$P(B) = P(B|IS^L)P(IS^L) + P(B|IS^H)P(IS^H) = pt_l + (1-p)t_h.$$

Given that the message is B, the probability that the informed Sender has low bias is given by:

$$P(IS^L|B) = \frac{P(IS^L)P(B|IS^L)}{P(B)} = \frac{pt_l}{pt_l+(1-p)t_h},$$

Given that the message is B, the probability that informed Sender has high bias is given by:

$$P(IS^H|B) = \frac{P(IS^H)P(B|IS^H)}{P(B)} = \frac{(1-p)t_h}{pt_l+(1-p)t_h}.$$

The Receiver's expected payoff from choosing action a is given by:

$$E[u_R|B] = E[u_R|IS^L,B]P(IS^L|B) + E[u_R|IS^H,B]P(IS^H|B)$$

$$= \frac{pt_l}{pt_l+(1-p)t_h}\int_0^{t_l} -\frac{(a-\theta)^2}{t_l}d\theta + \frac{(1-p)t_h}{pt_l+(1-p)t_h}\int_0^{t_h} -\frac{(a-\theta)^2}{t_h}d\theta$$

$$= -\frac{p\int_0^{t_l}(a^2-2a\theta+\theta^2)\,d\theta}{pt_l+(1-p)t_h} - \frac{(1-p)\int_0^{t_h}(a^2-2a\theta+\theta^2)\,d\theta}{pt_l+(1-p)t_h} =: uR.$$

To determine the Receiver's optimal choice of a, we invoke the first-order condition for optimality:

$$\partial uR/\partial a = 0$$

$$-2\left(\frac{p\int_0^{t_l}(a-\theta)\,d\theta}{pt_l+(1-p)t_h}+\frac{(1-p)\int_0^{t_h}(a-\theta)\,d\theta}{pt_l+(1-p)t_h}\right)=0$$

$$\frac{p}{pt_l+(1-p)t_h}\left[a\theta-0.5\theta^2\right]_0^{t_l}+\frac{(1-p)}{pt_l+(1-p)t_h}\left[a\theta-0.5\theta^2\right]_0^{t_h}=0$$

$$\frac{p(at_l-0.5{t_l}^2)+(1-p)(at_h-0.5{t_h}^2)}{pt_l+(1-p)t_h}=0$$

$$a_B=0.5\left(\frac{p({t_l}^2-{t_h}^2)+{t_h}^2}{p(t_l-t_h)+t_h}\right). \tag{15}$$

Senders' Strategies

The informed Senders individually set their thresholds such that they are indifferent between sending T and B. In equilibrium, this means that both of the following equations are satisfied by (t_l^*, t_h^*):

$$\begin{aligned}
&-\left(0.5\left(\frac{p({t_l}^2-{t_h}^2)+{t_h}^2-1}{p(t_l-t_h)+t_h-1}\right)-(t_l+b_l)\right)^2\\
&=-\left((t_l+b_l)-0.5\left(\frac{p({t_l}^2-{t_h}^2)+{t_h}^2}{p(t_l-t_h)+t_h}\right)\right)^2\\
&\quad-\left(0.5\left(\frac{p({t_l}^2-{t_h}^2)+{t_h}^2-1}{p(t_l-t_h)+t_h-1}\right)-(t_h+b_h)\right)^2\\
&=-\left((t_h+b_h)-0.5\left(\frac{p({t_l}^2-{t_h}^2)+{t_h}^2}{p(t_l-t_h)+t_h}\right)\right)^2.
\end{aligned} \tag{16}$$

Journal of Mechanism and Institution Design
ISSN: 2399-844X (Print), 2399-8458 (Online)
DOI:10.22574/jmid.2024.12.004

THE EXISTENCE OF EQUILIBRIUM FLOWS

Alfred Galichon

New York University, USA and Sciences Po, France
ag133@nyu.edu

Larry Samuelson

Yale University, USA
larry.samuelson@yale.edu

Lucas Vernet

Sciences Po and Banque de France, France
lucas.vernet@banque-france.fr

ABSTRACT

In this paper, we establish conditions for the existence of an equilibrium in the equilibrium flow problem studied by Galichon et al. (2024). The problem nests several classical economic models such as bipartite matching models, hedonic pricing models, shortest-path and minimum-cost flow problems, and time-dependent routing problems.

Keywords: Equilibrium flow, unified gross substitutes, matching.

JEL Classification Numbers: C6, D85.

1. INTRODUCTION

Galichon, Samuelson and Vernet introduced a class of problems, *equilibrium flow problems*, that nests several classical economic models such as bipartite matching models, hedonic pricing models, shortest-path and minimum-cost flow problems, and time-dependent routing problems (Galichon et al. 2024). This paper first provides a more detailed presentation and

We thank the editor and a referee for helpful comments. Galichon gratefully acknowledges funding from ERC grant CoG-866274

motivation for the equilibrium flow problem. Our main result then provides conditions sufficient to ensure the existence of an equilibrium flow.

The framework for the equilibrium flow problem is a network, consisting of a set of nodes and a collection of directed arcs. A subset of the nodes are source nodes, each characterized by a negative quantity that can be interpreted as a quantity that must be conveyed away from the source node. Another subset of nodes are destination nodes, each characterized by a positive quantity that can be interpreted as a quantity that must be conveyed to the destination node.

A price vector attaches a price to each node, and a connection function assigns a payoff to each arc as a function of the prices at the arc's origin and destination. The structure of the network is exogenous, and the quantities assigned to source and destination nodes are also typically exogenous, while prices are determined endogenously as part of an equilibrium specification.

An equilibrium flow is a specification of the prices assigned to the nodes and a specification of flows along the arcs with the properties that (i) all of the quantity assigned to each source node is transported away from the source node and each destination node receives its specified quantity, (ii) the prices are such that no arc exhibits a positive payoff, and (iii) no flow occurs on an arc with a negative payoff.

The equilibrium flow framework captures a variety of applications. If all nodes are either source nodes or destination nodes, and an arc runs from every source node to every destination node, then the equilibrium flow problem can be used to model one-to-one matching, either with transfers (as in Crawford & Knoer 1981 or Demange & Gale 1985) or without transfers (as in Gale & Shapley 1962), with prices taking the role of utilities. Interpreting source nodes as firms, destination nodes as consumers, and intermediate nodes as products produced by the firms, the equilibrium flow problem can model hedonic pricing (Rosen 1974; Chiappori et al. 2010; Ekeland et al. 2004). If there is a single source node and single destination node (but perhaps many intermediate nodes), the equilibrium flow problem generalizes the shortest path problem of Bellman (1958). Expanding to multiple sources and destinations gives the minimum-cost flow problem of optimal transport (Galichon 2018). Time dependent routing problems (Gendreau et al. 2015) are similarly a special case. We expect there to be other applications.

The equilibrium flow problem has two useful features. First, the framework requires no assumptions on the payoff functions other than the appropriate monotonicity conditions, obviating the need to work with transferable utility

or analogous linearity assumptions. Second, Galichon et al. (2024) show that the equilibrium correspondence, associating with each price vector the set of quantities consistent with an equilibrium flow, satisfies their condition of unified gross substitutes. The unified gross substitutes condition is a natural extension to correspondences of the familiar weak gross substitutes assumption for functions and is a natural strengthening of Kelso & Crawford's (1982) gross substitutes condition for correspondences. This characterization of the equilibrium correspondence in turn provides the foundation for comparative static results. Section 4 illustrates.

Section 2 presents the equilibrium flow problem, making precise this section's intuitive description.

Section 3 presents the existence argument that constitutes the contribution of this paper. The argument rests on two building blocks, each a variation on familiar ideas. First, there must exist a set of feasible flows capable of carrying the quantities from the source nodes to the destination nodes. We rely on a generalization of Hall's (1935) theorem to obtain a sufficient condition for feasibility. Second, we must be able to find prices such that flows occur only along zero-payoff arcs, with no arcs exhibiting positive payoffs. Here, we rely on a condition reminiscent of Rochet's (1987) cyclical monotonicity.

Section 4 presents an example of the applications of the equilibrium flow problem.

2. EQUILIBRIUM FLOW PROBLEM

This section expands on the presentation of the equilibrium flow problem in Galichon et al. (2024).

2.1. Formulation

Network. Consider a network $(\mathcal{Z}, \mathcal{A})$ where $\mathcal{Z}$ is a finite set of nodes and $\mathcal{A} \subseteq \mathcal{Z} \times \mathcal{Z}$ is the set of directed arcs. If $xy \in \mathcal{A}$, we say that xy is the arc from $x \in \mathcal{Z}$ to $y \in \mathcal{Z}$. We say that x is the starting point of the arc, while y is its end point. We assume that there is no arc in $\mathcal{A}$ whose starting point coincides with the end point, but make no other assumptions on the set of arcs at this point. For example, we allow the possibility that $\mathcal{A}$ contains the arc xy and also the arc yx.

We describe the network with an $|\mathcal{A}|\times|\mathcal{Z}|$ *arc-node incidence matrix* ∇, defined by letting, for $xy \in \mathcal{A}$ and $z \in \mathcal{Z}$

$$\nabla_{xy,z} = \mathbf{1}_{\{z=y\}} - \mathbf{1}_{\{z=x\}}.$$

Each arc thus contributes a pair of nonzero values to the matrix ∇, with $\nabla_{xy,z} = 1$ if xy is an arc ending at z (i.e., $y = z$), and $\nabla_{xy,z} = -1$ if xy is an arc beginning at z (i.e., $x = z$). Otherwise $\nabla_{xy,z} = 0$.

For example, when using the equilibrium flow problem to describe a bipartite matching problem in which all agents are matched, one would have a set of nodes representing sellers (or workers, or men, and so on). Each seller node would be connected by an arc to each element of a set of nodes representing buyers (or firms, or women, and so on), and there would be no other arcs. To expand the model to allow agents to go unmatched, one would add a "dummy" node, with appropriate arcs, interpreting an agent who matches with the dummy node as remaining unmatched. In a routing problem, one would have a collection of nodes representing traffic origins, another set of nodes representing destinations, and a collection of arcs and intermediate nodes describing the possible routes between origins and destinations. **Quantities and Flows.** Let $q \in \mathbb{R}^{\mathcal{Z}}$ with $\sum_{z\in\mathcal{Z}} q_z = 0$ attach a quantity to each node $z \in \mathcal{Z}$. If $q_z < 0$, then the net quantity $|q_z|$ must flow away from node z, while $q_z > 0$ indicates that the net quantity $|q_z|$ must flow into node z. We call q the vector of *quantities*. In many applications, we will have $q_z = 0$ for many nodes.

We let $\mu \in \mathbb{R}_+^{\mathcal{A}}$ be the vector of *internal flows* along arcs, so that μ_{xy} is the flow through arc xy. The feasibility condition connecting these notions is that, for any node $z \in \mathcal{Z}$, the total internal flow that arrives at z minus the total internal flow that leaves z equals the quantity at z, that is

$$\sum_{x:xz\in\mathcal{A}} \mu_{xz} - \sum_{y:zy\in\mathcal{A}} \mu_{zy} = q_z,$$

which we call the *mass balance equation*, and which can be rewritten as

$$\nabla^\intercal \mu = q. \tag{1}$$

In a matching application, without a dummy node, a quantity $q_z < 0$ would identify for each seller node z the quantity of good the node has for sale. A quantity $q_z > 0$ would identify the quantity to be purchased at buyer node z. The restriction $\sum_{z\in\mathcal{Z}} q_z = 0$ would capture the market balance

condition that the total quantities supplied and demanded are equal. The mass balance equation stipulates that each seller sells all of her good, and each buyer purchases all she desires. With the addition of a dummy node, one could capture an imbalance in the total quantities supplied and demanded, as well as the possibilities that some sellers do not sell all of their good, or some buyer demands go unfilled. In a routing problem, a quantity $q_z < 0$ would identify the number of departures from source node z, while $q_z > 0$ would identify the number of arrivals at destination node z. The mass balance equation indicates that every traveler leaves her source and arrives at a destination.

Prices and connection functions. To capture incentives, we let $p \in \mathbb{R}^{\mathcal{Z}}$ be a price vector, where we interpret p_z as the price at node z. If modeling a competitive economy, we could think of a trader as being able to purchase a unit of a commodity at node x (or in market x) at price p_x and then sell it at a node y (for which $xy \in \mathcal{A}$) at price p_y. If thinking about a bipartite matching problem, we could interpret p_x as the utility of the agent at node x and $-p_y$ as the utility of the agent at node y (with the negative sign explained in footnote 1).

Given the price at node y, there is a threshold value of the price at node x such that a trader in a competitive economy is indifferent between buying at price p_x to sell at price p_y, and not doing so. Equivalently, given the utility $v_y = -p_y$ in a matching market, there is a utility $u_x = p_x$ such that the agents at nodes x and y are just able to match and achieve the utilities u_x and v_y. To capture this, we assume there exists a function $G_{xy} : \mathbb{R} \to \mathbb{R}$, for each $xy \in \mathcal{A}$, with the following interpretation. If $p_x > G_{xy}(p_y)$, the purchase price at node x is excessive, and the trader in our competitive economy interpretation will not engage in the trade. On the contrary, if $p_x < G_{xy}(p_y)$, the purchase price is strictly below indifference level and positive payoff can be made from the trade on the arc xy. The value $G_{xy}(p_y)$ thus identifies the highest price at which a trader able to sell at price p_y at node y is willing to buy at node x. Alternatively, value $G_{xy}(p_y)$ identifies the highest utility agent x in a matching market can achieve when matching with agent y in a way that allows y to achieve utility $-p_y$. In keeping with these interpretations, we maintain the following throughout:

Assumption 1. *For each $xy \in \mathcal{A}$, the function $G_{xy} : \mathbb{R} \to \mathbb{R}$ is increasing and continuous, with limits $\lim_{p_y \to \infty} G_{xy}(p_y) = \infty$ and $\lim_{p_y \to -\infty} G_{xy}(p_y) = -\infty$.*

The function G_{xy} is called the *connection function*, as it connects the price at

the endpoint to the price at the starting point.[1]

We refer to a triple $(\mathcal{Z}, \mathcal{A}, G)$ as an equilibrium flow problem.

2.2. Equilibrium

The triple $(q, \mu, p) \in \mathbb{R}^{\mathcal{Z}} \times \mathbb{R}_{+}^{\mathcal{A}} \times \mathbb{R}^{\mathcal{Z}}$ is an equilibrium outcome if it satisfies three conditions. The first condition is the conservation of the flow given by the mass balance equation (1). In the case of a competitive economy, this is the statement that markets balance. In the matching interpretation, this is the statement that agents are matched into pairs.

The second condition is that there is no positive payoff on any arc, that is:

$$p_x \geq G_{xy}\left(p_y\right) \quad \forall xy \in \mathcal{A}. \tag{2}$$

In the interpretation as a competitive economy, this is the statement that excess profits will be dissipated by entry into the market. In the matching interpretation, a positive payoff ($p_x < G_{xy}\left(p_y\right)$) indicates that the utilities for the agents at nodes x and y lie inside the utility frontier characterizing their match.

The third condition is that arcs with negative payoffs carry no flow. Hence

$$\mu_{xy} > 0 \implies p_x \leq G_{xy}\left(p_y\right), \tag{3}$$

which combines with the no-positive-payoff requirement to yield

$$\mu_{xy} > 0 \implies p_x = G_{xy}\left(p_y\right).$$

This is a complementary slackness condition, which can be written

$$\sum_{xy \in \mathcal{A}} \mu_{xy}\left(p_x - G_{xy}\left(p_y\right)\right) = 0.$$

In the case of a competitive economy, a negative payoff would induce exit from the market. In case of a matching market, a negative payoff signifies that the equilibrium utilities of the agents at nodes x and y lie outside their utility frontier, with the complementary slackness condition in turn indicating that these agents must not be matched with each other.

In summary, we define:

[1] This is related to the idea of Galois connections, which explains the choice of the letter G; see Nöldeke & Samuelson (2018). In the matching market, we let the utility of the agent at node y be $-p_y$ so that it makes sense for G_{xy} to be increasing, since an increase in p_y is then a decrease in agent y's utility, allowing an increase agent x's utility p_x.

Definition 1 (**Equilibrium Outcome**). *The triple* $(q, \mu, p) \in \mathbb{R}^{\mathcal{Z}} \times \mathbb{R}_+^{\mathcal{A}} \times \mathbb{R}^{\mathcal{Z}}$ *is an equilibrium outcome when the following conditions are met:*

(i) $\nabla^\intercal \mu = q$
(ii) $p_x \geq G_{xy}\left(p_y\right) \quad \forall xy \in \mathcal{A}$
(iii) $\sum_{xy\in\mathcal{A}} \mu_{xy}\left(p_x - G_{xy}\left(p_y\right)\right) = 0.$

The first condition implies our structural requirement that $\sum_{z\in\mathcal{Z}} q_z = 0$.

Notice that if p satisfies condition (ii), then setting $q = 0$ and $\mu = 0$ ensures that the remaining conditions are satisfied. Indeed, if (q, μ, p) is a equilibrium outcome, then so is $(\lambda q, \lambda\mu, p)$ for any nonnegative scalar λ. Hence, there may be no equilibrium outcome (for example, there may be no p satisfying condition (ii)), but there will often be multiple equilibrium outcomes.

3. EXISTENCE OF EQUILIBRIUM

Our task is now to fix an equilibrium flow problem $(\mathcal{Z}, \mathcal{A}, G)$ and a vector of quantities q (with $\sum_{z\in\mathcal{Z}} q_z = 0$), and then ask whether there exists a price vector p and flow μ such that (q, μ, p) is an equilibrium outcome of the equilibrium flow problem.

3.1. The Sufficient Conditions

We will unsurprisingly require some sufficient conditions for the existence of an equilibrium outcome, which are developed in this section.

3.1.1. Feasible Flows

Suppose we have a particularly simple equilibrium flow problem with $\mathcal{Z} = \{x, y\}$ and $\mathcal{A} = \emptyset$, with quantities $q = (-1, 1)$. It is then immediately obvious that there is no equilibrium flow, as there are no arcs along which to carry mass from node x to node y. Our first task is thus to ensure that the collection of arcs is rich enough to ensure there exist flows satisfying the mass balance condition. To formulate this requirement, we introduce the idea of a retaining set.

For a subset B of $\mathcal{Z}$, an arc xy is said to be outward if $x \in B$ and $y \notin B$. A subset B of $\mathcal{Z}$ is called retaining if there is no arc outward from B. Hence,

subset B is retaining if and only if

$$\nabla \mathbf{1}_B \geq 0,$$

where $\mathbf{1}_B$ is a column vector of dimension $|\mathcal{Z}|$ identifying the nodes contained in the set B, and recalling that ∇ is the $|\mathcal{A}|\times|\mathcal{Z}|$ arc-node incidence matrix.

We state the feasibility condition in terms of retaining sets:

Assumption 2. *If B is a retaining set, then $q(B) = \sum_{z\in B} q_z \geq 0$.*

It is immediately clear that Assumption 2 is necessary for the existence of a flow satisfying the mass balance condition. If Assumption 2 fails, then there is a set (such as the set $\{x\}$ in the motivating example for this subsection) with the property that some quantity must flow away from this set, but there are no arcs to carry such flow. It is perhaps more surprising that this assumption suffices to ensure the existence of a slow satisfying mass balance:[2]

Lemma 1. *Let $q \in \mathbb{R}^{\mathcal{Z}}$ be such that $\sum_{z\in\mathcal{Z}} q_z = 0$. Then the following statements are equivalent.*

[1.1] There exists $\mu \in \mathbb{R}^{\mathcal{A}}_+$ such that $q = \nabla^\intercal \mu$.

[1.2] For each retaining set $B \subseteq \mathcal{Z}$, one has $q(B) \geq 0$.

Proof. That [1.1] implies [1.2] is straightforward: If there is no arc from a node in B to a node in $\mathcal{Z}\backslash B$, then mass can only enter into B. If a flow exists satisfying mass balance, it must then be that $q(B) \geq 0$.

To show the converse, we exploit Hoffman's (1960) circulation theorem. For any subset $B \subset \mathcal{Z}$, let $O(B) = \{xy \in \mathcal{A} : x \in B, y \notin B\}$, and $I(B) = \{xy \in \mathcal{A} : x \notin B, y \in B\}$ be respectively the set of outward and inward arcs of B. Hoffman's theorem states that, given $q \in \mathbb{R}^{\mathcal{Z}}$ with $\sum_{z\in\mathcal{Z}} q_z = 0$ and given the vectors $\underline{\mu}$ and $\overline{\mu}$ in $\mathbb{R}^{\mathcal{A}}_+$, there exists $\mu \in \mathbb{R}^{\mathcal{A}}_+$ such that

$$\nabla^\intercal \mu = q \qquad \text{and} \qquad \underline{\mu} \leq \mu \leq \overline{\mu}$$

if and only if for all subsets B of $\mathcal{Z}$, one has

$$\sum_{z\in B} q_z \leq \sum_{a\in I(B)} \overline{\mu}_a - \sum_{a\in O(B)} \underline{\mu}_a.$$

[2] For example, one might have speculated that sufficient conditions would also require a separate stipulation that if we have a set B with no inbound arcs, such as the set $\{y\}$ in our example, then $\sum_{z\in B} q_z \leq 0$.

This result is usually stated for vectors q that are identically equal to zero, but the adaptation to nonzero vectors q for which $\sum_{z\in\mathcal{Z}} q_z = 0$ is straightforward. With this result in hand, let $\overline{\mu}_a = +\infty$ and $\underline{\mu}_a = 0$ for each $a \in \mathcal{A}$, in which case Hoffman's condition becomes the requirement that for all sets $\tilde{B}$ that have no inbound links,

$$q(\tilde{B}) \leq 0.$$

We then argue that this condition is equivalent to the second condition in Lemma 1. Call a set with no inbound links *repelling*. The result then follows from noting that the set $B \subset \mathcal{Z}$ is retaining if and only if B^c is repelling, and for any set B,

$$q(B) + q(B^c) = 0. \qquad \square$$

Lemma 1 implies Hall's (1935) marriage theorem.[3] We make an additional, technical feasibility assumption.

Assumption 3. *For each $z \in \mathcal{Z}$, either $q_z > 0$ or there exists a path from z to a node $y \in \mathcal{Z}$ with $q_y > 0$.*

To see why we refer to this as a technical assumption, notice that it is automatically satisfied if $q_z > 0$, and is implied by Assumption 2 if $q_z < 0$. However, Assumption 2 is compatible with the existence of a node z with $q_z = 0$ that fails Assumption 3. In this case, it is impossible for an equilibrium to exhibit

[3] To see this, assume we are in the bipartite case to which Hall's marriage theorem applies. Hence, $\mathcal{Z} = \mathcal{X} \cup \mathcal{Y}$ and $\mathcal{A} \subseteq \mathcal{X} \times \mathcal{Y}$. From Lemma 1, we have that there exists a flow satisfying mass balance if and only if for all retaining sets B, one has $q(B) \geq 0$. That is, for all retaining sets B, one has

$$q(B \cap \mathcal{X}) + q(B \cap \mathcal{Y}) \geq 0.$$

Therefore for all subsets $B_{\mathcal{X}}$ of $\mathcal{X}$ and all subsets $B_{\mathcal{Y}}$ of $\mathcal{Y}$, we have

$$[B_{\mathcal{X}} \cup B_{\mathcal{Y}} \text{ retaining }] \Longrightarrow [q(B_{\mathcal{Y}}) \geq -q(B_{\mathcal{X}})]. \tag{4}$$

Now fix $B_{\mathcal{X}} \subset \mathcal{X}$ and let $B_{\mathcal{Y}}$ be the set of nodes y in $\mathcal{Y}$ with the property that there is an arc from a node in $B_{\mathcal{X}}$ to y. Then by construction $B_{\mathcal{X}} \cup B_{\mathcal{Y}}$ is retaining, and so from (4) mass balance is attainable if and only if

$$q(B_{\mathcal{Y}}) \geq -q(B_{\mathcal{X}}),$$

which is Hall's marriage theorem.

flow through node z, and we can without loss of generality eliminate node z from the network. Assumption 3 thus ensures that there are no obviously irrelevant nodes in the network. We could accommodate such nodes, but then would require another step in our equilibrium construction, to attach prices to such nodes.

3.1.2. Absence of Profitable Flows

Assumption 2 ensures that there exists a flow capable of moving quantities from their sources to their destinations. However, in equilibrium the flow must make use of only zero-payoff arcs and there must be no positive-payoff arcs.

To see how these payoff conditions could fail, let $\mathcal{Z} = \{x, y\}$ and let $\mathcal{A} = \{xy, yx\}$. Consider the quantities $q = \{-1, 1\}$. The only retaining set is $\mathcal{Z}$ itself, and so Assumption 2 is satisfied. Let the connection function be:

$$\begin{aligned} G_{xy}(p_y) &= p_y \\ G_{yx}(p_x) &= p_x + 1. \end{aligned}$$

Then there is no set of prices ensuring that no link exhibits a strictly positive payoff. A nonpositive payoff on link xy requires $p_x \geq p_y$, while a nonpositive payoff on link yx requires $p_y \geq p_x + 1$. In effect, the equilibrium flow problem admits a money pump (see Echenique et al. 2011 for a connection to revealed preference theory), with a buyer (in the competitive economy interpretation) able to generate unlimited payoffs by a continuing moving flow through the loop from x to y to x, and so on forever.

We accordingly require a condition on the connection function excluding such profitable loops. One obvious condition would be to simply preclude the existence of loops in the set of arcs, but that is stronger than we need.

We say that the collection of nodes $(z_0, z_1, \ldots z_{k+1})$ is a loop if it contains $k + 1$ distinct nodes with $z_\ell z_{\ell+1} \in \mathcal{A}$ for $\ell = 0, \ldots k$, and $z_0 = z_{k+1}$. We assume:

Assumption 4. *If* $(z_0, z_1, \ldots, z_{k+1})$ *is a loop, then* $p > G_{z_0 z_1} \circ G_{z_1 z_2} \circ \ldots \circ G_{z_k z_{k+1}}(p)$ *for all* p.

Assumption 4 ensures there is no way to achieve positive payoffs by moving flow through a loop, with no net change in the quantities. If such a loop existed, continual movement along this loop would constitute a money pump, giving

rise to infinite payoffs. Equivalently, if Assumption 4 fails, then it may be impossible to assign prices to nodes in such a way as to ensure the absence of positive-payoff arcs.

Conversely, if Assumption 4 holds, one can show that it is possible to assign prices to nodes so that no arc admits a positive payoff. The argument is by construction, following the lines of a construction that appears in the proof of Lemma 2 below. One first arbitrarily fixes the price at a single node, letting all other prices be $-\infty$. One then successively updates prices at nodes, setting the price at each node z just high enough to ensure that no arc emanating from z admits a positive payoff. Assumption 4 ensures that this process terminates in a specification of prices satisfying the no-positive-payoff requirement.

3.2. The Existence Result

Assumption 2 ensures that it is possible to transport the required mass from sources to destinations. Assumption 4 ensures that it is possible to attach prices to nodes so that there are no positive-payoff arcs. In equilibrium, we require some coordination between these two constructions, in that an equilibrium flow must transport the required mass from sources to targets using only zero-payoff arcs. It would accordingly be no surprise if additional conditions were required to ensure the existence of an equilibrium flow. However, these assumptions suffice:

Theorem 1. *Let the equilibrium flow problem $(\mathcal{Z}, \mathcal{A}, G)$ and quantity q satisfy Assumptions 1-4. Fix a node z with $q_z \neq 0$ and a price θ. Then there exists an equilibrium outcome with $p_z = \theta$.*

The node z and price θ are arbitrary (as long as $q_z \neq 0$). If we did not fix the price at some node, then we would have multiple equilibrium outcomes.[4] For example, Let $\mathcal{Z} = \{x, y\}$, $\mathcal{A} = \{xy\}$, $G_{xy}(p_y) = p_y$, and $q = (-1, 1)$. Then any outcome (q, μ, p) with $q = (-1, 1)$, $\mu_{xy} = 1$ and $p_x = p_y$ is an equilibrium outcome. This is the functional equivalent of the familiar observation that the conditions for competitive equilibrium in an exchange economy set relative but not absolute prices.

[4] At the end of Section 2.2, we noted that that if (q, μ, p) is an equilibrium outcome, then so is $(\lambda q, \lambda\mu, p)$ for any $\lambda > 0$. Here, we are noting that there will in general exist multiple equilibrium outcomes, exhibiting different prices, for a *fixed* quantity q.

3.3. Proof of Theorem 1

The proof of Theorem 1 associates the equilibrium flow problem with a bipartite matching problem, and then exploits existence results for the bipartite matching problem. This association may be useful for other purposes.

3.3.1. Equilibrium Outcomes and Bipartite Solutions

Given an equilibrium flow problem $(\mathcal{Z}, \mathcal{A}, G)$, define the set of source nodes as $\mathcal{X} = \{x \in \mathcal{Z} : q_x < 0\}$ and the set of destination nodes as $\mathcal{Y} = \{y \in \mathcal{Z} : q_y > 0\}$. Given nodes $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, we say that the sequence $(z_0, \ldots, z_{k+1})$ is a path from x to y if $z_0 = x$, $z_{k+1} = y$, and $z_\ell z_{\ell+1} \in \mathcal{A}$ for $\ell = 0, \ldots k$. Let $\tilde{\mathcal{A}}$ be the set of pairs $xy \in \mathcal{X} \times \mathcal{Y}$ for which such a path exists.

Given an equilibrium flow problem $(\mathcal{Z}, \mathcal{A}, G)$, we define an associated bipartite equilibrium flow problem $(\mathcal{X} \cup \mathcal{Y}, \tilde{\mathcal{A}}, \tilde{G})$, where the connection function $\tilde{G}$ is defined by letting, for $xy \in \tilde{\mathcal{A}}$,

$$\tilde{G}_{xy}\left(p_y\right) = \sup_{\substack{z_0, z_1, z_2, \ldots, z_{k+1} \\ \text{s.t. } z_0 = x,\ z_{k+1} = y \\ z_\ell z_{\ell+1} \in \mathcal{A}}} G_{xz_1} \circ G_{z_1 z_2} \circ \ldots \circ G_{z_k y}\left(p_y\right).$$

By definition of $\tilde{\mathcal{A}}$, the set over which we are taking the supremum is nonempty, and therefore we have necessarily $\tilde{G}_{xy}\left(p_y\right) > -\infty$. In principle, we cannot preclude the possibility that $\tilde{G}_{xy}\left(p_y\right) = +\infty$, which could happen with a sequence of paths visiting an increasing and unbounded number of nodes, in which case the path must contain a loop. However, it is straightforward to verify that Assumption 4 ensures that $\tilde{G}_{xy}(p_y)$ is bounded and the maximum is attained.

The bipartite equilibrium flow problem thus contains the source nodes and destination nodes of the original problem. There is an arc from a source node to a target node in the bipartite problem if there is a path between these two nodes in the original problem. The connection function in the bipartite problem is defined by letting $\tilde{G}_{xy}(p_y)$ be the composition of the connection functions along one of the paths from x to y in the original problem, namely the path in the original problem that maximizes $\tilde{G}_{xy}(p_y)$. Intuitively, we ask which path from x to y in the original problem maximizes the value of moving flow from x to y, and use this path to induce the function $\tilde{G}_{xy}$ in the bipartite problem.

Given a function q assigning quantities to the nodes in $\mathcal{Z}$, we simplify notation by also using q to denote the restriction of q to the nodes $\mathcal{X} \cup \mathcal{Y}$. Notice that in restricting q to $\mathcal{X} \cup \mathcal{Y}$, we exclude a node $z \in \mathcal{Z}$ if and only if $q_z = 0$.

Lemma 2. *Let Assumptions 1–4 hold. Then there exists an equilibrium outcome (q, μ, p) for the equilibrium flow problem $(\mathcal{Z}, \mathcal{A}, G)$ if and only if there exists an equilibrium outcome $(q, \tilde{\mu}, \tilde{p})$ for the associated bipartite equilibrium flow problem $(\mathcal{X} \cup \mathcal{Y}, \tilde{\mathcal{A}}, \tilde{G})$ such that p and $\tilde{p}$ agree on $\mathcal{X}$ and $\mathcal{Y}$.*

Proof. [Only if:] Let (q, μ, p) be an equilibrium outcome of the equilibrium flow problem $(\mathcal{Z}, \mathcal{A}, G)$. We derive an equilibrium outcome for the associated bipartite equilibrium flow problem with the required properties.

Let Π_{xy} be the set of paths in $\mathcal{A}$ beginning at x and ending at y. Let $\tilde{\mu}_{xy} = \sum_{\rho \in \Pi_{xy}} \mu^\rho$, where μ^ρ is the flow associated with path ρ.[5] It is straightforward from the mass balance condition for the equilibrium outcome (q, μ, p) that for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ we have

$$\begin{aligned} -\sum_{y \in \mathcal{Y}} \tilde{\mu}_{xy} &= q_x \\ \sum_{x \in \mathcal{X}} \tilde{\mu}_{xy} &= q_y. \end{aligned}$$

This ensures that q and $\tilde{\mu}$ satisfy the mass balance equation (1) in the bipartite equilibrium flow problem $(\mathcal{X} \cup \mathcal{Y}, \tilde{\mathcal{A}}, \tilde{G})$.

We next argue that for $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ with $xy \in \tilde{\mathcal{A}}$, we have

$$p_x \geq \tilde{G}_{xy}\left(p_y\right).$$

To see this, we note that for any given path $x = z_0, z_1, z_2, \ldots, z_{k+1} = y$ from x to y, we have (from the no-positive-payoff equilibrium condition (2)) $p_{z_\ell} \geq G_{z_\ell z_{\ell+1}}\left(p_{z_{\ell+1}}\right)$ and thus

$$p_x \geq G_{xz_1} \circ G_{z_1 z_2} \circ \ldots \circ G_{z_k y}\left(p_y\right).$$

Taking the supremum over paths from x to y, we have $p_x \geq \tilde{G}_{xy}\left(p_y\right)$. This ensures that $(q, \tilde{\mu}, \tilde{p})$ satisfies the equilibrium condition (2).

[5] This sum could conceivably be infinite, as there may be an infinite number of paths from x to y. However, all but finitely many of the paths from x to y must involve a loop, and Assumption 4 implies that an equilibrium outcome cannot induce a positive flow along a loop.

Finally, suppose we have $\tilde{\mu}_{xy} > 0$. Then there must exist a path $(z_0, z_1, z_2, \ldots, z_{k+1})$ from x $(= z_0)$ to y $(= z_{k+1})$ along which the flow is strictly positive. Thus, along this path $p_{z_\ell} = G_{z_\ell z_{\ell+1}}\left(p_{z_{\ell+1}}\right)$, and hence

$$p_x = G_{xz_1} \circ G_{z_1 z_2} \circ \ldots \circ G_{z_k y}\left(p_y\right).$$

Because $p_x \geq \tilde{G}_{xy}\left(p_y\right) \geq G_{xz_1} \circ G_{z_1 z_2} \circ \ldots \circ G_{z_k y}\left(p_y\right)$, it follows that equality holds everywhere and thus

$$p_x = \tilde{G}_{xy}\left(p_y\right).$$

This ensures that $(q, \tilde{\mu}, \tilde{p})$ satisfies the equilibrium condition (3). Therefore, letting $\tilde{p}$ be the restriction or p to $\mathcal{X} \cup \mathcal{Y}$, we have the equilibrium outcome $(q, \tilde{\mu}, \tilde{p})$ for the bipartite equilibrium flow problem $(\mathcal{X} \cup \mathcal{Y}, \tilde{\mathcal{A}}, \tilde{G})$, i.e., satisfying:

$$\begin{aligned} -\textstyle\sum_{y \in \mathcal{Y}} \tilde{\mu}_{xy} &= q_x \\ \textstyle\sum_{x \in \mathcal{X}} \tilde{\mu}_{xy} &= q_y \\ \tilde{p}_x &\geq \tilde{G}_{xy}\left(\tilde{p}_y\right) \\ \tilde{\mu}_{xy} > 0 &\implies \tilde{p}_x = \tilde{G}_{xy}\left(\tilde{p}_y\right). \end{aligned} \tag{5}$$

[If:] Suppose $(q, \tilde{\mu}, \tilde{p})$ is an equilibrium outcome of the bipartite equilibrium flow problem $(\mathcal{X} \cup \mathcal{Y}, \tilde{\mathcal{A}}, \tilde{G})$, and hence satisfies (5). We find an associated equilibrium (q, μ, p) of the equilibrium flow problem $(\mathcal{Z}, \mathcal{A}, G)$, where q retains its specification on $\mathcal{X} \cup \mathcal{Y}$ and is otherwise zero.

For each pair $xy \in \tilde{\mathcal{A}}$ with $\tilde{\mu}_{xy} > 0$, consider the (by construction nonempty) set of paths from x to y satisfying $\tilde{p}_x = \tilde{G}_{xy}(\tilde{p}_y)$. Choose any one of these paths, and let the flow μ route a flow of size $\tilde{\mu}_{xy}$ along this path. This ensures that μ satisfies the mass balance equilibrium condition (1). For an arc $zz' \in \mathcal{A}$, $\mu_{zz'}$ is defined as the sum of the flows along the selected paths that contain zz', for every origin-destination pair $xy \in \tilde{\mathcal{A}}$.

We now define an iterative procedure that generates the prices p. For the initial step, we let

$$p_z^0 = \begin{cases} \tilde{p}_z & \text{if } z \in \mathcal{Y} \\ -\infty & \text{otherwise} \end{cases},$$

with the induction step given by

$$p_z^{t+1} = \max\left\{p_z^t, \max_{y: zy \in \mathcal{A}}\left\{G_{zy}\left(p_y^t\right)\right\}\right\}.$$

We then complete the argument with the following lemma:

Lemma 3. *Let Assumptions 1–4 hold. For t sufficiently large (but finite), (q, μ, p^t) is an equilibrium outcome of the equilibrium flow problem $(\mathcal{Z}, \mathcal{A}, G)$.*

Proof. We establish this result in the following series of steps:

1. After a finite number of steps, $p_z^t > -\infty$ for all z.

   To see this, let $x \rightsquigarrow y$ denote a path from node x to node y, let $l(x \rightsquigarrow y)$ denote the number of arcs along this path, let $G_{z \rightsquigarrow y}(p_y)$ be the composition of the functions G along the arcs in this path, and let $d(x, y)$ be the number of arcs in the *shortest* path from x to y. Then we have,

   $$p_z^t = \max_{y \in \mathcal{Y}} \max_{z \rightsquigarrow y : l(z \rightsquigarrow y) \leq t} G_{z \rightsquigarrow y}(p_y),$$

   letting the inner maximum be $-\infty$ if $\{z \rightsquigarrow y : l(z \rightsquigarrow y) \leq t\}$ is empty. Hence, if there is a directed path of length at most k from z to $\mathcal{Y}$, then the price of z will be updated to a finite price (and never thereafter reduced) after at most k steps. Assumption 2 (for nodes z with $q_z > 0$) and Assumption 3 (for nodes z with $q_z = 0$) ensure that for every node $z \notin \mathcal{Y}$, there is a path from node z to $\mathcal{Y}$. As a result, all prices will be updated to finite values, ensuring $p_z^t > -\infty$ for all z, after $\max_z \min_{y \in \mathcal{Y}} d(z, y)$ steps.

2. The price p_y of a node $y \in \mathcal{Y}$ is never updated by the algorithm.

   Assume that one updates the price of node $y \in \mathcal{Y}$ at some step. Then there is a path $y \rightsquigarrow y'$ from y to some $y' \in \mathcal{Y}$ with $p_y < G_{y \rightsquigarrow y'}(p_{y'})$. Let $x \in \mathcal{X}$ be an element of $\mathcal{X}$ such that $\tilde{\mu}_{xy} > 0$. Then one has $p_x = \tilde{G}_{xy}(p_y)$, and hence $p_x < \tilde{G}_{xy} \circ G_{y \rightsquigarrow y'}(p_{y'}) = G_{x \rightsquigarrow y'}(p_{y'}) \leq \tilde{G}_{xy'}(p_{y'})$, a contradiction.

3. There exists a T such that $p^t = p^T$ for all $t \geq T$.

   If the price at node z is updated at step t, then it must be that

   $$p_z^t = \max_{y \in \mathcal{Y}} \max_{z \rightsquigarrow y : l(z \rightsquigarrow y) \leq t} G_{z \rightsquigarrow y}(p_y) > p_z^{t-1}$$

   for some node $y \in \mathcal{Y}$. Assumption 4 ensures that $\max_{z \rightsquigarrow y : l(z \rightsquigarrow y) \leq t} G_{z \rightsquigarrow y}(p_y)$ is achieved by path $z \rightsquigarrow y$ that contains no loop. There are a finite number of nodes $y \in \mathcal{Y}$ and for each pair of nodes $z \in \mathcal{Z}$ and $y \in \mathcal{Y}$ a finite

number of paths $z \rightsquigarrow y$ that contain no loops. This ensures that for each node $z \in \mathcal{Z}$, the price p_z is updated a finite number of times. Hence, there exists a step T after which no further updating occurs.

4. For a node $y \in \mathcal{Y}$, $p_y^T = \tilde{p}_y$. For $x \in \mathcal{X}$, $p_x^T = \tilde{p}_x$.

   The first statement follows from steps 2 and 3. For the second statement, if there exists a x such that $p_x^T \neq \tilde{p}_x$, then:

   - If $p_x^T > \tilde{p}_x$, then there exists a node $y \in \mathcal{Y}$ and a path $x \rightsquigarrow y$ such that $G_{x \rightsquigarrow y}\left(\tilde{p}_y\right) = p_x^T > \tilde{p}_x = \max_{y' \in \mathcal{Y}} \tilde{G}_{xy'}\left(\tilde{p}_{y'}\right) \geq G_{x \rightsquigarrow y}\left(\tilde{p}_y\right)$, a contradiction.
   - If $p_x^T < \tilde{p}_x$, then there is a node $y \in \mathcal{Y}$ and a path $x \rightsquigarrow y$ such that $\tilde{p}_x = G_{x \rightsquigarrow y}\left(p_y^T\right)$. Moving along path $x \rightsquigarrow y$, take the last z such that $p_z^T \neq \tilde{G}_{zy}\left(p_y^T\right)$, and let z' be its successor. One has by definition $p_z^T < \tilde{G}_{zy}\left(p_y^T\right) = G_{zz'}\left(\tilde{G}_{z'y}\left(p_y^T\right)\right) = G_{zz'}\left(p_{z'}^T\right)$, which contradicts the fact that the algorithm has reached stationarity.

5. We let the stationary prices p^T be our price vector p. The construction of the prices ensures that for every arc $zz' \in \mathcal{A}$, we have $p_z \geq G_{zz'}(p_{z'})$, and hence the no-positive-payoff equilibrium condition (2) holds. Moreover, if $\mu_{zz'} > 0$, then the arc zz' is contained in a path from source node x to destination node y for some $xy \in \tilde{\mathcal{A}}$ with $\tilde{\mu}_{xy} > 0$. We then have $p_x = \tilde{G}_{xy}(p_y)$ and hence $p_x = G_{x \rightsquigarrow y}(p_y)$, which suffices to ensure $p_z = G_{zz'}(p_{z'})$ and hence the equilibrium condition (3) is met. □

This completes the proof of Lemma 2. □

3.3.2. *Bipartite Existence Result*

The bipartite equilibrium flow problem $(\mathcal{X} \cup \mathcal{Y}, \tilde{\mathcal{A}}, \tilde{G})$ and the quantity q satisfy Hall's (1935) condition if, for every subset $\mathcal{B}$ of $\mathcal{X}$, we have

$$-\sum_{x \in B} q_x \leq \sum_{\{y \in \mathcal{Y} \,:\, \exists x \in B \text{ with } xy \in \tilde{\mathcal{A}}\}} q_y. \tag{6}$$

Intuitively, the set B is a set of nodes from which from which the total quantity $-\sum_{x \in B} q_x$ (≥ 0) must flow. The term on the right side sums over the set of

nodes to which this quantity might possibly flow, and the condition requires that the latter set has at least the capacity to receive the desired flow.

Lemma 4. *Let Assumptions 1, 3 and 4 hold. Fix a node z with $q_z \neq 0$ and a price $\theta \in \mathbb{R}$. If the bipartite equilibrium flow problem $(\mathcal{X} \cup \mathcal{Y}, \tilde{\mathcal{A}}, \tilde{G})$ and quantity q satisfy Hall's condition, then there exists an equilibrium outcome $(q, \tilde{u}, \tilde{p})$ for $(\mathcal{X} \cup \mathcal{Y}, \tilde{\mathcal{A}}, \tilde{G})$ with $\tilde{p}_z = \theta$.*

Proof. The proof first extends the set of arcs $\tilde{\mathcal{A}}$ to the set $\mathcal{X} \times \mathcal{Y}$ by introducing an arc for every $xy \in (\mathcal{X} \times \mathcal{Y}) \setminus \tilde{\mathcal{A}}$, with the connection function $\tilde{G}$ extended to such arcs (without introducing new notation) by letting

$$\tilde{G}_{xy}(p_y) = p_y - n$$

for some integer n. Our approach will be to establish the existence of an equilibrium outcome in the equilibrium flow problem created by added these additional arcs, and then to ensure that for sufficiently large n, flow along any such arc not in $\tilde{\mathcal{A}}$ is so unprofitable that the equilibrium outcome will not exhibit such flow. This will ensure that the outcome is also an equilibrium outcome of the bipartite equilibrium flow problem $(\mathcal{X} \cup \mathcal{Y}, \tilde{\mathcal{A}}, \tilde{G})$.

We refer to the bipartite equilibrium flow problem with the additional arcs as the expanded (bipartite, hereafter omitted from the name) equilibrium flow problem, denoted by $(\mathcal{X} \cup \mathcal{Y}, \mathcal{A}^E, \tilde{G})$. Notice that if an equilibrium flow in this expanded problem directs flow along an arc xy that is not contained in $\tilde{\mathcal{A}}$, then it must be that $p_x = p_y - n$.

Let $\left(q, \mu^E_{xy}, p\right)$ be an equilibrium outcome of the expanded equilibrium flow problem, i.e., let $\left(q, \mu^E_{xy}, p\right)$ satisfy

(i) -$\sum_y \mu^E_{xy} = q_x$ for all $x \in \mathcal{X}$ and $\sum_x \mu^E_{xy} = q_y$ for all $y \in \mathcal{Y}$
(ii) $p_x \geq \tilde{G}_{xy}\left(p_y\right)$ for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$
(iii) $\mu^E_{xy} > 0$ implies $p_x = \tilde{G}_{xy}\left(p_y\right)$, with $p_x = p_y - n$ if $xy \notin \tilde{\mathcal{A}}$.
(iv) $p_z = \theta$.

The existence of this outcome follows from Proposition 6 of Nöldeke and Samuelson. The key to applying this proposition is to note that the function $G_{xy}(p_y)$ is increasing, continuous, and as p_y varies, sweeps out the entire real line. Alternatively, one could first work with a regularized version of the equilibrium flow problem, following the lines of Galichon et al. (2019), for

which a constructive proof of existence can be given, and then exploit results from Galichon et al. (2015) to provide a limiting argument giving existence in the underlying equilibrium flow problem.

We begin with an intermediate result. Consider the arcs defined by

$$\mathcal{A}^* = \left\{xy \in \tilde{\mathcal{A}}\right\} \cup \left\{yx : x \in \mathcal{X}, y \in \mathcal{Y}, \mu^E_{xy} > 0\right\}$$

The set $\mathcal{A}^*$ thus adds to the set of arcs $\mathcal{A}$ from the bipartite equilibrium flow problem an arc yx from node y to node x whenever the arc xy is *not* contained in $\tilde{\mathcal{A}}$ but carries positive flow in the equilibrium outcome of the extended equilibrium flow problem. These added arcs may not appear in the extended equilibrium flow problem, nor will we use them to construct yet another equilibrium flow problem, but the set $\mathcal{A}^*$ is useful in formulating the following intermediate result:

Lemma 5. *Let Assumptions 1, 3 and 4 hold and let the bipartite equilibrium flow problem* $(\mathcal{X} \cup \mathcal{Y}, \tilde{\mathcal{A}}, \tilde{G})$ *and quantity* q *satisfy Hall's condition. Assume* $x^*y^* \notin \tilde{\mathcal{A}}$ *and* $\mu^E_{x^*y^*} > 0$. *Then there exists no partition of the set of nodes* $\mathcal{Z} = \mathcal{X} \cup \mathcal{Y}$ *into sets A and B such that* $y^* \in A$, $x^* \in B$ *and there are no arcs of* $\mathcal{A}^*$ *from B to A.*

The question is whether we can partition the set of nodes $\mathcal{Z}$ so that the set B containing node x^* is retaining (given the arcs $\mathcal{A}^*$) *and* excludes the node y^*. The set B must contain more than simply the node x^*, since there is some arc x^*y in $\mathcal{A}^*$. We can accordingly think of successively adding nodes to the set B in an effort to make it retaining. The proof uses the fact that $\mu^E_{x^*y^*} > 0$ is an equilibrium outcome of the extended equilibrium flow problem to show that this effort leads to a contradiction.

Proof. We suppose that a partition with the desired properties exists, and seek a contradiction. Let $A^{\mathcal{X}} = A \cap \mathcal{X}$ and $A^{\mathcal{Y}} = A \cap \mathcal{Y}$, and define $B^{\mathcal{X}}$ and $B^{\mathcal{Y}}$ in a similar fashion.

Hall's condition implies that the bipartite equilibrium flow problem has a flow satisfying mass balance. Because there are by assumption no arcs in $\mathcal{A}^*$ from $B^{\mathcal{X}}$ to $A^{\mathcal{Y}}$, this implies that $-q\left(A^{\mathcal{X}}\right) \geq q\left(A^{\mathcal{Y}}\right)$.

Similarly, by assumption there exists a flow μ^E satisfying the mass balance equation in the extended equilibrium flow problem. Because there are by assumption no arcs in $\mathcal{A}^*$ from $B^{\mathcal{Y}}$ to $A^{\mathcal{X}}$, this implies that $-q\left(B^{\mathcal{X}}\right) \geq q\left(B^{\mathcal{Y}}\right)$.

Summing the previous two inequalities yields $-q(\mathcal{X}) \geq q(\mathcal{Y})$, but as this holds as an equality, we get an equality in the two previous inequalities. Thus $-q\left(A^{\mathcal{X}}\right) = q\left(A^{\mathcal{Y}}\right)$ and $-q\left(B^{\mathcal{X}}\right) = q\left(B^{\mathcal{Y}}\right)$.

Now $-q\left(A^{\mathcal{X}}\right) = \sum_{x\in A^{\mathcal{X}}} \sum_{y\in\mathcal{Y}} \mu^E_{xy}$. In addition, for $x \in A^{\mathcal{X}}$ and $y \in B^{\mathcal{Y}}$, we have $\mu^E_{xy} = 0$, since otherwise there would be an arc of $\mathcal{A}^*$ from B to A. Hence

$$-q\left(A^{\mathcal{X}}\right) = \sum_{x\in A^{\mathcal{X}}} \sum_{y\in A^{\mathcal{Y}}} \mu^E_{xy}.$$

Because μ^E satisfies mass balance, we also have

$$q\left(A^{\mathcal{Y}}\right) = \sum_{x\in\mathcal{X}} \sum_{y\in A^{\mathcal{Y}}} \mu^E_{xy}.$$

Subtracting gives

$$0 = -q\left(A^{\mathcal{X}}\right) - q\left(A^{\mathcal{Y}}\right) = \sum_{x\in B^{\mathcal{X}}} \sum_{y\in A^{\mathcal{Y}}} \mu^E_{xy},$$

and hence for all $x \in B^{\mathcal{X}}$ and $y \in A^{\mathcal{Y}}$, $\mu^E_{xy} = 0$, which contradicts $x^* \in B^{\mathcal{X}}$, $y^* \in \mathcal{A}^Y$ and $\mu^E_{x^*y^*} > 0$. This contradiction proves the claim. □

The proof of Lemma 4 is then completed by the following lemma.

Lemma 6. *Let Assumptions 1, 3 and 4 hold and let the bipartite equilibrium flow problem $(\mathcal{X}\cup\mathcal{Y}, \tilde{\mathcal{A}}, \tilde{G})$ and quantity q satisfy Hall's condition. For n large enough, $\mu^E_{x^*y^*} > 0$ implies $x^*y^* \in \tilde{\mathcal{A}}$.*

Proof. The proof makes use of Lemma 5 to establish a contradiction. Suppose we have $x^* \in \mathcal{X}$ and $y^* \in \mathcal{Y}$ with $\mu^E_{x^*y^*} > 0$ and $x^*y^* \notin \tilde{\mathcal{A}}$. We now seek a partition of the set of nodes. We start by partitioning the set of nodes $\mathcal{Z}$ into the sets $A = \{y^*\}$ and $B = \mathcal{Z}\backslash\{y^*\}$. Then by Lemma 5 there must be an arc in $\mathcal{A}^*$ from B to A. Call z_1 the origin point of that arc, and form a new partition with $A = \{y^*, z_1\}$ and $B = \mathcal{Z}\backslash\{y^*, z_1\}$. Once again, as long as the new candidate for the set A does not contain x^*, by Lemma 5 there must be an arc in $\mathcal{A}^*$ from B to A. We continue in this fashion, successively adding nodes to our candidate set A, until encountering an arc in $\mathcal{A}^*$ from x^* to some z_k in the preceding iteration to A, at which point the node x^* is added to the updated set A. At this point, the elements of A along with their arcs, taken in

the reverse of the order in which they were added to A, constitute a path from x^* to y^*, given by

$$x^* = z_{k+1}, z_k, z_{k-1}, \ldots, z_2, z_1, z_0 = y^*.$$

By the construction of the path, any pair of successive nodes $z_{\ell+1}z_\ell$ from this path for which $\ell + 1$ is odd is an arc from a node in $\mathcal{X}$ to a node in $\mathcal{Y}$. This arc is included in the set of arcs $\mathcal{A}^*$ in the bipartite equilibrium flow problem, and so satisfies the no-positive-surplus equilibrium condition

$$p_{z_{\ell+1}} \geq \tilde{G}_{\ell+1,\ell}(p_{z_\ell}). \tag{7}$$

This includes the arc x^*z_k, as this must be an arc from a node in $\mathcal{X}$ (namely x^*) to a node in $\mathcal{Y}$, namely z_k, where k must be an even integer.

Again by the construction of the path, any pair of successive nodes $z_{\ell+1}z_\ell$ from this path for which $\ell + 1$ is even is an arc $z_{\ell+1}z_\ell \in \mathcal{A}^*\backslash\tilde{\mathcal{A}}$. Then by the construction of $\mathcal{A}^*$, the arc $z_\ell z_{\ell+1}$ (note the reversal of subscripts) carries a positive flow in the equilibrium outcome of the expanded equilibrium flow problem, and hence satisfies

$$p_{z_\ell} = p_{z_{\ell+1}} - n. \tag{8}$$

By assumption, we have $\mu^E_{x^*y^*} > 0$ with $x^*y^* \notin \tilde{\mathcal{A}}$, and hence we have

$$p_{x^*} = \tilde{G}_{x^*y^*}\left(p_{y^*}\right) = p_{y^*} - n. \tag{9}$$

Equation (9) shows that as n grows, $p_{y^*} - p_{x^*}$ diverges. We show that this gives a contradiction, establishing the result.

Write $p_{y^*} - p_{x^*}$ as

$$\begin{aligned}
p_{y^*} - p_{x^*} &= p_{y^*} - p_{z_1} && (p_{z_1} \geq \tilde{G}_{z_1y^*}(p_{y^*})) && (1)\\
&+ p_{z_1} - p_{z_2} && (p_{z_2} = p_{z_1} + n) && (2)\\
&+ p_{z_2} - p_{z_3} && (p_{z_3} \geq \tilde{G}_{z_3z_2}(p_{z_2})) && (3)\\
&+ p_{z_3} - p_{z_4} && (p_{z_4} = p_{z_3} + n) && (4)\\
&+ p_{z_4} - p_{z_5} && (p_{z_5} \geq \tilde{G}_{z_5z_4}(p_{z_3})) && (5)\\
&+ p_{z_5} - p_{z_6} && (p_{z_6} = p_{z_5} + n) && (6)\\
&\vdots\\
&+ p_{z_k} - p_{x^*} && (p_{x^*} \geq \tilde{G}_{x^*z_k}(p_{z_k})).
\end{aligned}$$

The terms with odd numbers correspond to arcs from $\mathcal{X}$ to $\mathcal{Y}$ that are contained in $\tilde{\mathcal{A}}$. (The final $\cdot$ is an odd-numbered term.) The terms with even numbers correspond to arcs from Y to X, contained in $\mathcal{A}^* \setminus \tilde{\mathcal{A}}$. In each case, we have included in parentheses the relationship between the prices at the two nodes of the corresponding arc, taken from (7) or (8).

We now examine a sequence along which n grows arbitrarily large, with $\mu^E_{x^*,y^*} > 0$ along this sequence. Suppose first that as n grows, p_{y^*} approaches a limit. This imposes a lower bound on p_{z_1} (from condition (1) in our expansion of $p_{y^*} - p_{x^*}$), which in turn imposes a lower bound on p_{z_2} (from condition (2) in the expansion), and so on, leading to the conclusion that p_{x^*} does not diverge to $-\infty$, a contradiction.

A similar argument can be constructed, beginning from the assumption that p_{x^*} approaches a limit as n grows large, with a similar construction ensuring that it cannot be the case that $p_{y^*} \to +\infty$, giving a contradiction.

It remains to preclude the possibilities that p_{x^*} and p_{y^*} both approach $+\infty$ or both approach $-\infty$. We present the argument for the case in which both approach $+\infty$, with the other case being analogous. If p_{x^*} and p_{y^*} both approach $+\infty$, then we can divide $\mathcal{X}$ into disjoint sets $\underline{\mathcal{X}}$ and $\overline{\mathcal{X}}$ and can divide $\mathcal{Y}$ into disjoint sets $\underline{\mathcal{Y}}$ and $\overline{\mathcal{Y}}$ such that the prices in $\overline{\mathcal{X}}$ and $\overline{\mathcal{Y}}$ diverge and the prices in $\underline{\mathcal{X}}$ and $\underline{\mathcal{Y}}$ do not diverge. The sets $\overline{\mathcal{X}}$ and $\overline{\mathcal{Y}}$ are nonempty, because they contain x^* and y^*. The set $\underline{\mathcal{X}} \cup \underline{\mathcal{Y}}$ is nonempty because it contains the node z whose price we have fixed. There must be no arcs in $\mathcal{A}$ from $\underline{\mathcal{X}}$ to $\overline{\mathcal{Y}}$ (since otherwise such an arc would generate positive payoff, a contradiction). There must exist no flow from nodes in $\overline{\mathcal{X}}$ to nodes in $\underline{\mathcal{Y}}$ (since these arcs earn negative payoff). Any flow from $\underline{\mathcal{X}}$ to $\underline{\mathcal{Y}}$ must occur along arcs in $\tilde{\mathcal{A}}$, since otherwise the arc must earn negative payoff. Hall's condition must then hold for the system whose nodes are $\underline{\mathcal{X}}$ and $\underline{\mathcal{Y}}$, and then must separately hold the system whose nodes are $\overline{\mathcal{X}}$ and $\overline{\mathcal{Y}}$.

We now repeat the previous existence argument separately for each of these systems, with one price $\underline{p}$ fixed at a node in the system whose nodes are $\underline{\mathcal{X}}$ and $\underline{\mathcal{Y}}$ and price $\overline{p}$ fixed for the system whose nodes are $\overline{\mathcal{X}}$ and $\overline{\mathcal{Y}}$. This gives us an equilibrium for each system separately. We cannot be sure that we have an equilibrium of the combined systems, since there may be arcs in $\tilde{\mathcal{A}}$ from $\overline{\mathcal{X}}$ to $\underline{\mathcal{Y}}$. However, if we set $\underline{p}$ sufficiently low and $\overline{p}$ sufficiently high, these arcs will carry no flow, and we will indeed have an equilibrium in each of the disjoint sets. This again either gives a contradiction, or we have prices in the

system whose nodes are $\underline{\mathcal{X}}$ and $\underline{\mathcal{Y}}$ that diverge. In the latter case, we repeat this procedure, until obtaining a contradiction. □

This completes the proof of Lemma 4. □

3.3.3. Closing the Ring

The proof of Theorem 1 is then completed by the following:

Lemma 7. *Let Assumptions 1, 3 and 4 hold. The equilibrium flow problem* $(\mathcal{Z}, \mathcal{A}, G)$ *and quantity* q *satisfy Assumption 2, that* $q(B) \geq 0$ *for all retaining* B*, if and only if the associated bipartite equilibrium flow problem* $(\mathcal{X} \cup \mathcal{Y}, \tilde{\mathcal{A}}, \tilde{G})$ *together with the quantity* q *satisfies Hall's marriage condition.*

Proof. [Only if:] Suppose the equilibrium flow problem $(\mathcal{Z}, \mathcal{A}, G)$ and quantity q are such that $q(B) \geq 0$ for all retaining sets. Suppose $\tilde{\mathcal{X}} \cup \tilde{\mathcal{Y}}$ is a subset of $\mathcal{X} \cup \mathcal{Y}$ for which, in the bipartite equilibrium flow problem $(\mathcal{X} \cup \mathcal{Y}, \tilde{\mathcal{A}}, \tilde{G})$, there are no arcs from nodes in $\tilde{\mathcal{X}}$ to nodes outside $\tilde{\mathcal{Y}}$. To establish Hall's condition, it suffices to show that $q(\tilde{\mathcal{X}} \cup \tilde{\mathcal{Y}}) \geq 0$ for any such set.

In the equilibrium flow problem $(\mathcal{Z}, \mathcal{A}, G)$, define a sequence of sets by letting $\mathcal{Z}_1 = \tilde{\mathcal{X}} \cup \tilde{\mathcal{Y}}$, letting $\mathcal{Z}_2$ be the union of $\mathcal{Z}_1$ and any node that is the endpoint of an arc whose origin lies in $\mathcal{Z}_1$; letting $\mathcal{Z}_3$ be the union of $\mathcal{Z}_2$ and any node that is the endpoint of an arc whose origin lies in $\mathcal{Z}_2$, and so on. This sequence terminates (since the set $\mathcal{Z}$ of nodes is finite) in a set $\mathcal{Z}^*$ that is retaining (otherwise the process would not terminate) and hence by assumption satisfies $q(\mathcal{Z}^*) \geq 0$. Moreover, $\mathcal{Z}^*$ contains no node z for which $q_z > 0$ that is not contained in $\tilde{\mathcal{Y}}$ (since otherwise there would be a link in the problem $(\mathcal{X} \cup \mathcal{Y}, \tilde{\mathcal{A}}, \tilde{G})$ from some node in $\tilde{\mathcal{X}}$ to a node outside $\tilde{\mathcal{Y}}$, a contradiction). Then we have

$$0 \leq q(\mathcal{Z}^*) \leq q(\tilde{\mathcal{X}} \cup \tilde{\mathcal{Y}}),$$

giving Hall's condition.

[If:] For the other direction, let $\tilde{\mathcal{Z}}$ be a retaining set in equilibrium flow problem $(\mathcal{Z}, \mathcal{A}, G)$. Then

$$q(\tilde{\mathcal{Z}}) = q((\mathcal{X} \cap \tilde{\mathcal{Z}}) \cup (\mathcal{Y} \cap \tilde{\mathcal{Z}}))$$

Because $\tilde{\mathcal{Z}}$ is retaining in the equilibrium flow problem $(\mathcal{Z}, \mathcal{A}, G)$, the set $(\mathcal{X} \cap \tilde{\mathcal{Z}}) \cup (\mathcal{Y} \cap \tilde{\mathcal{Z}})$ is retaining in the bipartite equilibrium flow problem $(\mathcal{X} \cup \mathcal{Y}, \tilde{\mathcal{A}}, \tilde{G})$. Hence, Hall's condition gives $q((\mathcal{X} \cap \tilde{\mathcal{Z}}) \cup (\mathcal{Y} \cap \tilde{\mathcal{Z}})) \geq 0$, establishing the needed result that $q(\tilde{\mathcal{Z}}) \geq 0$. □

4. DISCUSSION

Section 1 argued that a number of familiar problems can be formulated as special cases of the equilibrium flow problem. This section offers an illustration.

Consider the following one-to-one matching market. Let $\mathcal{X}$ be a finite set of types of workers and $\mathcal{Y}$ a finite set of types of firms. There are n_x workers of each type $x \in \mathcal{X}$, and m_y firms of each type $y \in \mathcal{Y}$. We can think of n_x as identifying the number of workers of type x, or we can think of n_x as the mass of a continuum of workers of type x in an infinite population, as may be appropriate when modeling a competitive market.

A match involves a worker of type x, a firm of type y, and a wage w_{xy}. Such a match generates utility $\mathcal{U}_{xy}\left(w_{xy}\right)$ for the worker (increasing in w_{xy}) and $\mathcal{V}_{xy}\left(w_{xy}\right)$ for the firm (decreasing in w_{xy}). We allow workers and firms to be unmatched, with an unmatched worker receiving utility $\mathcal{U}_{x0}$ and an unmatched firm receiving utility $\mathcal{V}_{0y}$. A matching is a pair (μ, w), where μ is a vector identifying the mass μ_{xy} of matches between workers of type x and firms of type y, for each $xy \in \mathcal{X} \times \mathcal{Y}$, and w is a vector specifying the wage w_{xy} attached to each pair $xy \in \mathcal{X} \times \mathcal{Y}$.

We can reformulate this matching problem as an example of the equilibrium flow problem. The sets of types of workers and firms become the nodes in the network, augmented by a dummy node used to allow the possibility that a worker or firm remains unmatched. The quantities attached to the nodes represent the quantities of workers and firms of each type. In particular, let

$$\begin{aligned}
\mathcal{Z} &= \mathcal{X} \cup \mathcal{Y} \cup \{0\} \\
\mathcal{A} &= ((\mathcal{X} \cup \{0\}) \times (\mathcal{Y} \cup \{0\})) \setminus \{(0,0)\} \\
q_z &= -n_z, \quad z \in \mathcal{X} \\
q_z &= m_z, \quad z \in \mathcal{Y} \\
q_0 &= \sum_{x \in \mathcal{X}} n_x - \sum_{y \in \mathcal{Y}} m_y.
\end{aligned}$$

There is thus a node for each type of worker, a node for each type of firm, and a dummy node denoted by 0. There is an arc from each worker node to each

firm node as well as to the dummy node. There is an arc from the dummy node to each firm node. Each node corresponding to a type of worker carries a negative quantity equal to the mass of workers of that type, while each node corresponding to a type of firm carries a positive quantity corresponding to the mass of that type. The dummy node carries the remaining mass, which ensures that the total $\sum_{z\in\mathcal{Z}} q_z$ equals zero. Because the quantity vector q identifies the negative of the quantity of each type of worker and the quantity of each type of firm, we can represent a match between a worker and a firm as a flow along the arc connecting the node containing that type of worker to the node representing the firm. In general, a flow along an arc signifies the mass of matches between the types residing at each node connected by the arc.

The price attached to a node identifies the utility of the agent at that node (if a worker) or the negative of the utility of the agent at that node (if a firm). Reminiscent of our treatment of quantities, the payoff vector p identifies the payoffs of workers and the *negative* of the payoffs of firms. An increase in the payoff p_y at firm node y thus corresponds to a smaller utility requirement from the firm at that node, making it more attractive for workers to traverse the arc terminating at the node. This allows us to write connection functions $G_{xy}(p_y)$ that are increasing functions. Formally, the connection function G describes the utility possibilities generated by a match, with

$$\begin{aligned} G_{xy}\left(p_y\right) &= \mathcal{U}_{xy} \circ \mathcal{V}_{xy}^{-1}\left(-p_y\right) \text{ for } x \in \mathcal{X}, y \in \mathcal{Y} \\ G_{x0}\left(p_0\right) &= \mathcal{U}_{x0} + p_0 \text{ for } x \in \mathcal{X} \\ G_{0y}\left(p_y\right) &= \mathcal{V}_{0y} + p_y \text{ for } y \in \mathcal{Y}. \end{aligned}$$

As required, $G_{xy}(p_y)$ is increasing in p_y.

We adopt the normalization $p_0 = 0$.[6] This ensures that if buyer x is unmatched, so that the equilibrium outcome involves flow along the arc $x0$, then the no-positive-payoff equilibrium condition ensures that $p_x = \mathcal{U}_{x0} + p_0 = \mathcal{U}_{x0}$, the payoff specified for an unmatched buyer. Similarly, if seller y is unmatched, then we have $p_0 = 0 = \mathcal{V}_{0y} + p_y$, and hence $-p_y = V_{0y}$.

[6] The existence result in Theorem 1 allows us to normalize the price at any node z for which $q_z \neq 0$. We can thus immediately apply Theorem 1 if the matching market is unbalanced, and hence the dummy node exhibits a nonzero quantity. However, if the market is balanced, then $q_0 = 0$. In this case, we can bring the equilibrium flow problem into the purview of Theorem 1 by introducing an additional buyer type $\hat{x}$ with $q_{\hat{x}} < 0$, with an arc from $\hat{x}$ only to the dummy node, so that this buyer must remain unmatched. To accommodate this additional type, we let the mass on the dummy node equal $-q_{\hat{x}}$ (> 0), and then can normalize $p_0 = 0$.

Remembering that p_y measures the negative of the seller's utility, we again have the seller receiving her unmatched payoff.

Following the standard definition, the matching (μ, w) is stable if

$$\sum_{y \in \mathcal{Y}} \mu_{xy} \leq n_x, \ \sum_{x \in \mathcal{X}} \mu_{xy} \leq m_y,$$

and

$$\begin{aligned}
\mu_{xy} > 0 \quad &\Longrightarrow \quad \left\{ \begin{array}{l} \mathcal{U}_{xy}(w_{xy}) = \max \left\{ \max_{\tilde{y} \in \mathcal{Y}} \mathcal{U}_{x\tilde{y}} \left(w_{x\tilde{y}} \right), \mathcal{U}_{x0} \right\} \\ \mathcal{V}_{xy}(w_{xy}) = \max \left\{ \max_{\tilde{x} \in \mathcal{X}} \mathcal{V}_{\tilde{x}y} \left(w_{\tilde{x}y} \right), \mathcal{V}_{0y} \right\} \end{array} \right\} \\
\sum_{y \in \mathcal{Y}} \mu_{xy} < n_x \quad &\Longrightarrow \quad \mathcal{U}_{x0} = \max \left\{ \max_{\tilde{y} \in \mathcal{Y}} \mathcal{U}_{x\tilde{y}} \left(w_{x\tilde{y}} \right), \mathcal{U}_{x0} \right\} \\
\sum_{x \in \mathcal{X}} \mu_{xy} < m_y \quad &\Longrightarrow \quad \mathcal{V}_{0y} = \max \left\{ \max_{\tilde{x} \in \mathcal{X}} \mathcal{V}_{\tilde{x}y} \left(w_{\tilde{x}y} \right), \mathcal{V}_{0y} \right\}
\end{aligned}$$

The first condition is the feasibility requirement that the quantity of each type of worker and firm that is matched is no more than the total quantity of this type in the market (with excess agents remaining unmatched). The second set of conditions is the stability requirement that no worker and firm can improve their utilities, either by matching with one another at some appropriate wage or by remaining unmatched.

Galichon et al. (2024, Lemma 6) prove the following:

Lemma 8. *The matching (μ, w) is stable in the matching market if and only if the associated outcome (q, μ, p) is an equilibrium outcome of the associated equilibrium flow problem.*

The proof is a relatively straightforward bookkeeping exercise, connecting the requirements for stability to their equivalents in the definition of an equilibrium flow.

To see how this result might be useful, we define the *equilibrium flow correspondence* $\mathcal{Q}$, which associates with each utility vector p the set of vectors q, i.e., the specifications of the sets of firms and workers, for which there exists an equilibrium flow outcome yielding the utilities specified by p. The point of Galichon et al. (2024) is to show that the equilibrium flow correspondence in general satisfies their condition of unified gross substitutes (Theorem 4), and hence has an inverse which is lattice valued (Corollary 1) and increasing (in the strong set order, Theorem 1). This implies, for example, that as the

number of a given type of firms increases, the set of equilibrium payoffs of all firms decreases (in the strong set order) while the set of equilibrium payoffs of all workers increases. An increase in the number of workers has the reverse effects.[7]

More generally, let $\mathcal{Q}(p)$ be a correspondence that maps a vector p, specifying the price attached to each node of the equilibrium flow network, to the set of quantities q with the property that if $q \in \mathcal{Q}(p)$, then there exists an equilibrium outcome (q, μ, p). Again, $\mathcal{Q}$ has an inverse which is lattice valued and increasing. In a hedonic pricing context, this implies that adding a consumer to the market increases (again, at least weakly, in the strong set order) the set of equilibrium prices received by all sellers in the market. In a time-dependent routing context, it implies that adding a new supplier to the production network accelerates the delivery times attained by all final users. We thus view the equilibrium flow problem as a potentially useful tool for characterizing equilibrium outcomes in a variety of settings.

References

Bellman, R. (1958). On a routing problem. *Quarterly of Applied Mathematics*, *16*(1), 87-90.

Chiappori, P.-A., McCann, R. J., & Nesheim, L. P. (2010). Hedonic price equilibria, stable matching, and optimal transport: Equivalence, topology and uniqueness. *Economic Theory*, *42*, 317-354.

Crawford, V. P., & Knoer, E. M. (1981). Job matching with heterogeneous firms and workers. *Econometrica*, *49*(2), 437-450.

Demange, G., & Gale, D. (1985). The strategy structure of two-sided matching markets. *Econometrica*, *53*(4), 873-888.

Echenique, F., Lee, S., & Shum, M. (2011). The money pump as a measure of revealed preference violations. *Journal of Political Economy*, *119*(6), 1201-1223.

Ekeland, I., Heckman, J., & Nesheim, L. P. (2004). Identification and estimation of hedonic models. *Journal of Political Economy*, *112*(1), 60-109.

Gale, D., & Shapley, L. (1962). College admissions and the stability of marriage. *American Mathematical Monthly*, *69*, 9-15.

[7] Demange & Gale (1985, Lemma 2 and Property 2) establish similar results for a model in which the quantities attached to the various nodes must take integer values, which we do not require. Dropping this requirement is useful in allowing one to accommodate a continuum of each type of agent, as is common in models of competitive matching (cf. Nöldeke & Samuelson 2024).

Galichon, A. (2018). *Optimal Transport Methods in Economics*. Princeton: Princeton University Press.

Galichon, A., Kominers, S. D., & Weber, S. (2015). The nonlinear Bernstein-Schrödinger equation in economics. In F. Nielsen & F. Barbaresco (Eds.), *Geometric science of information* (p. 51-59). New York: Springer.

Galichon, A., Kominers, S. D., & Weber, S. (2019). Costly concessions: An empirical framework for matching with imperfectly transferable utility. *Journal of Political Economy*, *127*(6), 2875-2925.

Galichon, A., Samuelson, L., & Vernet, L. (2024). Unified gross substitutes and inverse isotonicity for equilibrium problems. *Forthcoming in Theoretical Economics*.

Gendreau, M., Ghiani, G., & Guerriero, E. (2015). Time-dependent routing problems: A review. *Computers & Operations Research*, *64*, 189-197.

Hall, P. (1935). On representatives of subsets. *Journal of the London Mathematical Society*, *10*(1), 26-30.

Hoffman, A. J. (1960). Some recent applications of the theory of linear inequalities to extremal combinatorial analysis. In *Proceedings of symposia in applied mathematics* (p. 113-127).

Kelso, A. S., & Crawford, V. P. (1982). Job matching, coalition formation, and gross substitutes. *Econometrica*, *50*(6), 1483-1504.

Nöldeke, G., & Samuelson, L. (2018). The implementation duality. *Econometrica*, *86*(4), 1283-1384.

Nöldeke, G., & Samuelson, L. (2024). Investment and competitive matching. In Y.-K. Che, P. Chiappori, & B. Salanie (Eds.), *Handbook of Matching*. New York, Forthcoming: North Holland.

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

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, *82*(1), 34-55.

Journal of Mechanism and Institution Design
ISSN: 2399-844X (Print), 2399-8458 (Online)
DOI:10.22574/jmid.2024.12.005

RESPECT FOR IMPROVEMENTS AND COMPARATIVE STATICS IN MATCHING MARKETS

Scott Duke Kominers

Harvard University & a16z crypto, USA
kominers@fas.harvard.edu

ABSTRACT

One of the oldest results in the theory of two-sided matching is the *entry comparative static*, which shows that under the Gale–Shapley deferred acceptance algorithm, adding a new agent to one side of the market makes all the agents on the other side weakly better off. Here, we give a new proof of the entry comparative static, by way of a well-known property of deferred acceptance called *respect for improvements*. Our argument extends to yield comparative static results in more general settings, such as matching with slot-specific preferences.

Keywords: Matching, market entry, respect for improvements.

JEL Classification Numbers: C78, D47, D82.

I want to dedicate this article to Vincent Crawford – an extraordinary mentor, collaborator, and friend, who I learn from continuously (and also discretely). I appreciate the helpful comments of Péter Biró, Rachel Bayefsky, Vincent Crawford, Zoe DeStories, Piotr Dworczak, Federico Echenique, John William Hatfield, Emir Kamenica, Shengwu Li, Paul Milgrom, Michael Ostrovsky, Alvin Roth, Larry Samuelson, Tayfun Sönmez, Alexander Teytelboym, Bertan Turhan, Zaifu Yang, Bumin Yenmez, and especially Jiafeng Chen, Yunseo Choi, Jo Ellery, Ravi Jagadeesan, Shira Li, Jonathan Ma, Franklyn Wang, and other members of the Lab for Economic Design. I gratefully acknowledge the support of National Science Foundation grant SES-1459912 and the Ng Fund and the Mathematics in Economics Research Fund of the Harvard Center of Mathematical Sciences and Applications. Part of this work was conducted during the Simons Laufer Mathematical Sciences Institute Fall 2023 program on the Mathematics and Computer Science of Market and Mechanism Design, which was supported by the National Science Foundation under Grant No. DMS-1928930 and by the Alfred P. Sloan Foundation under grant G-2021-16778.

1. INTRODUCTION

In their seminal 1962 paper, Gale & Shapley introduced the two-sided matching problem: We are given sets of men and women, each with preferences over members of the opposite gender; we seek a *stable matching*, i.e., an assignment of partners such that no one finds his or her partner unacceptable, and no two agents mutually prefer each other to their assigned partners.

Gale & Shapley (1962) showed that stable matchings can be found via the following *deferred acceptance* algorithm, under which men propose marriage in sequence, and women defer accepting prospective partners until the full sequence of proposals has been completed.

> DEFERRED ACCEPTANCE ALGORITHM
>
> **Step** 1**.** Each man proposes to his first-choice woman. Each woman *"holds"* her best acceptable proposal (if any) and rejects all other proposals.
>
> **Step** $\ell \geq 2$**.** Each man who was rejected in the previous step proposes to his most-preferred woman to whom he has not yet proposed (if any). Each woman holds her best acceptable proposal among those made in this step and held from the previous step (if any) and rejects all others.
>
> If at any time no men are available to make proposals—that is, if all men not currently held have proposed to all women they find acceptable—then the algorithm terminates. The *(man-proposing) deferred acceptance outcome* is the matching that matches each woman to the man whose proposal she is holding (if any) at the end of the last step before the algorithm terminates.

Variants of deferred acceptance are by now applied in a number of real-world contexts, including medical residency matching (Roth 1984; Roth & Peranson 1999), school choice (Balinski & Sönmez 1999; Abdulkadiroğlu & Sönmez 2003; Pathak 2011), and the assignment of cadets to branches of military service (Sönmez & Switzer 2013; Sönmez 2013; Greenberg et al. 2024).[1]

[1] For surveys of this and related work in *market design*, see, e.g., Roth (2008, 2013, 2015) and Kominers et al. (2017).

One of the most classic results in the theory of stable matching is an *entry comparative static* that characterizes how the deferred acceptance outcome changes when a new woman is added to the market (Kelso & Crawford 1982; Gale & Sotomayor 1985; Roth & Sotomayor 1990; Crawford 1991). The exact form of the comparative static varies across settings, but the core result is that *under deferred acceptance, adding a new agent to one side of the market makes all agents on the other side of the market weakly better off* (see, e.g.,Blum et al. 1997; Blum & Rothblum 2002; Hatfield & Milgrom 2005; Biró et al. 2008; Ostrovsky 2008; Hatfield & Kominers 2013; Chambers & Yenmez 2017; Yenmez 2018; Fleiner et al. 2018).[2]

Proofs of the entry comparative static and its generalizations tend to rely either on meticulous inductive arguments (Kelso & Crawford 1982; Crawford 1991; Blum et al. 1997; Blum & Rothblum 2002; Hatfield & Milgrom 2005; Biró et al. 2008; Hatfield & Kominers 2013; Fleiner et al. 2018) or on broad results characterizing the structure of the set of stable matchings (Gale & Sotomayor 1985; Ostrovsky 2008; Chambers & Yenmez 2017).[3]

In this paper, we give a new, more concise proof of the comparative static by way of another classic result—specifically, the fact that deferred acceptance *respects (unambiguous) improvements* in the sense that making one agent more highly ranked in other agents' preferences improves that agent's deferred acceptance outcome (Balinski & Sönmez 1999; Choi et al. 2024).[4] We show how to transform the entry of a new agent on one side of the market into a simultaneous preference rank improvement for the agents on the other side; the comparative static result then follows directly from respect for improvements.

As we describe, our approach extends naturally to yield novel comparative static results in more general settings, such as the "slot-specific preferences" framework that Kominers & Sönmez (2016) introduced to model matching with diversity constraints. In particular, we show that expanding the capacity

[2] As we discuss in Section 6, we typically also obtain a sort of dual to this result: agents on the same side of the market as the entering agent are made weakly worse off. To our knowledge, however, the methods we use here only enable us to derive the comparative static for agents on the side of the market opposite to that of the entering agent.

[3] Additionally, Choi (2022) used the logical compactness–based methods introduced by Gonczarowski et al. (2024) to establish a version of the entry comparative static for markets with infinitely many agents.

[4] See also the work of Wang & Kominers (2020), which uses the theory of matching with contracts to show a particularly strong link between respect for improvements and strategy-proofness.

of a firm that has slot-specific preferences weakly improves employees' match outcomes. Then, again in the slot-specific preference setting, we obtain what—to the best of our knowledge—are the first comparative statics for how the set of stable matchings changes when new contracts with an agent become possible.

2. AN ILLUSTRATIVE EXAMPLE

We consider a market with three men $\{\mathsf{Alex}, \mathsf{Bob}, \mathsf{Charlie}\}$ and two women $\{\mathsf{Xena}, \mathsf{Yvette}\}$, who rank each other as follows (with the option $\emptyset$ representing being unmatched):

$$\begin{aligned} \mathsf{Alex} &: \mathsf{Yvette} \succ \mathsf{Xena} \succ \emptyset & \mathsf{Xena} &: \mathsf{Alex} \succ \mathsf{Bob} \succ \mathsf{Charlie} \succ \emptyset \\ \mathsf{Bob} &: \mathsf{Yvette} \succ \mathsf{Xena} \succ \emptyset & \mathsf{Yvette} &: \mathsf{Alex} \succ \mathsf{Bob} \succ \emptyset. \\ \mathsf{Charlie} &: \mathsf{Xena} \succ \mathsf{Yvette} \succ \emptyset \end{aligned} \tag{1}$$

If we run deferred acceptance on our example, we see that first Alex and Bob propose to Yvette, while Charlie proposes to Xena; then Yvette and Xena hold Alex's and Charlie's proposals, respectively. Thus Bob is rejected; he proposes to Xena in the second step, causing Xena to hold Bob and reject Charlie. Finally, Charlie proposes to Yvette but is rejected, leaving us with an outcome μ in which Alex is matched to Yvette and Bob is matched to Xena. By construction, this matching μ is stable—everyone who is matched has an acceptable partner, and no man and woman mutually prefer each other to their assigned partners.

Now, we imagine that a new woman, Zelda, is added to the market we considered earlier, and that Zelda has the same preferences over men as Yvette. We suppose furthermore that Zelda is especially popular[5] so that preferences are now given by

$$\begin{aligned} \mathsf{Alex} &: \mathsf{Zelda} \succ \mathsf{Yvette} \succ \mathsf{Xena} \succ \emptyset & \mathsf{Xena} &: \mathsf{Alex} \succ \mathsf{Bob} \succ \mathsf{Charlie} \succ \emptyset \\ \mathsf{Bob} &: \mathsf{Zelda} \succ \mathsf{Yvette} \succ \mathsf{Xena} \succ \emptyset & \mathsf{Yvette} &: \mathsf{Alex} \succ \mathsf{Bob} \succ \emptyset \\ \mathsf{Charlie} &: \mathsf{Zelda} \succ \mathsf{Xena} \succ \mathsf{Yvette} \succ \emptyset & \mathsf{Zelda} &: \mathsf{Alex} \succ \mathsf{Bob} \succ \emptyset. \end{aligned} \tag{2}$$

Deferred acceptance now produces a matching $\tilde{\mu}$ under which Alex, Bob, and Charlie are matched to Zelda, Yvette, and Xena, respectively. Under $\tilde{\mu}$

[5] Perhaps she has an entire videogame franchise to her name.

every man is matched to a partner he weakly prefers to his partner under μ; the entry comparative static implies that this sort of improvement should arise in general.

Now, to illustrate our approach to the comparative static, we recall that deferred acceptance *respects improvements* in the sense that making one agent more highly ranked in other agents' preferences improves that agent's deferred acceptance outcome. We show in particular that the entry of a new woman such as Zelda can be modeled as a (weak) increase in the preference rank of every man in the market. Indeed, we could have written our original market (1) in the form

$$\begin{array}{rlrl} \text{Alex} : & \text{Zelda} \succ \text{Yvette} \succ \text{Xena} \succ \emptyset & \text{Xena} : & \text{Alex} \succ \text{Bob} \succ \text{Charlie} \succ \emptyset \\ \text{Bob} : & \text{Zelda} \succ \text{Yvette} \succ \text{Xena} \succ \emptyset & \text{Yvette} : & \text{Alex} \succ \text{Bob} \succ \emptyset \\ \text{Charlie} : & \text{Zelda} \succ \text{Xena} \succ \text{Yvette} \succ \emptyset & \text{Zelda} : & \emptyset \succ \text{Alex} \succ \text{Bob}, \end{array} \tag{3}$$

with Zelda treated as if she were already present, but found all men unacceptable. (Note that in (3) we have given Zelda a ranking of unacceptable men consistent with her preferences over men in (2), even though she will end up assigned to "$\emptyset$" rather than to any of them.) Now, the deferred acceptance outcome under (3) corresponds directly to that under (1): Alex is matched to Yvette and Bob is matched to Xena, just as under μ. Yet the "entry" of Zelda obtained by transforming (3) to (2) represents a ranking improvement for all the men, since each man is now more highly ranked relative to $\emptyset$. This implies by the respect for improvements result that all the men's deferred acceptance outcomes must weakly improve—precisely the claim of the entry comparative static.

We show in the sequel that the argument just described can be formalized to prove the full entry comparative static for the marriage model. Moreover, the argument extends directly to yield comparative static results in more general settings.

3. THE MARRIAGE MODEL

We start by introducing the marriage model of Gale & Shapley (1962): There are finite sets M and W of *men* and *women*; we denote by $I \equiv M \cup W$ the set of *agents*. We assume that each man $m \in M$ has a complete, transitive, and strict *preference ordering* $\succ_m$ over $W \cup \{\emptyset\}$, where $\emptyset$ denotes an *outside*

option that represents the possibility of remaining unmatched. Similarly, each woman $w \in W$ has a complete, transitive, and strict *preference ordering* $\succ_w$ over $M \cup \{\emptyset\}$. For each agent $i \in I$, we denote the weak part of i's preferences by $\succcurlyeq_i$, so that if $j \succcurlyeq_i k$ then either $j \succ_i k$ or $j = k$. We use the convention that $\succ_{I'} \equiv (\succ_i)_{i \in I'}$; analogously, we write $\succcurlyeq_{I'} \equiv (\succcurlyeq_i)_{i \in I'}$.

3.1. Stable Matchings

A *(marriage) matching* is a map $\mu : I \to I \cup \{\emptyset\}$ from the set of agents to the set of agents plus the outside option, such that:

1. Under μ, each man is assigned to a woman or to the outside option—that is, $\mu(m) \in (W \cup \{\emptyset\})$ for each $m \in M$—and each woman is assigned to a man or to the outside option—$\mu(w) \in (M \cup \{\emptyset\})$ for each $w \in W$.

2. If a man $m \in M$ is assigned to a woman $w \in W$, then w is assigned to m—i.e., if $\mu(m) = w \in W$, then $\mu(w) = m$—and vice versa—if $\mu(w) = m \in M$, then $\mu(m) = w$.

We say that agent $i \in I$ is *matched to* $j \in I$ *under* μ if $\mu(i) = j$; we say that $i \in I$ is *unmatched under* μ if $\mu(i) = \emptyset$.

A matching μ is *individually rational* if no agent prefers the outside option to his or her assigned partner, i.e., if for all $i \in I$, we have that $\mu(i) \succcurlyeq_i \emptyset$. A matching μ is *unblocked* if there does not exist a man $m \in M$ and woman $w \in W$ who mutually prefer each other to their assigned match partners, i.e., such that $w \succ_m \mu(m)$ and $m \succ_w \mu(w)$. A matching μ is *stable* if it is both individually rational and unblocked.

The main result of Gale & Shapley (1962) shows that the deferred acceptance algorithm introduced in the Introduction always produces a stable outcome.

4. THE CLASSIC COMPARATIVE STATIC

Now, we show that the entry of a new woman $\tilde{w}$ to the market must make all the men weakly better off under (man-proposing) deferred acceptance. Abusing notation slightly, we extend our model to the set of agents $I \cup \{\tilde{w}\} = M \cup W \cup \{\tilde{w}\}$, writing $\tilde{\succ}$ and $\tilde{\mu}$, respectively, for preferences and matchings in

the expanded market. To reflect the idea that $\tilde{w}$ is a new entrant to an existing market, we require that the preference profile $\tilde{\succ}$ be consistent with $\succ$ on I:

$$\text{for all } i,j,k \in I, \text{ we have } j \tilde{\succ}_i k \iff j \succ_i k;$$

equivalently, for each man $m \in M$, the expanded preferences $\tilde{\succ}_m$ correspond to $\succ_m$ with $\tilde{w}$ added in somewhere, and for each woman $w \in W$, we have $\tilde{\succ}_w = \succ_w$.

Theorem 1 (Kelso & Crawford 1982; Gale & Sotomayor 1985; Roth & Sotomayor 1990)**.** *If μ^* is the outcome of man-proposing deferred acceptance in the market I and $\tilde{\mu}^*$ is the outcome of man-proposing deferred acceptance in the market $I \cup \{\tilde{w}\}$ that arises after the entry of woman $\tilde{w}$, then each man $m \in M$ (weakly) prefers his assignment under $\tilde{\mu}^*$ to his assignment under μ^*; that is,*

$$\tilde{\mu}^*(m) \tilde{\succcurlyeq}_m \mu^*(m).^{6} \tag{4}$$

One can prove Theorem 1 by explicitly tracking how the presence of $\tilde{w}$ affects each step of deferred acceptance (see, e.g., Crawford 1991; Chambers & Yenmez 2017). Another approach to Theorem 1 starts with the outcome of man-proposing deferred acceptance in the original market and then studies how the market re-equilibrates after $\tilde{w}$ enters (see, e.g., Kelso & Crawford 1982; Blum et al. 1997).[7]

Here, we give a novel proof of Theorem 1 that avoids the need to explicitly track either the algorithm or the path of adjustment. The key to our approach is a core property of deferred acceptance called *respect for (unambiguous) improvements*.

4.1. Respect for Improvements

For $w \in W$ and preference relations $\hat{\succ}_w$ and $\check{\succ}_w$, we say that $\hat{\succ}_w$ *unambiguously improves upon* $\check{\succ}_w$ *for* $m \in M$ if m is ranked (weakly) higher under $\hat{\succ}_w$ than under $\check{\succ}_w$; the relative rankings of all other men are left unchanged; and no man becomes unacceptable where he was previously acceptable. Formally:

[6] Theorem 1 is also a special case of the comparative static results of Crawford (1991); Hatfield & Milgrom (2005); Ostrovsky (2008); Hatfield & Kominers (2013); Chambers & Yenmez (2017); Yenmez (2018); Fleiner et al. (2018).

[7] Dworczak (2021) tracks similar re-equilibration dynamics in his *Deferred Acceptance with Compensation Chains* algorithms, which generalize deferred acceptance by allowing agents on both sides of the market to make proposals.

Definition 1. We say that $\hat{\succ}_w$ is an *unambiguous improvement over* $\check{\succ}_w$ *for* $m \in M$ if

- for all $m' \in ((M \setminus \{m\}) \cup \{\emptyset\})$, if $m \check{\succ}_w m'$ then $m \hat{\succ}_w m'$;
- for all $m', m'' \in (M \setminus \{m\})$, we have $m' \check{\succ}_w m''$ if and only if $m' \hat{\succ}_w m''$; and
- for all $m' \in (M \setminus \{m\})$, if $m' \check{\succ}_w \emptyset$ then $m' \hat{\succ}_w \emptyset$.

We say that a profile of the women's preferences $\hat{\succ}_W$ is an *unambiguous improvement over* $\check{\succ}_W$ *for* m if each $\hat{\succ}_w$ is an unambiguous improvement over $\check{\succ}_w$ for m.

Balinski & Sönmez (1999) showed that deferred acceptance *respects (unambiguous) improvements* in the sense that if $\hat{\succ}_W$ is an unambiguous improvement over $\check{\succ}_W$ for m, then m (weakly) prefers his man-proposing deferred acceptance outcome when the women's preferences are $\hat{\succ}_W$ to that when the women's preferences are $\check{\succ}_W$.[8]

Lemma 1. *If* $\hat{\succ}_W$ *is an unambiguous improvement over* $\check{\succ}_W$ *for* m*, and* $\hat{\mu}^*$ *and* $\check{\mu}^*$ *denote the man-proposing deferred acceptance outcomes under* $(\succ_M, \hat{\succ}_W)$ *and* $(\succ_M, \check{\succ}_W)$*, respectively, then* m *(weakly) prefers his or her assignment under* $\hat{\mu}^*$ *to that under* $\check{\mu}^*$*; that is,*

$$\hat{\mu}^*(m) \succcurlyeq_m \check{\mu}^*(m).$$

Respect for unambiguous improvements is an extremely natural property to desire in practice, as it means that agents have no incentive to try to lower their standing in others' preference relations.[9,10] But the respect for improvements condition is also subtle: If a man's rank just improves relative to the outside option $\emptyset$, then the second and third conditions of Definition 1 hold automatically. Thus a (strict) unambiguous improvement for one man can also

[8] As discussed by Choi et al. (2024), the form of the respect for unambiguous improvements condition stated in Definition 1 is slightly weaker than the version used by Balinski & Sönmez (1999); as a result, the statement of Lemma 1 is slightly stronger than the Balinski & Sönmez (1999) version of the respect for improvements result. But the essence of the result is the same—and formally, our Lemma 1 follows from Lemma 2, which in turn comes from Choi et al. (2024).

[9] Imagine the chaos if intentional self-sabotage were a standard feature of the marriage market!

[10] Moreover, respect for unambiguous improvements in some sense uniquely characterizes deferred acceptance (see Theorems 5 and 6 of Balinski & Sönmez 1999).

be a (strict) unambiguous improvement for some other man, so long as both men's positions only improve relative to the outside option $\emptyset$; this property allows us to derive an elegant and efficient proof of Theorem 1.[11]

4.2. Proof of Theorem 1

As in the statement of Theorem 1, we suppose that μ^* is the outcome of man-proposing deferred acceptance in the market I and $\tilde{\mu}^*$ is the outcome of man-proposing deferred acceptance in the market $I \cup \{\tilde{w}\}$.

We write the preferences of $\tilde{w}$ in the form

$$\succ_{\tilde{w}} : m_1 \succ_{\tilde{w}} \cdots \succ_{\tilde{w}} m_\ell \succ_{\tilde{w}} \emptyset \succ_{\tilde{w}} m_{\ell+1} \cdots \tag{5}$$

and let $\succ_{\tilde{w}}$ be the "null" preference relation for $\tilde{w}$ that is consistent with $\succ_{\tilde{w}}$ but treats all men as unacceptable:

$$\succ_{\tilde{w}} : \emptyset \succ_{\tilde{w}} m_1 \succ_{\tilde{w}} \cdots \succ_{\tilde{w}} m_\ell \succ_{\tilde{w}} m_{\ell+1} \cdots . \tag{6}$$

We let $\bar{\mu}^*$ be the outcome of man-proposing deferred acceptance in the market with agents $I \cup \{\tilde{w}\}$ and preference profile $(\succ_M, \succ_W, \succ_{\tilde{w}})$. As $\succ_W = \succ_W$ and each $\succ_m$ agrees with $\succ_m$ except for the inclusion of $\tilde{w}$, we see directly that

$$\bar{\mu}^*(i) = \mu^*(i) \tag{7}$$

for each $i \in I$. Indeed, as $\tilde{w}$ treats all men as unacceptable under $\succ_{\tilde{w}}$, deferred acceptance's outcome is unchanged if all men drop $\tilde{w}$ from their preference relations—and with that preference adjustment, the algorithm corresponds exactly to deferred acceptance in the market I, and so must yield an outcome that corresponds to μ^*.

Now, we note by inspecting (5) and (6) that $\succ_{\tilde{w}}$ is an unambiguous improvement over $\succ_{\tilde{w}}$ for each man $m \in M$. Hence, by Lemma 1, we see that each man $m \in M$ (weakly) prefers his assignment under $\tilde{\mu}^*$ to that under $\bar{\mu}^*$; that is,

$$\tilde{\mu}^*(m) \succcurlyeq_m \bar{\mu}^*(m) \tag{8}$$

for each man $m \in M$. Combining (7) and (8) shows the desired result, (4).

[11] As this intuition suggests, the argument does not require the full strength of the respect for improvements result—we use only respect for improvements relative to the outside option $\emptyset$.

4.3. Discussion

Our proof of Theorem 1 highlights that the key feature underlying Theorem 1 is that $\tilde{w}$'s entry simultaneously improves all the men's overall standing in the market by increasing competition among the women. This idea is closely related to a much earlier argument due to Gale & Sotomayor (1985). Like us, Gale & Sotomayor (1985) treated $\tilde{w}$'s entry as equivalent to transforming $\tilde{w}$'s preferences from a null relation to $\succ_{\tilde{w}}$. Unlike in our argument, however, Gale & Sotomayor (1985) derived the comparative static through appeal to a result showing that the set of stable outcomes has a lattice structure (see Roth & Sotomayor 1990, pp. 43–44). Approaching the argument in terms of respect for improvements instead is useful because it immediately suggests substantial generalizations, some of which are outside the scope of the lattice structure result—as we describe next.

5. A MORE GENERAL MODEL: MATCHING WITH SLOT-SPECIFIC PREFERENCES

We now extend Theorem 1 to a setting with *many-to-one matching*—that is, one in which employees on one side of the market may take multiple partners on the other side (Gale & Shapley 1962; Roth 1985; Roth & Sotomayor 1990). We furthermore generalize by allowing the matching process to determine not just who matches with whom but also *contracts* that specify terms of exchange like wages or hours worked (Crawford & Knoer 1981; Kelso & Crawford 1982; Hatfield & Milgrom 2005). Specifically, we work with the *slot-specific preference* structure introduced by Kominers & Sönmez (2016).

5.1. Intuition

The slot-specific preferences framework, which we describe formally in the next section, is a model of employee–firm matching in which each firm has a set of positions—*slots*—that can be assigned to different employees. Slots have their own (potentially independent) rankings over contracts. Within each firm, a linear order called the *order of precedence* determines the order in which slots are filled.

For example, we might imagine a firm with two slots, s^1 and s^2, one of which can be used to hire any worker at either a high (£$_H$) or low (£$_L$) salary,

while the second slot can only be used to hire workers at a lower salary:

$$s^1 : (i, £_L) \succ (i, £_H) \succ (j, £_L) \succ (j, £_H) \succ \cdots$$
$$s^2 : (i, £_L) \succ (j, £_L) \succ \cdots .$$

One special case of slot-specific preferences is when we make all the slots within a given firm identical, making that firm's preferences consistent with a single linear order over contracts and a maximum number of positions that can be filled (the *responsive preference* model of Roth 1985). More broadly, slot-specific preferences embed multiple models of affirmative action, such as using some slots to reserve positions for members of disadvantaged groups (Abdulkadiroğlu 2005; Kojima 2012; Hafalir et al. 2013; Kominers & Sönmez 2014; Dur et al. 2018, 2020).

5.2. Formal Model

We suppose that there is a set of *employees* E and a set of *firms* F, and a (finite) set of *contracts* X. Each contract $x \in X$ is between an employee $\mathsf{e}(x) \in E$ and firm $\mathsf{f}(x) \in F$, and may also specify additional "terms" of exchange drawn from a set T. Thus X may be considered a subset of $E \times F \times T$. We extend the notations $\mathsf{e}(\cdot)$ and $\mathsf{f}(\cdot)$ to sets of contracts $Y \subseteq X$ by setting $\mathsf{e}(Y) \equiv \cup_{y \in Y} \{\mathsf{e}(y)\}$ and $\mathsf{f}(Y) \equiv \cup_{y \in Y} \{\mathsf{f}(y)\}$. For $Y \subseteq X$, we denote $Y_e \equiv \{y \in Y : \mathsf{e}(y) = e\}$ and $Y_{E'} \equiv \cup_{e \in E'} Y_e$; analogously, we denote $Y_f \equiv \{y \in Y : \mathsf{f}(y) = f\}$ and $Y_{F'} \equiv \cup_{f \in F'} Y_f$.

Each employee $e \in E$ has a complete, transitive, and strict preference order $\succ_e$ (with weak order $\succcurlyeq_e$) over contracts in $X_e \cup \{\emptyset\}$, where, as before, $\emptyset$ is an outside option that represents remaining unmatched; we use the convention that $\emptyset \succ_e x$ for all $x \in X \setminus X_e$. We say that the contracts $x \in X$ for which $\emptyset \succ_e x$ are *unacceptable to* e. We denote the profile of all employees' preferences by $\succ_E$.

Each firm $f \in F$ has a set S_f of *slots*; each slot can be assigned up to one contract in X_f. Slots $s \in S_f$ have (linear) preference orders $\succ_s$ (with weak orders $\succcurlyeq_s$) over contracts in X_f. As with employees, we assume that each slot $s \in S_f$ ranks an outside option $\emptyset$ that represents remaining unassigned, and as with employees, we use the convention that $\emptyset \succ_s x$ if $x \in X \setminus X_f$. We set $S \equiv \cup_{f \in F} S_f$ and denote the profile of all slots' preferences by $\succ_S$.

Employees have *unit demand*, that is, they choose at most one contract from a set of contract offers. We assume also that employees always choose

the best available contract, so that the choice $C^e(Y)$ of an employee $e \in E$ from contract set $Y \subseteq X$ is the $\succ_e$-maximal element of Y_e (or the outside option if $\emptyset \succ_e y$ for all $y \in Y_e$).[12]

Meanwhile, firms $f \in F$ may be assigned as many as $q_f \equiv |S_f|$ contracts—one for each slot in S_f—but may hold no more than one contract with a given employee. We assume that for each $f \in F$, the slots in S_f are ordered according to a (linear) *order of precedence* $\rhd^f$. We denote $S_f \equiv \left\{s_f^1, \ldots, s_f^{q_f}\right\}$ with the understanding that $s_f^\ell \rhd^f s_f^{\ell+1}$ unless otherwise noted. The interpretation of $\rhd^f$ is that if $s \rhd^f s'$ then—whenever possible—firm f fills slot s before filling s'.

Formally, the choice $C^f(Y)$ of a firm $f \in F$ from contract set $Y \subseteq X$ is defined as follows:

- First, slot s_f^1 is assigned the contract x^1 that is $\succ_{s_f^1}$-maximal among contracts in Y.
- Then, slot s_f^2 is assigned the contract x^2 that is $\succ_{s_f^2}$-maximal among contracts in the set $Y \setminus Y_{\mathsf{e}(x^1)}$ of contracts in Y with employees other than $\mathsf{e}(x^1)$.
- This process continues in sequence, with each slot s_f^ℓ being assigned the contract x^ℓ that is $\succ_{s_f^\ell}$-maximal among contracts in the set

$$Y \setminus Y_{\mathsf{e}(\{x^1,\ldots,x^{\ell-1}\})}.$$

If no contract $x \in Y$ is assigned to slot $s_f^\ell \in S_f$ in the computation of $C^f(Y)$, then s_f^ℓ is assigned the null contract $\emptyset$.

5.3. Stable Outcomes

An *outcome* is a set of contracts $Y \subseteq X$ that is "feasible" in the sense that

- Y contains at most one contract for each employee, i.e., $|Y_e| \leq 1$ for each $e \in E$, and
- Y contains at most q_f contracts for each firm f, i.e., $|Y_e| \leq q_f$ for each $f \in F$.

[12] To simplify our exposition and notation, we treat individual contracts as interchangeable with singleton contract sets.

We say that an outcome Y is *stable* if it is

1. *individually rational*—$C^e(Y) = Y_e$ for all $e \in E$ and $C^f(Y) = Y_f$ for all $f \in F$—and
2. *unblocked*—there does not exist a firm $f \in F$ and *blocking set* $Z \neq C^f(Y)$ such that $Z = C^f(Y \cup Z)$ and $Z_e = C^e(Y \cup Z)$ for all $e \in \mathsf{e}(Z)$.

Note that if all employee–firm pairs may contract and the set T of contractual terms is trivial, then $X = E \times F$. If moreover all workers and firms have unit demand (i.e., if we always have $|C^e(Y)| \leq 1$ and $|C^f(Y)| \leq 1$), then we recover the marriage model from Section 3, where the man–woman pairs matched under some matching μ correspond to the pairs contained in an outcome Y.

5.4. Generalized Deferred Acceptance

Kominers & Sönmez (2016) showed that stable outcomes exist under slot-specific preferences, and can be found via the following *(employee-proposing) cumulative offer process* (Kelso & Crawford 1982; Hatfield & Milgrom 2005), which generalizes deferred acceptance:[13]

CUMULATIVE OFFER PROCESS

Step 1. Some employee $e^1 \in E$ proposes their most-preferred acceptable contract, $x^1 \in X_{e^1}$. Firm $\mathsf{f}(x^1)$ *holds* x^1 if $x^1 \in C^{\mathsf{f}(x^1)}(\{x^1\})$, and *rejects* x^1 otherwise. Set $A^2_{\mathsf{f}(x^1)} = \{x^1\}$, and set $A^2_{f'} = \varnothing$ for each $f' \neq \mathsf{f}(x^1)$; these are the sets of contracts *available* to firms at the beginning of Step 2.

Step $\ell \geq 2$. Some employee $e^\ell \in E$ for whom no contract is currently held by any firm proposes their most-preferred acceptable contract that has not yet been rejected, $x^\ell \in X_{e^\ell}$.

[13] The cumulative offer process generalizes deferred acceptance by (in principle) allowing firms to hold contracts they had rejected in earlier steps—although under slot-specific preferences, firms never actually use this extra degree of freedom (see, e.g., Hatfield et al. 2021). For consistency with Kominers & Sönmez (2016), we state the cumulative offer process with a single agent proposing in each step; Hirata & Kasuya (2014) have proven that in our setting this formulation is equivalent to one in which employees propose simultaneously, directly generalizing the version of deferred acceptance we presented in the Introduction.

> Firm $\mathsf{f}(x^\ell)$ holds the contracts in $C^{\mathsf{f}(x^\ell)}(A^\ell_{\mathsf{f}(x^\ell)} \cup \{x^\ell\})$ and rejects all other contracts in $A^\ell_{\mathsf{f}(x^\ell)} \cup \{x^\ell\}$; firms $f' \neq \mathsf{f}(x^\ell)$ continue to hold all contracts they held at the end of Step $\ell - 1$. Set $A^{\ell+1}_{\mathsf{f}(x^\ell)} = A^\ell_{\mathsf{f}(x^\ell)} \cup \{x^\ell\}$, and set $A^{\ell+1}_{f'} = A^\ell_{f'}$ for each $f' \neq \mathsf{f}(x^\ell)$.
>
> If at any time no employee is able to propose a new contract—that is, if all employees for whom no contracts are on hold have proposed all the contracts they find acceptable—then the algorithm terminates. The *outcome of the (employee-proposing) cumulative offer process* is the set of contracts held by firms at the end of the last step before the algorithm terminates.[14]

Although stable outcomes exist under slot-specific preferences, the set of stable outcomes under slot-specific preferences does not have lattice structure (see Kominers & Sönmez 2016; Hatfield & Kominers 2020). Consequently the approach Gale & Sotomayor (1985) used to prove Theorem 1 does not carry over to the slot-specific preference setting. Nevertheless, as we show next, our approach based on respect for improvements extends directly.

5.5. Respect for Improvements Under Slot-Specific Preferences

Following Kominers & Sönmez (2016) and Choi et al. (2024), $\hat{\succ}_S$ *is an unambiguous improvement over preference profile* $\check{\succ}_S$ *for* $e \in E$ if $\hat{\succ}_S$ is obtained from $\check{\succ}_S$ by raising the positions of some of e's contracts (at some slots) while leaving the relative preference orders of other employees' contracts unchanged (and not making any contracts that were previously acceptable at some slot unacceptable[15]). Formally:

Definition 2. We say that preference profile $\hat{\succ}_S$ is an *unambiguous improvement over* $\check{\succ}_S$ *for* $e \in E$ if for all slots $s \in S$:

[14] Here we refer to "the" outcome of the cumulative offer process because—at least under slot-specific preferences—the outcome is independent of the order proposals (see Kominers & Sönmez 2016, as well as Hirata & Kasuya 2014).

[15] This last condition—which corresponds to the third condition in Definition 2—was not included by Kominers & Sönmez (2016); however, as Choi et al. (2024) showed, correcting Kominers & Sönmez (2016), it is actually needed for the respect for improvements result to work as stated.

1. for all $x \in X_e$ and $y \in (X_{E\setminus\{e\}} \cup \{\emptyset\})$, if $x \check{\succ}_s y$, then $x \hat{\succ}_s y$;
2. for all $y, z \in X_{E\setminus\{e\}}$, $y \hat{\succ}_s z$ if and only if $y \check{\succ}_s z$; and
3. for all $y \in X_{E\setminus\{e\}}$, if $y \check{\succ}_s \emptyset$, then $y \hat{\succ}_s \emptyset$.

The cumulative offer process *respects unambiguous improvements* in the sense that if $\hat{\succ}_S$ is an unambiguous improvement over $\check{\succ}_S$ for e, then e (weakly) prefers their cumulative offer process outcome under $(\succ_E, \hat{\succ}_S)$ to that under $(\succ_E, \check{\succ}_S)$.

Lemma 2 (Kominers & Sönmez 2016; Choi et al. 2024)**.** *If $\hat{\succ}_S$ is an unambiguous improvement over $\check{\succ}_S$ for e, and $\hat{Y}^*$ and $\check{Y}^*$ denote the employee-proposing cumulative offer process outcomes under $(\succ_E, \hat{\succ}_S)$ and $(\succ_E, \check{\succ}_S)$, respectively, then e (weakly) prefers their assignment under $\hat{Y}^*$ to that under $\check{Y}^*$; that is,*

$$\hat{Y}^*_e \succcurlyeq_e \check{Y}^*_e.$$

5.6. Comparative Statics

Lemma 2 immediately implies several comparative statics for markets with slot-specific preferences—generalizing Theorem 1—through a version of the argument we presented in Section 4.2.

5.6.1. Expanding Capacity

First, we consider what happens to the cumulative offer process outcome when we add a slot $\tilde{s}$ at firm f. Abusing notation again, we extend our slot-specific preference model to the set of slots $S \cup \{\tilde{s}\}$, writing $\tilde{\succ}_S$ and $\tilde{Y}$, respectively, for slot preferences and outcomes in the expanded market ($\tilde{s}$ can appear anywhere in the precedence order). To reflect the idea that $\tilde{s}$ is a new addition to an existing market, we require that preferences of slots in S are unchanged: $\tilde{\succ}_S = \succ_S$. (Note that here we may also leave the employee preferences $\succ_E$ unchanged, as we have not in any way changed the set of firms F or the set of contracts X.)

By construction, adding the new slot $\tilde{s}$ impacts the market exactly the way that adding a new woman impacted the market in Section 4.2: it is as if we raised the position of all agents' contracts at a slot that previously treated all contracts as unacceptable, holding all other slots' preferences fixed. Hence,

adding $\tilde{s}$ results in an unambiguous improvement for all agents; this implies the following comparative static.

Theorem 2. *If Y^* is the outcome of the employee-proposing cumulative offer process in the market with the set of slots S and $\tilde{Y}^*$ is the outcome of the employee-proposing cumulative offer process in the market with the set of slots $S \cup \{\tilde{s}\}$, then each employee $e \in E$ (weakly) prefers their assignment under $\tilde{Y}^*$ to their assignment under Y^*; that is,*

$$\tilde{Y}_e^* \succcurlyeq_e Y_e^*.$$

Theorem 2 generalizes Theorem 1, as in the slot-specific preferences framework we can model entry of a new woman into a marriage market as adding a single slot at a firm that previously had none. Theorem 2 also has practical implications: Perhaps most naturally, the result means that under (employee-proposing) deferred acceptance, expanding the number of available positions at a firm always works in employees' favor. Moreover, by applying Theorem 2 iteratively (adding one slot at a time), we obtain the conclusion of Theorem 2 for the entry of a wholly new employer to the market.

5.6.2. Adding Contracts

A further generalization of Theorem 1 shows that adding new contracts at the bottom of some slots' preference orders (that is, right before the null contracts $\emptyset$) again results in improved outcomes for all employees.

Suppose that we introduce new contracts $\tilde{X}$, yielding the contract set $X \cup \tilde{X}$. We write $\tilde{\succ}$ and $\tilde{Y}$ for preferences and outcomes in the expanded market, assuming that $y \tilde{\succ}_r y'$ if and only if $y \succ_r y'$ for all $r \in E \cup S$ and $y, y' \in X$—reflecting the idea that adding $\tilde{x}$ does not affect agents' or slots' preferences over pre-existing contract options. If $y \tilde{\succ}_s \tilde{x}$ for all $s \in S$, $y \in X$, and $\tilde{x} \in \tilde{X}$, then, just as in the argument we used in Section 4.2, we may interpret $\tilde{\succ}_S$ as an unambiguous improvement over $\succ_S$ under the contract set $X \cup \tilde{X}$ by imagining that $\succ_S$ ranks all the contracts in $\tilde{X}$ as unacceptable—with relative ranking consistent with $\tilde{\succ}_S$—so that $\tilde{\succ}_S$ can be obtained from $\succ_S$ by raising the position of contracts in $\tilde{X}$ relative to the outside option. Thus, we have the following comparative static by Lemma 2.

Theorem 3. *If Y^* is the outcome of the employee-proposing cumulative offer process in the market with the set of contracts X and $\tilde{Y}^*$ is the outcome of*

the employee-proposing cumulative offer process in the market with the set of contracts $X \cup \tilde{X}$ (with $y \tilde{\succ}_s \tilde{x}$ for all $s \in S$, $y \in X$, and $\tilde{x} \in \tilde{X}$), then each employee $e \in E$ (weakly) prefers their assignment under $\tilde{Y}^*$ to their assignment under Y^*; that is,

$$\tilde{Y}^*_e \tilde{\succcurlyeq}_e Y^*_e.$$

Lastly, we note that adding new contracts for a single employee $e \in E$ *anywhere* in slots' preferences also results in an unambiguous improvement—and hence ensures that e will be better off under the cumulative offer process.[16]

Indeed, suppose that we introduce a new contract $\tilde{x}$, yielding the contract set $X \cup \{\tilde{x}\}$. We write $\tilde{\succ}$ and $\tilde{Y}$ for preferences and outcomes in the expanded market, assuming that $y \tilde{\succ}_r y'$ if and only if $y \succ_r y'$ for all $r \in E \cup S$ and $y, y' \in X$—again reflecting the idea that adding $\tilde{x}$ does not affect agents' or slots' preferences over pre-existing contract options. By construction, adding the new contract $\tilde{x}$ results in an unambiguous improvement for $\mathsf{e}(\tilde{x})$: employee $\mathsf{e}(\tilde{x})$ has weakly–higher ranked contracts at every slot, while relative rankings of all other contracts are unchanged; this implies the following comparative static by Lemma 2.

Theorem 4. *If Y^* is the outcome of the employee-proposing cumulative offer process in the market with the set of contracts X and $\tilde{Y}^*$ is the outcome of the employee-proposing cumulative offer process in the market with the set of contracts $X \cup \{\tilde{x}\}$, then $\mathsf{e}(\tilde{x})$ (weakly) prefers their assignment under $\tilde{Y}^*$ to their assignment under Y^*; that is,*

$$\tilde{Y}^*_{\mathsf{e}(\tilde{x})} \tilde{\succcurlyeq}_{\mathsf{e}(\tilde{x})} Y^*_{\mathsf{e}(\tilde{x})}.$$

5.7. Discussion

Our use of the respect for improvements result (Lemma 2) here is similar to an early application of Kominers & Sönmez (2014) for the slot-specific preferences framework. Indeed, Kominers & Sönmez (2014) used respect for improvements under slot-specific preferences to show that guaranteeing slots at a school for minorities (Hafalir et al. 2013) improves welfare relative to simply capping the number of majority students allowed to attend that school (Kojima 2012). The crux of the Kominers & Sönmez (2014) argument consists

[16] We cannot in general add contracts at arbitrary positions for multiple agents, however, as the resulting change in preferences might not be an unambiguous improvement.

of observing that converting a quota slot into a reserve slot is an unambiguous improvement for all majority students, as it corresponds to raising majority students' positions relative to the null option at each quota slot (while still ranking majority students below minority students at those slots); this is analogous to the way we apply Lemma 2.

Meanwhile, Chambers & Yenmez (2017) showed a result analogous to Theorem 2 for ways of "expanding" choice rules that satisfy a regularity condition called *path-independence* (Aizerman & Malishevski 1981). Yenmez (2018) extended the Chambers & Yenmez (2017) result still further to cover choice rules that can be modified—or in the language of Hatfield & Kominers (2020), *completed*—in ways that make them path-independent (see also Echenique & Yenmez 2015; Kamada & Kojima 2015). Slot-specific preferences are not path-independent, in general, but *do* have path-independent completions (see Hatfield & Kominers 2020). However, the additions of slots and contracts considered in Theorems 2 to 4 are *not* expansions in the sense of Chambers & Yenmez (2017) and Yenmez (2018); hence, our results here extend Chambers & Yenmez's (2017) and Yenmez's (2018) comparative statics to new types of transformations. To our knowledge, we are the first to prove comparative statics for the addition of new contracts to the market.[17]

6. CONCLUSION

Using the *respect for improvements* property of deferred acceptance, we have developed a new method of proving entry comparative statics in matching markets. Our method generalizes readily to any matching setting for which a respect for improvements result is known: We illustrated one such generalization in Section 5, where we used respect for improvements to show comparative statics for matching settings with slot-specific preferences.[18,19]

[17] That said, a few comparative statics have been shown previously for structured changes to the full set of contracts, such as adjustments to the level at which transfers between agents are taxed (see, e.g., Dupuy et al. 2020).

[18] Likewise, we can extend the entry comparative static to some matching settings under weakened substitutability conditions using the respect for improvements results of Afacan (2017, 2022).

[19] It also seems possible that variations of the standard respect for improvements concept, such as Choi & Li's (2024) *respect for improvements with manipulations*, would give rise to corresponding versions of the entry comparative static.

We note, however, that others have found more refined comparative static results through appeal to structural results for the set of stable matchings. Most common is a sort of dual to Theorem 1, showing that entry of a new woman to the market makes all the other women (weakly) worse off (see, e.g., Gale & Sotomayor 1985; Roth & Sotomayor 1990; Crawford 1991). Others have given more precise characterizations of which agents are helped and/or harmed by other agents' entry (see, e.g., Romm 2014), or shown comparative statics for the full set of stable matchings, rather than just for the deferred acceptance outcome (see, e.g., Blum et al. 1997; Chambers & Yenmez 2017). It is unclear whether our respect for improvements–based approach can also be used to prove the refinements just described. At a minimum, such an argument would seem to require a sharper version of the respect for improvements theorem. If we may pun a bit, let us leave such further "improvements" for future work.

References

Abdulkadiroğlu, A. (2005). College admissions with affirmative action. *International Journal of Game Theory*, *33*, 535–549.

Abdulkadiroğlu, A., & Sönmez, T. (2003). School choice: A mechanism design approach. *American Economic Review*, *93*(3), 729–747.

Afacan, M. O. (2017). Some further properties of the cumulative offer process. *Games and Economic Behavior*, *104*, 656–665.

Afacan, M. O. (2022). Corrigendum to "Some further properties of the cumulative offer process" [Games Econ. Beh. 104 (2017) 656–665]. *Games and Economic Behavior*, *133*, 352–354.

Aizerman, M. A., & Malishevski, A. V. (1981). General theory of best variants choice: Some aspects. *IEEE Transactions on Automatic Control*, *26*(5), 1030–1040.

Balinski, M., & Sönmez, T. (1999). A tale of two mechanisms: Student placement. *Journal of Economic Theory*, *84*(1), 73–94.

Biró, P., Cechlárová, K., & Fleiner, T. (2008). The dynamics of stable matchings and half-matchings for the stable marriage and roommates problems. *International Journal of Game Theory*, *36*(3-4), 333–352.

Blum, Y., Roth, A. E., & Rothblum, U. G. (1997). Vacancy chains and equilibration in senior-level labor markets. *Journal of Economic Theory*, *76*(2), 362–411.

Blum, Y., & Rothblum, U. G. (2002). "Timing is everything" and marital bliss. *Journal of Economic Theory*, *103*(2), 429–443.

Chambers, C. P., & Yenmez, M. B. (2017). Choice and matching. *American Economic Journal: Microeconomics*, *9*(3), 126–147.

Choi, Y. (2022). *On two-sided matching in infinite markets.* (Harvard University Working Paper)

Choi, Y., Kominers, S. D., & Sönmez, T. (2024). *Matching with slot-specific priorities: Theory – corrigendum.* (In Preparation)

Choi, Y., & Li, S. (2024). *Respect for improvements with manipulations.* (Harvard University Working Paper)

Crawford, V. P. (1991). Comparative statics in matching markets. *Journal of Economic Theory*, *54*(2), 389–400.

Crawford, V. P., & Knoer, E. M. (1981). Job matching with heterogeneous firms and workers. *Econometrica*, *49*(2), 437–450.

Dupuy, A., Galichon, A., Jaffe, S., & Kominers, S. D. (2020). Taxation in matching markets. *International Economic Review*, *61*(4), 1591–1634.

Dur, U., Kominers, S. D., Pathak, P. A., & Sönmez, T. (2018). Reserve design: Unintended consequences and the demise of Boston's walk zones. *Journal of Political Economy*, *126*(6), 2457–2479.

Dur, U., Pathak, P. A., & Sönmez, T. (2020). Explicit vs. statistical targeting in affirmative action: Theory and evidence from Chicago's exam schools. *Journal of Economic Theory*, *187*, 104996.

Dworczak, P. (2021). Deferred acceptance with compensation chains. *Operations Research*, *69*(2), 456–468.

Echenique, F., & Yenmez, M. B. (2015). How to control controlled school choice. *American Economic Review*, *105*(8), 2679–2694.

Fleiner, T., Jankó, Z., Tamura, A., & Teytelboym, A. (2018). *Trading networks with bilateral contracts.* (Oxford University Working Paper)

Gale, D., & Shapley, L. S. (1962). College admissions and the stability of marriage. *American Mathematical Monthly*, *69*(1), 9–15.

Gale, D., & Sotomayor, M. A. O. (1985). Some remarks on the stable matching problem. *Discrete Applied Mathematics*, *11*(3), 223–232.

Gonczarowski, Y. A., Kominers, S. D., & Shorrer, R. I. (2024). To infinity and beyond: A general framework for scaling economic theories. *Forthcoming in Theoretical Economics*.

Greenberg, K., Pathak, P. A., & Sönmez, T. (2024). Redesigning the US Army's branching process: A case study in minimalist market design. *American Economic Review*, *114*(4), 1070–1106.

Hafalir, I. E., Yenmez, M. B., & Yildirim, M. A. (2013). Effective affirmative action in school choice. *Theoretical Economics*, *8*(2), 325–363.

Hatfield, J. W., & Kominers, S. D. (2013). Vacancies in supply chain networks. *Economics letters*, *119*(3), 354–357.

Hatfield, J. W., & Kominers, S. D. (2020). *Hidden substitutes.* (Harvard University

Working Paper)
Hatfield, J. W., Kominers, S. D., & Westkamp, A. (2021). Stability, strategy-proofness, and cumulative offer mechanisms. *Review of Economic Studies*, *88*(3), 1457–1502.
Hatfield, J. W., & Milgrom, P. (2005). Matching with contracts. *American Economic Review*, *95*(4), 913–935.
Hirata, D., & Kasuya, Y. (2014). Cumulative offer process is order-independent. *Economics Letters*, *124*(1), 37–40.
Kamada, Y., & Kojima, F. (2015). Efficient matching under distributional constraints: Theory and applications. *American Economic Review*, *105*, 67–99.
Kelso, A. S., Jr., & Crawford, V. P. (1982). Job matching, coalition formation, and gross substitutes. *Econometrica*, *50*(6), 1483–1504.
Kojima, F. (2012). School choice: Impossibilities for affirmative action. *Games and Economic Behavior*, *75*, 685–693.
Kominers, S. D., & Sönmez, T. (2014). *Designing for diversity in matching.* (Harvard University Working Paper)
Kominers, S. D., & Sönmez, T. (2016). Matching with slot-specific priorities: Theory. *Theoretical Economics*, *11*(2), 683–710.
Kominers, S. D., Teytelboym, A., & Crawford, V. P. (2017). An invitation to market design. *Oxford Review of Economic Policy*, *33*(4), 541–571.
Ostrovsky, M. (2008). Stability in supply chain networks. *American Economic Review*, *98*(3), 897–923.
Pathak, P. A. (2011). The mechanism design approach to student assignment. *Annual Review of Economics*, *3*, 513–536.
Romm, A. (2014). Implications of capacity reduction and entry in many-to-one stable matching. *Social Choice and Welfare*, *43*(4), 851–875.
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. (1985). The college admissions problem is not equivalent to the marriage problem. *Journal of Economic Theory*, *36*, 277–288.
Roth, A. E. (2008). Deferred acceptance algorithms: History, theory, practice, and open questions. *International Journal of Game Theory*, *36*(3-4), 537–569.
Roth, A. E. (2013). What have we learned from market design? In N. Vulkan, A. E. Roth, & Z. Neeman (Eds.), *The Handbook of Market Design* (pp. 7–50). Oxford University Press.
Roth, A. E. (2015). *Who Gets What – And Why: The New Economics of Matchmaking and Market Design*. Houghton Mifflin Harcourt.
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. (1990). *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Cambridge University Press.

Sönmez, T. (2013). Bidding for army career specialties: Improving the ROTC branching mechanism. *Journal of Political Economy*, *121*(1), 186–219.

Sönmez, T., & Switzer, T. B. (2013). Matching with (branch-of-choice) contracts at United States Military Academy. *Econometrica*, *81*(2), 451–488.

Wang, F., & Kominers, S. D. (2020). *Respect for improvements via matching with contracts.* (Harvard University Working Paper)

Yenmez, M. B. (2018). A college admissions clearinghouse. *Journal of Economic Theory*, *176*, 859–885.

Journal of Mechanism and Institution Design
ISSN: 2399-844X (Print), 2399-8458 (Online)
DOI:10.22574/jmid.2024.12.006

AN EFFICIENT AND GENERAL ASCENDING MENU AUCTION UNDER BUDGET CONSTRAINTS

Zaifu Yang

University of York, United Kingdom
zaifu.yang@york.ac.uk

Jingsheng Yu

Wuhan University, China
yujingsheng@whu.edu.cn

ABSTRACT

An auctioneer wishes to sell multiple heterogeneous indivisible items to several bidders. Every bidder can demand several items, have complex preferences and faces a hard budget constraint. In this setting, Walrasian equilibria may fail to exist. We propose an ascending menu auction that always yields an efficient allocation of items, which is not only in the core but also strongly Pareto efficient. Furthermore, the auction finds a strong core allocation with a fully efficient assignment of items when bidders are not budget constrained.

Keywords: Dynamic menu auction, core, budget constraints.

JEL Classification Numbers: D44.

1. INTRODUCTION

This paper addresses the problem of how to auction n heterogeneous indivisible goods to m bidders who face hard budget constraints. Every bidder wants to buy several goods and has personal valuations on his interested bundles of goods. Valuations can be very general, irregular and complex. Some

This paper has improved and superseded an unpublished manuscript of Yang & Yu (2018). We are grateful to many colleagues and participants for their helpful comments at various seminars and conferences. Jingsheng Yu is supported by National Natural Science Foundation of China (Grants 72103165).

bidders may view the goods as substitutes, while the others may view them as complements. In this setting, Walrasian equilibria are not guaranteed to exist so it is not possible to follow the traditional approach of using market-clearing prices to allocate goods. We aim to develop a new dynamic auction to overcome the nonexistence problem of market-clearing prices and still achieve an efficient market outcome.

Auctions are powerful market mechanisms and have been long and widely used by both private and public sectors to carry out a broad range of economic activities; see e.g., Klemperer (2004), Milgrom (2004, 2017), and Krishna (2010). A key assumption in auction theory has been that all potential bidders do not face any budget constraints. Under this assumption, over the last 40 years, a number of efficient dynamic auctions have been proposed, such as Crawford & Knoer (1981), Demange et al. (1986), and Andersson & Erlanson (2013), for the cases in which there are multiple items but each bidder demands at most one item, and Kelso & Crawford (1982), Gul & Stacchetti (2000), Milgrom (2000), Ausubel (2004, 2006), Milgrom & Strulovici (2009), Sun & Yang (2009, 2014), and Shioura & Yang (2015), for the cases in which there are multiple items and each may demand several items. Unfortunately, in reality, buyers may not always have enough cash or credit to buy the goods that they want to buy. It is well-known that budget constraints pose a serious obstacle to the efficient allocation of resources as pointed out by Che & Gale (1998), Laffont & Robert (1996), and Maskin (2000). Besides budget constraints, the current model also faces complex valuations going beyond those identified by Kelso & Crawford (1982), Danilov et al. (2001), Fujishige & Yang (2002), Sun & Yang (2006), Baldwin & Klemperer (2019), and Baldwin et al. (2023). Complex valuations can also make Walrasian equilibrium unattainable.

In this paper, we develop an ascending menu auction that can handle both complex and irregular preferences and hard budget constraints. In spite of complex preferences, budget constraints and non-transferable utilities, we prove that the auction always finds an efficient allocation of items which is not only in the core but also strongly Pareto efficient. Furthermore, it locates a strong core allocation with a fully efficient assignment of goods when bidders face no budget constraints. Roughly speaking, every bidder initially sets his bids on all his interested bundles according to his valuations and budget and reports his initial bids to the auctioneer. Then the auctioneer selects a provisional assignment based on the reported bids from all bidders and her own valuations over all bundles. If a bidder gets no item from the provisional assignment

and can make new bids (such a bidder is called a provisional loser), then at least one of such bidders reduces his target utility and all other bidders keep their target utilities unchanged. Every bidder then updates his bids on his interested bundles and reports his updated bids. The auctioneer chooses again a provisional assignment based on currently reported bids and her own valuations. This process stops at the time when there is no provisional loser anymore. The auction contains a unique tie-breaking rule.

Another salient feature of this new auction is that we generalize a long-established pricing rule: a seller will sell her good only if the price is no less than her valuation or reserve price. Because our current model allows the seller to have any kind of valuation on every bundle of items, we generalize this conventional rule in a way that a bundle can be sold only if its price is no less than its marginal value to the seller. We prove that at every core allocation the price of every sold bundle is at least equal to its marginal value to the seller when no condition is imposed on the seller's valuations, and that at every core allocation the price of every sold bundle is at least equal to the seller's valuation when her valuations are super-additive or when all items are sold out.

The current auction generalizes those of Bernheim & Whinston (1986), in particular, Ausubel & Milgrom (2002), Milgrom (2007), Mishra & Parkes (2007), Day & Milgrom (2008), and Erdil & Klemperer (2010) from their settings without budget constraints to the current setting with budget constraints. This article also connects with several recent papers on the assignment markets with budget constraints.[1] In such markets, every bidder demands at most one item or every firm hires at most one worker and every worker works for at most one firm. Aggarwal et al. (2009) examine an assignment model with budget constraints under which a Walrasian equilibrium exists. Ashlagi et al. (2010) investigate a position auction model with budget constraints, a special assignment model.[2] In their Section 8, Ausubel & Milgrom (2002) also discuss a specific auction model with budget constraints and present a mechanism for finding a core allocation. They require that bidders have strict preferences over a finite number of choices. Their auction does not work for the assignment models, or the current one, in which bidders have a continuum of choices and

[1] The well-known assignment markets without budget constraints have been studied by Koopmans & Beckmann (1957), Shapley & Shubik (1971), Crawford & Knoer (1981), and Demange et al. (1986).

[2] See Dobzinski et al. (2012) for an impossibility result when several identical items are sold to budget constrained bidders each of whom may demand multiple items.

can be indifferent between different choices.

The following four papers deal with the assignment models with budget constraints that may fail to achieve Walrasian equilibria. Talman & Yang (2015) propose a dynamic auction that finds a core allocation. Laan van der & Yang (2016) introduce an ascending auction to locate a constrained equilibrium. Herings & Zhou (2022) propose an algorithm for finding a quantity-constrained competitive equilibrium. In these two papers, both constrained equilibria need not be a core allocation. Yang & Yu (2022) introduce a dynamic auction which generates a core allocation and also induces bidders to bid truthfully. The other literature on auctions concerning one or two goods with budget constraints includes Zheng (2001), Hafalir et al. (2012), Benoit & Krishna (2001), Pitchik (2009), and Pai & Vohra (2014) among others.

The rest of the paper is organized as follows. Section 2 presents the model and basic concepts. Section 3 introduces and analyzes the auction. Section 4 concludes.

2. THE MODEL

An auctioneer (seller) wishes to sell a set of heterogeneous indivisible goods (items) $N = \{a, b, \dots\}$ to a group of potential bidders $M = \{1, 2, \dots\}$. Let 0 represent the seller (she) and let $M_0 = M \cup \{0\}$ stand for all agents in the market. Let 2^N denote the collection of all subsets of N, and an element $A \in 2^N$ is called a bundle of goods. For every bidder $i \in M$, let $\mathcal{F}^i \subseteq 2^N$ represent the family of bundles in which he is interested. We assume that consuming nothing can be one of his options, i.e., $\emptyset \in \mathcal{F}^i$. Bidder i attaches a monetary value (units of money) to each interested bundle of items, namely, each bidder i has a valuation function $v^i : \mathcal{F}^i \to \mathbb{Z}_+$ with $v^i(\emptyset) = 0$. Every bidder i is endowed with an amount $m^i \in \mathbb{Z}_+$ of money. We say that bidder i is *budget constrained* if $m^i < \max_{A \in \mathcal{F}^i} v^i(A)$, that is, the valuation of bidder i for some bundle exceeds what he can afford. Otherwise, bidder i is not budget constrained or faces no budget constraint. The seller has a valuation $v^0(A) \in \mathbb{Z}_+$ for every bundle $A \in 2^N$ and an initial income $m^0 = 0$. Observe that for the seller we have $\mathcal{F}^0 = \{A \mid A \subseteq N\}$. Here money will be treated as a perfectly divisible commodity. Because all $v^i(A)$ and m^i are integers, we can actually prove that there will be an integral solution by using our proposed auctions and the size of price adjustment in the auctions can be 1.

A bundle $A \in 2^N$ can be sold only if its price $p(A)$ is no less than its

marginal value to the seller, i.e., $p(A) \geq v^0(A \cup B) - v^0(B)$, where B is the bundle that will not be sold but kept by the seller. This pricing rule is a natural and practical generalization of the traditional one that the seller is willing to sell her good only if its price is not below her reserve price. This new rule is needed because we allow the seller to have a very general pattern of valuations on her bundles of items. In the literature, it is typically assumed that the seller has a zero reserve price on every item or bundle of items.

The following assumptions are imposed upon the model:

(A1) *Private Values*: Every bidder $i \in M$ knows privately his own feasible bundles $\mathcal{F}^i$, valuation function v^i, and budget m^i.

(A2) *Quasilinear Utility*: For any bidder $i \in M$, if he pays $p(A)$ in exchange for bundle $A \in \mathcal{F}^i$, he gets utility of $v^i(A) - p(A)$ for $p(A) \leq m^i$ and utility of $-\infty$ for $p(A) > m^i$.

(A3) *Monotonicity of the Seller Values*: The seller's valuation function v^0 is weakly increasing with $v^0(\emptyset) = 0$, i.e., $v^0(A) \leq v^0(B)$ for $A \subseteq B \subseteq N$.

We use $(v^i, \mathcal{F}^i, m^i)_{i \in M_0}$ to represent this auction model.

An assignment of items in N is a partition $\pi = (\pi(i))_{i \in M_0}$ of items among all agents in M_0 such that $\pi(i) \cap \pi(j) = \emptyset$ for all $i \neq j$ and $\bigcup_{i \in M_0} \pi(i) = N$. Note that $\pi(i) = \emptyset$ is allowed. An assignment π assigns the bundle $\pi(i)$ to agent i. If $\pi(0) \neq \emptyset$, then the bundle $\pi(0)$ is not sold and thus stays with the seller. An assignment π is *feasible* if $\pi(i) \in \mathcal{F}^i$ for every bidder $i \in M$. Let $\bar{\mathcal{A}}$ denote the family of all assignments and $\mathcal{A} \subseteq \bar{\mathcal{A}}$ the family of all feasible assignments. A feasible assignment π is *fully efficient* if for every feasible assignment ρ, we have

$$\sum_{i \in M_0} v^i(\pi(i)) \geq \sum_{i \in M_0} v^i(\rho(i)). \tag{1}$$

Given a fully efficient assignment π^*, let $w(M_0) = \sum_{i \in M_0} v^i(\pi^*(i))$. We call $w(M_0)$ *the potential market value*. Clearly, this value is the same for any fully efficient assignment. Moreover, for any subset of agents $S \subset M_0$ which includes the seller, we use $\mathcal{A}^S$ to denote the family of all assignments $\rho \in \bar{\mathcal{A}}$ which satisfy $\rho(i) = \emptyset$ for every bidder $i \in M_0 \setminus S$. Then the value of the coalition S is $w(S) = \max_{\rho \in \mathcal{A}^S} \sum_{i \in S} v^i(\rho(i))$.

A vector $r = (r^i)_{i \in M_0}$ is a feasible income distribution if $\sum_{i \in M_0} r^i = \sum_{i \in M_0} m^i$ and $r^i \geq 0$ for all $i \in M_0$. A pair (π, r) of a feasible assignment π

and a feasible income distribution r is called an allocation. At (π, r), agent $i \in M_0$ receives bundle $\pi(i)$ and holds r^i a total amount of income. Then the net utility that the agent achieves is given by

$$u^i(\pi, r) = v^i(\pi(i)) - (m^i - r^i).^3$$

When there is no budget constraint, the Walrasian equilibrium has been the most widely used solution for auction and equilibrium models and Walrasian equilibrium or market-clearing prices are used in auction design.

Given a price vector $p = (p^j)_{j \in N}$ which specifies a price for each item, the set of affordable bundles of bidder i who faces a budget m^i is represented by $\tilde{\mathcal{F}}^i = \{B \in \mathcal{F}^i \mid \sum_{j \in B} p^j \leq m^i\}$. The demand set of bidder i given $p = (p^j)_{j \in N}$ is defined by

$$D^i(p) = \left\{ A \in \tilde{\mathcal{F}}^i \mid v^i(A) - \sum_{j \in A} p^j = \max_{B \in \tilde{\mathcal{F}}^i} \left(v^i(B) - \sum_{j \in B} p^j \right) \right\}.$$

So $D^i(p)$ is the collection of his most preferred affordable bundles at prices p.

Definition 1. *A* Walrasian equilibrium *is a pair (p, π) of prices p and assignment π such that $\pi(i) \in D^i(p)$ for every bidder $i \in M$ and $v^0(\pi(0)) = \sum_{j \in \pi(0)} p^j$ for the seller.*

At equilibrium, every bidder gets his best bundle at the prices within his budget and the price of the unsold bundle is equal to the seller's valuation of the bundle.

If (p, π) is a Walrasian equilibrium, then p is called an equilibrium or market-clearing price vector and π a Walrasian equilibrium allocation. It is well-known from Koopmans & Beckmann (1957) and Kelso & Crawford (1982) that there will be at least one Walrasian equilibrium when all indivisible goods are substitutes and traders face no budget constraints.

It is known that when buyers are budget constrained or preferences are very general and complex, Walrasian equilibrium cannot be guaranteed to exist. For the current model, we have to adopt a more general solution: the core. The notion of core has been widely used in general equilibrium theory and cooperative game theory; see e.g., Gillies (1953), Debreu & Scarf (1963),

[3] The seller's utility is also $u^0(\pi, r) = v^0(\pi(0)) - (m^0 - r^0) = v^0(\pi(0)) + \sum_{i \in M}(m^i - r^i)$.

Scarf (1967), Shapley & Scarf (1974), and Predtetchinski & Herings (2004). We now introduce this concept for a non-transferable utility n-person game.

An allocation (π, r) is *individually rational* if every agent $i \in M_0$ achieves no less utility than they would achieve if they stood alone, i.e., $u^i(\pi, r) \geq 0$ for every $i \in M$ and $u^0(\pi, r) \geq v^0(N)$ for the seller. An allocation (π, r) is *Pareto efficient* if there does not exist another allocation (ρ, t) such that $u^i(\pi, r) < u^i(\rho, t)$ for all $i \in M_0$; otherwise, we say that (π, r) is strongly Pareto dominated by (ρ, t). An allocation (π, r) is *strongly Pareto efficient* if there does not exist another allocation (ρ, t) such that $u^i(\pi, r) \leq u^i(\rho, t)$ for all $i \in M_0$ with at least one strict inequality; otherwise, we say that (π, r) is Pareto dominated by (ρ, t). A nonempty subset $S \subseteq M_0$ is called a viable coalition if S consists of either the seller with any number of bidders or a single bidder. Given a viable coalition S, an allocation (ρ^S, t) is feasible for S, if ρ^S is in $\mathcal{A}^S$ such that $\rho^S(i) = \emptyset$ for every bidder $i \in M_0 \setminus S$, and $t = (t^i)_{i \in M_0}$ is a feasible income distribution such that $\sum_{i \in S} t^i = \sum_{i \in S} m^i$ and $t^i = m^i$ for every bidder $i \in M_0 \setminus S$. An allocation (π, r) is blocked by a viable coalition S if there exists a feasible allocation (ρ^S, t) such that $u^i(\rho^S, t) > u^i(\pi, r)$ for all $i \in S$; the allocation (π, r) is weakly blocked by a viable coalition S if there exists a feasible allocation (ρ^S, t) such that $u^i(\rho^S, t) \geq u^i(\pi, r)$ for all $i \in S$ and with at least one strict inequality.

Definition 2. *An allocation (π, r) is in the* core *and is called a core allocation if it is not blocked by any coalition. It is in the* strong core *and is called a strong core allocation if it cannot be weakly blocked by any coalition.*

Clearly, every core allocation is Pareto efficient and every strong core allocation is strongly Pareto efficient. Note that in the current model utilities cannot be fully transferable because of budget constraints and indivisibility.

The following example shows that when bidders are budget constrained, the core is not empty but the strong core is empty.

Example 1. *There are three items $\{a, b, c\}$ and three bidders $\{1, 2, 3\}$. Each bidder's interested bundles, valuations, and budget are given in Table 1. For ease of notation, we express bundle $\{a, b\}$ as ab.*

Case 1: Consider the case of $(m^1, m^2, m^3) = (4, 4, 6)$. In this case no bidder is budget constrained. We will show that the market has a nonempty strong core. The market has a unique fully efficient assignment $\pi^* = (\emptyset, \emptyset, b, ac)$, which assigns b to bidder 2 and $\{a, c\}$ to bidder 3. By giving a feasible income

Table 1: Valuations and budgets.

Bidders	$\mathcal{F}^i$	$v^i(\cdot)$	m^i	m^i
1	$\{ab, ac, \emptyset\}$	$v^1(ab) = 4$, $v^1(ac) = 4$, $v^1(\emptyset) = 0$	4	3
2	$\{b, ab, \emptyset\}$	$v^2(b) = 3$, $v^2(ab) = 4$, $v^2(\emptyset) = 0$	4	2
3	$\{c, ac, \emptyset\}$	$v^3(c) = 4$, $v^3(ac) = 6$, $v^3(\emptyset) = 0$	6	2

distribution $r^* = (9, 4, 1, 0)$, we construct a strong core allocation (π^*, r^*). To show this, see the utility distribution $(u^0, u^1, u^2, u^3) = (9, 0, 0, 0)$, which satisfies $\sum_{i \in M_0} u^i = 9 = w(M_0)$ and $\sum_{i \in S} u^i \geq w(S)$ for any proper subset $S \subset \{0, 1, 2, 3\}$.[4]

Case 2: Consider the case of $(m^1, m^2, m^3) = (3, 2, 2)$. In this case, all bidders are budget constrained. We will show that the strong core is empty but the core is not empty. Since any wasteful allocation, in which the seller holds a bundle unsold, can be Pareto improved by assigning the unsold object(s) to the right person without changing the income distribution, so we focus on the four non-wasteful assignments: $\pi^* = (\emptyset, \emptyset, b, ac)$, $\pi' = (\emptyset, ac, b, \emptyset)$, $\pi'' = (\emptyset, ab, \emptyset, c)$, and $\pi''' = (\emptyset, \emptyset, ab, c)$.

We first consider the fully efficient assignment $\pi^* = (\emptyset, \emptyset, b, ac)$. For any feasible income distribution $r = (7 - r^1 - r^2 - r^3, r^1, r^2, r^3)$, individual rationality implies that $r^1 \geq m^1$. (π^*, r) is weakly blocked by coalition $\{0, 1, 2\}$, and the blocking allocation is $(\pi', t(r))$ with $t(r) = (8 - r^1 - r^2, r^1 - 3, r^2, 2)$ and

$$
\begin{aligned}
u^0(\pi', t(r)) &= 8 - r^1 - r^2 > 7 - r^1 - r^2 - r^3 = u^0(\pi^*, r), \\
u^1(\pi', t(r)) &= v^1(ac) + r^1 - 3 - m^1 > 0 = u^1(\pi^*, r), \\
u^2(\pi', t(r)) &= v^2(b) + r^2 - m^2 = v^2(b) + r^2 - m^2 = u^2(\pi^*, r).
\end{aligned}
$$

In a similar way, we can show that any individually rational allocation (π', r) is weakly blocked by coalition $\{0, 1, 3\}$ and the proposed allocation is $\big(\pi'', (7 - r^1 - r^3, r^1, 2, r^3 - 2)\big)$. The individually rational allocation (π'', r) will be weakly blocked by coalition $\{0, 1, 2\}$ with the proposed allocation $\big(\pi', (7 - r^1 - r^2, r^1, r^2 - 2, 2)\big)$. Finally, (π''', r) will be weakly blocked by coalition $\{0, 1, 3\}$ with allocation $\big(\pi'', (8 - r^1 - r^2, r^1 - 3, 2, r^3)\big)$.

[4] The values of all the coalitions are listed here: $w(0) = w(1) = w(2) = w(3) = w(12) = w(13) = w(23) = w(123) = 0$, $w(01) = 4$, $w(02) = 4$, $w(03) = 6$, $w(012) = 7$, $w(013) = 8$, $w(023) = 9$, and $w(0123) = 9$.

By contrast, the core is not empty. In fact, $(\pi', (5,0,0,2))$ and $(\pi'', (5,0,2,0))$ are both in the core. Because there is always a bidder receiving nothing and thus holding his budget in line with individual rationality, the seller cannot get more than an income of 5. Then no coalition can improve the seller's revenue based on the two given core allocations.

The following example shows that even if bidders face no budget constraints, complex preferences can still fail to achieve Walrasian equilibrium.

Example 2. *There are two items* $\{a, b\}$ *and two bidders* $\{1, 2\}$. *Table 2 shows every bidder's budget and valuations.*

Table 2: Valuations and budgets.

Bidders	$\mathcal{F}^i$	$v^i(\cdot)$	m^i
1	$\{a, b, ab, \emptyset\}$	$v^1(a) = 1, v^1(b) = 2, v^1(ab) = 5, v^1(\emptyset) = 0$	7
2	$\{a, b, ab, \emptyset\}$	$v^2(a) = 4, v^2(b) = 4, v^2(ab) = 5, v^2(\emptyset) = 0$	9

In this example, no bidder is budget constrained. Bidder 1 views items a and b as complements but bidder 2 views them as substitutes. The market has a unique fully efficient assignment $\pi^* = (\emptyset, b, a)$. Suppose to the contrary that the market has a Walrasian equilibrium. Then there exist supporting prices p^a and p^b for π^*. For bidder 1, we have $2 - p^b \geq 5 - p^a - p^b$ which implies $p^a \geq 3$. We also have $p^b \leq 2$. For bidder 2, we have $4 - p^a \geq 4 - p^b$ which implies $p^b \geq p^a$, contradicting $p^a \geq 3$ and $p^b \leq 2$. This shows that the market has no Walrasian equilibrium. Nevertheless, the market has a strong core allocation at which bidder 1 gets item b and pays 2 and bidder 2 gets item a and pays 4.

We will establish the following core existence theorem for our auction model.

Theorem 1. *There exists at least one core allocation in the auction model* $(v^i, \mathcal{F}^i, m^i)_{i \in M_0}$.

In the next section we will propose a dynamic auction mechanism which finds a strongly Pareto efficient core allocation, not just a core allocation, thus giving a constructive proof of a stronger version of the theorem.

3. MAIN RESULTS

This section first introduces the new dynamic auction and then illustrates it and discusses its properties.

3.1. A New Dynamic Auction

Before introducing the new dynamic auction, we make some preliminary remarks. At each round k of the auction, every bidder i will update his bids on his feasible bundles and his target utility $\hat{u}^i_k$ as follows. For every feasible bundle $A \in \mathcal{F}^i$, he finds the intermediate bidding price

$$\hat{p}^i(A \mid \hat{u}^i_k) = v^i(A) - \hat{u}^i_k.$$

If $\hat{p}^i(A \mid \hat{u}^i_k) < 0$, he will not bid for A in this round. If $0 \leq \hat{p}^i(A \mid \hat{u}^i_k) \leq m^i$, he proposes an offer of buying bundle A with price $p^i_k(A) = \hat{p}^i(A \mid \hat{u}^i_k)$. And if $m^i < \hat{p}^i(A \mid \hat{u}^i_k)$, he proposes to buy bundle A with price $p^i_k(A) = m^i$. The bidder treats the empty bundle $\emptyset$ as a normal bundle. If bidder i is required by the auctioneer to submit new offers, he makes new offers by reducing his target utility. We define the decrement of target utility, $\Delta^i(\hat{u}^i_k)$, as the minimal integer that leads to new offers. Formally,

$$\Delta^i(\hat{u}^i_k) = \min\{d \in \mathbb{Z}_{++} \mid \hat{p}^i(A|\hat{u}^i_k - d) \in [0, m^i] \text{ for some } A \in \mathcal{F}^i\}.$$

The decrement is usually one but can be larger than one when the bidding price of some bundle reaches the bidder's budget. We use $x = (i, A, p) \in I \times 2^N \times \mathbb{R}_+$ to denote bidder i's offer of buying bundle A at price p. Given an offer $x = (i, A, p)$, let $x_M = i$ denote the proposer, $x_B = A$ the bundle, and $x_P = p$ the bidding price.

In each round, the auctioneer uses every bidder's reported bids to calculate feasible assignments that maximize her revenues. Because the auctioneer can use only the revealed information such as offer prices and bid bundles of every bidder $i \in M$, she will set the price of every nonempty bundle for which he has not bid at $-\infty$. Now we describe the new dynamic auction mechanism in detail.

Initialization: Set $k = 1$ and go to the Bidding stage.

Bidding stage: If it is the first round $k = 1$, every bidder $i \in M$ sets his initial target utility $\hat{u}^i_1 = \max_{A \in \mathcal{F}^i} v^i(A)$ and the intermediate bidding

price $\hat{p}^i_k(A \mid \hat{u}^i_k) = v^i(A) - \hat{u}^i_k$ for every feasible bundle $A \in \mathcal{F}^i$, and makes the offer $\left(i, A, \min\{\hat{p}^i_k(A \mid \hat{u}^i_k), m^i\}\right)$ and sets his price $p^i_k(A) = \min\{\hat{p}^i_k(A \mid \hat{u}^i_k), m^i\}$ on A only if $\hat{p}^i_k(A \mid \hat{u}^i_k) \geq 0$. Let X^i_k denote bidder i's offer at round k. Go to the Assigning stage. In any other round $k \geq 2$, if bidder i is required to submit new offers, he updates his target utility $\hat{u}^i_k = \hat{u}^i_{k-1} - \Delta^i(\hat{u}^i_{k-1})$, sets the intermediate bidding price $\hat{p}^i_k(A \mid \hat{u}^i_k) = v^i(A) - \hat{u}^i_k$ for every feasible bundle $A \in \mathcal{F}^i$, and makes the offer $\left(i, A, \min\{\hat{p}^i_k(A \mid \hat{u}^i_k), m^i\}\right)$ and sets his price $p^i_k(A) = \min\{\hat{p}^i_k(A \mid \hat{u}^i_k), m^i\}$ on A only if $\hat{p}^i_k(A \mid \hat{u}^i_k) \geq 0$. Let X^i_k denote bidder i's offer at round k. Every other bidder $j \neq i$ keeps his target utility and offer unchanged $\hat{u}^j_k = \hat{u}^j_{k-1}$ and $X^j_k = X^j_{k-1}$.

Go to the Assigning stage.

Assigning stage: The auctioneer keeps the record of every bidder $i's$ bidding history $\{X^i_\kappa\}_{1 \leq \kappa \leq k}$. For every bidder $i \in M$, the auctioneer sets the price of every nonempty bundle A for which he has not bid until the current round as $p^i_k(A) = -\infty$ and sets the price of the empty bundle for which he has not bid as $p^i_k(\emptyset) = 0$. Then the auctioneer calculates i's first-bidding-round $f^i_k(A)$ of the current offer on his bid bundle A by setting

$$f^i_k(A) = \min\{\kappa \in \{1, \cdots, k\} \mid \exists x \in X^i_\kappa \text{ such that } x_B = A \text{ and } x_P = p^i_k(A)\}$$

and by setting $f^i_k(A) = \infty$ for any bundle A for which he has not bid until the current round k. The auctioneer records $f_k(\pi) = \left(f^i_k(\pi(i))\right)_{i \in M}$ as the first-bidding-round vector of every assignment $\pi \in \bar{\mathcal{A}}$ at round k.

Based on the current bidding prices $P_k = (p^i_k)_{i \in M_0}$ with $p^0_k(A) = v^0(A)$ for every $A \in 2^N$, the auctioneer finds all optimal assignments that maximize her revenues by solving the problem:

$$C(P_k) = \left\{\pi \in \bar{\mathcal{A}} \mid \pi = \arg\max_{\rho \in \bar{\mathcal{A}}} \Big(\sum_{i \in M_0} p^i_k(\rho(i)) \Big) \right\}. \tag{2}$$

If there is only one assignment in $C(P_k)$, the auctioneer chooses the unique assignment as the provisional assignment π_k. If $C(P_k)$ contains

more than one assignment, the auctioneer chooses $\pi_k \in C(P_k)$ such that there does not exist another assignment $\rho \in C(P_k)$ with $f_k(\rho) \leq f_k(\pi_k)$.[5]

Go to the Continuation or termination.

Continuation or termination: At the provisional assignment π_k, if $f_k^i(\pi_k(i)) < \infty$ for all bidders $i \in M$, the auction stops with the provisional assignment π_k. Then every bidder $i \in M$ receives bundle $\pi_k(i)$ and pays his bidding price $p_k^i(\pi_k(i))$. The seller keeps bundle $\pi_k(0)$ and receives the payments for the sold bundles. Otherwise, the auctioneer randomly selects one bidder with $f_k^i(\pi_k(i)) = \infty$ (which means that he is assigned the empty bundle $\emptyset$ not proposed by him) and asks him to submit new offers at next round.

Set $k = k + 1$ and go back to the Bidding stage.

Several comments are in order. First, for every bidder the price of the empty bundle is always zero regardless of whether the bidder has bid on it or not. Second, if a bidder has bid on a nonempty bundle, the price of the bundle will not decrease. Third, the tie-breaking rule of using the functions $f_k(\cdot)$ plays an indispensable role in finding a strongly Pareto efficient core allocation, not just a core allocation. Note that when bidders face no budget constraints, no tie-breaking rule is needed, because in this case the core coincides with the strong core. Lastly, target utility is also used in Day & Milgrom (2008) for a model without budget constraints. Concerning the auction, we have the following observation.

Proposition 1. *If π is an optimal solution to Problem (2), it must be a feasible assignment, i.e., $\pi \in \mathcal{A}$. Furthermore, for any bidder $i \in M$, if $\pi(i)$ is not empty, then bidder i must have bid for it.*

Proof. Note that for any bidder $i \in M$ and any nonempty bundle A, if the bidder has no interest in the bundle, he will not bid on the bundle, resulting in the bundle's price being $-\infty$. Moreover, the price of the empty set $\emptyset$ is always zero. It is clear that when the auctioneer tries to find an assignment maximizing her revenues by solving Problem (2), she will definitely not choose any nonempty bundle for which no bidders have bid. The conclusion follows immediately. □

[5] The inequality of the vectors means that $f_k^i(\rho(i)) \leq f_k^i(\pi_k(i))$ for every bidder and with strict inequality for at least one bidder.

3.2. Illustration of the Proposed Auction

To have a better understanding of the proposed auction mechanism, we illustrate it through the following Example:

Example 3. *There are three items* $\{a, b, c\}$ *and three bidders* $\{1, 2, 3\}$. *The seller values every bundle at* 0. *Table 3 shows every bidder's budget and valuations.*

Table 3: Valuations and budgets.

Bidders	$\mathcal{F}^i$	$v^i(\cdot)$	m^i
1	$\{a, b, \emptyset\}$	$v^1(a) = 6, v^1(b) = 2, v^1(\emptyset) = 0$	2
2	$\{c, \emptyset\}$	$v^2(c) = 5, v^2(\emptyset) = 0$	2
3	$\{abc, \emptyset\}$	$v^3(abc) = 4, v^3(\emptyset) = 0$	4

Table 4 collects the data generated by the proposed auction mechanism and shows that the auction finds the core allocation. Let's take a closer look at the auction process. With respect to the prices in round 1, all the feasible assignments yield zero revenue for the auctioneer. However, the auctioneer should choose an assignment with minimal first-bidding-round vector, so she cannot choose the no-sale assignment $\bar{\rho} = (abc, \emptyset, \emptyset, \emptyset)$ since $f_1(\bar{\rho}) = (\infty, \infty, \infty)$ is larger than $f_1(\pi_1) = (1, 1, \infty)$. At round 2, bidder 3 is required to propose a new offer. He reduces his target utility by 1 and makes the offer $(3, abc, 1)$. Recall that the offer means that bidder 3 is willing to buy the bundle abc with price 1. Then, π_2 is the unique feasible assignment, maximizing the seller's revenue, so it is chosen regardless of the first bidding times. Turn to round 7 in which bidder 1 is required to propose new offer. If bidder 1 reduces his target utility by 1 as usual, then the two intermediate prices of his interested bundles both fall outside the feasible interval (i.e., $\hat{p}^1(a|3) = 3 > 2 = m^1$ and $\hat{p}^1(b|3) = -1 < 0$). So bidder 1 reduces his target by 2 and proposes an offer $(1, b, 0)$. In addition, his previous offer $(1, a, 2)$ is kept unchanged, and the first round of bids is recorded as 4. At round 12, there are three feasible assignments yielding the highest revenue, they are $\pi_{12} = (\emptyset, a, c, \emptyset)$, $\pi_{11} = (\emptyset, \emptyset, \emptyset, abc)$, and $\pi' = (\emptyset, b, c, \emptyset)$. The three first-bidding-round vectors are $f_{12}(\pi_{12}) = (4, 12, 9)$, $f_{12}(\pi_{11}) = (11, \infty, 9)$, and $f_{12}(\pi') = (11, 12, 9)$. The auctioneer selects π_{12} and the auction stops because it has the minimal first-bidding-round vector and no bidder makes any

new bid. Observe that π_{12} is the unique strongly Pareto efficient allocation. At the final allocation, bidder 1 gets a and pays 2, bidder 2 gets c and pays 2, bidder 3 gets nothing and pays nothing, and the seller gets 4 and keeps b. It is easy to verify that this allocation π_{12} is a core allocation, creating a social value of 11. As will be shown, the current auction will always find a strongly Pareto efficient core allocation. Let us now briefly compare the current auction with that of Ausubel & Milgrom (2002). Their auction cannot guarantee to find the outcome given by our auction but can end up with the outcome which assigns abc to bidder 3 who pays 4, generating a social value of 4.

Table 4: Illustration of the proposed auction mechanism via Example 3.

k	$\hat{u}_k^1$	$p_k^1(a,b,\emptyset)$ $f_t^1(a,b,\emptyset)$	$\hat{u}_k^2$	$p_k^2(c,\emptyset)$ $f_t^2(c,\emptyset)$	$\hat{u}_k^3$	$p_k^1(abc,\emptyset)$ $f_t^3(abc,\emptyset)$	$\pi_k(0,1,2,3)$ $f_k(\pi_k)$	New Offer Required
1	6	$(0,-,-)$ $(1,\infty,\infty)$	5	$(0,-)$ $(1,\infty)$	4	$(0,-)$ $(1,\infty)$	$(b,a,c,\emptyset)$ $(1,1,\infty)$	3
2	6	$(0,-,-)$ $(1,\infty,\infty)$	5	$(0,-)$ $(1,\infty)$	3	$(1,-)$ $(2,\infty)$	$(\emptyset,\emptyset,\emptyset,abc)$ $(\infty,\infty,2)$	1
3	5	$(1,-,-)$ $(3,\infty,\infty)$	5	$(0,-)$ $(1,\infty)$	3	$(1,-)$ $(2,\infty)$	$(\emptyset,\emptyset,\emptyset,abc)$ $(\infty,\infty,2)$	1
4	4	$(2,-,-)$ $(4,\infty,\infty)$	5	$(0,-)$ $(1,\infty)$	3	$(1,-)$ $(2,\infty)$	$(b,a,c,\emptyset)$ $(4,1,\infty)$	3
5	4	$(2,-,-)$ $(4,\infty,\infty)$	5	$(0,-)$ $(1,\infty)$	2	$(2,-)$ $(5,\infty)$	$(b,a,c,\emptyset)$ $(4,1,\infty)$	3
6	4	$(2,-,-)$ $(4,\infty,\infty)$	5	$(0,-)$ $(1,\infty)$	1	$(3,-)$ $(6,\infty)$	$(\emptyset,\emptyset,\emptyset,abc)$ $(\infty,\infty,6)$	1
7	2	$(2,0,-)$ $(4,7,\infty)$	5	$(0,-)$ $(1,\infty)$	1	$(3,-)$ $(6,\infty)$	$(\emptyset,\emptyset,\emptyset,abc)$ $(\infty,\infty,6)$	2
8	2	$(2,0,-)$ $(4,7,\infty)$	4	$(1,-)$ $(8,\infty)$	1	$(3,-)$ $(6,\infty)$	$(b,a,c,\emptyset)$ $(4,8,\infty)$	3
9	2	$(2,0,-)$ $(4,7,\infty)$	4	$(1,-)$ $(8,\infty)$	0	$(4,0)$ $(9,9)$	$(\emptyset,\emptyset,\emptyset,abc)$ $(\infty,\infty,9)$	1
10	1	$(2,1,-)$ $(4,10,\infty)$	4	$(1,-)$ $(8,\infty)$	0	$(4,0)$ $(9,9)$	$(\emptyset,\emptyset,\emptyset,abc)$ $(\infty,\infty,9)$	1
11	0	$(2,2,0)$ $(4,11,11)$	4	$(1,-)$ $(8,\infty)$	0	$(4,0)$ $(9,9)$	$(\emptyset,\emptyset,\emptyset,abc)$ $(11,\infty,9)$	2
12	0	$(2,2,0)$ $(4,11,11)$	3	$(2,-)$ $(12,\infty)$	0	$(4,0)$ $(9,9)$	$(b,a,c,\emptyset)$ $(4,12,9)$	2

3.3. Properties of the Proposed Auction

This section will demonstrate several properties of the proposed auction mechanism.

Lemma 1. *The proposed auction terminates in a finite number of rounds with each bidder's target utility no less than his budget.*

Proof. Observe that, for every bidder $i \in M$, his initial target value $\hat{u}_1^i$ is an integer $\max_{A\in\mathcal{F}^i} v^i(A) \geq 0$. His target utility decreases by an integer if he is required to submit new offers. And 0 is also an integer and the offer $(i, \emptyset, 0)$ will be made if bidder i sets his target utility at 0. He cannot set the target utility below 0. When $\hat{u}_k^i = 0$, he has proposed all of his feasible bundles including $\emptyset$, so at least one of his offers is chosen in the feasible provisional assignment. Thus the bidder will stop reducing target utility when he reaches his budget. Therefore, the auction must terminate in a finite number of rounds. □

In the auction process, every bidder i's bidding price p_k^i is an increasing function of k. Because $\hat{u}_k^i \geq 0$ for all k, by the pricing formula his bidding price for every feasible bundle will never be higher than his valuation of the bundle nor above his budget, i.e., $p_k^i(A) \leq \min\{m^i, v^i(A)\}$ for all k and $A \in \mathcal{F}^i$.

Suppose the auction stops at round K. The final assignment is π_K, and the corresponding income distribution is $r^0 = \sum_{i\in M} p_K^i(\pi_K(i))$ for the seller, and $r^i = m^i - p_K^i(\pi_K(i))$ for every bidder $i \in M$. Bidder $i \in M$ is said to be a loser if his assigned bundle $\pi_K(i)$ is empty; otherwise, he is a winner. Let (π_K, r) denote the final outcome generated by the auction. Clearly, a loser gets no item and pays nothing.

The next result shows that the price of every sold bundle by the auction is at least as high as its marginal value to the seller.

Lemma 2. *At the outcome (π_K, r) generated by the proposed auction, every bidder $i \in M$ pays a price for his assigned bundle no less than its marginal value to the seller, i.e.,*

$$p_K^i(\pi_k(i)) \geq v^0(\pi_K(0) \cup \pi_K(i)) - v^0(\pi_K(0)).$$

Proof. It follows immediately from the Assigning stage of the auction that the seller is maximizing her revenue. □

Recall that we do not impose any specific condition upon the valuation functions of bidders and the seller. The next result shows that when the seller's valuation function is super-additive, the price of every sold bundle by the auction is at least as high as the seller's reserve price of the bundle.

Corollary 1. *When the seller's valuation function v^0 is super-additive, i.e., $v^0(A \cup B) \geq v^0(A) + v^0(B)$ for any two disjoint sets A and B of items, then at the outcome (π_K, r) generated by the proposed auction, every bidder $i \in M$ pays a price for his assigned bundle no less than the seller's reserve price, i.e., $p^i_K(\pi_K(i)) \geq v^0(\pi_K(i))$.*

Proof. It follows from Lemma 2 that

$$\begin{aligned} p^i_K(\pi_K(i)) &\geq v^0(\pi_K(0) \cup \pi_K(i)) - v^0(\pi_K(0)) \\ &\geq v^0(\pi_K(0)) + v^0(\pi_K(i)) - v^0(\pi_K(0)) = v^0(\pi_K(i)). \end{aligned}$$

□

We have another desirable property.

Corollary 2. *When all items are sold to the bidders by the auction, every bidder $i \in M$ pays a price for his assigned bundle no less than the seller's reserve price, i.e., $p^i_K(\pi_K(i)) \geq v^0(\pi_K(i))$.*

Proof. It follows from $\pi_K(0) = \emptyset$ and

$$p^i_K(\pi_k(i)) \geq v^0(\pi_K(0) \cup \pi_K(i)) - v^0(\pi_K(0)).$$

that we have $p^i_K(\pi_k(i)) \geq v^0(\pi_K(i))$ for all $i \in M$. □

The following lemma shows that the outcome given by the proposed auction is individually rational.

Lemma 3. *The outcome (π_K, r) generated by the proposed auction is individually rational and gives every bidder i at least his target utility $\hat{u}^i_K$.*

Proof. Observe that bidder i is a loser, i.e., his assigned bundle $\pi_K(i)$ is empty, only if $(i, \emptyset, 0)$ is in his offer set and thus his target utility is $\hat{u}^i_K = 0$. In this case, he achieves utility $u^i(\pi_K, r) = v^i(\emptyset) - p^i_K(\emptyset) = 0$. We prove that each winner i achieves his final utility $u^i(\pi_K, r) \geq \hat{u}^i_K$. To see this, let $A = \pi_K(i)$ denote i's assigned bundle, and his final bidding price on A is $p^i_K(A) = \min\{v^i(A) - \hat{u}^i_K, m^i\}$. We need to consider the following two

cases. Firstly, if $p^i_K(A) = v^i(A) - \hat{u}^i_K \leq m^i$ holds, then bidder i's utility is $u^i(\pi_K, r) = v^i(A) - p^i_K(A) = \hat{u}^i_K$. Secondly, if $p^i_K(A) = m^i \leq v^i(A) - \hat{u}^i_K$ holds, then bidder i's utility equals $u^i(\pi_K, r) = v^i(A) - m^i \geq \hat{u}^i_K$.

For the seller, consider the no-sale assignment $\bar{\rho}$ such that $\bar{\rho}(0) = N$ and $\bar{\rho}(i) = \emptyset$ for all $i \in M$. It is feasible (i.e., $\bar{\rho} \in \mathcal{A}$) and gives her the utility of $\sum_{i\in M_0} p^i_K(\bar{\rho}(i)) = p^0_K(N) = v^0(N)$. The seller's optimal choice guarantees her rationality, $u^0(\pi_K, r) = \max_{\rho\in\mathcal{A}} \sum_{i\in M_0} p^i_K(\rho(i)) \geq v^0(N)$. □

We are ready to establish the first key result concerning the auction: the outcome generated by the auction is not only a core allocation but also strongly Pareto efficient. This also gives a constructive proof of Theorem 1.

Theorem 2. *The outcome (π_K, r) yielded by the proposed auction is not only in the core but also strongly Pareto efficient.*

Proof. We first show that the outcome is a core allocation. By Lemma 3, the outcome (π_K, r) is individually rational. Suppose to the contrary that (π_K, r) is not in the core. Then there exists a coalition S consisting of the seller and at least one bidder with an allocation (ρ^S, t) such that $u^i(\rho^S, t) > u^i(\pi_K, r)$ for all $i \in S$. Without loss of generality, we can assume that $t^i \leq m^i$ for all $i \in S \setminus \{0\}$. To see this, suppose that there is a nonempty subset of bidders $J \subseteq S \setminus \{0\}$ such that $t^j > m^j$ for all $j \in J$. We can construct a new blocking coalition $I = S \setminus J$ who blocks (π_K, r) via an allocation (ρ^I, τ) defined as follows. Let every agent in I keep the same bundle and the same income as at (ρ^S, t), then we have $u^i(\rho^I, \tau) = u^i(\rho^S, t) > u^i(\pi_K, r)$. The seller $0 \in I$ takes back the bundles and transfers assigned to agents in J. Then we have $\rho^I(0) \supset \rho^S(0)$ and $v^0(\rho^I(0)) \geq v^0(\rho^S(0))$ by the monotonicity of the seller values. We also have $\tau^0 = \sum_{i\in I\setminus\{0\}}(m^i - t^i) > \sum_{i\in I\setminus\{0\}}(m^i - t^i) + \sum_{i\in J}(m^i - t^i) = t^0$. So we get $u^0(\rho^I, \tau) > u^0(\rho^S, t) > u^0(\pi_K, r)$.

For each bidder in the coalition $i \in S \setminus \{0\}$, if he is a winner of the auction, then his final target utility in the auction is lower than the improved utility, i.e., $\hat{u}^i_K \leq u^i(\pi_K, r) < u^i(\rho^S, t) = v^i(\rho^S(i)) + t^i - m^i$. This implies that he sets a higher intermediate price on $\rho^S(i)$, i.e., $\hat{p}^i_K(\rho^S(i)) = v^i(\rho^S(i)) - \hat{u}^i_K > m^i - t^i \geq 0$, and proposes a higher bidding price on $\rho^S(i)$, i.e.,

$$p^i_K(\rho^S(i)) = \min\{m^i, \hat{p}^i_K(\rho^S(i))\} \geq m^i - t^i.$$

If $i \in S \setminus \{0\}$ is a loser of the auction, he is also improved in the coalition, i.e., $0 = u^i(\pi_K, r) < u^i(\rho^S, t) = v^i(\rho^S(i)) + t^i - m^i$. It follows that $v^i(\rho^S(i)) > m^i - t^i$.

The loser sets his final target utility $\hat{u}^i_K = 0$, so his offering price on bundle $\rho^S(i)$ is

$$p^i_K(\rho^S(i)) = \min\{m^i, v^i(\rho^S(i))\} \geq m^i - t^i.$$

Then for the seller we have

$$\begin{aligned}\sum_{i\in S} p^i_K(\rho^S(i)) &\geq p^0_K(\rho^S(0)) + \sum_{i\in S\setminus\{0\}} (m^i - t^i)\\ &= u^0(\rho^S, t) > u^0(\pi_K, r) = \sum_{i\in M_0} p^i_K(\pi_K(i)).\end{aligned}$$

This contradicts the fact that π_K maximizes the seller's revenue based on the final bidding prices P_K. Conclusively, (π_K, r) is in the core.

We then show that (π_K, r) is strongly Pareto efficient. Suppose on the contrary that (π_K, r) is Pareto dominated by an allocation (ρ, t) such that $u^i(\rho, t) \geq u^i(\pi_K, r)$ for all $i \in M_0$ with at least one strict inequality. Because every bidder is (weakly) better off at (ρ, t), similar to the coalition members in above proof, at the final round of the auction each bidder $i \in M$ sets a comparatively lower target utility $\hat{u}^i_K \leq u^i(\rho, t) = v^i(\rho(i)) + t^i - m^i$ and a higher intermediate price $\hat{p}^i_K(\rho(i)) = v^i(\rho(i)) - \hat{u}^i_K \geq m^i - t^i$. If $t^i \leq m^i$, we have that $\hat{p}^i_K(\rho(i)) \geq m^i - t^i \geq 0$ and $p^i_K(\rho(i)) = \min\{m^i, \hat{p}^i_K(\rho(i))\} \geq m^i - t^i$; if $t^i > m^i$, it is obvious that $p^i_K(\rho(i)) \geq 0 > m^i - t^i$. Anyhow, we have $p^i_K(\rho(i)) \geq m^i - t^i$. The seller is also (weakly) better off, so we have

$$\begin{aligned}\sum_{i\in M_0} p^i_K(\rho(i)) &\geq p^0_K(\rho(0)) + \sum_{i\in M}(m^i - t^i) = u^0(\rho, t)\\ &\geq u^0(\pi_K, r) = \sum_{i\in M_0} p^i_K(\pi_K(i)) \geq \sum_{i\in M_0} p^i_K(\rho(i)).\end{aligned}$$

The last inequality comes from the fact that π_K maximizes seller's utility on prices P_K. So we have $\sum_{i\in M_0} p^i_K(\pi_K(i)) = \sum_{i\in M_0} p^i_K(\rho(i))$, and $p^i_K(\rho(i)) = m^i - t^i$ for all $i \in M$. These imply that, for each bidder $i \in M$, his final offering price on bundle $\rho^S(i)$ is $m^i - t^i$, (i.e., $(i, \rho^S(i), m^i - t^i) \in X^i_K$); for the auctioneer, ρ yields the highest revenue as π_K does (i.e., $\rho \in C(P_K)$). Note that if i strictly prefers $(i, \rho(i), m^i - t^i)$ to $(i, \pi_K(i), p^i_K(\pi_K(i)))$, then it holds that $f^i_K(\rho(i)) < f^i_K(\pi(i))$; if i is indifferent between $(i, \rho(i), m^i - t^i)$ and $(i, \pi_K(i), p^i_K(\pi_K(i)))$, then it holds that $f^i_K(\rho(i)) = f^i_K(\pi(i))$. So the allocation (ρ, t) Pareto dominates the auction outcome (π_K, r) means that $f_K(\rho) \leq f_K(\pi_K)$. It contradicts the choice rule of Assigning stage. □

The following example shows an interesting aspect of the proposed auction in that it finds a strongly Pareto efficient core allocation, which is not in the strong core, even though the strong core is not empty. More surprisingly, the assignment of the found strongly Pareto efficient core allocation is fully efficient, but the assignment of the unique strong core allocation is not fully efficient.

Example 4. *There are two items a and b, and two bidders* 1 *and* 2*. Valuations and budgets of bidders are given in Table 5. The seller values every bundle at* 0*.*

Table 5: Valuation and budget with two items.

Bidder	$\emptyset$	Item a	Item b	Package ab	Budget
1	0	2	–	10	5
2	0	–	5	6	5

Table 6 shows the bidding process of the dynamic auction.

Table 6: The bidding process of the dynamic auction via Example 5.

k	$\hat{u}_k^1$	$p_k^1(a, ab, \emptyset)$ $f_t^1(a, ab, \emptyset)$	$\hat{u}_k^2$	$p_k^2(b, ab, \emptyset)$ $f_t^2(b, ab, \emptyset)$	$\pi_k(0, 1, 2, 3)$ $f_k(\pi_k)$	New Offer Required
1	10	$(-, 0, -)$ $(\infty, 1, \infty)$	6	$(-, 0, -)$ $(\infty, 1, \infty)$	$(\emptyset, ab, \emptyset)$ $(1, \infty)$	2
2	10	$(-, 0, -)$ $(\infty, 1, \infty)$	5	$(0, 1, -)$ $(2, 2, \infty)$	$(\emptyset, \emptyset, ab)$ $(\infty, 2)$	1
3	9	$(-, 1, -)$ $(\infty, 3, \infty)$	5	$(0, 1, -)$ $(2, 2, \infty)$	$(\emptyset, ab, \emptyset)$ $(3, \infty)$	2
4	9	$(-, 1, -)$ $(\infty, 3, \infty)$	4	$(1, 2, -)$ $(4, 4, \infty)$	$(\emptyset, \emptyset, ab)$ $(\infty, 4)$	1
⋮						
11	5	$(-, 5, -)$ $(\infty, 11, \infty)$	1	$(4, 5, -)$ $(10, 10, \infty)$	$(\emptyset, ab, \emptyset)$ $(11, \infty)$	2
12	5	$(-, 5, -)$ $(\infty, 11, \infty)$	0	$(5, 5, 0)$ $(12, 10, 12)$	$(\emptyset, ab, \emptyset)$ $(11, 12)$	-

The auction ends with the allocation $(\pi_{12}, r_{12}) = ((\emptyset, ab, \emptyset), (5, 0, 5))$. This outcome assigns the two items to bidder 1, maximizes the seller's revenue and the total nontransferable utility and yields utilities $(u^0, u^1, u^2) = (5, 5, 0)$. The assignment is fully efficient. The allocation is in the core but not in the strong core. To see this, consider the coalition $\{0, 2\}$ with a feasible allocation $\rho^S = (\emptyset, \emptyset, ab)$ and a feasible income distribution $t = (5, 5, 0)$. The new allocation yields utilities $(5, 0, 1)$ weakly blocking (π_{12}, r_{12}). Actually, allocations (π_{12}, r_{12}) and (ρ, t) weakly block each other.

To show that the strong core is not empty, consider the allocation

$$(\pi, r) = ((\emptyset, a, b), (6, 4, 0)).$$

It offers utilities $(u^0, u^1, u^2) = (6, 1, 0)$. Clearly, it is individually rational. We show why it cannot be weakly blocked by any coalition. The first viable coalition is $S = \{0, 1\}$. In this, bidder 1 cannot pay more than 5 and thus the seller's revenue decreases. The second viable coalition $S = \{0, 2\}$ faces a similar problem. In it, bidder 2 cannot pay more than 5 and so the seller gets less. The final viable coalition is $S = \{0, 1, 2\}$. In this case, each bidder pays money to sustain the seller's revenue no less than 6; in return, each bidder gets one item. The only possible assignment is $\pi = (\emptyset, a, b)$. Using the same assignment, any redistribution of income favors one but also hurts someone else in the coalition.

Observe that the strong core assignment π is not fully efficient. In contrast, the assignment π_k found by the proposed auction is fully efficient.

The above example shows that under budget constraints, the auction can always find a strongly Pareto efficient core allocation but may not guarantee to locate a strong core allocation even if it exists. This does not contradict Theorem 3, which says that the proposed auction can always find a strong core allocation when no bidder is budget constrained.

We have proved that the proposed auction always finds a strongly Pareto efficient core allocation no matter whether bidders are budget constrained or not. We have also shown that when bidders are budget constrained, the auction cannot guarantee to find a strong core allocation even if it exists. So we have to accept some loss of market efficiency. This raises an important question as to whether the auction can find a strong core allocation when bidders are not budget constrained. Our next result establishes the equivalence between the core and the strong core when no bidders are budget constrained.

Lemma 4. *When no bidder faces a budget constraint, an allocation (π, r) is in the core if and only if it is in the strong core.*

Proof. The 'if' part is obvious, so we prove the 'only if' part. Suppose on the contrary that (π, r) is in the core but not in the strong core. By the definition, (π, r) is individually rational and cannot be blocked by any single agent. Then there would exist a viable coalition $S \subseteq M_0$ and an implementable pair (ρ^S, t) such that $u^i(\rho^S, t) \geq u^i(\pi, r)$ for all $i \in S$ with at least one strict inequality.

Let $j \in S$ be one of the agents being strictly improved, i.e. $u^j(\rho^S, t) > u^j(\pi, r)$. Define $\Delta = u^j(\rho^S, t) - u^j(\pi, r) > 0$. Define a new income distribution τ by

$$\tau^i = \begin{cases} t^j - \frac{|S|-1}{|S|}\Delta, & \text{if } i = j; \\ t^i + \frac{1}{|S|}\Delta, & \text{if } i \in S \setminus \{j\}; \\ t^i, & \text{if } i \in M \setminus S. \end{cases}$$

For every $i \in M \setminus S$, $\tau^i = t^i = m^i$ is feasible. For every $i \in S \setminus \{j\}$, $\tau^i > t^i > 0$ is also feasible. Let's consider the feasibility for agent j. If j is a bidder, then we have

$$\begin{aligned} t^j - \frac{|S|-1}{|S|}\Delta &= v^j(\pi(j)) + r^j - v^j(\rho^S(j)) + \frac{1}{|S|}\Delta \\ &\geq m^j - v^j(\rho^S(j)) + \frac{1}{|S|}\Delta \geq \frac{1}{|S|}\Delta. \end{aligned}$$

The first inequality arises because (π, r) is individually rational such that $v^j(\pi(j)) + r^j \geq m^i$; the second inequality is because bidder j is not budget constrained such that $m^j - v^j(\rho^S(j)) \geq 0$. If j is the seller, we have a similar condition

$$\begin{aligned} t^0 - \frac{|S|-1}{|S|}\Delta &= v^0(\pi(0)) + r^0 - v^j(\rho^S(0)) + \frac{1}{|S|}\Delta \\ &\geq v^0(N) - v^0(\rho^S(0)) + \frac{1}{|S|}\Delta \geq \frac{1}{|S|}\Delta. \end{aligned}$$

The first inequality arises because (π, r) is individually rational for the seller; the second inequality is attributed to the monotonicity of the seller's values. Anyhow, $\tau^j > 0$ is also feasible for agent j.

Then consider the implementable allocation (ρ^S, τ). For agent j, we have $u^j(\rho^S, \tau) = u^j(\rho^S, t) - \frac{|S|-1}{|S|}\Delta = u^j(\pi, r) + \frac{1}{|S|}\Delta$. For every other agent $i \in S \setminus \{j\}$, we have $u^i(\rho^S, \tau) = u^i(\rho^S, t) + \frac{1}{|S|}\Delta \geq u^i(\pi, r) + \frac{1}{|S|}\Delta$. We

reach a contradiction that (π, r) is blocked by the coalition S with allocation (ρ^S, τ). □

As an implication of Theorem 2 and Lemma 4, the proposed auction will always generate a strong core allocation when bidders face no budget constraints. Recall that a strong core allocation must be strongly Pareto efficient. The next result shows that the assignment of every strong core allocation is fully efficient.

Lemma 5. *When no bidder faces a budget constraint, the assignment π of every core allocation (π, r) is fully efficient.*

Proof. Suppose that (π, r) is in the core but its assignment π is not fully efficient. Then there must be another assignment $\rho \in \mathcal{A}$ such that $\sum_{i\in M_0} v^i(\rho(i)) > \sum_{i\in M_0} v^i(\pi(i))$. Define $\Delta = \sum_{i\in M_0} v^i(\rho(i)) - \sum_{i\in M_0} v^i(\pi(i))$. Clearly, we have $\Delta > 0$. Now consider the income distribution $t = (t^i)_{i\in M_0}$ with

$$t^i = v^i(\pi(i)) + r^i - v^i(\rho(i)) + \frac{\Delta}{|M_0|} \text{ for every agent } i \in M_0.$$

The individual rationality of (π, r) implies that $v^i(\pi(i)) + r^i - m^i \geq 0$. That bidder i is not budget constrained implies that $m^i \geq v^i(\rho(i))$. So we have $t^i > 0$ for every agent $i \in M_0$. What's more, we have

$$\sum_{i\in M_0} t^i = \sum_{i\in M_0} v^i(\pi(i)) + \sum_{i\in M_0} r^i - \sum_{i\in M_0} v^i(\rho(i)) + \Delta = \sum_{i\in M_0} r^i = \sum_{i\in M_0} m^i.$$

In sum, the income distribution t is feasible. It is easy to see that $u^i(\rho, t) = u^i(\pi, r) + \frac{1}{|M_0|}\Delta > u^i(\pi, r)$. We have a contradiction that (π, r) is blocked by the coalition of all agents M_0 with allocation (ρ, t). □

We have the following theorem.

Theorem 3. *When no bidder faces a budget constraint, the outcome (π_k, r) yielded by the proposed auction is in the strong core and the assignment π_k is fully efficient.*

4. CONCLUDING REMARKS

In this paper we have proposed an efficient dynamic menu auction for a general market where bidders can demand several items and have general and complex preferences and also face hard budget constraints. The auction always finds a strongly Pareto efficient core allocation.[6] It returns with a strong core allocation when no bidders are budget constrained. The current auction can be potentially used to tackle challenging and complex resource allocation problems involving significant indivisibility, heterogeneity in preferences and a shortage of financial resources. This study leaves a challenging open question of whether it is possible to design an efficient and strategy-proof dynamic auction which induces bidders to bid truthfully and yields an efficient outcome for the market studied here. Yang & Yu (2022) have shown this possibility for the assignment market.

References

Aggarwal, G., Muthukrishnan, S., Pál, D., & Pál, M. (2009). General auction mechanism for search advertising. In *Proceedings of the 18th International Conference on World Wide Web - WWW '09* (p. 241). Madrid, Spain: ACM Press.

Andersson, T., & Erlanson, A. (2013). Multi-item Vickrey–English–Dutch auctions. *Games and Economic Behavior*, *81*, 116–129.

Ashlagi, I., Braverman, M., Hassidim, A., Lavi, R., & Tennenholtz, M. (2010). Position auctions with budgets: Existence and uniqueness. *The B.E. Journal of Theoretical Economics*, *10*(1), 1–21.

Ausubel, L. M. (2004). An efficient ascending-bid auction for multiple objects. *American Economic Review*, *94*(5), 1452–1475.

[6] PS: After we completed the revision of the current article, we became aware of three recent papers. These authors have taken very different approaches from Yang & Yu (2018) and the current article. Gul et al. (2024) study a market with multiple indivisible goods and budget constrained buyers. Every buyer possesses gross-substitutes preferences and may demand several goods. The market may not have any Walrasian equilibrium because of budget constraints. They demonstrate the existence of a competitive and efficient probability distribution over feasible allocations of goods by using lotteries. Nguyen & Vohra (2024) examine a similar model and show the existence of social-approximate equilibria which permit price vectors with small excess demands. Nguyen & Teytelboym (2024) investigate another model and propose sealed-bid combinatorial auction formats which allow bidders to express their budgets. They prove the existence of Walrasian equilibrium if one can adjust budgets or the supply of goods.

Ausubel, L. M. (2006). An efficient dynamic auction for heterogeneous commodities. *American Economic Review*, *96*(3), 602–629.

Ausubel, L. M., & Milgrom, P. R. (2002). Ascending auctions with package bidding. *The B.E. Journal of Theoretical Economics*, *1*(1), Article 1.

Baldwin, E., Jagadeesan, R., Klemperer, P., & Teytelboym, A. (2023). The equilibrium existence duality. *Journal of Political Economy*, *131*, 1440–1476.

Baldwin, E., & Klemperer, P. (2019). Understanding preferences:"demand types", and the existence of equilibrium with indivisibilities. *Econometrica*, *87*(3), 867–932.

Benoit, J.-P., & Krishna, V. (2001). Multiple-object auctions with budget constrained bidders. *Review of Economic Studies*, *68*(1), 155–179.

Bernheim, D., & Whinston, M. (1986). Menu auctions, resource allocations and economic influence. *Quarterly Journal of Economics*, *101*, 1–31.

Che, Y.-K., & Gale, I. (1998). Standard auctions with financially constrained bidders. *Review of Economic Studies*, *65*(1), 1–21.

Crawford, V. P., & Knoer, E. M. (1981). Job matching with heterogeneous firms and workers. *Econometrica*, *49*(2), 437–450.

Danilov, V., Koshevoy, G., & Murota, K. (2001). Discrete convexity and equilibrium in economies with indivisible goods and money. *Mathematical Social Science*, *41*, 251–273.

Day, R., & Milgrom, P. (2008). Core-selecting package auctions. *International Journal of Game Theory*, *36*(3-4), 393–407.

Debreu, G., & Scarf, H. (1963). A limit theorem on the core of an economy. *International Economic Review*, *4*(3), 235–246.

Demange, G., Gale, D., & Sotomayor, M. (1986). Multi-item auctions. *Journal of Political Economy*, *94*(4), 863–872.

Dobzinski, S., Lavi, R., & Nisan, N. (2012). Multi-unit auctions with budget limits. *Games and Economic Behavior*, *74*(2), 486–503.

Erdil, A., & Klemperer, P. (2010). A new payment rule for core-selecting package auctions. *Journal of the European Economic Association*, *8*(2-3), 537–547.

Fujishige, S., & Yang, Z. (2002). Existence of an equilibrium in a general competitive exchange economy with indivisible goods and money. *Annals of Economics and Finance*, *3*, 135–147.

Gillies, D. B. (1953). *Some theorems on n-person games*. Princeton University.

Gul, F., Pesendorfer, W., & Zhang, M. (2024). Efficient allocation of indivisible goods in pseudomarkets with constraints. *Journal of Political Economy*, *132*, 3708–3736.

Gul, F., & Stacchetti, E. (2000). The English auction with differentiated commodities. *Journal of Economic Theory*, *92*(1), 66–95.

Hafalir, I. E., Ravi, R., & Sayedi, A. (2012). A near Pareto optimal auction with budget constraints. *Games and Economic Behavior*, *74*(2), 699–708.

Herings, P. J.-J., & Zhou, Y. (2022). Competitive equilibria in matching models with financial constraints. *International Economic Review*, *63*(2), 777–802.

Kelso, A. J. S., & Crawford, V. P. (1982). Job matching, coalition formation, and gross substitutes. *Econometrica*, *50*, 1483–1504.

Klemperer, P. (2004). *Auctions: Theory and Practice*. Princeton: Princeton University Press.

Koopmans, T., & Beckmann, M. (1957). Assignment problems and the location of economic activities. *Econometrica*, *25*(1), 53–76.

Krishna, V. (2010). *Auction Theory* (2nd ed.). Burlington, MA: Academic Press/Elsevier.

Laan van der, G., & Yang, Z. (2016). An ascending multi-item auction with financially constrained bidders. *Journal of Mechanism and Institution Design*, *1*(1), 107–147.

Laffont, J.-J., & Robert, J. (1996). Optimal auction with financially constrained buyers. *Economics Letters*, *52*(2), 181–186.

Maskin, E. (2000). Auctions, development, and privatization: Efficient auctions with liquidity-constrained buyers. *European Economic Review*, *44*(4-6), 667–681.

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

Milgrom, P. (2004). *Putting Auction Theory to Work*. Cambridge, UK ; New York: Cambridge University Press.

Milgrom, P. (2007). Package auctions and exchanges. *Econometrica*, *75*(4), 935–965.

Milgrom, P. (2017). *Discovering Prices: Auction Design in Markets with Complex Constraints*. Columbia University Press.

Milgrom, P., & Strulovici, B. (2009). Substitute goods, auctions, and equilibrium. *Journal of Economic Theory*, 212–247.

Mishra, D., & Parkes, D. (2007). Ascending price Vickrey auctions for general valuations. *Journal of Economic Theory*, *132*(1), 335–366.

Nguyen, T., & Teytelboym, A. (2024). Approximate combinatorial auctions with budgets. *Unpublished Manuscript, University of Oxford*.

Nguyen, T., & Vohra, R. (2024). (Near-)substitute preferences and equilibria with indivisibilities. *Forthcoming in Journal of Political Economy*.

Pai, M. M., & Vohra, R. (2014). Optimal auctions with financially constrained buyers. *Journal of Economic Theory*, *150*, 383–425.

Pitchik, C. (2009). Budget-constrained sequential auctions with incomplete information. *Games and Economic Behavior*, *66*(2), 928–949.

Predtetchinski, A., & Herings, P. (2004). A necessary and sufficient condition for non-emptiness of the core of a non-transferable utility game. *Journal of Economic Theory*, *116*(1), 84–92.

Scarf, H. (1967). The core of an N person game. *Econometrica*, *35*(1), 50–69.

Shapley, L., & Scarf, H. (1974). On cores and indivisibility. *Journal of Mathematical Economics*, *1*(1), 23–37.

Shapley, L., & Shubik, M. (1971). The assignment game I: The core. *International Journal of Game Theory*, *1*(1), 111–130.

Shioura, A., & Yang, Z. (2015). Equilibrium, auction, generalized gross substittes and complements. *Journal of the Operations Research Society of Japan*, 61–103.

Sun, N., & Yang, Z. (2006). Equilibria and indivisibilities: Gross substitutes and complements. *Econometrica*, *74*(5), 1385–1402.

Sun, N., & Yang, Z. (2009). A double-track adjustment process for discrete markets with substitutes and complements. *Econometrica*, *77*(3), 933–952.

Sun, N., & Yang, Z. (2014). An efficient and incentive compatible dynamic auction for multiple complements. *Journal of Political Economy*, *122*(2), 422–466.

Talman, A. J. J., & Yang, Z. (2015). An efficient multi-item dynamic auction with budget constrained bidders. *International Journal of Game Theory*, *44*(3), 769–784.

Yang, Z., & Yu, J. (2018). Efficient ascending menu auctions with budget constrained bidders. *Working Paper Series No. 18/09, DERS, University of York*.

Yang, Z., & Yu, J. (2022). An efficient and strategy-proof multi-item ascending auction under financial constraints. *Working Paper Series No.* 22/02*, DERS, University of York*.

Zheng, C. (2001). High bids and broke winners. *Journal of Economic Theory*, *100*(1), 129–171.

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

Papers should be submitted electronically in PDF to the Journal of Mechanism and Institution Design through the website `http://www.mechanism-design.org/`.

University of York, University of Klagenfurt, Southwestern University of Economics and Finance, 2024

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.